BRNO UNIVERSITY OF TECHNOLOGY

Faculty of Electrical Engineering and Communication

MASTER'S THESIS

BRNO, 2020

Bc. ZUZANA VAŠÍČKOVÁ

BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING
ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DETECTION AND EVALUATION OF DISTORTED FRAMES IN RETINAL IMAGE DATA
DETEKCE A HODNOCENÍ ZKRESLENÝCH SNÍMKŮ V OBRAZOVÝCH SEKVENCÍCH

MASTER'S THESIS
DIPLOMOVÁ PRÁCE

AUTHOR AUTOR PRÁCE	Bc. Zuzana Vašíčková
ADVISOR VEDOUCÍ PRÁCE	doc. Ing. Radim Kolář, Ph.D.

BRNO 2020

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

Master's Thesis

Master's study program **Biomedical Engineering and Bioinformatics**

Department of Biomedical Engineering

Student: Bc. Zuzana Vašíčková ***ID:*** 186699

Year of study: 2 ***Academic year:*** 2019/20

TITLE OF THESIS:

Detection and evaluation of distorted frames in retinal image data

INSTRUCTION:

1) Describe the properties and acquisition of retinal image sequences acquired by experimental fundus camera. 2) Describe the different distortions that may occur during acquisition. 3) Describe approaches to assessment of the image quality. 4) Design a methodology for detecting distorted images in retinal image sequences. 4) Implement this method, test on real sequences and evaluate. 5) Modify the method in order to achieve the best results. 6) Design a metric that will evaluate the quality of individual frames in the sequence and also quality of entire sequence. 7) Test the method on the supplied data, evaluate and discuss the results.

RECOMMENDED LITERATURE:

[1] LeCun, Y. et al., “Deep learning”, Nature, 521 (7553), pp. 436–444, 2015.

[2] Tornow, R. et al. „Time resolved quantitative inter-eye comparison of cardiac cycle induced blood volume changes in the human retina“, Biomedical Optical Express, 9(12), pp. 6237-6254, 2018.

Date of project specification: 3.2.2020 ***Deadline for submission:*** 29.5.2020

Supervisor: doc. Ing. Radim Kolář, Ph.D.

prof. Ing. Ivo Provazník, Ph.D.
Chair of study program board

ABSTRACT

The master's thesis deals with detection and evaluation of distorted frames in retinal image data. The theoretical part contains brief summary of eye anatomy and methods for image quality assessment generally, and also particularly on retinal images. The practical part is carried out in programming language Python. It contains preprocessing of the available retinal images in order to create an appropriate dataset. Further a method for evaluation of three types of blur in distorted retinal images is proposed, specifically Inception-ResNet-v2 model. This method is not feasible and thus another method consisting of two steps is designed - classification of the type of blur and subsequently evaluation of the particular blur level. Filtered Fourier spectrum is used to classify the type of blur and features extracted by ResNet50 serve as the input for regression model. This method is further extended with initial step of detection of blurred frames in retinal sequences.

KEYWORDS

retina, image quality assessment, image blur, deep learning, feature extraction, regression, ResNet50, retinal sequences

ABSTRAKT

Diplomová práca sa zaoberá detekciou a hodnotením skreslených snímok v retinálnych obrazových dátach. Teoretická časť obsahuje stručné zhrnutie anatómie oka a metód hodnotenia kvality obrazov všeobecne, ako aj konkrétne hodnotenie retinálnych obrazov. Praktická časť bola vypracovaná v programovacom jazyku Python. Obsahuje predspracovanie dostupných retinálnych obrazov za účelom vytvorenia vhodného datasetu. Ďalej je navrhnutá metóda hodnotenia troch typov šumu v skreslených retinálnych obrazoch, presnejšie pomocou Inception-ResNet-v2 modelu. Táto metóda nebola prijateľná a navrhnutá bola teda iná metóda pozostávajúca z dvoch krokov - klasifikácie typu šumu a následného hodnotenia úrovne daného šumu. Pre klasifikáciu typu šumu bolo využité filtrované Fourierove spektrum a na hodnotenie obrazu boli využité príznaky extrahované pomocou ResNet50, ktoré vstupovali do regresného modelu. Táto metóda bola ďalej rozšírená ešte o krok detekcie zašumených snímok v retinálnych sekvenciách.

KLÍČOVÁ SLOVA

sietnica, hodnotenie kvality obrazov, skreslenie obrazu, hlboké učenie, extrakcia príznakov, regresia, ResNet50, retinálne sekvencie

VAŠÍČKOVÁ, Zuzana. *Detection and evaluation of distorted frames in retinal image data*. BRNO, 2020, 63 p. Master's Thesis. Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Biomedical Engineering. Advised by doc. Ing. Radim Kolář, Ph.D.

Vysázeno pomocí balíčku `thesis` verze 3.03; `http://latex.feec.vutbr.cz`

ROZŠÍŘENÝ ABSTRAKT

Diplomová práca sa zaoberá navrhnutím metódy hodnotenia kvality obrazov. Dôraz sa kladie na skreslenie vyskytujúce sa v retinálnych snímkach, a to skreslenie šumom spôsobeným pohybom, šumom spôsobeným rozostrením a šumom spôsobeným dlhou expozíciou. Uvedené typy šumov sú charakteristické pre spracovávané retinálne snímky. V kapitolách 2 a 3 sú popísané metódy hodnotenia obrazov a skreslenia obrazov, ako všeobecne, tak aj konkrétne pre retinálne snímky. Kapitola 4 sa však už zaoberá predspracovaním dostupných retinálnych dát s cieľom navrhnutia sady dát simulovaných skreslených obrazov. Ďalšie sekcie tejto kapitoly popisujú navrhnuté a testované metódy hodnotenia skreslenia obrazov. Pôvodne bola navrhnutá metóda zostavená tak, aby hodnotila rozsah troch typov skreslenia naraz a vzhľadom k tomu bol vytvorený dataset so snímkami, na ktorých sú simulované všetky tri typy šumu naraz. Zvolenou metódou bola sieť InceptionResNet-v2 implementovaná pomocou open-source knižnice neurónových sietí Keras. Keďže táto metóda nebola vhodná, bola navrhnutá ďalšia metóda. Tá je založená na klasifikácii skreslenia do troch tried - pohybové skreslenie, skreslenie spôsobené rozostrením a skreslenie spôsobené dlhou expozíciou, ktoré boli všetky predom simulované. Bola vytvorená nová vhodná sada dát. Tentokrát bola každá snímka poškodená iba jedným typom šumu. Zo snímok bolo získané normalizované logaritmizované Fourierove spektrum, ktoré bolo následne upravené Laplaceovým operátorom. Takto upravené spektrum slúžilo ako vstup do trojvrstvovej neurónovej siete, ktorej výstupom bol typ šumu. Po kroku klasifikácie nasleduje hodnotenie úrovne skreslenia daného typu pomocou regresného modelu. Ako vstup do regresného modelu sa využívajú príznaky extrahované pomocou siete ResNet50 z poslednej vrstvy. Tieto príznaky sú agregované do troch vektorov, a to podľa priemeru a odchýlky, momentov a kvantilu. Posledná metóda dosahuje úspešnosť klasifikácie typu šumu na simulovaných skreslených dátach 0,92 a na skutočných skreslených dátach 0,46. Na simulovaných dátach dosahuje následný odhad úrovne šumu spôsobeného pohybom SROCC 0,96, PLCC 0,97, RMSE 6,61 a R2 skóre 0.94. Odhad úrovne šumu spôsobeného rozostrením dosahuje hodnoty SROCC 0,72, PLCC 0,77, RMSE 9,57 a R2 skóre 0.58. Odhad úrovne šumu spôsobeného dlhou expozíciou dosahuje SROCC 0,50, PLCC 0,66, RMSE 12,96 a R2 skóre 0,43. Aby bolo možné vyhodnotiť kvalitu retinálnych sekvencií, bol pridaný ďalší krok detekcie šumu. Pri detekcii šumu sa taktiež ako vstup do modelu logistickej regresie využívajú extrahované agregované príznaky. Úspešnosť detekcie šumu je 0.46, je však potrebné poznamenať, že samotné predpovedné pravdepodobnosti zodpovedajú faktu, že zašumené snímky boli určené ručne a často bolo otázne, či danú snímku považovať ešte za ostrú, alebo za skreslenú. Konečná predikovaná kvalita retinálnych sekvencií zodpovedá subjektívnemu vnemu ich kvality, aj keď navrhovaná metóda je prísnejšia ako subjektívne hodnotenie.

DECLARATION

I declare that I have written the Master's Thesis titled "Detection and evaluation of distorted frames in retinal image data" independently, under the guidance of the advisor and using exclusively the technical references and other sources of information cited in the thesis and listed in the comprehensive bibliography at the end of the thesis.

As the author I furthermore declare that, with respect to the creation of this Master's Thesis, I have not infringed any copyright or violated anyone's personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation § 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll., Section 2, Head VI, Part 4.

