Xi'an Jiaotong-Liverpool University
西交利物浦大学

Xi'an Jiaotong-Liverpool University

Department of Computer Science and Software Engineering

Predict the Impact of Visual Distortion on Medical Images

Author:
Yuhao Sun (1611049)

Supervisor:
Dr. Gabriela Mogos
Assessor:
Prof. Ka Lok Man

A dissertation submitted

in partial fulfillment of the requirements for the degree of

BSc Information and Computing Science

June 4, 2020

Abstract

Medical imaging techniques are increasingly used in the medical community, such as hospitals and healthcare institutions. However, similar to most images, the distortion (blurry, contrast- and noise-distorted) possibly happens in the medical imaging area. To predict the impact of visual distortion on medical images, a four-step methodology has been designed to find if the current Objective Image Quality Assessment (IQA) mathematical models can perform as good as human eyes. Our work has led us to the conclusion that Subjective and Objective IQA can result in similar results at this stage. The evidence from the project also suggests Perception based Image Quality Evaluator (PIQE) is a relatively efficient mathematical model of No-Reference IQA (NR-IQA). Compared to PIQE, we have found an innovative result which both Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) and Naturalness Image Quality Evaluator (NIQE) could not well perform under the context of medical imaging, especially X-ray CT scans in our case; however, BRISQUE and NIQE are considered as two of best NR-IQA models. The results of the project also indicate some minor findings in clinical and computational disciplines. Finally, evaluation and suggestions on future work on this topic have been provided.

Key words: Distortion, Medical Imaging, Image Quality Assessment (IQA), Subjective IQA, No-Reference IQA (NR-IQA), Perception based Image Quality Evaluator (PIQE).

Acknowledgements

The work is carried out by Yuhao Sun and submitted in partial fulfilment of the requirements for the degree of BSc Information and Computing Science. It is supervised by Dr Gabriela Mogos, Department of Computer Science and Software Engineering at Xi'an Jiaotong-Liverpool University (XJTLU).

We gratefully acknowledge Dr Gabriela Mogos for her valuable suggestions and discussions as project supervisor during the whole project process. We also appreciate the assessment and suggestions from project assessor Prof Ka Lok Man, and the initial guidance from Dr Lucie Lévêque.

Finally, the dissertation, project and project team pay tribute to the doctors and hospital participated in the project, and all medical personnel fighting COVID-19.

We hope this dissertation will not be the end of the academic thinking, and we hope the previous sentence is not just a hope.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Medical imaging techniques are increasingly used in the medical community, such as hospitals and healthcare institutions. The doctors, including clinicians and radiologists, can view the internal construction of the particular organ or tissue of the patient, so that confirm the corresponding diagnosis and treatment suggestions to the patient. However, similar to most images, the distortion (blurry, contrast- and noise-distorted) possibly happens in the medical imaging area [1]. The medical images with distortion potentially affect the decisions of doctors at the clinical level, which may cause the misdiagnosis, missed-diagnosis, or other inaccurate judgements. In this project, we aim to predict the impact of visual distortion on medical images. Accurately, by using both Subjective and Objective Image Quality Assessments, determine if the results of current mainstream mathematical models can perceive the quality of medical images as good as human eyes. Additionally, we try to find a relatively efficient mathematical model which can be used in the medical imaging area or industry in the future. In spite of the fact that the limited number of observers in Subjective Image Quality Assessment, a mathematical model, Perception based Image Quality Evaluator (PIQE) has been proved its efficiency under medical imaging context in our project. This dissertation demonstrates the full process to discover the answers to the above topic.

Chapter 2

Background

The objective of this chapter is to demonstrate the background of the project. When you have read this chapter, you will:

- have been introduced the glossary of terms appeared frequently in this project other than computer science discipline;
- have recalled the previous literature emphasised in the Specification and Design Report [2];
- have been introduced the literature of Medical Imaging and Image Quality Assessment in a further and detailed way, including the limitations of previous research; and also a step-by-step definition to the mathematical models used in Objective Image Quality Assessment;
- have fully understood the research questions, project requirements and the comparison between the solution produced in project and existing approaches.

2.1 Glossary of Terms

The project, Predict the Impact of Visual Distortion on Medical Images, is an interdisciplinary topic combining with the subjects of computer science and medicine. In addition to computer science, knowledge of medicine, especially medical imaging, is necessary. However, the professions in the computer science area potentially are unfamiliar to specific appeared terms. This glossary will introduce the terms used frequently in the project so that you can get a better understanding of this project, as below:

- Medical Imaging

  A technique and process of creating the visual representation of certain organs or tissues of internal human bodies under medical intervention and clinical analysis circumstances [3]. It includes the common types of X-ray Computed Tomography (X-ray CT), Magnetic Resonance Imaging (MRI), Nuclear Medicine Imaging, Ultrasound Imaging, but not limited to these [3]. Medical Images are the pictures generated by medical imaging machines, such as X-ray CT.

- Lesion

  A lesion is an abnormal change of an organ or tissue in a living being due to the injury or damage [4].

- CT

  Computed Tomography, i.e. CT, is one of the medical imaging techniques used in the clinical area. With the ideas of geometry in mathematics, according to the rotation around a fixed axis, a 3-dimensional volume can be generated with a series of 2-dimensional images for a specific object, usually the human bodies [5]. The common type of CT is X-Ray CT and usually, the term "CT" means X-Ray CT. However, it is not merely limited to it. In the project, all medical images in the dataset are X-Ray CT scans.

- Slice and Volume

  A slice is one 2-dimensional image. A volume is one 3-dimensional image collection made of a series of 2-dimensional slices. The doctors can continuously adjust the layers of the current slice to view the 3-dimensional effect of the object. The relationship between "slice" and "volume" is, for example, we can say "this is a 330-slice volume CT".

- Other frequent abbreviations

  - **IQA** Image Quality Assessment
  - **Subjective IQA** Subjective Image Quality Assessment
  - **Objective IQA** Objective Image Quality Assessment
  - **NR-IQA** No-Reference Image Quality Assessment
  - **BRISQUE** Blind/Referenceless Image Spatial Quality Evaluator
  - **NIQE** Naturalness Image Quality Evaluator
  - **PIQE** Perception based Image Quality Evaluator

We are aware that it may be not enough for you to understand everything above fully with one-time reading. Hence, we encourage you to come back to this part again once you have the difficulty of understanding specific terms in future reading.

2.2 Medical Imaging

Previous studies have emphasised that the people paid increasing attention to medical imaging techniques since the 1890s' the discovery of X-rays [6]. In these decades, the strength of medical imaging has been improving, and it has widely used in the medical community, especially in hospitals. From current perspectives, classic medical imaging techniques include X-ray CT, Magnetic Resonance Imaging (MRI), Ultrasound Imaging, and Radionuclide Imaging, which belong to the radiology branch [7]. Other well-known medical imaging techniques include cardiology, pathology, ophthalmology [8].

Medical images generated by medical imaging techniques can vividly show the internal construction of a particular organ or tissue regarding a patient [7]. Experienced radiologists or clinicians can find the possibly suspicious lesions through observing the images. According to the suitable judgements, proper diagnosis and treatment for the patient can be preliminary confirmed. In summary, medical imaging techniques assist doctors in the clinical area by two processes between doctors and images, which are visual perception and cognition [8].

Recent years, medical imaging is increasingly studied by computer science professionals. They fit computational intelligence techniques into medical imaging area to assist doctors in analysing the uncommon situation of patients and trying to find possible solutions. Neural networks, evolutionary optimisation techniques, and wound inflammation by colour analysis improve the medical imaging area gradually advanced [9]. Additionally, towards different medical imaging technique and specific organs or tissues of human bodies, the research on particular branches become a target of computer science professionals. For examples, image-guided lung biopsy, ultrasound imaging for knee osteoarthritis detection, and virtual surgery [10]. The development of medical imaging technique has turned many unsolved cases in the medical community from impossible to solved instances.

As mentioned earlier, two processes visual perception and cognition influence doctors to confirm the diagnosis decisions and treatment decision through the medical images generated by medical imaging techniques. However, the process is not always precise and exact. Subjective influential factors, such as the difference of room lighting environment and the images display devices; Objective influential factors, such as distortion on medical images, both types of influential factors are not negligible in the processes [8]. Instead, the improper measures to subjective and objective influential factors possibly result in the contrary results on diagnosis compared to the actual situation of the patient. Image Quality Assessment is necessary to involve in this to ensure the high-quality of medical images. The following section introduces Image Quality Assessment.

2.3 Image Quality Assessment

In computer science discipline, Image Quality Assessment is an extremely essential approach to measure the quality of an image. Image Quality Assessment (IQA) consists of two sub-assessments, which are Subjective Image Quality Assessment (Subjective IQA) and Objective Image Quality Assessment (Objective IQA) [1]. The full classification of IQA refers to Figure 2.1.

2.3.1 Subjective Image Quality Assessment

Prior research suggests that Subjective IQA is the most reliable approach to measure the quality of an image as the eventual users of most multimedia applications are human beings [11]. Specifically, in Subjective IQA, the selected observers are required to decide the quality of the given images, most of the time within a fixed period [11]. The existing studies have examined two main categories are under Subjective IQA, including [1, 11]:

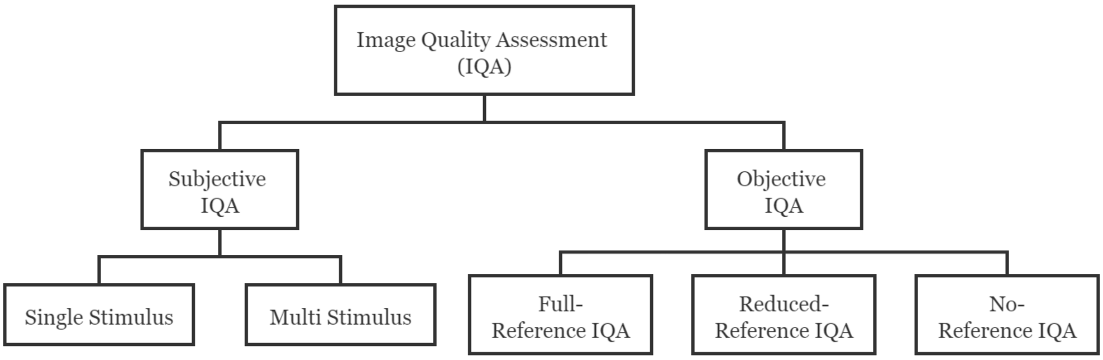


Figure 2.1: The Full Classification of IQA

- Single Stimulus Methods;
- Multi Stimulus Methods.

The most significant difference between the two above methods is the number of stimuli [1]. In Single Stimulus Methods, only one incentive will be provided to observers, i.e. only test image. In Multi Stimulus Methods, two incentives will be presented to observers, which include a reference image and corresponding test image. Now please imagine a scenario, when a person sees a picture in the smartphone and is required to judge the quality of the picture, this is a Single Stimulus Methods; when a person sees two pictures and has been told one of two is a high-quality image and is required to judge the quality of another picture, this is a Multi Stimulus Methods. For both two methods, besides the above difference, the human observers are required to categorise the quality of the test image according to a 5-point Likert scale, within a fixed time [12].

In our project, we select Single Stimulus Methods for the approach in Subjective IQA. We have explained why we chose this method in Section 2.5.2.

A further question is how do we decrease the limitations and drawbacks Subjective IQA brought by. Although we can expect from the name of this approach that the results of this approach will be highly subjective, it is exactly important to control the negative impact of influential factors including system, context, and human influential factors [13], which usually include system, context and human influential factors [14]. System influential factors are the methods we will adopt, in this case, are Single Stimulus Methods. Context influential factors include the viewing conditions when process the assessments. Human influential factors are the level of personal emotional states. The key to ensuring the success of the assessment is to decrease the negative effects brought by these influential factors to the maximum extent.

Additionally, many studies appear consistent with that Subjective IQA is an expensive and time-consuming approach [1, 11]. Indeed, the human observers are necessarily required in Subjective IQA causes the approach expensive; it needs time for observers to process the assessment causes the method time-consuming. Even so, Subjective IQA plays an essential role in IQA research, as the ultimate users of it are entirely human beings [1].

2.3.2 Objective Image Quality Assessment

Objective IQA is supposed to return the numerical score results generated by mathematical models, which should perform similarly with human observers [11]. Objective IQA can be divided into the three categories, which are Full-Reference Image Quality Assessment (FR-IQA), Reduced-Reference Image Quality Assessment (RR-IQA), and No-Reference Image Quality Assessment (NR-IQA), according to the availability of reference images [15]. Like their names, FR-IQA means reference images are available; RR-IQA means reference images are partially available, and NR-IQA means there are no reference images [15, 16]. The descriptions of each type of Objective IQA are as below:

- Full-Reference Image Quality Assessment (FR-IQA)

  FR-IQA means reference images are available entirely, which can be deemed as "high-quality" or "distortion-free". Under FR-IQA, the algorithms firstly perceive the reference images and then score the test images. Famous mathematical metrics of FR-IQA are Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR), in which PSNR is a lower complexity version converted by MSE [17]. Different from the above two models, the classic model Structural Similarity (SSIM) constructs the structure of the image and thus compare the similarity of the reference image and corresponding test image [18]. SSIM is frequently adopted in the research as the results of it are have a more considerable resemblance to the results generated by human eyes, compared to MSE and PSNR [17, 18]. In addition to MSE and SSIM, [19], [20], [21] and [22] are other FR-IQA models.

- Reduced-Reference Image Quality Assessment (RR-IQA)

  RR-IQA means reference images are partially available, somewhere between FR-IQA and NR-IQA. For example, there are some watermarks on the images [23]. In the 1990s', without current advanced technology, it always happened that the full image could not be extracted from the videos on multimedia communication networks; instead, some features of the image, which is where RR-IQA was derived from [24]. RR-IQA is less frequently used than FR-IQA and NR-IQA. There are three types of RR-IQA methods, including the models of source images, the models of distortion of captured images and the models of Human Visual System [23]. [25], [26], [27], [28] and [29] are some famous mathematical models of RR-IQA covering above three types.

- No-Reference Image Quality Assessment (NR-IQA)

  NR-IQA means reference images are unavailable entirely, which the image quality merely can be judged by corresponding test image. NR-IQA is matching most of the circumstances that happened in reality. Now imagine a scenario, you have a picture on your phone and are required to judge the quality of it - this is precisely NR-IQA. The significant feature of NR-IQA is there is no reference images but test images. The objective of NR-IQA models aims to

construct mathematical models which can perceive the quality of images automatically and similar to the results generated by human eyes to the maximum extent [30]. NR-IQA is more complicated than FR-IQA and RR-IQA, as the models need to consider many unexpected distortion types [30]. The human can judge the quality of the image without the reference image is because the brains are sufficiently knowledgeable to store a lot of information which tells them how a good-quality picture should look like [11, 15].
Most commonly adopted NR-IQA models include Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [31] and Naturalness Image Quality Evaluator (NIQE) [32]. In this project, we also used another mathematical model, Perception based Image Quality Evaluator (PIQE) [33]. A further literature review on NR-IQA is in Section 2.4. [34], [35], [36], [37], and [38] are other mathematical models of NR-IQA.