BRNO ....

author's signature

ACKNOWLEDGEMENT

I would like to thank the supervisor of the master's thesis doc. Ing.Radim Kolář, Ph.D. for his professional mentoring, consultations, patience, and thoughtful suggestions concerning the diploma thesis.

BRNO ....

author's signature

Contents

List of Figures

List of Tables

Introduction

Retinal image data are a very important tool in the diagnosis of retinal diseases. To make the correct diagnosis it is necessary to acquire images of the best possible quality, as every detail might be important. The importance of the details in the retinal images is also linked to other algorithms aiming for automation of the diagnosis. To properly segment, and classify the data, the image must have the best possible quality, otherwise it can cause errors, and lead to a false diagnosis or results.

Retinal image data are specific in the type of image corruption, and to provide a functional method detecting and evaluating the distorted frames it is crucial to understand the basic anatomy of the eye, and the means of capturing the retinal image data, which is provided in the first chapter.

The idea of this thesis is to provide a reliable method for detection and evaluation of distorted frames. In chapter 2 are described general methods of IQA (image quality assessment), while chapter 3 is dedicated to a specific type of IQA, image blur identification, providing a summary of applied methods.

Chapter 4 contains proposed methods of image blur evaluation. Two datasets are described in this chapter. One contains images distorted by three types of blur at the same time and the other one contains images distorted only by one type of blur at once. Firstly a method that would be able to evaluate the images with combined blur is proposed. The results of this method suggest that the problem is too complex for such a small dataset. Hence the second dataset is used, firstly to propose a method of identifying the image blur and secondly to evaluate the level of the identified image blur. Normalized logarithmic Fourier spectrum is used to classify the type of blur with neural network. To further evaluate the particular image blur level is two regression methods are compared, Inception-ResNet-v2 modified for regression and ResNet50 extraction of features that are regressed by general regression neural network (GRNN) and partial least squares regression (PLSR).

The results of method proposed in chapter 4 are described in chapter 5. Evaluation of real retinal image sequences is also provided in this chapter.

1 Eye

The main purpose of the eye is to focus light on the retina in order to process chemical reaction into an electrical signal, which can be further propagated through the optic nerve to provide a visual information. To achieve this goal, the eye is a sphere consisting of several different elements (Fig. 1.1), each ensuring a different function. The geometrical (optical) axis passing through the center of the cornea and the lens and intersecting sclera at the posterior pole, is the axis about which the eye is rotated using the eye muscles. The visual axis does not run concurrently with the geometrical axis, as it is passing through the fovea, the central point of image focusing in the retina. The horizontal (transverse) diameter of the eye is 23.4 mm and the analogous vertical diameter (intersecting the geometrical axis) is about 23 mm and it is usually less than the horizontal diameter. [3], [4]

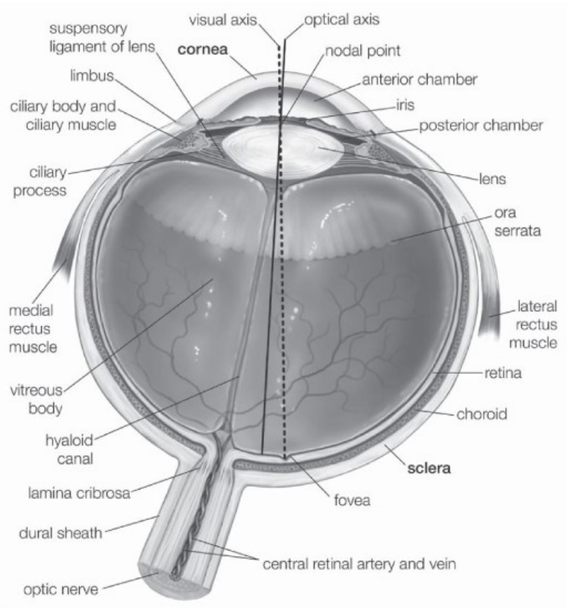


Fig. 1.1: Horizontal cross section of the human eye. [4]

The sclera is a membrane formed by protein fibres arranged into ribbons (*lamellae*) and pierced by numerous vessels. In thickness the sclera is about 0.3 to 0.8 mm thick, being thinnest at the connection of the extraocular muscles. The sclera turns through the corneoscleral junction into the cornea. The cornea is a thin convex-concave lens, protected by a layer of tears on the outside and surrounded by aqueous humor on the inside. The cornea is about 0.52 mm thick in the middle, flattening towards the edges. It has complex structure, comprising of several layers. Uvea, covering the inside of the sclera, consists of two segments. The posterior layer of uvea is the choroid, which includes many blood vessels with a connective tissue. The anterior layer of the uvea is thickened and forms the ciliary body, that is a part of the ciliary muscle, essential to change the shape of the lens in order to focus. While the muscle contracts, the lens assumes a rounded shape, bringing the near objects on the retinal surface and vice-versa. The ciliary muscle is connected to the retina through the *ora serrata*. As the retina is an important part of the thesis, it is in detail described later. From the base of the ciliary muscle expands the iris, a thin contractile tissue with a central hole, the pupil. Its posterior layer is a heavily pigmented tissue, making it light-tight and the anterior layer is formed by cells making up dilator iridis muscle. The dilator iridis muscle is connected to the sphincter iridis or sphincter pupillae at the border of the pupil. The dilator and sphincetr muscles are antagonists, adjusting the size of the pupil according to nerve impulses. The iris partially lays on the lens, pushing forward the pupil, forming a truncated cone, allowing the light to enter only through the pupil. The crystalline lens, or only lens, are transparent, biconvex, and semisolid elements, changing shape to facilitate focus. The shape is changed by the suspensory ligament, as its anterior leaf originates at the anterior surface of the lens and runs towards the ciliary body. The major portion of the eye consist of the vitreous. It has a spherical shape and it is transparent, colorless and gel-like. [3]

The eye can be divided into three chambers, the anterior chamber, the posterior chamber, and the vitreous chamber. The anterior chamber is bordered by the sclera, the lens, the ciliary body, the posterior surface of the cornea and the anterior surface of the iris. It is filled with aqueous humor, providing nutrition and also flow for metabolic waste. The aqueous humor leaves the anterior chamber through the canal of Schlemm. The posterior chamber is filled with aqueous humor too, being bordered by the posterior of the iris, lens, the anterior of the vitreous body and the ciliary body. The vitreous chamber is behind the lens and it is filled with vitreous body, as mentioned before. [3]

1.1 Types of Eye Movement

As it has been long known, the eyes are not still during the fixation, even if the movements go on unnoticed. The fixation is interrupted by rapid jerks in a irregular intervals. It is important to understand the eye movements during the fixation as they might result into undesirable consequences in retinal sequences. Three main types of eye movement occur in humans during fixation - tremor, microsaccades, and drifts. [5], [6]

Tremor is described as wave-like irregular movements of high frequency of 30-17/sec and small deviation, approximately 0.34 arcmin. It is the smallest of all eye movements and therefore it is also difficult to record. Tremor is believed to be independent in the two eyes. [6], [5] Flicks, or microsaccades, are small, fast motions of eyes occurring approximately every second with deviation od several arcmins during voluntary fixation, taking place simultaneously in both eyes. According to statistical studies, it seems that microsaccades act as a corrective saccade that corrects displacement outside of central fovea after a previous saccade or drift. The occurence, direction, and amplitude of microsaccades are statistically related to previous displacements cause by drifts or previous microsaccades. [6], [5]

Slow irregular drifts occurs simultaneously with tremor and between microsaccades and their diversion of 6 arcmin. During drifts the object can move across a dozen photoreceptors. Drifts are conjugate and non-conjugate.

1.2 Retina

The retina is a part of the eyeball, which is lining the inner layer of the shell of the bulbus. The main purpose of the retina is to convert light energy to chemical energy and subsequently to electrical energy to further process visual information. [3], [7], [8]

The retina is a complex and delicate tissue, being quite thin, as it is approximately 0.5 to 0.1 mm thick, while it reaches its maximal thickness around the optic disc. The retina contains the fovea, the site of the most accurate vision, and the macula. [3], [7], [8]

Regarding the structure of the retina, it consists of ten specific layers (Fig. 1.2), subsequently from inside to outward:

1. inner limiting membrane
2. nerve fiber layer
3. ganglion cell layer
4. inner plexiform layer
5. inner nuclear layer
6. outer plexiform inner
7. outer nuclear layer
8. outer limiting membrane
9. photoreceptor layer
10. RPE (retinal pigment epithelium)

In general four main layers are recognised - RPE next to the choroid, layer of rods and cones, layer of neurons (bipolar nerve cells) and layer of ganglion cells. [3] [4]

The photoreceptor layer protected by the RPE contains rods and cones which are light sensitive. The outer nuclear layer contains the nuclei of the rods and cones

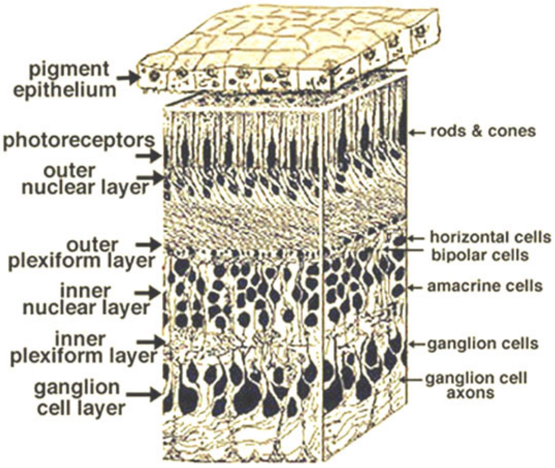


Fig. 4. 3-D block of a portion of human retina.

Fig. 1.2: 3D block with a portion of human retina. [8]

and the inner nuclear layer contains the nuclei and cell bodies of bipolar cells. The ganglion layer consists of the ganglion cells. The neurons from nuclear layers make their connections through outer and inner plexiform layers. In the outer plexiform layer the rods and cones are connecting to the dendrites of the bipolar cells, while in the inner plexiform layer are connected axons of the bipolar cells and the dendrites of ganglion cells. Thus the light sensitive rods and cones process light input and the induced chemical changes in the rods and cones are further transmitted as electrical message to the bipolar cells and further to the ganglion cells through mentioned layers. The changes are afterwards passed along the optic nerve along the axons of the ganglion cells from the eye into the higher regions of the brain. [4]

The photoreceptor layer, rods and cones, are sensory neurons converting light energy to chemical energy, in order to provide a signal interpretable by the nervous system. The rods are highly sensitive photoreceptors, reporting any photon absences, regulating perception of size, shape and brightness. The cones differ from rods morphologically, functionally and in the number, as there is less cones than rods in the human eye. While cones are less sensitive than rods and their provide briefer light response, they ensure color vision with improved temporal resolution. [4], [7]

The RPE is a single sheet of cells containing pigment granules, that act as a restraining barrier preventing the diffusion of the blood material between the blood supply of the choroid and the light/sensitive outer segments of the photoreceptors. The pigmented cells also absorb scattered light, improving the quality of the optical system.[4], [8]

Fovea centralis is a localized region of the retina, near the optic axis. The main advantage of the fovea is the absence of inner layers of the retina, resulting into a depressed foveal pit. This is the place of the accurate vision, the eyes are being directed towards the objects of interest in a way that their images fall on this place. Thus the centre of the lens, the object and the centre of the macula lutea are in a straight line. In the centre of the fovea are only cones, while rods appear again on the boarders. The fovea is sometimes called also macula lutea, a circular patch characterized by yellow pigment in nervous layers. The macula lutea is a small vague area stretching over the whole central retina. On the other hand the retina also contains a so-called blind spot at the place of the the optic nerve head. [4]

The OD (Optic disc), or the optic nerve head is a place on the retina without photoreceptors. It corresponds to the blind spot and it there is no light detection. It is not possible to perceive blind spot while both eyes are opened and if one of them is closed, the brain fills in the missing information. The OD can be seen with an ophatlmoscope in the back of an eye on the nasal side of the macula lutea. It is oval and approximately 1.5 mm in diameter. The OD is a place of optic nerve head, however, it is also the entry point into the eye for blood vessels to the retina. The evaluation of the OD is necessary in diagnosis of glaucoma and other nerve diseases resulting in vision impairment. [4]

1.3 Blood Supply of the Retina

The retina is supplied both directly and indirectly. The inner layers of the retina receive blood supply through the CRA (central retinal artery), running out of the optic nerve head and branching on the surface of the OD, generally forming four branches in the human retina. Further the arterial intraretinal branches then supply three layers of capillary networks, the RPCs (radial peripapillary capillaries), an inner, and outer layer of capillaries. The RPCs are the most superficial layer, running in the inner part of the nerve fiber layer. They anastomose with each other and other deeper capillaries. [6], [8], [9]

However, the retina may be also directly nourished by a branch of a ciliary artery in the OD. When the ciliary artery branches are present in the OD, they are forming a circle with possible anastomoses of the CRA. The retinal capillaries eventually coalesce into retinal venules, that are deeper to the arterioles. These venules drain the blood back into main veins, the retinal veins, eventually draining into the CRV (central retinal vein), which empties into the superior ophtalmic vein. [6], [8], [9]

The choroid supplies the retina, explicitly the outer layers of the retina. The choroid has the greatest blood flow (85% of the total eye blood flow) and provides nutrience and maintenance of the retinal photoreceptors. The choroidal arteries arise from the ciliary arteries and branches around the OD. The venules drain into veins running anterior towards the equator of the eyeball and eventually drain into the vortex veins (one or two), each draining each of the four quadrants of the eyeball. [6], [8], [9]

1.4 Experimental Videoophtalmoscope

In general the VO (videoophtalmoscope) (Fig. 1.4) is similar to a fundus camera (Fig. 1.3). The main difference from the fundus camera is the inverted illumination-acquisition approach of the VO. In the standard fundus cameras is the fundus illuminated by a ring on the cornea, while the central are remains for observation. On the other hand, in this case is for the illumination used the central part of the pupil (the entrance pupil), and for the imaging is used the rest of the pupil (the exit pupil). As a result, the employed light source is a simple LED source, with the wavelength of 575 nm. The LED light source is centered of the optical axis and shines the lights towards the eye. The image detector in this case is a CMOS camera, that takes short video sequences with 25 fps, a resolution of 1936x1216 pixels. The sequence is centered on the optic nerve head and is of rectangular view of 1000x770, as the interest of the final image is the optic nerve head. One pixel of the recorded image is 9,3 μm on the retina and a visual angle of .2 arcmin. [2], [10]

To acquire the image an ophtalmoscopic lens of 4O D is used, which later forms aerial image of the retina in the image plane, later re-imaged by system of two lenses with the field of view being 20° × 15°. [2], [10]

Examples of several fundus retinal images are provided in Fig. 1.5 - 1.8.

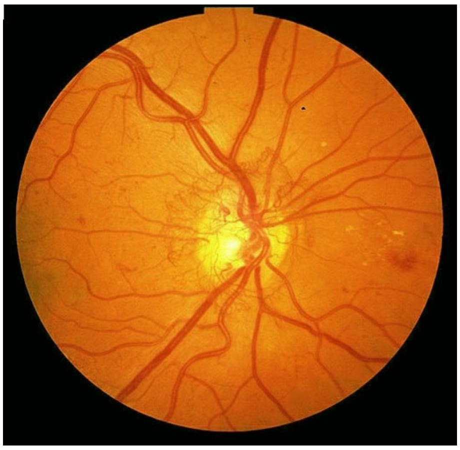


Fig. 1.3: A view of the fundus retinal image [8] .

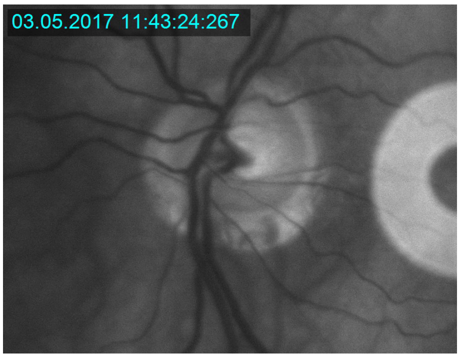


Fig. 1.4: A view of the VO retinal image.

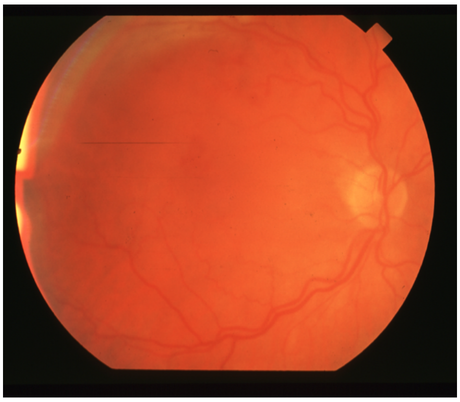


Fig. 1.5: An example of a blurred retinal image [11].

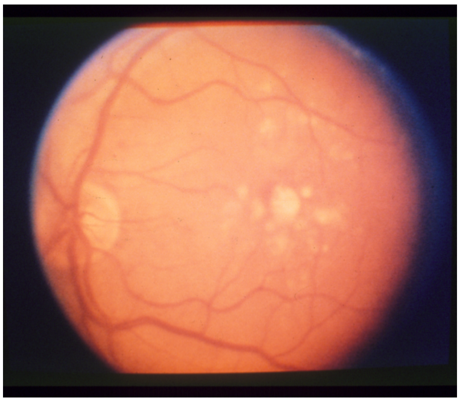


Fig. 1.6: An example of a blurred retinal image [11].

Fig. 1.7: An example of a clear retinal image [11].