In our project, we select NR-IQA as the approach for Objective IQA. We have explained why we have chosen this specific method in Section 2.5.2. A further literature review on NR-IQA demonstrates in Section 2.4.

A closer look at the literature on the algorithms of Objective IQA, however, also reveals several limitations and drawbacks. One of the key constraints is the adaptability of the algorithms, which we have discussed in Section 5.3.2 of the Specification and Design Report [2]. Because some of the algorithms [31, 32], as mentioned earlier, have been tested with merely natural images, we cannot only predict their performances with medical images. Thus, the potential failure of the algorithms will be within our expectations, and in this case, we will consider alternatives, and the project will be more stringent. This project addresses the distortion appeared in medical images, so far lacking in the scientific literature. To fill this literature gap, more work is necessary for the algorithms of Objective IQA, and this project identifies the performances of preceding algorithms under the medical images environment.

2.4 Mathematical Models of No-Reference IQA

2.4.1 BRISQUE

Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) is one of the well-known NR-IQA mathematical models. It extracts Natural Scene Statistics (NSS) of the image of normalised luminance coefficients to perceive the existence of distortion [31]. To normalise, a normalisation approach named Mean Substracted Contrast Normalization (MSCN) will be adopted [31]. The calculation of MSCN coefficients can refer to Equation 2.1.

$$\widehat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{2.1}$$

To take a further look on Equation 2.1, $I(i,j)$ stands for the image intensity at the pixel (i,j). $\mu(i,j)$ and $\sigma(i,j)$ respectively means local mean field (Gaussian blur of the image) and local variance field (Gaussian Blur of the square of the difference of original image and μ) [31]. The equation is also a core in most of NR-IQA mathematical models, and used in NIQE and PIQE.

BRISQUE is a simple mathematical model and thus has a relatively lower complexity of computation [31]. The score generated by BRISQUE usually ranges from 0 to 100. Same to all NR-IQA models, the lower score means the higher image quality and vice versa.

An image scored by BRISQUE will usually follow the below methodology [31]:

1. Extract NSS: Calculate MSCN coefficients;
2. Extract NSS: Calculate the pairwise products by using MSCN coefficients to find neighbourhood relationships (Horizontal, Vertical, Left Diagonal, Right Diagonal);
3. Calculate Feature Vectors: Fit the MSCN image to Generalized Gaussian Distribution (GGD), which can get the first two elements of feature vectors (size of 36-by-1);
4. Calculate Feature Vectors: Fit the pairwise products images to Asymmetric Generalized Gaussian Distribution (AGGD), which can get all other 16 elements of feature vectors;
5. Repeat: Downsize another image as 50% as original one and repeat all same steps above to generate other 18 elements;
6. Score: With training by datasets in Support Vector Machine, by the feature vectors calculated above, the ultimate score of BRISQUE can be generated for the image.

2.4.2 NIQE

Naturalness Image Quality Evaluator (NIQE) is another well-known NR-IQA mathematical model. Similar to BRISQUE, it is a "quality-aware" collection of statistical features based on NSS model [32]. The undistorted images in the nature are the targets of these features. However, different from BRISQUE, NIQE uses measurable deviations from statistical regularities observed in natural images to perceive the quality of images, without prior knowledge to the human-rated distorted images [32].

With the evaluation by Mittal and et al, NIQE has performed extremely well and beyond their expectations [32]. Same to all NR-IQA models, the lower score means the higher image quality and vice versa.

An image scored by NIQE will usually follow the below methodology [32]:

1. Extract NSS: Calculate MSCN coefficients;
2. Patch Selection: Divide the image to $P \times P$ patches, calculate the average variance of the patches, and choose those patches which contain the richest information;

3. Characterising Image Patches: Calculate the feature vectors by GGD and AGGD to get 18 elements, and to yield another 18 elements by downsizing with a factor of 2 (same to the Step 3, 4, 5 in BRISQUE, Section 2.4.1);

4. Score: Fit the 36 features to multivariate Gaussian (MVG) model and compare this with a natural MVG model, then the ultimate score of NIQE can be generated for the image.

2.4.3 PIQE

Perception based Image Quality Evaluator (PIQE) is a novel NR-IQA mathematical model. Different from BRISQUE, PIQE can perceive the quality of the distorted images without any training data; however, currently, most of NR-IQA models are based on supervised learning, which personal opinions involved [33]. To predict the quality of the image, PIQE extracts the local features and focus on the significant spatial regions perceived, to mimic the human behaviours to the maximum extent [33]. The complexity of the model is relatively low.

PIQE also adopts a 5-point Likert scale to give the quality of the image. Instead of five numerical numbers, it uses "Excellent", "Good", "Fair", "Poor" and "Bad" to stand for the quality level [33]. In each level, the score range is $[0, 20]$, $[21, 35]$, $[36, 50]$, $[51, 80]$ and $[81, 100]$ respectively. The levels of the quality and corresponding score ranges are generated based on the dataset in LIVE Image Quality Assessment Database [33, 39], which can be deemed as reliable. Please note the size of ranges in each level is not exactly same. Similar to the BRISQUE and NIQE, the lower score means the higher quality and vice versa.

An image scored by PIQE will usually follow the below methodology [33]:

1. For each image pixel, MSCN coefficient will be computed;

2. Divide the image into blocks with a size of 16-by-16;

3. According to the variance of the value computed in the Step 1 (MSCN coefficient), the high spatially active blocks will be confirmed;

4. According to the MSCN coefficients in each block, decide the distortion level of noise and blocking artifacts;

5. According to a suitable threshold, the blocks will be classified as distorted blocks (further classified based on distorted types) and undistorted blocks;

6. Among those distorted blocks classified in the previous step, "noticeableArtifactsMask" and "noiseMask" will be generated automatically as from distorted blocks with blocking artifacts and with noise respectively;

7. The average score of all distorted blocks will be the ultimate score of PIQE for this image. The corresponding level of quality can be found on the aforementioned scale with this PIQE score.

2.5 Project

2.5.1 Objectives

The objectives of this project are to find the answers relevant to the distortion of medical imaging techniques in a computer science discipline, as follows:

- To confirm whether the results of Subjective Image Quality Assessment (Subjective IQA) and Objective Image Quality Assessment (Objective IQA) are similar or not, and to what extent;
- To find a mathematical model in No-Reference Image Quality Assessment (NR-IQA) which can perform relatively well in medical imaging discipline.

2.5.2 Choose Methods

In Subjective IQA, there are two methods (Single Stimulus Methods and Multi Stimulus Methods). In Objective IQA, there are three methods (FR-IQA, RR-IQA, and NR-IQA). However, to conduct Subjective IQA and Objective IQA do not mean it is necessary to perform all of the methods under each IQA. According to the project topic, a suitable method for each IQA should be chosen.

In this project, we will predict the impact of visual distortion on medical images, and thus the research object should be medical images. In Subjective IQA, we aim to find suitable doctor observes to judge the quality of medical images, similar to what they actually experience when they are working. Under such a circumstance, Single Stimulus Methods in Subjective IQA should be correctly chosen, as there is no any reference images for doctors to judge. To ensure the same standard in Objective IQA that there is also no reference images, we need to choose NR-IQA. Eventually, Single Stimulus Methods in Subjective IQA and NR-IQA in Objective IQA have been adopted in our project. We will only consider and discuss these two methods mentioned above in all future.

2.5.3 Solution Comparison

The results of the project have demonstrated the poor performances of two mathematical models BRISQUE and NIQE in the medical imaging area, especially for X-ray CT. This is an extremely novel result according to the previous literature reviews. We have also found another mathematical model named PIQE, performing as desirable as human beings in X-ray CT. However, all of the three models have claimed they can work well in natural images.

Chapter 3

Design

The objective of this chapter is to demonstrate the ideas of design for the project. When you have read this chapter, you will:

- have been introduced the framework of methodology design;
- have been introduced the complete design for whole project methodology, including Dataset Construction, Subjective and Objective Image Quality Assessments and Comparative Analysis;
- have been demonstrated the necessary revisions and justifications compared to the Specification and Design Report [2].

3.1 Overview

In the project, four steps include Dataset Construction, Subjective IQA, Objective IQA, and Comparative Analysis will be conducted in a logical order. However, before the first step of Dataset Construction, the Pre-Assessment Questionnaire will be first going, which is a revision compared to previous plans. With the completion of Comparative Analysis, we will test the methodology designed by using several testing cases to ensure our results received are without errors. Please refer to Figure 3.1 for the flowchart of the project.

3.2 Methodology

3.2.1 Dataset Construction

The first step of the methodology is to choose a suitable and project-tailored dataset. Before the actions, we are fully aware that an unsuitable dataset may fail the project, and we need to pay more attention to the construction of the dataset. It is necessary to consider what characteristics of this project-tailored dataset will contain. In summary, we should thoroughly examine the following aspects when constructing the dataset:

- Image files

  "Image files" here means the visual perception in two-dimensional form, preferably PNG/JPEG files. A direct reason is that some of the image datasets are

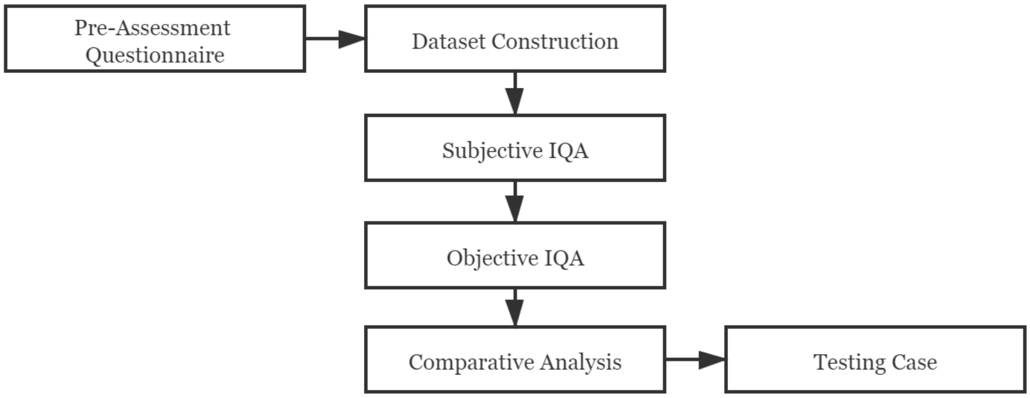


Figure 3.1: Overview of The Project

merely shown as numerical data, under the subject of the computer science context. However, only statistical data will not be acceptable because it will be impossible for human observers to assess the image quality by reading a sequence of numbers in Subjective Image Quality Assessment.

- Distortion-free

  The selected medical images should be high-quality. No distortion exists in the image. These images are then regarded as reference images, and therefore can be processed to other types of distortion images later on.

- Currency

  The selected medical images should be fresh, which means they were not created "long time" ago. Here sounds vague, but the medical images created within five years (i.e., since 2015) can be categorised as a suitable currency. An important reason to adopt the time range "within five years" is: for each personal profile on Google Scholar, it only shows the number of "All Citations" and "Since 2015 Citations", to respectively demonstrate all paper citations and recent active citations of the author.

- Annotations

  When a lesion appears in a medical image, we at least need to know the specific location and the size of the lesion. Some guidance on the judgements of lesions from the medical professionals will be instrumental. The selected medical images should contain the essential such annotations. An example of annotated image can refer to Figure 3.2. The part enclosed by the green frame is a lesion.

With the completion of gathering the medical images, we decide to process the original images to other types of distorted images. Several similar images can be analysed within a group in later assessments. We have planned to select 20 original ("high-quality") images and process them to other 30 distorted images. The distortion types we process to depend on the frequency of types that happened in the actual work of doctors. Therefore, the eventual project dataset will consist of 50 medical images totally.

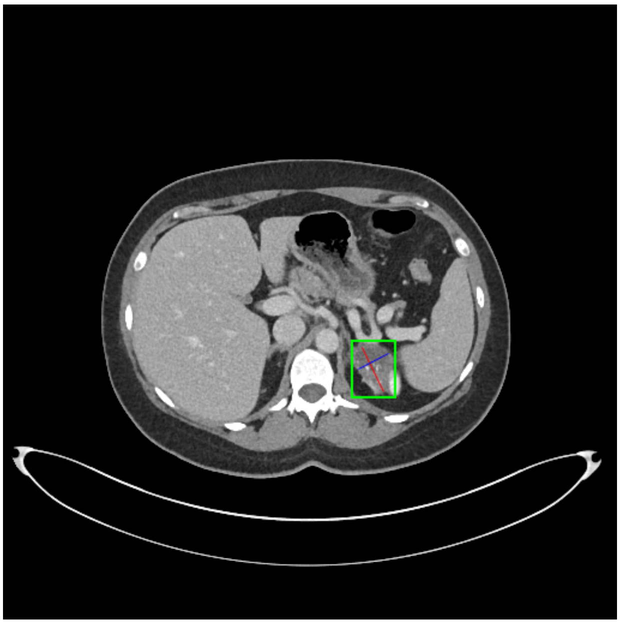


Figure 3.2: An Annotated Medical Image

3.2.2 Subjective Image Quality Assessment

Pre-Assessment Questionnaire

Before the formal assessment, we would like to analyse the project topic from the perspective of medical workers. As this topic is based on medical and clinical environments, the idea from the medical areas other than computer science seems to play a shared essential role.

Therefore, a questionnaire is necessary. The questionnaire will be divided into two parts around "the distortion" topic: personal background information and medical imaging related questions. We will firstly ask several questions about position and experience, then ask them questions about the distortion according to their views. The number of items should be controlled between eight and twelve, and most of them should be MCQ instead of sentence answers. Most importantly, the questionnaire cannot be designed for answering longer than five minutes as failure to do so may result in a negative mental effect on takers, such as pressure.

The respondents should be either clinical doctors or radiologists and currently are working at a valid hospital or healthcare institution, which is because radiologists and clinicians are supporting each other in the clinical area [40], and the eventual decisions from clinicians possibly will be changed after the investigations and discussions from radiologists [41]. The questionnaire will be created, and the data will be collected via the university survey website, Survey @ XJTLU (survey.xjtlu.edu.cn, powered by LimeSurvey).

To summarise above, the questionnaire should be designed as following:

- Create and collect via Survey @ XJTLU;
- Focus on the answers from clinical doctors and radiologists only;

- Answer the questions related to the distortion on medical images;
- Design the limited number of questions, and most of them are MCQs; thus can complete it within a short period;
- The respondents will fully understand where their data are used and they have the rights to withdraw their data after completed it.

The further details, the contents and the results of the questionnaire, please refer to the next chapter, Realisation.

Formal Assessment

In the formal assessment of Subjective IQA, the observers should judge the quality of the medical image in the dataset. For each medical image, the observers are required to view the medical images provided and then answer four questions relevant to the quality of the medical image, including:

1. Can you get useful information from this image?
2. Do you think there is at least a lesion in this image?
3. To what extent do you think the quality of this image is good enough for you to get the above answer?
4. To what extent do you think the quality of this image is similar to the images you actually encountered during your work?