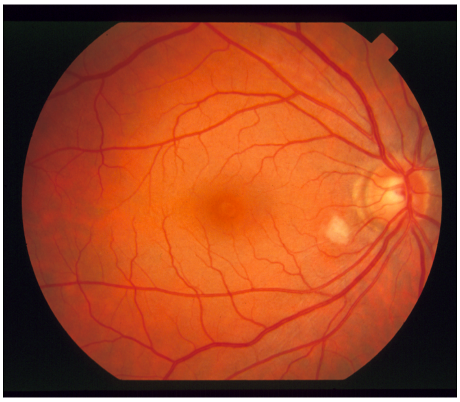


Fig. 1.8: An example of a clear retinal image [11].

2 Image Quality

To define image quality it is necessary to define the purpose of image itself. Images are produced in order to display visual information, that is considered to be useful. Unfortunately it is not possible to capture an image without any imperfections, especially due to technical challenges in the imaging and capturing process. Therefore the final image is only a degraded version of the scene. [12]

The captured image can be degraded in many ways, such as noise, geometrical degradation, illumination and color imperfections, and blur, the latter being one of the main topics of this thesis. [12]

2.1 IQA algorithms

It is necessary to determine the quality of an image in order to be able to use the image for diagnosis. Due to this requirement the need for the development of automated IQA algorithms has arisen. These algorithms often employ two basic different approaches and can be differentiated into four groups of algorithms. The first approach is focused on the use of generic image quality parameters, being sharpness, contrast, illumination, and others. On the other hand, the second approach is based on the structuring image quality parameters, being the location of the main anatomical features within the image and the vascular structure. [13]

The first group of the IQA algorithms uses the first approach, the second group uses the second approach, while the third group combines both of the approaches and the fourth group are deep learning based methods. However, these approaches also differ in the availability of reference information. They can have full reference (utilizes the information from the original image), reduced reference (utilizes some features of the original image), and no reference (does not utilize any information from the original image). It is understandable that no reference algorithms are the most required ones, as they do not need any original image. These algorithms often focus on one group of image distortions, for example image blur. [13] [14]

2.1.1 Generic Image Quality Based Algorithms

The earliest method for IQA of the retinal image quality selects a set of retinal images representing the ideal sharp image. To assess the image quality of the evaluated image it is required to obtain intensity histogram and to further convolve the histogram of the evaluated image with the histogram of the ideal image. The intensity histogram was used based on the observation, that in ideal images the values of brightness, contrast and SNR (Signal-to-noise ratio) are in relationship with intensity histogram. Based on the result of the convolution the evaluated image is given a value of quality index Q, which is then normalized, the best result being 1, representing the maximum possible quality, and the worst result being 0, representing the minimal possible quality. [13], [15], [16]

However, another study of IQA of the retinal image originated from the previously mentioned approach had realized that even bad quality images can have intensity histograms similar enough to the histograms of the selected ideal images. Thus they had decided to extend this technique by including another criterion defining an ideal image - the distribution (histogram) of the edge magnitudes. [13], [15], [17]

Other studies managed to create quality indicators based on many other generic features, such as illumination, image sharpness, color, luminance, contrast and other. The main advantage of this approach are its reduced computational demands. [13]

2.1.2 Structuring Image Quality Based Algorithms

One proposed algorithm based on the structural parameters is focused on the area of the vessel segmentation. It is necessary to firstly extract blood vessels using two Gaussian-shaped kernels rotating through 12 angles with increment of 15°, direction dependent recursive region growing algorithm, and threshold used to classify pixels as vessels or otherwise. Therefore the image quality score is in this case the count of vessel pixels. A threshold is used to classify images as gradable or ungradable. [13], [18]

Another approach employs the calculation of the contrast and quantity of visible blood vessels in the macula and also compares the contrast between the fovea region and the background. It is necessary to detect the optic nerve head and the fovea. The vessel tree is segmented and skeletonized afterwards. The calculation is made from the circular area around the detected fovea (radius of two optic discs).The primary parameter is the presence of small vessels in the area. Three other parameters of the vessels in the macula region also contribute to the image quality - the distance from the fovea, and again the contrast and the quantity. After all of these parameters are gathered, an overall vascular metric, being the first indicator of the image quality, is defined as the sum of the ratio of the number of pixels in the segment multiplied by the contrast of vessels in the segment to the contrast with the local background retina. The second indicator of the retinal image quality is the contrast of the fovea to the background. Both of these indicators are combined by multiplication into the final image quality measure, classifying the images into five categories using empirically determined thresholds. [13], [15] Another possible approach is to use a set of filters and clustering to describe the structure of the image. The output of one set of filters is a response vector, characterizing a property of the pixel. The pixel is then clustered by k-means clustering according to the distance of several given response vectors from several sets of filters. Therefore the pixels of one group have similar structural properties. This can be used to define ideal images and measuring the image quality by comparing the ideal clusters to the clusters of the analyzed image. [19]

These approaches are typically used to determine the image quality in regard to a specific use of the retinal image for diagnostic purposes. [15]

2.1.3 Deep Learning Based Algorithms

Introduction to Deep Learning

Deep learning networks use multiple processing layers to find abstract features on the given data. Therefore the main advantage in comparison with the previously mentioned algorithms is the lack of hand-crafted features, usually typical only for a specific dataset. Deep learning algorithms are able to find hidden or very complex features. When no hand-crafted features are used, the input is the whole information, in this case an image. One of the deep learning networks is a CNN (convolutional neural network) illustrated in Fig. 2.1. The main principle of this network is to use an input image and process the image through several convolutional layers consisting of several convolutional filters (kernels) with their respective weights. The output of each layer are feature maps, the number of feature maps depends on the number of filters in one layer. The outputs of one layer, and image consisting of the feature maps, is feed-forwarded into the next layer and the process continues through all of the convolutional layers. Another important part of the CNN is a max-pooling layer, to avoid over-fitting and to reduce the computational demands. The last layers are the fully connected layers followed by typically a softmax layer, aiming to classify the data according to the features extracted in the previous layers. [1],[20]

Regarding the application of the CNN for the purposes of the medical applications, three main options are available - training a new CNN from scratch, use pre-trained architectures such as AlexNet, VGG-Net, and others to extract features, or use large datasets for unsupervised pre-training. [21]

Applied methods

Most of the available articles regarding the use of deep learning algorithms for IQA differ essentially in the use of one of the previously mentioned techniques, the used architecture, and the dataset.

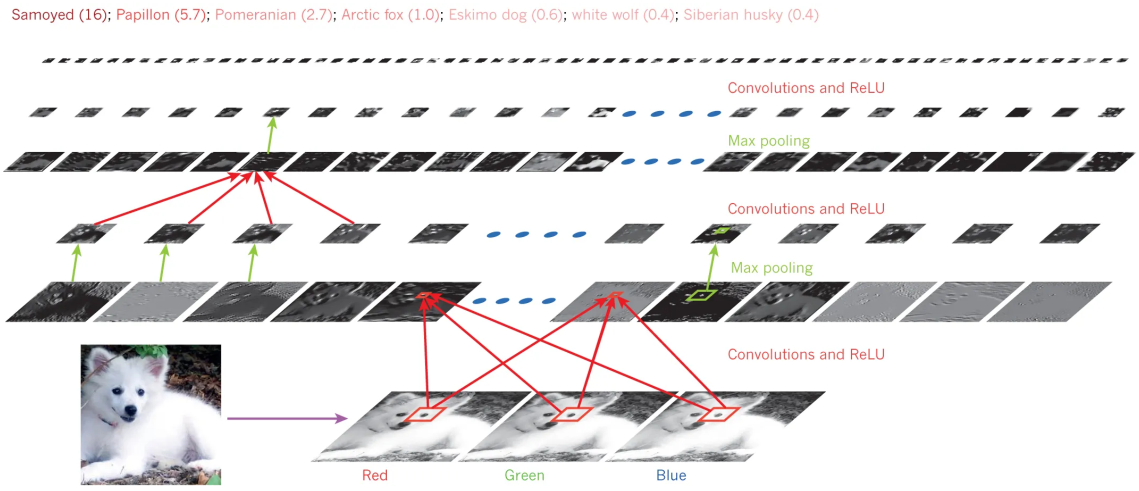


Fig. 2.1: Illustration of a CNN [1].

Mahapatra et al. used saliency maps combined with supervised CNN of 5 layers of convolution and max-pooling layers. Three fully connected layers with 4000, 2000, and 1000 nodes follow the last convolution layer. The last layer is the softmax resulting in the calculation of probabilities of each class. The used neurons employed ReLUs (Rectified Linear Unit)s, which allows to complete the training faster. [22]

Another method proposed by Sun et al. used a CNN pre-trained on a large dataset and fine-tune it. There are several choices of pre-trained CNNs, such as AlexNet,GoogLeNet, VGG-16, and ResNet-50, differing in the number of complexity of used layers. In this case two pre-trained CNNs had been chosen - GoogLeNet, VGG-16. Both of these networks had been trained at the same time and the total loss was a total of the loss of each network, allowing the backward propagation method to broadcast the classifier gradients to all networks. [23]

Fengli et al. decided to fuse features from saliency map and from fine-tuned CNNs and according to the selected features classify the retinal images by multi-kernel SVM (support vector machine). The CNN fine-tuned and used in this particular algorithm is Alexnet architecture, except its last layer, as the features are fused with other features and classified by SVM. [20]

Alexnet CNN architecture modified for two-class classification is used again in a study by Saha et al. The Caffe reference model had been used to initialise the networks and the network was fine-tuned using the retinal images from fundus camera. [21]

Based on the fact that it is difficult to train deep CNNs as the training requires a large dataset, which is often unavailable to restricted resources in medical applications, Tennakoon et al. decided to investigate two approaches - shallow CNN trained from scratch and pre-trained filters to extract features and use them in different classifiers. The proposed shallow CNN consists of three convolutional layers, each immediately followed by rectified linear activations and max-pooling layers. The results are fed to two consecutive fully connected layers and their output is classified using the softmax layer. As the pre-trained filters had been used Alexnet architecture and the output from the last fully connected layer. These features were used to train four different classifiers - single layer Neural Network, linear SVM, Boosted trees, and k-Nearest Neighbour. [24]

Tozatto et al. used the Inception v3. To adapt the pre-trained model to the IQA problem resulting in only two classes, the last three fully connected layers were replaced by three layers with 1024, 512, and 2 units, the last layer being the softmax layer. [25]

3 Image Blur

Image blur is the major source of image degradation, caused by several different phenomena. It is a very common form of bandwidth reduction of the ideal sharp image. It is mostly undesired, as important regions often become less sharp.[12], [26], [27]

Image blur is a concern of image restoration, also referred to as image deblurring. The aim of image reconstruction is to estimate uncorrupted image, often by performing the inverse function, however it is necessary to know the characteristics of the process causing the corruption of an image, which is of course not possible in real life. Therefore it is essential to perform a step of blur identification, to estimate the characteristics of the blurring process. This step is followed by the image restoration itself and together are known as blind image deconvolution. [12]

In this master's thesis the main goal is the blur identification as part of the IQA of retinal sequences. Further below are described sections dedicated to certain types of image blur present in retinal sequences and to selected image blur identification algorithms, in order to provide a proposal of the blur identification algorithm of retinal images and sequences.

3.1 Types of Image Blur

All images affected by any type of image blur exhibit similar characteristics, including lowpass smoothing of the original image. Thus the visual perception of the blurred image is degraded, as edges become less sharp, visible. [28]

Atmospheric Turbulence Blur
Can be modeled as Gaussian PSF (point spread function) (Fig. 3.1) due to long-term exposure through the atmosphere as:

$$h(i,j) = Kexp(-\frac{(i^2)+(j^2))}{2\sigma^2}) \tag{3.1}$$

where K is a normalizing constant ensuring that blur is a unit of volume, and σ^2 is the variance that determines the severity of the blur.[28]

Motion Blur
Is in images caused by movement of the camera or the objects. It is an average of neighbouring pixels in the direction of the movement (Fig. 3.2), for example for horizontal motion as:

$$h(i,j) = \frac{1}{L}, if(-\frac{L}{2}) < i < (\frac{L}{2}) \tag{3.2}$$

$$h(i) = 0, otherwise \tag{3.3}$$

where L is the length of motion. [28]

Out-of-Focus Blur

A simple de-focusing blur (Fig. 3.3) is defined as:

$$h(i) = \frac{1}{\pi R^2}, if \sqrt{i^2 + j^2} < R \tag{3.4}$$

$$h(i) = 0, otherwise \tag{3.5}$$

It is find in a lot of imagining systems and it is a uniform intensity distribution within a circular disc. [28]

Long Exposure Blur in Retinal Images

Besides usual commonly recognised types of image blur arising in natural photos, images of the retinal sequences have also other disadvantages. Due to long exposure the process of sensing the retinal sequences the eye often moves and the movement results in soft duplication of a part of the image around its original area, as seen in Fig. 3.4. [28], [29]

$$y = \sum_{i=1}^{i} \sum_{j=1}^{j} s.x(i,j) + (1-s)x(i-m, j-n) \tag{3.6}$$

where y is the blurred image, x is the original image, m and n is the constant offset of the original image causing the blur, i and j is the position, and s is the parameter determining the intensity in the original and moved position too.

3.2 Blur Identification Algorithms

The published methods are also based on either hand-crafted features or on the concept of deep learning networks, similarly to IQA methods.

Most of the features in the available works are focused on edge manipulation, as presence of edges seems to be a very good indicator of image blur, because edges are in general poorly visible in blurred images. In Pires et al. [30] is outside of area descriptor proposed also visual dictionary descriptor. The image is characterized by so-called points of interest, which capture inconsistencies or discontinuities. Random points of interest found by feature detector had been selected from the ideal image group and from the blurred image group to represent them. The interest points in the examined image are mapped to the representative interest points. Three other blur descriptors are introduced to improve the no reference method - blurring, sharpening,

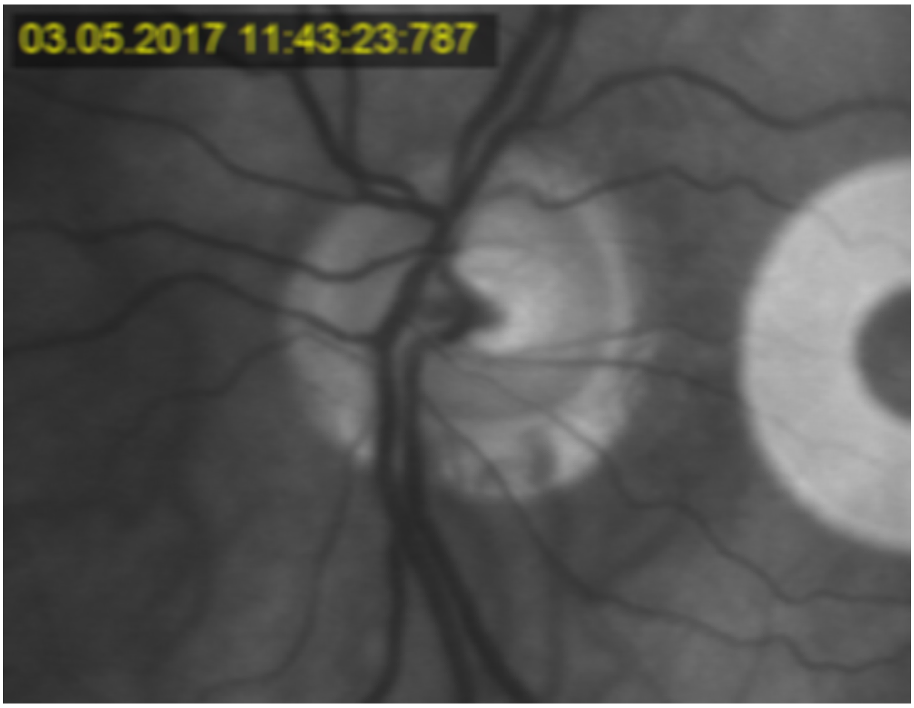


Fig. 3.1: An example of the Gaussian blur affecting a retinal image.

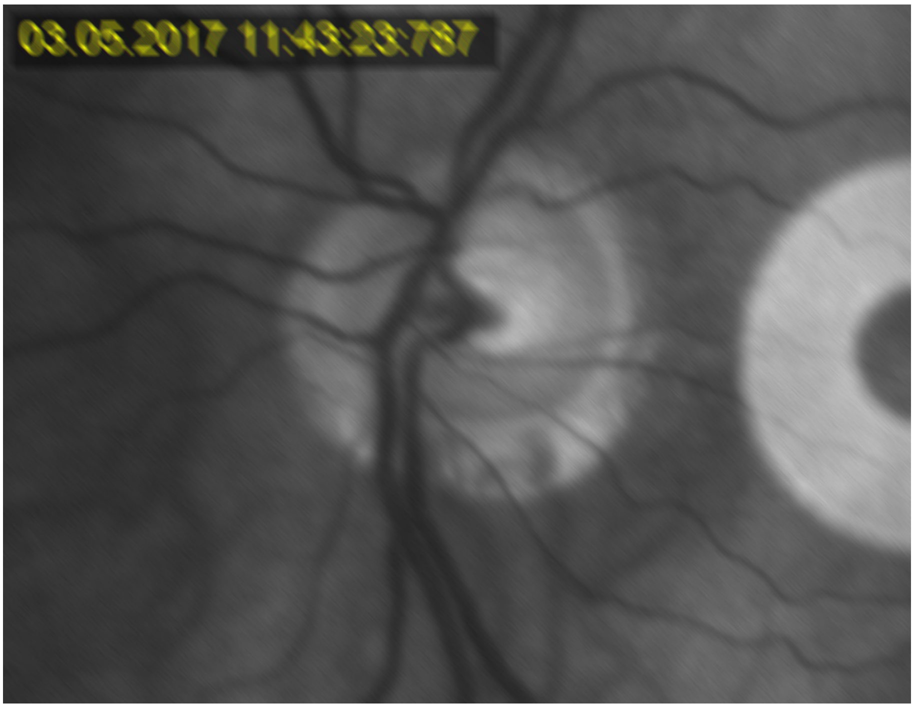


Fig. 3.2: An example of the motion blur affecting a retinal image.

and blurring + sharpening, based on the assumption that blurring will not worsen an already blurred image, sharpening will not enhance an already sharp image. The fusion and final classification was performed by SVM or concatenation. The datasets

Fig. 3.3: An example of the out-of-focus blur affecting a retinal image.