Questions 1 and 2 are Yes/No type questions. Questions 3 and 4 are scoring type questions based on a 5-point Likert scale. On each page, the observers are able to observe one specific medical image in detail and then answer four questions aforementioned attached. They are allowed to click "Next Page" and view the next medical image after they have made decisions for the current image. Before the start, all observers have been told the terms of the Subjective IQA and then are required to sign the consent forms. They also need to answer several questions regarding the preferences of the project.

We will visit approximately ten doctors and expect three participants involve in Subjective IQA, in which two clinicians and one radiologist. We take the results of one clinician and one radiologist into consideration in Subjective IQA as well as a result from another clinician will be used in testing and evaluation. Considering the potential adverse effects brought by influential factors, we took the following actions: 1) to choose a sunny day as the assessment day and ask several questions related to personal emotions (prevent lousy mood caused by the weather or other reasons); 2) to set all participants to have the assessment in the same day (ensure that most environmental factors are consistent); 3) to adopt the places and computer displays which the participants usually use for daily work, in the assessment.

To illustrate the third action in the above further, this was designed to acquire the results as much similar as when the clinicians and radiologists work in a normal situation, instead of holding an assessment in a traditional way, such as using the same facilities and under the same hall. Specifically in this project, most of the time, clinicians are in the clinic rooms or departmental office when they are observing the medical images shown on the computer displays; however, radiologists are usually using the more high-quality and tailored displays when they process the medical images in office. Our Subjective IQA conducted in such a way. The approaches related to decreasing the negative effects of influential factors in Subjective IQA followed the suggestions provided by International Telecommunication Union [42].

3.2.3 Objective Image Quality Assessment

For Objective IQA, Image Processing Toolbox in MATLAB [43] will be used to result in the scores for the aforementioned mathematical models. There are seven sections in the toolbox. Correctly, we will use functions in the four parts, including "Import, Export, and Conversion", "Display and Exploration", "Image Filtering and Enhancement" and especially "Image Segmentation and Analysis". For each mathematical model, three steps will be processed and demonstrated in a step-by-step basis below:

- Image Quality Score

  All medical images in FYPDataset have been scored by the corresponding mathematical model, i.e., BRISQUE and NIQE, in MATLAB. In all No-Reference IQA, the lower score stands for the better quality. The score can range from 0 (the best quality) to 100 (the least quality).

- Initial Check

  For the scores produced by the model aforementioned, an initial check is compulsory. The aim of the initial check is to make sure the results are logical. In the initial check, the quality of the original image should be better than distorted images. One of the potential reasonable circumstances is the score of the original image is lower than the score of the distorted image for the same original image set. Outliers should be removed carefully. As demonstrated earlier, there are 50 medical images in FYPDataset, including 20 original medical images, which means we need to conduct the initial checks for all 20 groups.

- Confirm the Ultimate Result of Objective IQA

  We compare the results among all mathematical models used and then choose the best performance model. The corresponding assessment scores are regarded as the Objective IQA results, which will be used in the final step, Comparative Analysis.

3.2.4 Comparative Analysis

The final step of the process of the project is to analyse the experiment data received before comparatively. Thus this section summarises the significant and minor findings in the whole project. We have compared the results from Subjective IQA (including Pre-Assessment Questionnaire) and Objective IQA. According to the properties of the analysis, the results will be divided into three subsections, which are:

Table 3.1: An Example of Subjective and Objective IQA Matching

-	**Scores**	
	Subjective IQA	**Objective IQA**
Original	53	5
Blurry	52	4
Contrast	46	3

Table 3.2: An Example of Subjective and Objective IQA Mismatching

-	**Scores**	
	Subjective IQA	**Objective IQA**
Original	46	5
Blurry	52	4
Contrast	53	3

- Main Results

  In this section, the findings relevant to the purposes and objectives of the project mentioned in the beginning of the dissertation, i.e. in Section 2.5.1, will be demonstrated.

- Clinical Results

  In this section, the finding relevant to the clinical area (e.g. clinicians and radiologists) will be demonstrated.

- Computational Results

  In this section, the findings relevant to the mathematical models and computing discipline will be demonstrated.

Both Clinical Results and Computational Results are extremely important in the project, in addition to the results to answer the questions raised in research purposes.

We will analyse the results from Subjective IQA and Objective IQA in a group basis, which means the similar medical images will be analysed together. A group of similar medical images include the original "high-quality" image and other distorted images processed by this original one. We will consider the results of Subjective IQA as references. The two IQA will be considered as same/similar when: for the same medical image group, for example, the Subjective IQA results for images quality are "original image" > "blurry image" > "contrast-distorted image"; and Objective IQA also can return the same relationship between images. An example of Subjective and Objective IQA matching can refer to Table 3.1, and a mismatching example can refer to Table 3.2. Please note that the higher scores in Subjective IQA stand for the higher quality of images, and the lower scores in Objective IQA stand for the higher quality of images.

3.3 Revisions

Compared to the original plans in the Specification and Design Report, we had amended several places before the formal conduct of realising the project, as below:

- Source changed from the first-hand (hospital) to secondary-hand (online), and this causes the ethical application has been changed from LRR to NRR (details relevant to ethical application see Section 7.3);
- Added Pre-Assessment Questionnaire component in Subjective IQA;
- Prepared an extra mathematical model for IQA, i.e. PIQE;
- Analysed the results from the main purposes, clinical and computational perspectives.

The first amendment to the project is necessary according to the current situation of the outbreak of COVID-19. Additionally, all left amendments to the project are aiming for the analysis of the further perspective, the alternatives and the in-depth results.

Chapter 4

Realisation

The objective of this chapter is to demonstrate how the ideas of design introduced in the previous chapter have been implemented successfully on a step-by-step basis in the project. When you have read this chapter, you will:

- have been introduced the flow of the whole realisation procedure;
- have been demonstrated the process of the four-step methodology, including Dataset Construction, Subjective and Objective Image Quality Assessments, and Comparative Analysis;
- have been demonstrated the testing samples and approaches to make sure the previous results have not affected by the outliers, errors and coincidences;
- have been introduced the encountered problems and provided solutions during the whole realisation process;
- have been introduced the assumptions towards to results.

4.1 Overview

As demonstrated in the previous chapter, in the project, four steps include Dataset Construction, Subjective IQA, Objective IQA, and Comparative Analysis will be conducted in a logical order. A full and detail overview of the flowchart for the project, please refer to Figure 4.1.

4.2 Methodology

4.2.1 Dataset Construction

According to the design ideas in Section 3.2.1, we construct the dataset of the project firstly. A repository on GitHub with 4.2K starred, named Medical Data for Machine Learning, covers a significant number of sources on current medical images [44]. Andrew L. Beam creates the repository ("beamandrew", his GitHub username), an Assistant Professor of Harvard School of Public Health, who has received 1,434 citations according to Google Scholar [45]. As a consequence, at the initial stage, the information from this repository is considered as reliable.

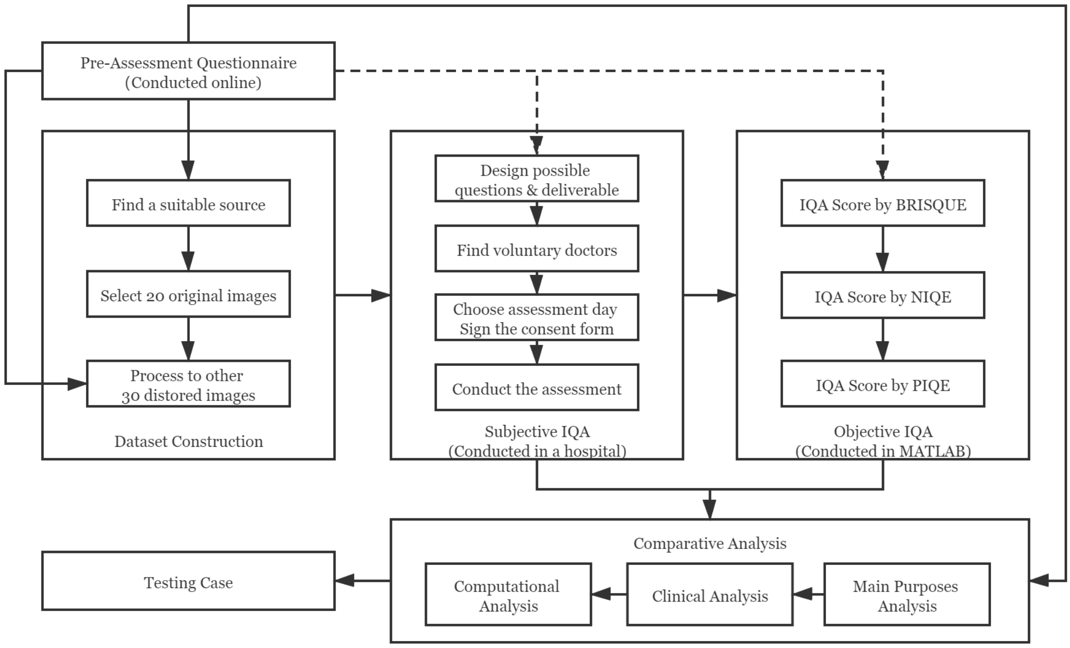


Figure 4.1: Full Overview of The Project

As mentioned before, the repository provides numerous sources of medical images. Indeed, it was a time-consuming process to analyse all sources and decide whether the images from the source match the characteristics of the assumed dataset we have discussed in Section 3.2.1. To find valuable medical images, we have retrieved each provided source, one-by-one. This process spent around a month. We concentrated on the datasets which:

- follow the prescribed characteristics of the dataset;
- are released by trustworthy organisations, such as some universities and the famous laboratories focusing on medical imaging technique.

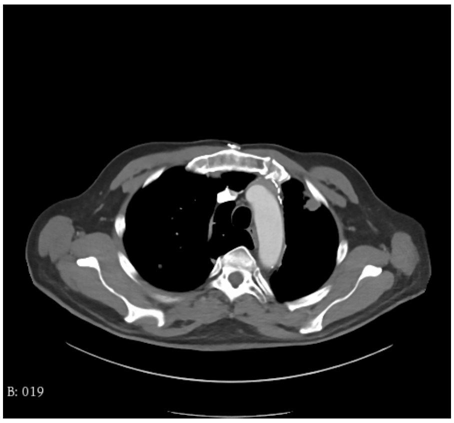
NIH, National Institutes of Health, has always been providing a large number of CT images publicly to assist the professional community in detecting the lesions of patients accurately [46], and these images can be downloaded from online by anyone via Box [47]. A current dataset released by NIH, named DeepLesion, covers 32,120 CT slices of 4,427 patients in total [48]. The last update was 26 April 2019, meaning this dataset possibly can be one of the latest medical images datasets around the world.

The new dataset we made for this project, named FYPDataset, is made up of 50 medical images in total. It was a significant problem that which of these 50 medical images should be gathered into FYPDataset. We took the statistics of the Pre-Assessment Questionnaire in Subjective Image Quality Assessment into consideration. The question 3 of medical imaging related group, "If you have ever encountered the distortion of medical images, what kind(s) of following possible distortion was it (were they)?" has received the following answers: 76.47% of respondents chose "Blurry", 43.14% chose "Contrast is too high/low", and 31.37%

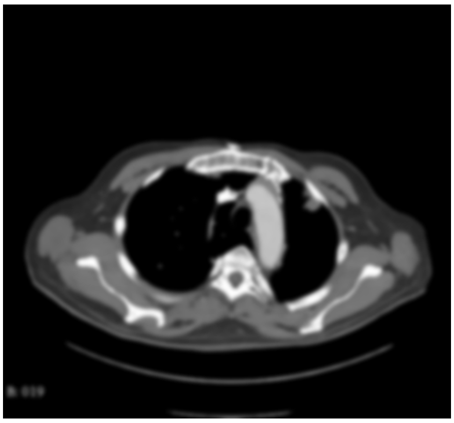
chose "Obscured by unknown reasons". Here we can take this as a frequency issue, in which the most common type of distortion is blurry, then is contrast-related and finally is obscured-related. Therefore, including the original image set, the blurred image set, the contrast problem image set, and the obscured problem image set, four kinds of medical images are gathered into FYPDataset. The number of each image set decreases along with the sequence aforementioned.

Eventually, FYPDataset contains four components, including 20 original medical images, 17 medical images as blur distorted, 9 medical images as contrast distorted, and 4 medical images as noise distorted. All 50 medical images in FYPDataset can be found in Appendix A.

A sample set of processed images can refer to Figure 4.2. Four medical images target to the same original medical image (a). Image (b) is blurry distorted with a Gaussian filter of the standard deviation of 2.25. Image (c) is contrast distorted with a grayscale [0.2, 0.4]. Image (d) is noise distorted with the type of 'Salt & Pepper' with 0.02 noise density. (a), (b) and (c) are three images in FYPDataset. Other distorted images with processed in FYPDataset are all similar to the images in this set.

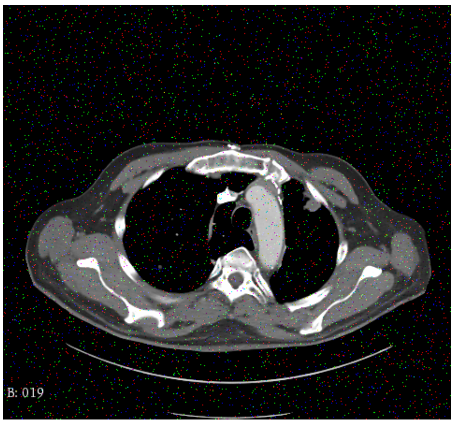


(a) Original



(b) Distortion: Blurry

(c) Distortion: Contrast



(d) Distortion: Noise

Figure 4.2: Sample Set of Processed Images

A real example of how the original medical images have been processed as "blurry" images via MATLAB is shown as below. The code segments for other types of distortion processing can refer to Appendix B.1.

```
% archive a medical image from local folder
medical_image_original = imread('FYP_Dataset\Processed\M61_143.png');

% process the image to blurred images
% by using Gaussian filters with different standard deviation
medical_image_blur_1 = imgaussfilt(medical_image_original,0.75);
medical_image_blur_2 = imgaussfilt(medical_image_original,1.5);
medical_image_blur_3 = imgaussfilt(medical_image_original,2.25);

% save the processed blurred images
imwrite(medical_image_blur_1, 'FYP_Dataset\Distortion\...blur075.png');
imwrite(medical_image_blur_2, 'FYP_Dataset\Distortion\...blur150.png');
imwrite(medical_image_blur_3, 'FYP_Dataset\Distortion\...blur225.png');
```

4.2.2 Subjective Image Quality Assessment

Pre-Assessment Questionnaire

We made a questionnaire consisting of four questions on personal background information and seven questions on medical imaging related. The full text of the questionnaire can be found in Appendix C. We referred to the book *Developing a Questionnaire* when we designed the questions and analysed the collected data [49]. Some options settings relevant to the medical area adopted the suggestions of an experienced clinical doctor—for example, question 3 of personal background information.