Fig. 3.4: An example of the long exposure blur affecting a retinal image.

used in the methods are DR1 and DR2 annotated by medical specialists. The area under the curve is in 95.5%. [30]

Wang et al. [14] decided to focus on the gradient of and image and evaluate the data using an extreme learning machine. By the use of Prewitt filters a gradient distribution is estimated, which is used to gain a feature vector, which serves as an input to the extreme learning machine, producing the final score. The gradient distribution is modeled based on the computation of the gradient magnitude. The extreme machine learning is a learning algorithm for a single layer feed forward network. Several datasets were used, including TID2008, CSIQ, LIVE, and IVC. Three performace criteria of the regression are evaluated. Test reached on the TID2008 datast results SROCC 0.9272, PLCC 0.9236, RMSE 0.4525, on the CSIQ dataset SROCC 0.9608, PLCC 0.9611, RMSE 0.0789, on the LIVE dataset SROCC 0.9672, PLCC 0.9660, RMSE 4.9393, and on the IVC dataset SROCC 0.9392, PLCC 0.9631, and RMSE 0.2609.[14]

Another work given to the study of image blur identification on retinal images is proposed by Chernomorets et al. [31], requiring a preprocessing in form of a rough vessel segmentation. After the vessel segmentation step, they were skeletonized and assorted. From the remaining skeletons was created a vessel profile, which serves to estimate the edge profile. The edge profile is further normalized by the edge amplitude in order to obtain adequate results. The edge width was obtained as a width of normalized edge and the results were scaled again by inverse original edge amplitudes. As a final parameter was selected the median value of the weighted edge widths. The employed database was the DRIVE database and no performance measure was provided. [31]

Deep Learning Based Algorithms

Numerous techniques employing the deep learning networks for image blur identification are available. Its goal is to improve the robustness of the algorithm and also try to catch similar features as the human eye compared to the standard machine learning based algorithms.

To identify the image blur, for instance local blur, Gaussian blur, haze blur, out-of-focus blur, are often used CNNs, for example Kang et al. used image patches as input to a CNN. The image is divided into patches to use them as an input for the CNN and the final quality estimation is calculated as an average of the patch scores. The CNN consists of five layers, the first layer being a convolutional layer with 50 kernels using ReLU, followed by a pooling layer reducing each of 50 feature maps of size 26x26 to only the maximum and the minimum. Two fully connected layers follow the pooling layer, each has 800 nodes. The last layer is a simple linear regression resulting in the score. The method used LIVE and TID2008 datasets. Their research concludes that the number of kernels leads to better performance, but the kernel size does not seem to affect the performance. The patch size also seems to be an important parameter, as according to the results it seems that bigger patch size leads to better results. The test database was the TID2008 dataset and the performance criteria were SROCC 0.920, and PLCC 0.903. [32]

Inspired by performance of Alexnet in other image processing tasks Wang et al. [33] decided to employ Simplified-Fast-Alexnet to classify four blur types - haze blur, Gaussian blur, defocus blur, and motion blur. The Simplified-Fast-Alexnet contains six hidden layers including five convolutional layers and one fully connected layer. The first two fully connected layers were removed along with the dropout, which makes the network prone to overfitting. Output of each convolutional layer is compressed by 0.5 to ensure simplified model and avoiding parameter redundancy. Batch normalization layers were added to replace the dropout method. The training dataset was The Oxford building dataset and Caltech 101 dataset, while the testing dataset was Berkeley dataset with accuracy 96.99%, and Pascal VOC 2007 dataset with accuracy 93.75 %.. [33]

Another method by Williams et al. proposes to use CNNs in order to distinguish between blur types. The employed CNN was a comprised of a modified version of LeNet CNN. In the beginning two convolutional layers using ReLU as the activation function to introduce nonlinearity. The convolutional layers extracting feature maps are followed by a max-pooling layer to reduce the dimensionality of the feature maps. To refrain from overfitting a dropout layer is used wit dropping probability of 0.25. The first dropout layer is again followed by two convolutional layers, max-pooling layer and another dropout layer with probability of 0.25. This is followed by a fully connected layer, again followed by another droupout layer with dropout probability 0.5 and the final layer is a second fully connected layer. The final activation function is a softmax function, classifying the input image into labels based on the training. The used dataset was in this case the Messidor dataset of 1200 eye colour fundus images and their own sharp data. The reached accuracy was 77 %. [34]

Zhang et al. proposes their own unified ABC-FuseNet, a deep neural network combining the attention map (A), blur map (B), and content feature map (C). The blur map estimation is built on top of Inception-V2. They removed the downsampling operation and replaced the regular convolution with dilated convolutions in order to preserve the resolution of feature map to detect blur even in small regions. The high level features were combined with the low-level features of the first layer, again to keep the resolution. A pyramide pooling module, combining the local and global clues, was adapted to detect blur also in homogeneous regions or multiple scale objects. They further classify the blur desirability using the ResNet-50 to extract semantic feature map and create a attention map to find important regions in the image. All these learn maps are fused and fed to a light classifier to estimate the category. For blur map estimation they applied sigmoid function on the last layer and calculated the L2 loss between the estimated blur map and the ground truth blur map. They used SmartBlur dataset, to train the network, and a public blur detection dataset CUHK to test the network, with test accuracy of classification 81.4 %. [35]

Shaode et al. uses a shallow CNN layer that consists of one feature layer made up of convolutional filtering and average pooling, retrieving a representation of sharpness, that is an input to a multilayer perceptron, general regression neural network, and support vector regression. The input of the CNN is grey-scale image with local contrast normalization, in order to remove local mean displacements and to normalize the local variance. According to the results the best performing network is support vector regression. It is clear that CNN in combination with other networks provides better results. As the baseline database was used LIVE-II database. CNN only reached overall SROCC 0.8852 and KROCC 0.9184, CNN+GRNN reached overall SROCC 0.9283 and KROCC 0.9377, and CNN+SVR reached overall SROCC 0.9310 and KROCC 0.9435. [36]

4 Proposed Method for Image Quality Assessment

According to the summary of the techniques used in the field of IQA and image blur identification, it is clear that the methods based on deep learning lead to the most robust results applicable for large datasets. This is an interest to this master's thesis too, therefore the further image quality assessment is with emphasis on pre-trained CNN architectures such as Alexnet that has been previously used in a simplified version to classify blur types in images in Wang et al. [33], LeNet that has been used to classify image blur in retinal images in Williams et al. [34], Inception-V2 that was modified to estimate a blur map in Zhang et al. [35], aiding in blur detection. Another option to detect blur is to use a shallow CNN, which also seems to perform quite well, as the first convolutional layers extract mainly features regarding edges and edges seem to be an important feature in respect to image blur according to previous research. This chapter is focused on identifying and evaluating the image blur with use of deep learning and the proposed methods are implemented in programming language Python using open source libraries Keras, Tensorflow and PyTorch.

4.1 Dataset

As it is apparent from the previous chapter addressing the problems of IQA and image blur identification, to use a method based on deep learning it is necessary to provide the network with large dataset of images.

A representative dataset including the types of blur in retinal sequences mentioned in the chapter 2 was prepared. Individual retinal images were sampled from the retinal sequences, each with time duration 10 s and frequency 25 fps while the sharpest images subjectively chosen from the sequence were used as the base image. The resolution of each image is 1000x770 pixels and the images were cropped to size 600x470 pixels in order to exclude the time tag and light artifacts on the image edges. A python algorithm modeling three types of image blur (motion, out-of-focus, and long exposure) was applied to each previously chosen representative sharp image.

The motion blur was modeled in four different directions - horizontal, vertical, diagonal from the left bottom corner to the right top corner, and diagonal from the left top corner to the right bottom corner of the image. The size of the kernel was randomly chosen between 1 and 30 pixels, and the direction was randomly chosen too. To blur the image with the randomly created kernel was used a function *filter2D* from the *cv2* library. An example of the motion blur image is in the Fig. 4.1.

The out-of-focus blur also had several randomly chosen parameters. First of all two different approaches were implemented in modelling of the out-of-focus blur. Firstly it was modelled as Gaussian blur, with random parameters being kernel size (height and width) generated between 1 a 30 and σx and σy generated between 1 and 20, an example of the blurred image is in the Fig. 4.2. Few conditions ensuring that kernel height and width are positive and even numbers are provided too. A function *GaussianBlur* from the *cv2* library was used to blur the image with the randomly created kernel and sigma values.

Second approach was to use a kernel of random size again (kernel height and width between 1 and 30), but this time the kernel contained a circle to model the blur according to equation (4) and (5). To convolve the kernel with the original image a function *convolve2d* from the *cv2* library and an example is in the Fig. 4.3.

While modelling the long exposure blur several randomly chosen parameters were used - number of offset rows and columns (between 1 and 30) and the direction - up, down, left, and right, and also random value of alpha (between 0.1-0.35). A copy of the original image was created and the copy was cropped by a number of rows and columns, according to the randomly chosen offset. The cropped image was then blended with a function *blend* from the *PIL* library with an α parameter randomly chosen between 0.1 and 0.35 - giving alpha values $1 - a$ to the whole original image and alpha values a to the cropped image blended on top of the original image.

Randomly generated values of the type of blur were saved as one number in range 0-100, 100 for the biggest values (extremely blurred images) and 0 for the smallest value (sharp images) along with the blurred image, in order to serve as the ground truth for the images in the further IQA.

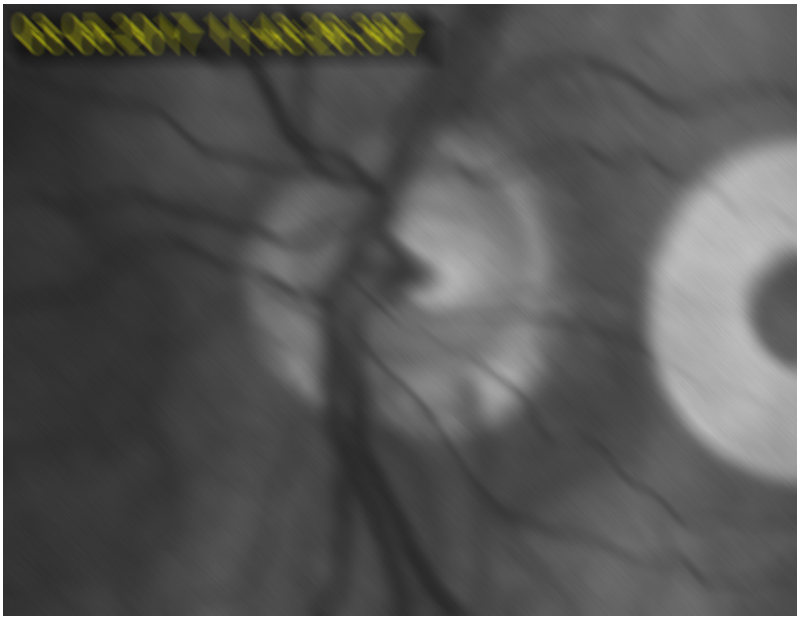


Fig. 4.1: An example of modeled motion blur, the kernel size 27, and the direction of the movement diagonal from left top corner to right corner.

Fig. 4.2: An example of modeled Gaussian blur, kernel width was 27 and kernel height was 12, σx was 5 and σy was 1.

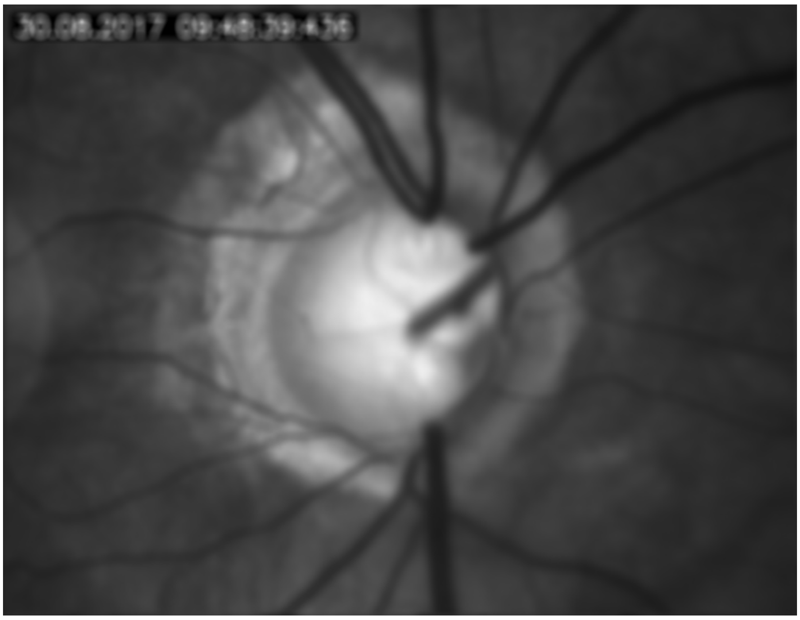


Fig. 4.3: An example of modeled out-of-focus blur, kernel height was 9 and kernel width was 13.

Fig. 4.4: An example of modeled long exposure blur, number of offset rows was 19 and columns 2, the direction of the movement was towards the upper left corner, and the alpha value was approximately 0.2979.

4.2 Direct Blur Identification and Evaluation

Initially the intention of this master's thesis was to provide a solution to identify and assess the extent of the blur distortion in retinal images. As the previously mentioned types of blur, characteristic of retinal images, usually occur jointly in a distorted frame, the ideal solution would be to find a method that would be able to evaluate them together.

The dataset for this task consisted of 300 images, each image distorted with every type of blur but to a different extent. The dataset was prepared according to the method mentioned in the section 4.1.

Based on the previous research of blur identification algorithms in chapter 3.2 few networks were considered for this task beside simple CNN layer, one of them being Inception-v2 or ResNet50. According to the benefits of these networks for this task network Inception-ResNet-v2 was taken into consideration to directly identify and evaluate blur in images.

4.2.1 Inception-ResNet-v2

Inception-ResNet-v2 is a result of a combination of two ideas - residual connections and Inception architecture. The intention was to replace filter connections of Inception architecture with residual connections, which seem to be necessary for training very deep architectures. The resulting network is therefore computationally efficient while being able to use benefits of residual learning such as improvement of the training speed. Inception-ResNet-v2 is a hybrid Inception version with improved recognition performance. Full Inception-ResNet-v2 architecture in simplified view is shown in Fig. 4.5. [37]

Loss Function

The employed loss function was mean absolute error (MAE) and the aim was to reduce this loss to minimum, such that the predicted values are as close to the ground truth values as possible. The loss function is the mean of the absolute error of the given batch, according to the formula:

$$MAE = \frac{\sum\limits_{i=1}^{n} |y_i - x_i|}{n}, \tag{4.1}$$

where y_i is the prediction an x_i is the label. [38]

Optimizer

The optimizer used in the implementation of Inception-ResNet-v2 was Adam. Adam is an algorithm for gradient-based optimization of stochastic optimization. The main idea is adaptive moment estimation - this method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients. It is supposed to combine AdaGrad, and RMSProp methods, combining their benefits, such as AdaGrad's ability to deal with sparse

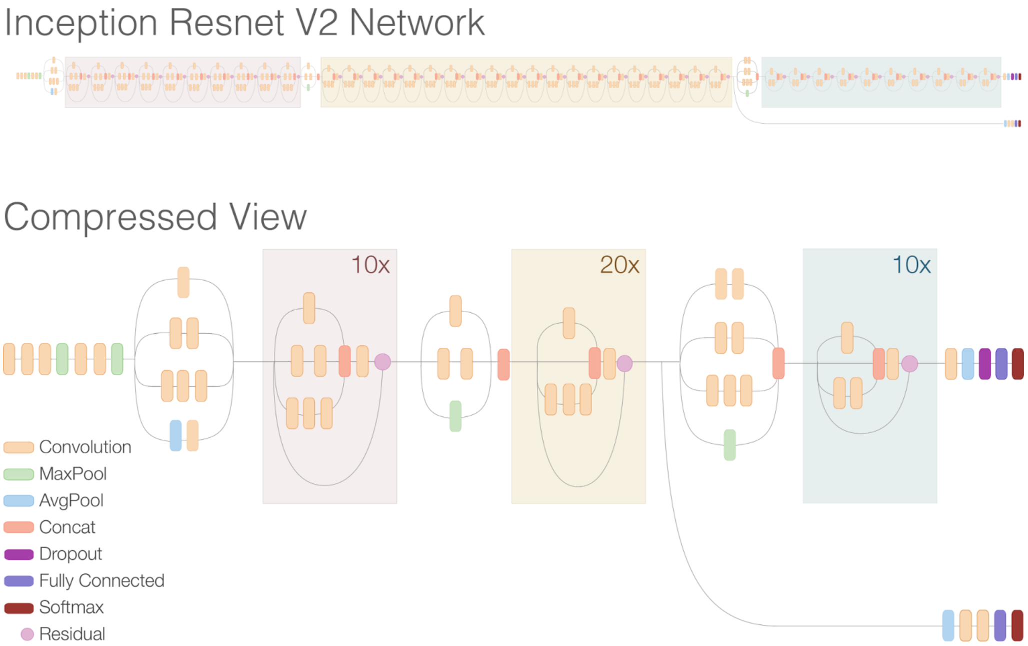


Fig. 4.5: A simplified compressed view of Inception-ResNet-v2 network with the full view on top of the figure. [39]

gradients and RMSProp's ability to deal with non-stationary objectives. This optimizer is particularly aimed at machine learning problems with large datasets, and/or high-dimensional spaces. [38],[40]

Algorithm of the Adam method is first of all based on the computation of gradient, i.e. the vector of partial derivatives of the objective function in a certain timestamp. The algorithm updates moving averages of the gradient and the squared gradient. The moving averages are estimates of the 1st moment (the mean) and the 2nd moment (the uncentered variance) of the gradient. The exponential decay of these moving averages is controlled by the hyper-parameters β_1, β_2 within the interval [0,1). The moments are described in equation 4.2 and 4.3 and the final update is described in equation 4.4. [38],[40]

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \tag{4.2}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t}, \tag{4.3}$$

where m_t is the 1st moment, v_t is the 2nd moment and β_1^t β_2^t are previously described hyper-parameters.

$$\theta_t = \theta_{t-1} - \frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\epsilon}, \tag{4.4}$$

where θ_t is the final parameter, computed from θ_{t-1}, which is the previous update, $\hat{m_t}$ and $\hat{v_t}$ are the 1st and 2nd moment, and ϵ is the numerical stability constant.