The full statistics of all questionnaire respondents can be found in Appendix C. Here come some critical results of full statistics. Totally, 59 questionnaires returned, and 51 returned questionnaires are eventually deemed as valid. In these 51 valid questionnaires, all respondents are currently working at a hospital or healthcare institution, in which 84% of them are clinicians, and 16% of them are radiologists. 90% of respondents regard their level of professional experience is equal to or higher than the average of all the same or similar professionals.

92% of respondents claim that they have ever encountered the problem of medical images distortion, and 98% of these respondents say this is not a frequent phenomenon, including 78% chooses the level of least frequency. The most common happened distortion for respondents is blurry, 76% determines. Some adverse effects can be brought due to the distortion, including "Missed diagnosis" (63%) and "Inaccurate judgements of disease severity" (63%). In the last, 86% of all respondents think the distortion on medical images is a severe problem for the medical community.

The respondents of radiologists are dramatically less than the respondents of clinicians, which is not beyond our expectation; as the number of radiologists is indeed considerably less than the number of clinicians, in most hospitals/healthcare institutions. Regarding the questionnaire results, we decided to pay more attention

Table 4.1: Subjective IQA Observers

Observer Index	Position	Experience (/yrs)	Time Cost (/mins)	Data Usage
001	Clinician	>13	31	Subjective
002	Clinician	3	40	Testing
003	Radiologist	>13	23	Subjective

to this distortion type, blurry. The results of the frequent types of distortion have helped us construct the FYPDataset, as demonstrated in Section 4.2.1. Additionally, their (the respondents') high evaluation of their level of experience illustrates that the credibility of their responses is relatively high.

The statistics have given us initial ideas on the distortion of medical images from the perspective of medical workers. We will analyse some of the statistics of the questionnaire in the section Comparative Analysis of this chapter further. Following these, the formal assessment then executes. The candidates for formal assessment have been selected from the respondents who left their contact information in the last question of the questionnaire.

Formal Assessment

This section demonstrates the official conduct of Subjective IQA, as most of the parts we discussed in Section 3.2.2 regarding Subjective IQA were actually pointing to this formal assessment (i.e. the current section), instead of the previous questionnaire.

As demonstrated in the step Dataset Construction, there are 50 medical images in the FYPDataset. The observers should give the decisions on qualities of all 50 medical images, based on their own experience. Additionally, we set a maximum time limit for the assessment, 50 minutes, which were designed in such a way as following the principles of Single Stimulus Methods. For each medical image, there are four questions relevant to the image quality. Totally there are 200 questions.

Three doctors eventually involved in the formal assessment of Subjective IQA, in which two of them are clinicians and one of them is a radiologist. In the beginning, the participants have been fully introduced the contents of Subjective IQA and what specific operations they would do during the assessment. On the condition that the participants did not have any further questions, they then signed the consent forms. The electronic copy of the consent form was also sent to the mailbox of each participant for back-up with the successful completion of the assessment. More importantly, all of them have been informed that they have the right to withdraw all their own experimental data before May 31, 2020, if they are unsatisfied with the research or have other personal reasons, which has been clearly stated on the consent form.

The full details of three observers in the formal assessment of Subjective IQA please refer to Table 4.1. The data collected for Subjective IQA can be referred to

Appendix D.1.

4.2.3 Objective Image Quality Assessment

In the course of the experiment, Objective IQA has always played the same important role. Based on the demonstration in Section 3.2.3, the whole process of Objective IQA conducted within the software MATLAB. The toolbox of MATLAB, Image Processing Toolbox, provides an extremely comprehensive and powerful set of mathematical models and algorithms for image processing, several subsections including but not limited to Geometric Transformation and Image Registration, Image Filtering and Enhancement, Image Segmentation and Analysis, and Deep Learning for Image Processing. We concentrated on the Image Quality part in subsection Image Segmentation and Analysis.

As illustrated before, for each mathematical model, three steps have been processed and will be demonstrated in a step-by-step basis, including "Result Image Quality Score", "Initial Check" and "Confirm the Ultimate Result of Objective IQA". All statistical data can be found in Appendix D.2.

BRISQUE

The mathematical model BRISQUE has been conducted firstly. In the paper [31], Mittal and others have proposed and provided a software release of BRISQUE in MATLAB. In addition to this, as mentioned before, we also use the functions (especially "brisque()") in Image Processing Toolbox provided by MATLAB.

The following code snippet is how we use BRISQUE algorithm to score the fifty medical images in MATLAB (file "BRISQUE.m", Appendix B.2.1):

```
% record the 50 medical images folder
image_folder = 'C:\Users\sunyu\Desktop\FYP_Dataset\Test\';
image_dir = dir([image_folder '*.png']);

% read all image files in this folder
% and score them by using BRISQUE algorithm
for i = 1:length(image_dir)
    image = imread([image_folder image_dir(i).name]);
    brisque_image = brisque(image);
    fprintf('%c',image_dir(i).name);
    fprintf('BRISQUE score for this image is: %0.4f.\n',brisque_image)
end
```

For the following step, the initial check can give us an idea of whether these results are desirable or not. In the initial check for BRISQUE, only the BRISQUE scores in 15 medical images of 6 groups are within our expectations, i.e., 30%. In this circumstance, the IQA scores mostly range from 40 to 55, thus there is no much difference even the quality of one image is lower or higher. However, on another hand, the quality of most original images is regarded as being lower than the quality of corresponding distorted images, according to the IQA scores generated by BRISQUE. Obviously, it is not logical.

Taking a group of results as an example, the original image named as "F51_030", two other images are distorted images. The types of the distortion respectively are "contrast-issue with a grayscale range [0.1 0.4]", and "blurry with Gaussian filter with a standard deviation of 0.5". In this case, it sounds like the original image will receive the lowest IQA score. However, the scores were 50.6976 (original), 49.9166 (contrast-distorted), 50.1219 (blurry). Three scores are very close but it still can be recognised that the contrast-distorted image is the best quality from the model's perspective. This happens frequently in the results of BRISQUE and it is wrong.

Less than one-third of the results qualified reveals a significant flaw that happened by practising BRISQUE in medical imaging discipline. Therefore, practising other models is extremely necessary.

NIQE

Following with BRISQUE, the mathematical model NIQE has then been conducted. Similar to BRISQUE above, Mittal and others have proposed and provided a software release of NIQE in November 2012 [32]. By using the functions in Image Processing Toolbox provided by MATLAB, specifically "niqe()", the experiment has done smoothly.

The following code snippet is how we use NIQE algorithm to score the fifty medical images in MATLAB: (file "NIQE.m", Appendix B.2.2):

```
% record the 50 medical images folder
image_folder = 'C:\Users\sunyu\Desktop\FYP_Dataset\Test\';
image_dir = dir([image_folder '*.png']);

% read all image files in this folder
% and score them by using NIQE algorithm
for i = 1:length(image_dir)
    image = imread([image_folder image_dir(i).name]);
    niqe_image = niqe(image);
    fprintf('%c',image_dir(i).name);
    fprintf('NIQE score for this image is: %0.4f.\n',niqe_image)
end
```

Similar to BRISQUE above, the initial check can give us an idea of whether these results are desirable or not. In the initial check for NIQE, disappointingly, this time only the NIQE scores in 14 medical images of 6 groups are within our expectations, i.e. 28%, which is slightly lower than the results of BRISQUE. Approximately one-quarter of the results qualified reveals a significant flaw that happened by practising another mathematical model, NIQE, in medical imaging discipline.

Recall the initial check results of BRISQUE and NIQE, which were 30% and 28% respectively; both are profoundly different from what we originally assumed.

We decide to recheck the whole process of Objective IQA firstly before going to the next phase. Two steps include 1) carefully check the MATLAB files and debug the corresponding codes in case of the potential errors; 2) practice two mathematical models, BRISQUE and NIQE, by using 20 natural images and other 20 unused medical images (not in FYPDataset) separately.

As we read the documentation several times, with careful debugs, we did not find potential errors in the codes shown above. This proved a possible error caused by the type of images. By using 20 natural images and 20 other medical images in BRISQUE and NIQE respectively, we surprisingly found that the results of most natural images were logical (i.e., the quality of the original image should be better than the quality of distorted ones, and the IQA score of the original image should be lower than the scores of distorted ones); however, either BRISQUE or NIQE, the results of medical images were still similar to the results of the previous experiments. In summary, we can assume that BRISQUE and NIQE may not perform well, especially in medical imaging discipline, at least in our case, the medical images of X-ray CT.

Due to the reason that both the results of BRISQUE and NIQE were significantly lower than our expectations, another alternative mathematical model should be necessarily considered, which is named as PIQE, as we discussed in Sections 2.3.2 and 3.2.3.

PIQE

The mathematical model NIQE has eventually been conducted. This time we have used the function "piqe()" in the same toolbox in MATLAB as before.

The following code snippet is how we use PIQE algorithm to score the fifty medical images in MATLAB: (file "PIQE.m", Appendix B.2.3):

```
% record the 50 medical images folder
image_folder = 'C:\Users\sunyu\Desktop\FYP_Dataset\Test\';
image_dir = dir([image_folder '*.png']);

% read all image files in this folder
% and score them by using PIQE algorithm
for i = 1:length(image_dir)
    image = imread([image_folder image_dir(i).name]);
    piqe_image = piqe(image);
    fprintf('%c',image_dir(i).name);
    fprintf('PIQE score for this image is: %0.4f.\n',piqe_image)
end
```

With the scores generated by the alternative model PIQE, the initial check firstly has been processed. Surprisingly, 41 medical images in 16 groups out of 50 images are within our expectation, i.e. 82%. This result is dramatically different from the results of previous models. Compare to the results in BRISQUE (30%) and NIQE (28%); it reveals the success of practising the PIQE model in medical imaging discipline. Observe the results further; we can already see the apparent difference of scores between original images and distorted ones, which prove the relatively good efficiency of the PIQE model. The results of PIQE will be deemed as the ultimate experiment results of Objective IQA in the project.

4.2.4 Comparative Analysis

Following the design ideas for Comparative Analysis in Section 3.2.4, we have compared the data received from the previous step, including Subjective IQA and Objective IQA. The results will be demonstrated into three parts according to their disciplines, which are Main Results, Clinical Results, and Computational Results, as below:

Main Results

1. The experiment results of Subjective IQA and Objective IQA were similar, with 80% similarity.

2. The mathematical model, PIQE (Perception based Image Quality Evaluator) has performed well in our experiments, which was under the medical imaging context.

Clinical Results

3. Clinicians and radiologists could have the same or similar opinions on the quality of medical images, for most of time.

4. Both clinicians and radiologists were highly sensitive to the changes of contrast value of medical images in the experiments.

5. Within a certain range of standard deviation for Gaussian filter, the distortion of blurry in medical images possibly might not cause the negative effects to doctors.

6. The experiment results in Subjective IQA were similar to the results in Pre-Assessment Questionnaire.

Computational Results

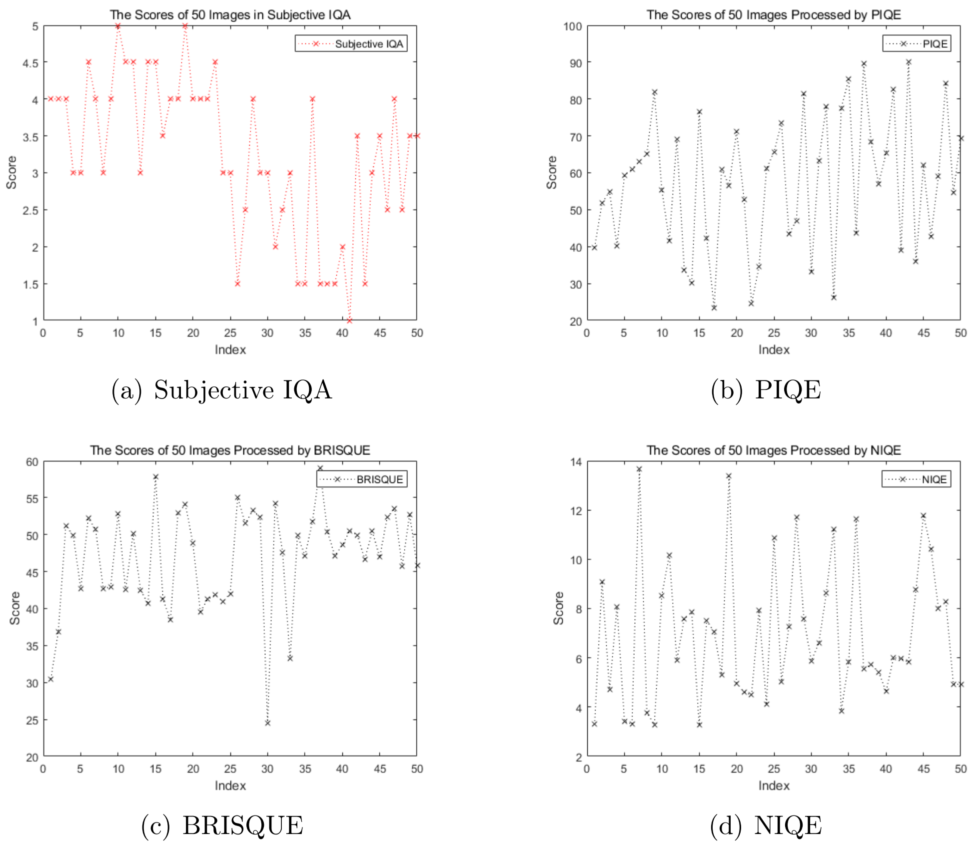
7. In medical imaging discipline, BRISQUE and NIQE models (no custom) did not have good performances when assessed the quality of X-ray Computed Tomography (CT) scans.

8. In the range (0.5, 1.0) for Gaussian filter with a standard deviation, the blurry distortion may not result in a negative effect to doctors.

9. In a certain range of grayscale for contrast values, the quality of the medical images may possibly improve to some extent. Specifically, the range for "low_in" in the grayscale = (0.1, 0.3) and the range for "high_in" in the grayscale = (0.6, 0.8).

In summary, all objectives of the project have been achieved successfully. From the results, it is clear that the mathematical model used in Objective IQA can predict the perceived image quality automatically as good as the results from human beings in Subjective IQA, which is within our expectations. A further novel finding is that two mathematical models in No-Reference IQA, BRISQUE and NIQE, cannot perform well in X-ray CT scans; however, these two models have always been regarded as two of best No-Reference IQA models within this decade.

The plots to the IQA scores of Subjective IQA and Objective IQA (PIQE, BRISQUE, and NIQE) can refer to Figure 4.3. The full scores of Subjective IQA please refer to Appendix D.1; the full scores of Objective IQA please refer to Appendix D.2.



(a) Subjective IQA

(b) PIQE

(c) BRISQUE

(d) NIQE

Figure 4.3: Score Plots of Subjective and Objective IQA

4.3 Testing

In order to investigate the results above have been produced without potential errors and coincidences, testing is necessary to conduct. The methodology can be maintained as same as before, with the reason that this has been evaluated by the supervisor and us in the Specification and Design Report [2]. However, the testing conduct with the new data samples which have never been used in the project.

For the testing sample of Subjective IQA and Objective IQA, we will first construct a new testing dataset of 20 medical images, and then practice IQA with the medical images in this dataset. In this new dataset, it consists of 8 original medical images, 7 medical images as blur distorted, 4 medical images as contrast distorted, and 1 medical image as noise distorted, which follows the proportion of distortion types in the FYPDataset. In Subjective IQA, we will adopt the result statistics of the second observers (index is 002), as shown in Table 4.1. In Objective IQA, the same three mathematical models will all be tested, especially PIQE.

We have analysed the results of Subjective IQA and Objective IQA comparatively, as to how we have done before. Unsurprisingly, the same outcomes covering the main results, clinical results, and computational results can be generated. Broadly speaking, the testing results highlight that there is no significant error, flaw, or coincidence in our previous experiments. Our results shown in the section Comparative Analysis are in a relatively fair way.

4.4 Problems and Solutions

During the process of project realisation, we indeed have encountered several questions, which some of them were extremely tricky to be solved. However, these problems were not obstacles; they pushed us to find more advanced and precise results for the project. The following is the list of the problems we have encountered, and the corresponding solutions we provided. Please note the severity of the problem decreases as the sequence decreases.

- Inadequacy of BRISQUE and NIQE

  As we demonstrated before, the Subjective IQA results from BRISQUE and NIQE were lower than our expectations. Although we have discussed the potential risks of adaptability of BRISQUE and NIQE in the Specification and Design Report [2], two mathematical models can well handle the natural pictures according to the previous literature review. At that time, we could not imagine that these two above models have such bad performances in medical imaging discipline. However, we still decided to try to find another alternative model for the project before the official conduct of the realisation process, in case of the worst situation. We paid more attention to the mathematical models on IQA, and thus we have done the research to another possible model, PIQE. The eventual result proved that our assumption was correct.

- Unavailability of intended first-hand source

  Another significant problem is that the intended source for medical images was unavailable anymore. Although the project is closely based on the knowledge of computer science, it is necessary that a suitable hospital and medical workers involved such as the construction of the project dataset. As the plan, we aimed to adopt a set of medical images from the first-tier hospital in Shenzhen, which made our project more intriguing and novel. However, this has turned to a severe problem after late January 2020 due to the outbreak of coronavirus COVID-19. In order to decrease the workload of the hospital, we immediately changed the plans for data construction, from the hospital to online sources.

This transformation increased our workload; however, we expect the outcomes remain as high-quality as before.

- Lack of observers in Subjective IQA

  Due to the same reason above, the outbreak of coronavirus, it was tricky for us to find several suitable doctors to join our project. On one hand, most of the doctors were working on the front-line to fight for coronavirus and they only had limited time available; for another, to ensure the judgements from doctors are high-quality, the experienced doctors with a higher position were preferred. However, many of them were busy with the treatments for patients with coronavirus. It sounds like a negative loop, but it did happen to our project. As Subjective IQA takes an essential role in IQA, we could not simply alter it to another form of quality assessment. We firstly made a Pre-Assessment Questionnaire to ask for some general ideas on the topic from the doctors' perspective. In the last question of the questionnaire, we proposed our action on Subjective IQA and encouraged the doctors with interests to provide us their contact method. 9 doctors out of 51 respondents have provided their information, which was a desirable result for us. Limited to time, current location (some of them are not currently in Shenzhen), types and experience of doctors, we finally chose three of them, as the observers of Subjective IQA, as shown in Table 4.1.

4.5 Assumptions

In Section 4.2.4, the results of comparative analysis have been demonstrated. The second result, also one of the main findings is, the mathematical model PIQE could perform well in our case. Two famous models in IQA area, BRISQUE and NIQE, both showed terrible performances. However, it is a thoughtful question to figure out what reasons caused in this phenomenon. For this question, we have supposed the following assumptions:

- Calculation of Blocks

  Calculation of blocks possibly increases the efficiency of mathematical models overall. PIQE divides the whole image to blocks, and then extracts local features of each block for predicting quality, and to mimic human behaviour; it estimates quality only from perceptually significant spatial regions. BRISQUE tends to perceive the quality based on the whole image, while although NIQE divides the image into patches, it also tends to the entire image from a macro perspective. Calculation of blocks like PIQE possibly can increase the accuracy to the judgement to the quality of images to some extent.

- No Human Opinions

  The current human opinions and intervention under the context of X-ray CT scans (the type of medical imaging technique in our case) possibly are not highly beneficial to train the mathematical model for the IQA. BRISQUE and NIQE require knowledge about anticipated distortions in the form of training examples and corresponding human opinions. Specifically, in BRISQUE, a set human opinion collection will be adopted to train the model, and in PIQE a natural MVG model which contains some existed knowledge will be used in the final step. However, PIQE does not need human intervention to score the images. Of course, it is crucial to involve Subjective IQA in IQA. We also believe a better form of human opinions and intervention can improve the efficiency of IQA mathematical models. How to fit human views and interference into the model with a more suitable way is what we chase for in future work, which can be more complicated than we thought.

- The Particularity of Medical Images

  Medical images are extremely different from natural images (such as scenery images). Under the context of X-ray CT scans in our case, the colours only can be white, black and the colours between the two. However, most of the natural images in this era are colourful, instead of in a grayscale. Additionally, most of the time, the contents in CT scans are also not possibly appeared in natural images. The particularity of medical images made BRISQUE and NIQE could not handle them well, as they have learnt the image quality knowledge from natural images mostly.

The above assumptions are targeting to why PIQE could perform dramatically better than BRISQUE and NIQE under the context of X-ray CT scans. At this stage, we cannot surely confirm our assumptions are exactly precise. However, we can confirm we are fully aware of the objectives and aims for future research on this topic. We also have the determinations to do in-depth research relevant to IQA under medical imaging context in the future. The upcoming study will be planned this summer, focusing on proving the assumptions raised earlier.

Chapter 5

Evaluation

The objective of this chapter is to demonstrate how the project has been evaluated and in what kinds of approaches. When you have read this chapter, you will:

- have been introduced and demonstrated the evaluation approaches used for the project, including the specific criteria and evaluation results;
- have been listed the strength and weakness of the project;
- have been introduced the topics of future works suggested by the project team.

5.1 Evaluation Approaches

The process of evaluation is mandatory in scientific research, and it also provides the last chance to analyse each component and find possible errors in the project. In the evaluation of this project, we adopted some guidance and criteria raised by Posavac in the book *Program evaluation: Methods and case studies* [50]. In the process, we evaluate firstly from the start point of the project and go increasingly deeper. One point that should be illustrated here is, the evaluation has been processing since the start of the project, i.e. September 2019. We would like to fastly amend the unsuitable places through continuous evaluation instead of evaluating everything till the end of the project, which potentially causes a project catastrophe and it will be impossible to correct mistakes.

The process of evaluation consists of five steps: evaluation to objectives, evaluation to Subjective IQA, evaluation to Objective IQA, evaluation to results, and integrated evaluation. The details of the evaluation are as below:

- Evaluation to Objectives: Assess, and examine the process

  The objectives of the project have been always monitored by the supervisor of the project. Additionally, we also discussed the feasibility of the objectives in the Specification and Design Report. Before the official conduct, we made a Pre-Assessment Questionnaire. The reason to do so is we would like to hear the voices from the doctor's perspective, as this a topic close to the medical community. With all of the above, we believe our objectives are achievable and always on the way. The more we understand the research aims and purposes, the less we will deviate them during the research, and it is exactly what we did.

- Evaluation to Subjective IQA: Professional feedback

  For Subjective IQA, we documented the detailed information of the assessment, including the specific experiment date, the time, the weather on that day, the doctor observers, the signed consent form, the moods check of observers, the time cost of the experiment, and the comments from the observers. To conduct more same assessments may sound time-consuming and impossible to do at this extraordinary period; however, we checked this information several times to ensure the influential factors did not affect or with little to the results of Subjective IQA. The fact proved that it was wise to record all relevant information even those information looked non-relevant at that time. We can check it whenever we are suspicious of a particular place.

- Evaluation to Objective IQA: Test the efficiency of all mathematical models

  For Objective IQA, the efficiency of the selected mathematical models should be well considered. In case of possible errors and outliers of previous results, other images should also be processed by the chosen models. We used natural images and other medical images as new datasets. As a result, we found BRISQUE and NIQE still could not perform well in the medical imaging CT area but could perform well in natural images. As illustrated in the Specification and Design Report, we found out an efficient algorithm prior to the official conduct of Objective IQA. Thus the poor performances of BRISQUE and NIQE did not cause an extremely negative effect on our project.

- Evaluation to Results: Measure the outcomes and impacts

  The results can be a focus in a project, also in evaluation. Recall the results in Section, three types of results include the main results, clinical results, and computational results. Thus, the evaluation for this part can be divided into three parts, accordingly. Based on a different level of complexity, the main results as the elementary level can be first to be evaluated. In the main results, based on previous literature, we can say the results were within the expectation. For the clinical results, we provided our results to all involved doctors and other doctors who have never involved in the project. Our results are highly agreed by the doctors, and we consider this is a positive result for the evaluation. The evaluation of the computational results can be the most challenging one, as there is a limited reference for us to compare with. Even though those numerical statistics are hard to be evaluated, we applied the data to the specific values into the medical images and then showed them to the doctors again. This time we also received the agreement from doctors. Overall, all results can be deemed as trustworthy and reliable.

- Integrated Evaluation: Objectives, Realisation, and Results

  Efficiency is also an important factor to evaluate a project, or we say cost. A good project should eventually also be evaluated in an integrated way, including the whole process of the research. The cost in the research includes the resource of data, the time cost, the participant involvement, and the financial cost. We chose to adopt the online data, which means our resource of data is at a low cost. However, we believe the level of cost is relatively high in the time cost, the participant involvement, and the financial cost. There are two reasons resulting in this circumstance. The first one is the characteristics of Subjective IQA, which are time-consuming and finance-consuming. However, indeed, it is necessary. The second reason is we have encountered some difficult issues in Objective IQA when conducting the mathematical model of BRISQUE and NIQE. We were wondering why the outcomes generated from these two models were not within our expectations and it did waste some time to process the PIQE. Overall, with the final desirable outcomes, the efficiency of the whole project can be deemed as a level of "Middle".

5.2 Strength and Weakness

5.2.1 Strength

FYPDataset, the core data of the project, is the strength of our research. From the beginning of the project, we put a lot of attention to the construction of the dataset. We believe a project-tailored and suitable dataset is a key to scientific research. We are fully aware of the transformation from the first-hand source to secondary-hand source, but it does not mean our expected outcomes will not be achieved. As demonstrated earlier, we set detailed rules to select the source and ultimately chose a dataset from NIH, a trustworthy organisation in the United States. As a matter of fact, the quality of our results improved by this transformation.

5.2.2 Weakness

One weakness is the limited number of observers in Subjective IQA. For general subjective assessments, the number of respondents has been always emphasised. A truth is: the results are increasingly close to the facts along with the number of respondents increases. The results averaged by copious respondents will weaken the effects of artificial mistakes and outliers potentially, which means the results can be more trustworthy and reliable. This can be demonstrated vividly by the performances of the Subjective IQA.

Only three doctors have involved in the formal assessment of Subjective IQA, including two clinicians and one radiologist. Among their responses, for one same medical image, three doctors marked the quality of the image as "1 - Worst", "3 - Middle", and "5 - Best" respectively. We could not simply take the average "3" as the final result; instead, by analysing their judgements on the lesion for the same image, we removed the outlier given by one doctor. Indeed, three doctors are not enough to result in an absolutely correct answer. However, due to the characteristics of Subjective IQA which are time-consuming and finance-consuming, and the current situation, this is the best result what we can do. It is worth mentioning that, similar to the above sample, we prepared sufficient actions during the project (such as ask the questions relevant to lesions) to still ensure the efficiency of the results from these three doctors in case of possible outliers and errors.

Different from the formal assessment of Subjective IQA, the results of the Pre-Assessment Questionnaire are more reliable. The final results consist of the responses from 51 doctors. This questionnaire is relatively easier to finish, as the doctors can fill in within five minutes wherever and whenever on the condition they have one mobile phone or computer. The questionnaire involved some basic opinions regarding medical imaging distortion. Even though the process of answering was without supervision and intervention by our team, the results still can be deemed as desirable. Even so, we did not change the original plans of the formal assessment, which will be conducted on a face-to-face basis and in the hospital. We believe this is most similar to the real circumstance that happened in working.

5.3 Future Works

5.3.1 Natural Distortion

The FYPDataset in the project consists of original medical images and the corresponding distorted images. Please note these distorted images are all artificial made by MATLAB, in this case, as demonstrated in the section of methodology, Dataset Construction. This approach of adding common distortion artificially is adopted by most of the previous relevant literature. As a matter of fact, the forms of image data are various. Meanwhile, the multiple layers of distortion possible can occur during the process of generation, transmission, and handling, which is impossible to simulate all by just a computer and certain software [51]. Therefore, the distorted image simply processed by MATLAB cannot represent the scenarios that happened in real applications perfectly. Due to the restriction of the source dataset, the extensibility of the mathematical models has been restricted in the same way [51]. Consequently, it is significantly essential to focus on the research regarding natural distortions for IQA, in the next phase.

5.3.2 Specific Scenes

In this project, the results reveal that recognised mathematical models for IQA may "crash" under the specific context. It also means further experimental tests and studies are needed to find these specific situations, so that to improve the performances of existed mathematical models. We are fully aware that it is extremely tough to design a mathematical model of IQA for research usage; however, we are confident that our research will serve as a base for future studies on the image distortion problem under different scenes. For example, the image degradation caused by tone-mapping of HDR pictures [52]. Future work should concentrate on enhancing the ability to handle the image distortion issues under different situations, as well as developing a more advanced mathematical model for No-Reference IQA for the full-aspects situations, which is also a vital issue for future research on IQA.

Chapter 6

Learning Points

The objective of this chapter is to demonstrate what learning points we have achieved from this project.

In ten months since September 2019, our project Predict the Impact of Visual Distortion on Medical Images, successfully complete, and has caused the positive results. In addition to the results in topic discipline, we also improved the individual comprehensive abilities academically. Specifically, the learning outcomes are as below:

- Specify a research problem and plan it

  Well begun is half done. The most tricky thing to do for research is to establish a clear and substantial research problem, which will lead the researchers along the whole process. For the project topic, medical imaging, we had never involved before. We chose it out of interest, to a large extent. In the first two months, we were sometimes confused about specific subsections of the topic due to the in-depth profession of IQA and medical disciplines. However, thanks to the project supervisor Dr Mogos, who always discussed the subject with us and made us have an increasingly better understanding of the topic, even she is not specialising in this area. She was always patient in answering some question raised by our team, yet sometimes those question seems simple (but we were confused). Prof Man and Dr Lévêque also raised useful suggestions on improvement and helped a lot in the initial stage of the project.

  Personally, my efforts in this project are vital to complete another half of the work. Out of interest, I read relevant references to have a better understanding of IQA and medical imaging. Therefore, I was fully aware of what I should do and what ultimate outcomes I expect. As demonstrated in the Specification and Design Report [2], I proposed a detailed plan to conduct the project. I realised a good proposal is essential as you may refer to this report later on to prevent the deviation of the future work. The fact also proved my opinion.

- Time-management

  For a ten months project, proper time management is necessary. We had created a Gantt Chart for the project in the Specification and Design Report [2]. However, the outbreak of COVID-19 made us reconsider a new Gantt Chart. The relationship between each process let us realise the logical order to conduct. It is also significant to master the ability to process the emergency during the project.

- Write a high-quality literature review

  To successfully conduct the research, we did sufficient on the previous work of the topic. We read relevant papers on IQA and medical imaging, especially the database of IEEE. It was a dreary process to read a lot of references; however, we decided first to read the paper on a summary of the topic which could give us a brief idea of the whole structure. As the property of the project, which is concentrating on research, we paid more attention to the background introduction and literature review. With the completion of the literature review written, we can locate the needed source precisely and summarise them as our contents.

- Design a methodology for the project

  In the project methodology, we have adopted a four-step approach. This is a novel methodology and we created this according to the previous papers. It is a good example for us to learn how will we make a suitable methodology for a project in the future, as a precise methodology will possibly lead researchers to success; however, a wrong methodology will definitely result in failure.

- Implement and Evaluate the project

  To implement a project accurately is also what we have learnt. In our project, we have to apply four steps with a logical order, which means we need to schedule our timetable well and implement each step successfully. We once met the problems in Objective IQA component, also illustrated above. However, we found the alternative so that the obstacle in Objective IQA did not bother the conduct to other steps. We also thoroughly evaluated the project, including the pros and cons regarding it. The future work in evaluation also pushes us to explore in-depth knowledge in medical imaging technique in the future.

- Structure and write a dissertation

  The most significant learning point should be structuring and writing a dissertation for the project. Combing with the structure of the provided dissertation, we modified some chapters on the necessary basis, according to our project. In order to format the dissertation properly, we typed all our work in LaTeX instead of writing in the Microsoft Word. We believe this will make our dissertation more beautiful, and also the bookmark tab on the left column gives the readers convenience to go to the chapter they would like to read. Additionally, the hyperlinks in the dissertation are all highlighted and the readers are easy to internal or external resources.

Chapter 7

Professional Issues

The objective of this chapter is to demonstrate the professional issues relevant to the project. When you have read this chapter, you will:

- know how the project is relevant to British Computer Society Code of Practice;
- know how the project is relevant to British Computer Society Code of Conduct;
- have been introduced the procedure for Ethical Assessment of the project.

7.1 Code of Practice

The project is obeying the British Computer Society Code of Practice during the whole research process [53]. According to the Practice, it ensures that the information system practitioners maintain the acceptable standards of competence in information system practice. The Practice provides the "Scope" of applicable people, "Responsibilities of Information Systems Practitioners", "Authorisation", and "Development and Maintenance" of the Practice.

The project observes the highest standard of the Practice in information systems. During the project, we are always passionate about improving the knowledge and skills of medical imaging and image quality assessment areas, which are closely relevant to the project.

For further illustration of relevant principles, please refer to the second subsection of the current chapter, Code of Conduct.

7.2 Code of Conduct

The project is obeying the British Computer Society Code of Conduct released on 5 September 2001 (Version 2.0) during the whole research process [54]. According to the Conduct, the contents can be summarised as four categories, including "The Public Interest", "Duty to Relevant Authority", "Duty to The Profession", and "Professional Competence and Integrity". We will discuss how the project obeys the Conduct based on four categories aforementioned, as below:

- The Public Interest

  As we are fully aware that the project involves the medical images, even though they are anonymous, we still process these medical images and corresponding information in a very careful way. All involved participants receive fair treatment without individual discrimination and have the rights to quit the project and withdraw all recorded data when they are not satisfied or due to other reasons. Most significantly, the project goes without any offer of bribery or inducement, and with following the legislation issued by the government of current location, i.e., the People's Republic of China.

- Duty to Relevant Authority

  As an essential part, the project has been approved by the relevant ethical committee of the university. Currently, the project has been monitoring by this committee, and can be suspended or stopped if any unexpected situations, especially those negative ones, occur in the future. For more details, please refer to Section 7.3 of the current chapter, Ethics.

  In addition to this, the project is also supervised by Dr. Gabriela Mogos and the Department of Computer Science and Software Engineering.

- Duty to The Profession

  We understand our words here or elsewhere may result in various consequences to others. Thus, we express our statements in a helpful, careful, and honest way.

- Professional Competence and Integrity

  As medical imaging technique is one of interdisciplinary directions combing with computer science, we are keen on upgrading the professional knowledge and skills in this specific area all the time. For the background research on this topic, it is still ongoing since the first day we chose the topic. We encountered tough problems sometimes, and we would not publish a vague solution before we did enough amount of relevant research. However, as meanwhile, we try our best to control most of the components in the project are processing within the level of competence that we possess. Finally, we accept the responsibility for the whole project and the potential effects of it.

7.3 Ethics

According to the Policy on Ethical Conduct in Research of XJTLU [55], it is mandatory for each research to process the ethical assessment and maintain "the highest standards of ethical conduct". In order to supervise the involved research, relevant responsible also need to submit ethical applications and get the approval from the University Ethical Committee (UEC, apply via ethics@xjtlu.edu.cn). The research can be eventually assessed as the following three kinds under different circumstances:

- Negligible Risk Research (NRR)
- Low Risk Research (LRR)

- Standard Risk Research (SRR)

This project is eventually assessed as a NRR, and received the official approval from the UEC on 24th March 2020 by email correspondence. We informed the full information relevant to the project when made the NRR application to the Committee, without misrepresenting or withholding the information. The project will be deemed as a NRR when two essential components and at least one optional component of NRR meet. In this project, it meets totally five characteristics of NRR:

- Essential components:
  1. All involved participants are non-vulnerable groups;
  2. All involved participants can provide consent forms;
- Optional components:
  1. All involved information can be accessed publicly;
  2. All medical images are the secondary use of anonymous information, and cannot result identifiable information;
  3. The collection of sensitive, private or health data is not involved in quality assessment project, i.e., in this project.

In the Specification and Design Report submitted in December 2019 [2], the project was assessed as an LRR. We would prefer to use first-hand medical images that could be provided from a hospital in Shenzhen China, as the fresh resources made the project unique. Additionally, other than the medical images, both source hospital and corresponding patients need to sign the consent form. Under these conditions, the project was correctly deemed as an LRR. However, the outbreak of COVID-19 since January 2020 around the world made the situation more serious. Almost all hospitals and healthcare institutions were concentrating on the increasing number of patients with coronavirus. We sincerely appreciate the devotion from them. To decrease their workloads, we decided to use secondary public resources. Under the updated conditions, the project is as an LRR currently. However, it does not mean that our project will not result in the same expected outcome. We paid more attention to selecting the suitable medical images, as demonstrated before.

Chapter 8

Conclusion

Our work has led us to the conclusion that Subjective and Objective Image Quality Assessment (IQA) can result in similar results at this stage. The evidence from the project also suggests Perception based Image Quality Evaluator (PIQE) is a relatively good mathematical model of No-Reference Image Quality Assessment (NR-IQA). We have found an innovative result which both Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) and Naturalness Image Quality Evaluator (NIQE) could not perform well under the context of medical imaging, compared to PIQE, especially X-ray CT scans in our case; however, BRISQUE and NIQE are considered as two of best NR-IQA models. For this result, we have supposed several assumptions, and further work will investigate the in-depth reasons to cause this phenomenon. The results of the project also indicate some minor findings in clinical and computational disciplines. Even though there are limitations due to the limited number of observers in Subjective IQA, we adopt the suitable approaches to ensure the removal of coincidences, errors and outliers. This project provides a backbone towards enhancing the future equipment of medical imaging techniques on the problem of distortion, which could conceivably lead to the significant improvement of the medical imaging area. We hope that further tests and studies will confirm our findings and work on more advanced mathematical models for NR-IQA under medical imaging context.

Bibliography

[1] L. Lévêque, H. Liu, S. Baraković, J. B. Husić, M. Martini, M. Outtas, L. Zhang, A. Kumcu, L. Platisa, R. Rodrigues *et al.*, "On the subjective assessment of the perceived quality of medical images and videos," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2018, pp. 1–6.

[2] Y. Sun, "Predict the impact of visual distortion on medical images," Xi'an Jiaotong-Liverpool University, Suzhou China, Final Year Project (CSE305): Specification and Design Report 1611049, 2019.

[3] P. Suetens, *Fundamentals of medical imaging.* Cambridge university press, 2017.

[4] "Lesion...what does the doctor mean?" https://www.medicinenet.com/script/main/art.asp?articlekey=9695, July 2004, (Accessed on 05/08/2020).

[5] G. T. Herman, *Fundamentals of computerized tomography: image reconstruction from projections.* Springer Science & Business Media, 2009.

[6] J. Beutel, H. L. Kundel, and R. L. Van Metter, *Handbook of medical imaging.* Spie Press, 2000, vol. 1.

[7] K. K. Shung, M. Smith, and B. M. Tsui, *Principles of medical imaging.* Academic Press, 2012.

[8] E. A. Krupinski, "Current perspectives in medical image perception," *Attention, Perception, & Psychophysics*, vol. 72, no. 5, pp. 1205–1217, 2010.

[9] G. Schaefer, A. Hassanien, and J. Jiang, *Computational intelligence in medical imaging: techniques and applications.* CRC press, 2009.

[10] K. W. Lai and D. E. O. Dewi, *Medical Imaging Technology: Reviews and Computational Applications.* Springer, 2015.

[11] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," *arXiv preprint arXiv:1406.7799*, 2014.

[12] B. Series, "Methodology for the subjective assessment of the quality of television pictures," *Recommendation ITU-R BT*, pp. 500–13, 2012.

[13] U. Reiter, K. Brunnstrom, and K. De Moor, "Quality of experience: advanced concepts applications and methods," *T-Lab Series in Telecommunication Services*, pp. 55–72, 2014.

[14] P. Le Callet, S. Möller, A. Perkis *et al.*, "Qualinet white paper on definitions of quality of experience," *European network on quality of experience in multimedia systems and services (COST Action IC 1003)*, vol. 3, no. 2012, 2012.

[15] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, no. 1, pp. 1–156, 2006.

[16] V. Azad and P. Sharma, "A review on objective image quality assessment techniques," *International Journal of Emerging Engineering Research and Technology*, vol. 2, no. 5, pp. 188–192, 2014.

[17] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 2366–2369.

[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[19] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.

[20] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on image processing*, vol. 15, no. 2, pp. 430–444, 2006.

[21] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of electronic imaging*, vol. 19, no. 1, p. 011006, 2010.

[22] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[23] Z. Wang and A. C. Bovik, "Reduced-and no-reference image quality assessment," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 29–40, 2011.

[24] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, "Objective video quality assessment system based on human perception," in *Human vision, visual processing, and digital display IV*, vol. 1913. International Society for Optics and Photonics, 1993, pp. 15–26.

[25] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE transactions on image processing*, vol. 21, no. 8, pp. 3378–3389, 2012.

[26] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE transactions on image processing*, vol. 15, no. 6, pp. 1680–1689, 2006.

[27] S. Wolf and M. H. Pinson, "Spatial-temporal distortion metric for in-service quality monitoring of any digital video system," in *Multimedia Systems and Applications II*, vol. 3845. International Society for Optics and Photonics, 1999, pp. 266–277.

[28] K. Chono, Y.-C. Lin, D. Varodayan, Y. Miyamoto, and B. Girod, "Reduced-reference image quality assessment using distributed source coding," in *2008 IEEE International Conference on Multimedia and Expo*. IEEE, 2008, pp. 609–612.

[29] M. Carnec, P. Le Callet, and D. Barba, "Visual features for image quality assessment with reduced reference," in *IEEE International Conference on Image Processing 2005*, vol. 1. IEEE, 2005, pp. I–421.

[30] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3129–3138, 2012.

[31] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[32] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[33] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *2015 Twenty First National Conference on Communications (NCC)*. IEEE, 2015, pp. 1–6.

[34] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: Jpeg2000," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1918–1927, 2005.

[35] L. Liang, S. Wang, J. Chen, S. Ma, D. Zhao, and W. Gao, "No-reference perceptual image quality metric using gradient profiles for jpeg2000," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 502–516, 2010.

[36] T. Brandão and M. P. Queluz, "No-reference image quality assessment based on dct domain statistics," *Signal processing*, vol. 88, no. 4, pp. 822–833, 2008.

[37] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Proceedings. International Conference on Image Processing*, vol. 1. IEEE, 2002, pp. I–I.

[38] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb)," *IEEE transactions on image processing*, vol. 18, no. 4, pp. 717–728, 2009.

[39] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2 (2005)," *URL http://live. ece. utexas. edu/research/quality*, 2005.

[40] L. Dalla Palma, F. Stacul, S. Meduri, and J. T. Geitung, "Relationships between radiologists and clinicians: results from three surveys," *Clinical radiology*, vol. 55, no. 8, pp. 602–605, 2000.

[41] J. M. Bosmans, L. Peremans, A. M. De Schepper, P. O. Duyck, and P. M. Parizel, "How do referring clinicians want radiologists to report? suggestions from the cover survey," *Insights into imaging*, vol. 2, no. 5, pp. 577–584, 2011.

[42] R. I.-R. BT, "Specifications and alignment procedures for setting of brightness and contrast of displays."

[43] *Image Processing Toolbox^{TM} Reference.* MathWorks, 2019.

[44] beamandrew, "beamandrew/medical-data," https://github.com/beamandrew/medical-data, Feb 2020, (Accessed on 05/12/2020).

[45] "Andrew l. beam - google scholar citations," https://scholar.google.com/citations?hl=en&user=SgHOsrsAAAAJ, (Accessed on 05/12/2020).

[46] "Nih clinical center releases dataset of 32,000 ct images | national institutes of health (nih)," https://www.nih.gov/news-events/news-releases/nih-clinical-center-releases-dataset-32000-ct-images, July 2018, (Accessed on 05/11/2020).

[47] "Deeplesion - box," https://nihcc.app.box.com/v/DeepLesion, May 2019, (Accessed on 05/11/2020).

[48] K. Yan, X. Wang, L. Lu, and R. M. Summers, "Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *Journal of medical imaging*, vol. 5, no. 3, p. 036501, 2018.

[49] B. Gillham, *Developing a questionnaire.* A&C Black, 2008.

[50] E. J. Posavac, *Program evaluation: Methods and case studies.* Routledge, 2015.

[51] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372–387, 2015.

[52] D. Kundu, D. Ghadiyaram, A. C. Bovik, and B. L. Evans, "Large-scale crowdsourced study for tone-mapped hdr pictures," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4725–4740, 2017.

[53] *British Computer Society Code of Practice*, January 1996.

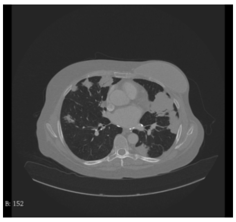
[54] *British Computer Society Code of Conduct*, September 2001.

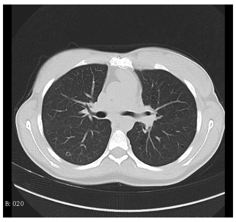
[55] XJTLU, "Final year project regulations: Procedures for the ethical assessment of undergraduate final year projects involving human research," February 2016.

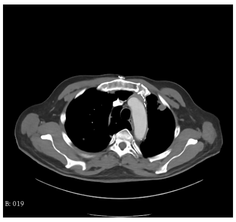
Appendix A

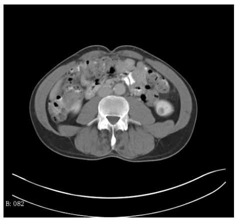
FYPDataset (Project Dataset)

The following images are 50 images in FYPDataset, which the dataset project has used all of them. The full images and relevant information collection file can be accessed and downloaded via the link below: FYPDataset | Powered by Box


(a) 1

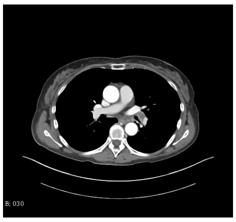

(b) 2

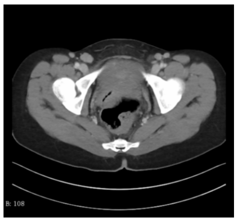

(c) 3

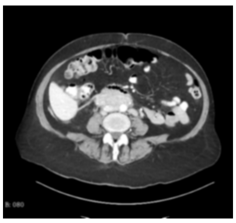

(d) 4


(e) 5


(f) 6


(g) 7


(h) 8


(i) 9

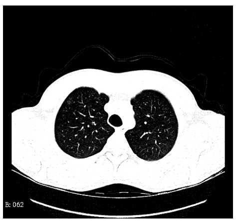

(j) 10


(k) 11


(l) 12

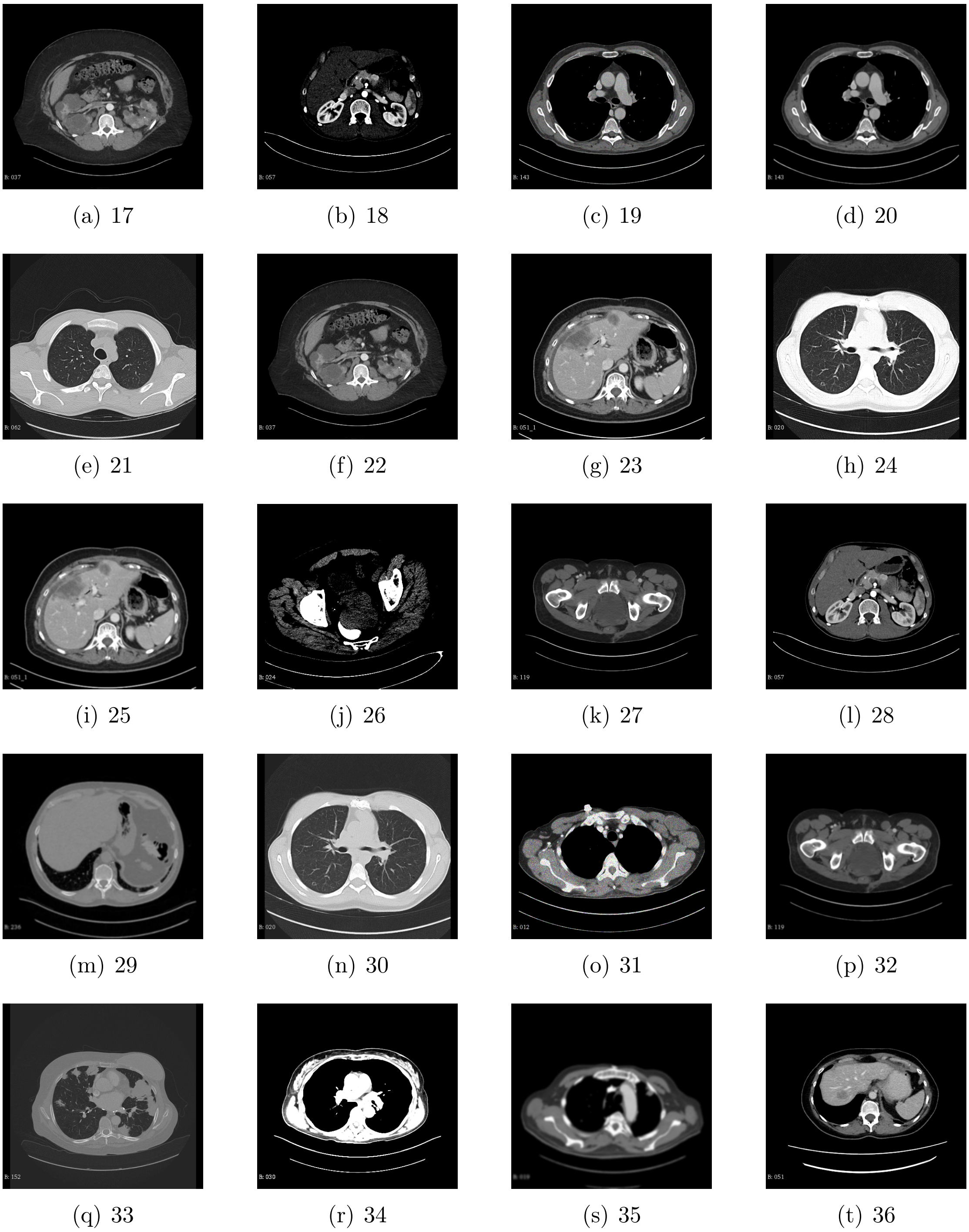

(m) 13


(n) 14


(o) 15


(p) 16

Figure A.1: Images 1-16 in FYPDataset



(a) 17 (b) 18 (c) 19 (d) 20

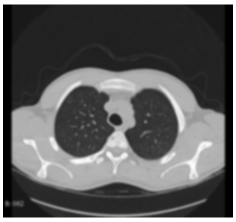
(e) 21 (f) 22 (g) 23 (h) 24

(i) 25 (j) 26 (k) 27 (l) 28

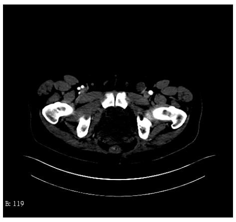
(m) 29 (n) 30 (o) 31 (p) 32

(q) 33 (r) 34 (s) 35 (t) 36

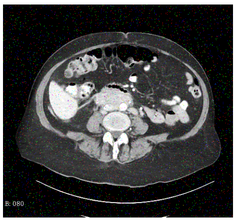
Figure A.2: Images 17-36 in FYPDataset



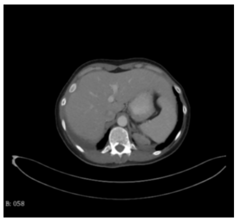
(a) 37



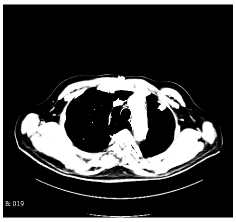
(b) 38



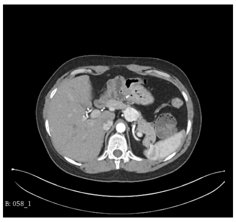
(c) 39



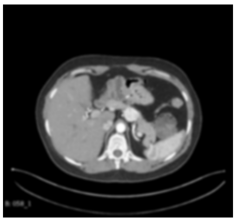
(d) 40



(e) 41

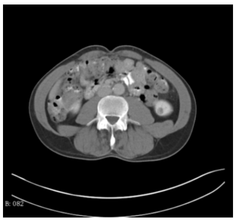


(f) 42



(g) 43

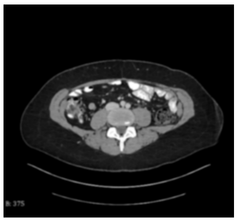
(h) 44



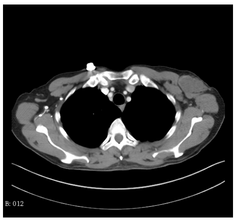
(i) 45

(j) 46

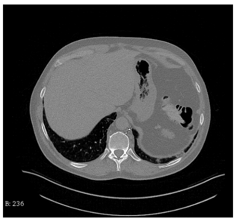
(k) 47



(l) 48



(m) 49



(n) 50

Figure A.3: Images 37-50 in FYPDataset

Appendix B

Code Listing

All codes have been coded and tested in MATLAB (R2019b).

B.1 Processing Images

B.1.1 Process to Blur-distorted Images

MATLAB file (blur.m):

```
% read the corresponding medical image
medical_image_original = imread('C:\Users\sunyu\Desktop\FYP_Dataset\
    Processed\M61_024.png');

% process to the blurry distortion images
% with Gaussian filter of standard deviation of 0.5, 1.0, 1.5
medical_image_blur_1 = imgaussfilt(medical_image_original,0.5);
medical_image_blur_2 = imgaussfilt(medical_image_original,1.0);
medical_image_blur_3 = imgaussfilt(medical_image_original,1.5);

% save the processed images
imwrite(medical_image_blur_1,'C:\Users\sunyu\Desktop\FYP_Dataset\
    Distortion\M61_024_blur050.png');
imwrite(medical_image_blur_2,'C:\Users\sunyu\Desktop\FYP_Dataset\
    Distortion\M61_024_blur100.png');
imwrite(medical_image_blur_3,'C:\Users\sunyu\Desktop\FYP_Dataset\
    Distortion\M61_024_blur150.png');

% read the corresponding medical image
medical_image_original = imread('C:\Users\sunyu\Desktop\FYP_Dataset\
    Processed\M61_143.png');

% process to the blurry distortion images
% with Gaussian filter of standard deviation of 0.75, 1.5, 2.25
medical_image_blur_1 = imgaussfilt(medical_image_original,0.75);
medical_image_blur_2 = imgaussfilt(medical_image_original,1.5);
medical_image_blur_3 = imgaussfilt(medical_image_original,2.25);

% save the processed images
imwrite(medical_image_blur_1,'C:\Users\sunyu\Desktop\FYP_Dataset\
    Distortion\M61_143_blur075.png');
imwrite(medical_image_blur_2,'C:\Users\sunyu\Desktop\FYP_Dataset\
    Distortion\M61_143_blur150.png');
```

```
imwrite(medical_image_blur_3, 'C:\Users\sunyu\Desktop\FYP_Dataset\
    Distortion\M61_143_blur225.png');
```

B.1.2 Process to Contrast-distorted Images

MATLAB file (contrast.m):

```
% read the corresponding medical image
medical_image_original = imread('C:\Users\sunyu\Desktop\FYP_Dataset\
    Processed\M61_024.png');

% process to the contrast distortion images
% with [low_in high_in] contrast limits of
% [0.2 0.8] [0.2 0.6] [0.2 0.4]
% [0.1 0.8] [0.1 0.6] [0.1 0.4]
% [0.3 0.8] [0.3 0.6] [0.3 0.4]
medical_image_contrast = imadjust(medical_image_original,[0.2 0.8],[]);
medical_image_contrast = imadjust(medical_image_original,[0.2 0.6],[]);
medical_image_contrast = imadjust(medical_image_original,[0.2 0.4],[]);
medical_image_contrast = imadjust(medical_image_original,[0.1 0.8],[]);
medical_image_contrast = imadjust(medical_image_original,[0.1 0.6],[]);
medical_image_contrast = imadjust(medical_image_original,[0.1 0.4],[]);
medical_image_contrast = imadjust(medical_image_original,[0.3 0.8],[]);
medical_image_contrast = imadjust(medical_image_original,[0.3 0.6],[]);
medical_image_contrast = imadjust(medical_image_original,[0.3 0.4],[]);

% save the processed images
imwrite(medical_image_contrast, 'C:\Users\sunyu\Desktop\FYP_Dataset\
    Distortion\Contrast\M61_024_0304.png');
```

B.1.3 Process to Noise-distorted Images

MATLAB file (noise.m):

```
% read the corresponding medical image
medical_image_original = imread('C:\Users\sunyu\Desktop\FYP_Dataset\
    Processed\M51_019.png');

% process to the noisy distortion images
% with the noisy types and intensity (when necessary) of
% Gaussian (0.01); 'Salt & Pepper' (0.02); Speckle; Poisson
medical_image_noise_1 = imnoise(medical_image_original,'gaussian',0.01)
    ;
medical_image_noise_1 = imnoise(medical_image_original,'salt & pepper'
    ,0.02);
medical_image_noise_1 = imnoise(medical_image_original,'speckle');
medical_image_noise_1 = imnoise(medical_image_original,'poisson');

% save the processed images
imwrite(medical_image_noise_1,'C:\Users\sunyu\Desktop\M51_019_noise.png
    ');
```

B.2 Mathematical Models of No-Reference IQA

B.2.1 BRISQUE

MATLAB file (BRISQUE.m):

```
% record the 50 medical images folder
image_folder = 'C:\Users\sunyu\Desktop\FYP_Dataset\Test\';
image_dir = dir([image_folder '*.png']);

% read all image files in this folder
% and score them by using BRISQUE algorithm
for i = 1:length(image_dir)
    image = imread([image_folder image_dir(i).name]);
    brisque_image = brisque(image);
    fprintf('%c',image_dir(i).name);
    fprintf('BRISQUE score for this image is: %0.4f.\n',brisque_image)
end
```

B.2.2 NIQE

MATLAB file (NIQE.m):

```
% record the 50 medical images folder
image_folder = 'C:\Users\sunyu\Desktop\FYP_Dataset\Test\';
image_dir = dir([image_folder '*.png']);

% read all image files in this folder
% and score them by using NIQE algorithm
for i = 1:length(image_dir)
    image = imread([image_folder image_dir(i).name]);
    niqe_image = niqe(image);
    fprintf('%c',image_dir(i).name);
    fprintf('NIQE score for this image is: %0.4f.\n',niqe_image)
end
```

B.2.3 PIQE

MATLAB file (PIQE.m):

```
% record the 50 medical images folder
image_folder = 'C:\Users\sunyu\Desktop\FYP_Dataset\Test\';
image_dir = dir([image_folder '*.png']);

% read all image files in this folder
% and score them by using PIQE algorithm
for i = 1:length(image_dir)
    image = imread([image_folder image_dir(i).name]);
    piqe_image = piqe(image);
    fprintf('%c',image_dir(i).name);
    fprintf('PIQE score for this image is: %0.4f.\n',piqe_image)
end
```

Appendix C

Pre-Assessment Questionnaire

Questionnaire on The Distortion of Medical Images

Welcome to fill out this questionnaire.

This questionnaire is about the distortion of medical images. No recognisable and sensitive personal identity information needs to be entered. Estimated time to complete: 3-5 minutes.

During the filling process, if any of the given choices do not meet your own circumstances, you can simply quit from this questionnaire, and thank you for your participation.

Please note that successfully completing this questionnaire will default to agreeing that all information entered will be used anonymously in scientific research projects. You have the right to withdraw all data you entered before May 31, 2020.

Who am I?

I am Yuhao Sun, a final year student of BSc Information and Computing Science from Xi'an Jiaotong-Liverpool University and the University of Liverpool. I am currently focusing on my Final Year Project, "Predict the Impact of Visual Distortion on Medical Images". Thank you for participating in this survey.

You will receive my contact information with the completion of this questionnaire.

There are 11 questions totally in this questionnaire.

Personal Background Information

1. Are you a hospital / healthcare institution worker?
   - Yes
   - No

2. Which of the following is your position at this institution? If both two are involved, please select the one that suits you best.

   - Clinician
   - Radiologist

3. How long have you worked in the above selected position?

   - Less than 3 years
   - Between 3 and 8 years
   - Between 8 and 13 years
   - More than 13 years

4. What do you think is your level of experience in all staff of the same / similar fields?

   - 1
   - 2
   - 3
   - 4
   - 5

   NB. 1 is the most lacking experience and 5 is the most experienced

Medical Imaging Related

1. Have you ever encountered the problem of medical images distortion?

   - Yes
   - No

2. If you have ever encountered the distortion of medical images, how often does this happen?

   - 1
   - 2
   - 3
   - 4
   - 5

   NB. 1 is the least frequent and 5 is the most frequent.

3. If you have ever encountered the distortion of medical images, what kind(s) of following possible distortion was it (were they)?

   - Contrast is too high / low

- Blurry
- Obscured by unknown reasons
- Compression
- Other, please specify

4. If you have ever encountered the distortion of medical images, what kind(s) of following negative effects have brought you?
    - Disease misdiagnosis
    - Missed diagnosis
    - Inaccurate judgment of disease severity
    - Later treatment plan is inaccurate
    - Other, please specify
5. When there is a problem with the image, are you more inclined to actively communicate with the relevant clinician / radiologist?
    - Yes
    - No
6. Do you think that the distortion on medical images is a serious problem for the medical community?
    - Yes
    - No
7. Would you like our team to talk with you further? If you are willing, please leave your contact information here, or please simply skip this question.

Thanks for your participation! If you are willing to talk to us further, please refer to the following contact information. We appreciate your contribution to our project.

Table C.1: Statistics from All Questionnaire Respondents

Group[a]	Question No.[b]	Options and Statistics (Percentage)				
1	1	**Yes**		**No**		-
		51 (100%)		0		-
	2	**Clinician**		**Radiologist**		-
		43 (84.31%)		8 (15.69%)		-
	3	**<3**[c]	**<8, >=3**	**<13, >=8**	**>=13**	-
		10 (19.61%)	11 (21.57%)	13 (25.49%)	17 (33.33%)	-
	4	**L1**[d]	**L2**	**L3**	**L4**	**L5**
		3 (5.88%)	2 (3.92%)	18 (35.29%)	23 (45.10%)	5 (9.80%)
2	1	**Yes**		**No**		-
		47 (92.16%)		4 (7.84%)		-
	2	**L1**	**L2**	**L3**	**L4**	**L5**
		18 (40.00%)	17 (37.78%)	9 (20.00%)	1 (2.22%)	0
	3	**Cont.**[e]	**Blur.**	**Obsc.**	**Comp.**	**Other**
		22 (43.14%)	39 (76.47%)	16 (31.37%)	8 (15.69%)	6 (11.76%)
	4	**R1**[f]	**R2**	**R3**	**R4**	**Other**
		28 (54.90%)	32 (62.75%)	32 (62.75%)	18 (35.29%)	1 (1.96%)
	5	**L1**	**L2**	**L3**	**L4**	**L5**
		5 (9.80%)	4 (7.84%)	5 (9.80%)	11 (21.57%)	26 (50.98%)
	6	**Yes**		**No**		-
		44 (86.27%)		7 (13.73%)		-
	7	***Provided***[g]		***Skipped***		-
		9 (17.65%)		42 (82.35%)		-

[a] Group 1 is about personal background information, Group 2 is about medical imaging related;
[b] Please refer to the full questionnaire for the question number;
[c] <3 stands for less than 3 years, (<8, >=13) stands for between 3 and 8 years; so on so forth;
[d] L1 and L5 stand for the least level and most level; from L1 to L5, the level is increasing;
[e] "Cont.", "Blur.", "Obsc.", and "Comp." stand for "Contrast is too high / low", "Blurry", "Obscured by unknown reasons", and "Compression" respectively;
[f] R1, R2, R3, and R4 stand for "Disease misdiagnosis", "Missed diagnosis", "Inaccurate judgment of disease severity", and "Later treatment plan is inaccurate" respectively;
[g] Provided means the respondents have filled out valid answers, i.e., contact information.

Appendix D

IQA Scores

D.1 Subjective IQA Scores

The detail information of observers (observer index, position, experience, time cost and data usage) please refer to Table 4.1.

Table D.1: The Quality Scores from Three Doctors

Image Set	**Index**	**File Name**	**Score[a]**			**Average[c]**
			001[b]	**002**	**003**	**(001+003)**
1	2	F21_020	3	2	5	4
	24	F21_020_0108	4	3	2	3
	30	F21_020_blur050	3	3	3	3
2	11	F21_108	4	3	5	4.5
	8	F21_108_blur075	3	3	3	3
3	46	F31_058	3	3	2	2.5
	10	F31_058_0106	5	3	5	5
	40	F31_058_blur100	2	2	2	2
4	44	F31_375	2	3	4	3
	48	F31_375_blur150	2	1	3	2.5
5	36	F41_051	3	3	5	4
6	42	F41_058	3	2	4	3.5
	47	F41_058_0208	3	2	5	4
	43	F41_058_blur225	2	1	1	1.5
7	7	F51_030	3	4	5	4
	34	F51_030_0104	1	2	2	1.5
	12	F51_030_blur050	4	4	5	4.5
8	33	F51_152	3	3	3	3
	1	F51_152_blur075	4	4	4	4
9	23	F61_051	4	3	5	4.5
	25	F61_051_blur100	3	2	3	3
	5	F61_051_Gaussian001	3	2	3	3
10	14	F61_080	4	2	5	4.5
	9	F61_080_blur150	3	1	5	4
	39	F61_080_Salt002	1	3	2	1.5
11	28	M21_057	4	3	4	4

	18	M21_057_0308	4	1	4	4
12	21	M21_062	4	3	4	4
	15	M21_062_0206	4	3	5	4.5
	37	M21_062_blur225	2	1	1	1.5
13	22	M31_037	4	2	4	4
	17	M31_037_blur050	3	2	5	4
14	4	M31_082	2	3	4	3
	45	M31_082_blur075	4	3	3	3.5
15	27	M41_119	4	3	1	2.5
	38	M41_119_0306	2	2	1	1.5
	32	M41_119_blur100	3	3	2	2.5
16	50	M41_236	3	3	4	3.5
	29	M41_236_blur150	3	1	3	3
17	49	M51_012	3	4	4	3.5
	31	M51_012_Speckle	2	2	2	2
18	3	M51_019	3	5	5	4
	41	M51_019_0204	1	2	1	1
	35	M51_019_blur225	2	1	1	1.5
19	13	M61_024	4	2	2	3
	26	M61_024_0304	2	1	1	1.5
	16	M61_024_blur050	5	2	2	3.5
20	6	M61_143	5	4	4	4.5
	20	M61_143_blur075	4	4	4	4
	19	M61_143_Poisson	5	3	5	5

[a] The score ranges from 1-5. The higher score, the higher quality;
[b] The index of doctor observers, refer to Table 4.1;
[c] The average score between observer 001 and 003, which has been used in the final results of Subjective IQA; the results of observer 002 was for testing part. Please refer to Table 4.1.
Correspond to the question: "To what extent do you think the quality of this image is good enough for you to get the above answer?"

Table D.2: The Similarity Scores from Three Doctors

Image Set	Index	File Name	Doctor 001	Doctor 002	Doctor 003	Average (001+003)
1	2	F21_020	4	2	5	4.5
	24	F21_020_0108	4	2	4	4
	30	F21_020_blur050	5	3	4	4.5
2	11	F21_108	4	3	5	4.5
	8	F21_108_blur075	4	3	2	3
3	46	F31_058	4	4	2	3
	10	F31_058_0106	5	4	5	5
	40	F31_058_blur100	3	2	3	3
4	44	F31_375	3	3	5	4
	48	F31_375_blur150	3	1	3	3
5	36	F41_051	5	3	5	5
6	42	F41_058	4	3	5	4.5
	47	F41_058_0208	4	2	5	4.5
	43	F41_058_blur225	3	1	3	3
7	7	F51_030	4	5	5	4.5
	34	F51_030_0104	2	2	2	2
	12	F51_030_blur050	5	4	5	5
8	33	F51_152	4	3	3	3.5
	1	F51_152_blur075	4	4	4	4
9	23	F61_051	4	3	5	4.5
	25	F61_051_blur100	5	2	4	4.5
	5	F61_051_Gaussian001	4	1	2	3
10	14	F61_080	5	2	5	5
	9	F61_080_blur150	5	1	5	5
	39	F61_080_Salt002	2	1	1	1.5
11	28	M21_057	4	3	5	4.5
	18	M21_057_0308	4	1	4	4
12	21	M21_062	4	3	3	3.5
	15	M21_062_0206	5	3	5	5
	37	M21_062_blur225	3	1	2	2.5
13	22	M31_037	3	2	4	3.5
	17	M31_037_blur050	4	2	5	4.5
14	4	M31_082	2	3	4	3
	45	M31_082_blur075	4	3	3	3.5
15	27	M41_119	4	3	2	3
	38	M41_119_0306	4	2	1	2.5
	32	M41_119_blur100	4	2	2	3
16	50	M41_236	4	3	4	4
	29	M41_236_blur150	4	1	4	4
17	49	M51_012	4	4	4	4
	31	M51_012_Speckle	3	2	3	3
18	3	M51_019	4	4	5	4.5
	41	M51_019_0204	2	1	1	1.5

	35	M51_019_blur225	3	1	2	2.5
19	13	M61_024	4	3	2	3
	26	M61_024_0304	2	1	2	2
	16	M61_024_blur050	5	3	2	3.5
20	6	M61_143	5	5	4	4.5
	20	M61_143_blur075	5	4	4	4.5
	19	M61_143_Poisson	5	3	5	5

Correspond to the question: "To what extent do you think the quality of this image is similar to the images you actually encountered during your work?"

D.2 Objective IQA Scores

Table D.3: The Scores for Three Mathematical Models

Image Set[a]	**Index**[b]	**File Name**[c]	**Score**[d]		
			BRISQUE	**NIQE**	**PIQE**
1	2	F21_020	36.8952	9.0878	51.7995
	24	F21_020_0108	40.9377	4.1091	61.1027
	30	F21_020_blur050	24.4362	5.8725	33.141
2	11	F21_108	42.5105	10.181	41.5581
	8	F21_108_blur075	42.7389	3.7654	65.2058
3	46	F31_058	52.3563	10.4238	42.7665
	10	F31_058_0106	52.8035	8.5358	55.42
	40	F31_058_blur100	48.5791	4.6395	65.3726
4	44	F31_375	50.4416	8.7737	35.9122
	48	F31_375_blur150	45.7001	8.2816	84.279
5	36	F41_051	51.7832	11.652	43.6362
6	42	F41_058	49.9559	5.9715	39.1052
	47	F41_058_0208	53.4997	8.0132	59.138
	43	F41_058_blur225	46.6561	5.827	90.0916
7	7	F51_030	50.6976	13.661	63.0897
	34	F51_030_0104	49.9166	3.8505	77.4766
	12	F51_030_blur050	50.1219	5.9049	69.1788
8	33	F51_152	33.1948	11.2343	26.082
	1	F51_152_blur075	30.4491	3.3007	39.7778
9	23	F61_051	41.8556	7.9137	34.4682
	25	F61_051_blur100	41.9974	10.863	65.5903
	5	F61_051_Gaussian001	42.6734	3.4227	59.2951
10	14	F61_080	40.6593	7.8448	30.2314
	9	F61_080_blur150	42.886	3.2939	81.9241
	39	F61_080_Salt002	47.1559	5.4157	56.9831
11	28	M21_057	53.2535	11.7042	47.0507
	18	M21_057_0308	52.9108	5.323	61.0339
12	21	M21_062	39.5836	4.6132	52.806
	15	M21_062_0206	57.878	3.2627	76.6391
	37	M21_062_blur225	58.9596	5.5571	89.6082
13	22	M31_037	41.2682	4.489	24.5025
	17	M31_037_blur050	38.4304	7.0544	23.3956
14	4	M31_082	49.8673	8.0557	40.2033
	45	M31_082_blur075	46.9585	11.7883	61.983
15	27	M41_119	51.4924	7.2616	43.4012
	38	M41_119_0306	50.4268	5.7233	68.2964
	32	M41_119_blur100	47.6113	8.6235	78.032
16	50	M41_236	45.7949	4.9243	69.319
	29	M41_236_blur150	52.3684	7.5833	81.4706
17	49	M51_012	52.7311	4.9314	54.7029
	31	M51_012_Speckle	54.2039	6.5946	63.294

18	3	M51_019	51.2514	4.7243	54.944
	41	M51_019_0204	50.5402	6.0035	82.5793
	35	M51_019_blur225	47.1502	5.825	85.3924
19	13	M61_024	42.4791	7.5811	33.7487
	26	M61_024_0304	55.0349	5.0354	73.5668
	16	M61_024_blur050	41.2651	7.5104	42.2571
20	6	M61_143	52.2119	3.3072	60.967
	20	M61_143_blur075	48.8722	4.9695	71.161
	19	M61_143_Poisson	54.077	13.3875	56.5819

[a] Image Set ranges from 1 to 20, almost all sets contain 2-3 medical images;
[b] Index ranges from 1 to 50;
[c] File Name provides information about the image. The letters before the first "_" stand for the gender and age group (F is female, M is male; 21 is the age group 21-30 so on and so forth). The letters before the second "_" and after the first "_" mean the index in the original folder. The letters after the second "_" stand for the types of the distortion (four digits such as "0108" mean the contrast distorted with the grayscale [0.1 0.8]; "blur050" means the blurry with the Gaussian filter with a standard deviation 0.50; Gaussian001/Salt002/Speckle/Poisson means the type of noisy distorted and corresponding density). If there is no second "_" thus the image is the original one (no distortion). For example, "F21_020_blur050" means the patient is a female who ages between 21 and 30; the image is distorted with the blurry "0.50"; [d] The Score ranges from 0-100. The lower scores, the better quality.