Training

Keras model of Inception-ResNet-v2 network was used. As the blur estimation is a regression problem, the last layer was modified in order to use the network for regression. In this case one fully connected layer with 1024 units and another fully connected layer with 3 units were added on top of Inception-ResNet-v2 model. Both layers were with reLu activation, as the intention was to evaluate three different types of blur continuously.

The input images were preprocessed according to the requirements, all the input images had zero mean, unit standard deviation and were scaled between -1 and 1. The output values should describe the blur distortion of the certain image. Smaller values suggest little or no blur while bigger values suggest considerable distortion of the image.

The network was initially trained for 25 epochs with Adam optimizer. The learning rate was set at 0.0001, while β_1 was 0.9, β_2 was 0.99 and ϵ was $1 \times 10^{(-8)}$ and no pre-trained weights were used. Unfortunately, assumably due to insufficiently large dataset the trained network predicted only constant value without regard to the input image.

Based on this understanding weights pre-trained on imagenet dataset were used as initial weights, with expectations that it might lead to better results. This network was trained again with learning rate 0.0001, β_1 0.9, β_2 0.99, and ϵ $1 \times 10^{(-8)}$. As expected, network trained in this manner performed better. Unfortunately, as seen in Fig. 4.6, the network started to over-fit quite soon, already after 9th epoch at the validation loss value 18.05, while the training loss kept lowering all the way 7.97, while also starting to stagnate after 22nd epoch.

Thus, another approach of training the Inception-ResNet-v2 with pre-trained weights was covered. First of all only the top layers were trained for 3 epochs, while the other layers were frozen. After 3 epochs all layers except the first two were unfrozen again and the network was subsequently trained for another 25 epochs. The parameters of the optimizer remained as in the previous examples. This approach does not seem fortunate as even the training loss keeps lowering quite slowly. As seen in Fig. 4.7 the validation loss seems to jump around, occasionally reaching the minimal loss of 15.54 at the 9th epoch, also indicating over-fitting as the network is apparently sensitive to changes in the data.

When network seems to be over-fitting it is favorable to introduce regularization, in this case for example in the form of dropout or early stopping. Even in the previous examples a from of early stopping was introduced, as two versions of

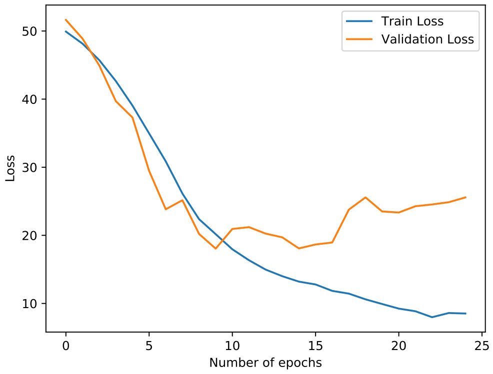


Fig. 4.6: Training and validation loss of Inception-ResNet-v2 without adjusted parameters with pre-trained weights.

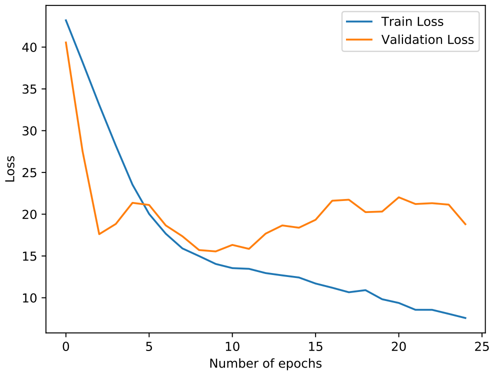


Fig. 4.7: Training and validation loss of Inception-ResNet-v2 with frozen layers.

networks were saved during the training - one at the end of the training, and the other one was saved only if the validation loss was improved. The last modification of this network was to add another dropout layer right before the output layer. The dropout layer randomly throws away part of the extracted features, in this case 0.3. This network was trained for 45 epoch all together, as it was expected that both the validation an training loss will lower more slowly. As it is clear from Fig. 4.8, even the addition of dropout layer did not eliminate the over-fitting problem.

4.2.2 Other proposed methods

Along with the Inception-ResNet-v2 other networks were trained too, for example VGG16, and a modification of VGG16. However, these networks were severely under-fitted to the given problem, plausibly due to large number of parameters. While Inception-ResNet-v2 has 55,873,736 parameters, VGG16 has 138,357,544 pa-

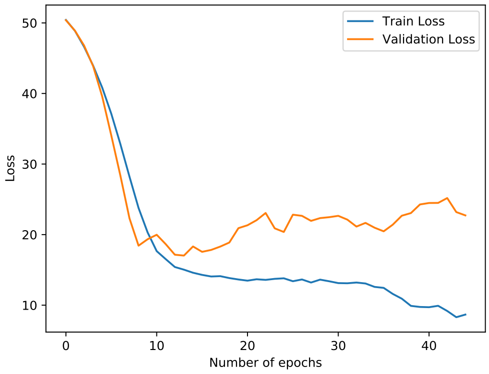


Fig. 4.8: Training and validation loss of Inception-ResNet-v2 with additional dropout layer and ϵ 1x10^{-4}.

rameters, which might be the reason why the VGG16 network did not show almost any signs of learning even after 200 epochs, and both the training and the validation loss changed only moderately. [38]

4.3 Image Blur Identification Followed by Blur Evaluation

In reference to the previous chapter, the proposed method of image blur evaluation was modified in the following manner. First of all the image must be subjected to analysis of the presence of image blur and identification of its type, and based on the results the image is directed to network estimating the distortion by the identified type of the image blur.

To identify the type of blur a Fourier spectrum of the image is computed by Fourier transform, adjusted to normalized logarithmic spectrum in the next step, according to the following equation:

$$log(|G(\mathbf{u})|) = \frac{log(|G(\mathbf{u})|) - log(|G_{min}|)}{log(|G_{max}| - log(|G_{min}|}, \tag{4.5}$$

where $G(\mathbf{u})$ represents the spectrum, $G_{min} = min_u(G(\mathbf{u}))$ is the minimum of the spectrum , $G_{max} = max_u(G(\mathbf{u}))$ is the maximum of the spectrum. The patterns observed as e.g. in Fig. 4.9-4.10 intuitively represent the blur. To simplify the classification, the spectrum was filtered by Laplace operator, using function *filters.laplace* from *skimage* library. Filtered spectrum was then flattened into a vector that serves as an input into classification network. This network consists of three layers with 10, 7, and 3 units, ReLu activation function between them and a softmax function in the end. The optimizer used in the training is Adam with leaning rate 0.001, loss function is categorical crossentropy which excepts the labels one-hot encoded and computes the crossentropy between the given classes. The model was trained for 20 epochs but only the model with best validation loss was saved. This process is shown in Fig. 4.11. [38] [41]

Fig. 4.9: Image distorted by motion blur.

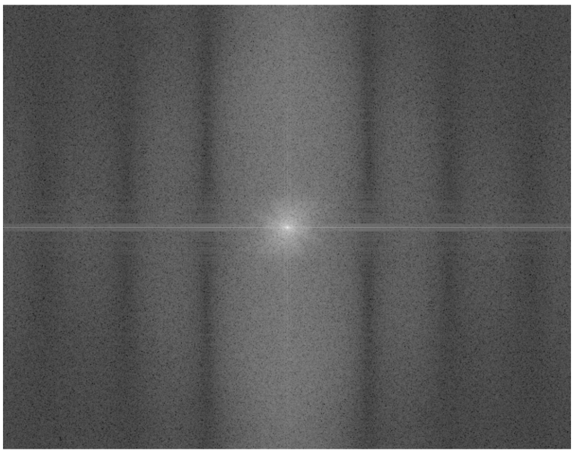


Fig. 4.10: Normalized image spectrum.

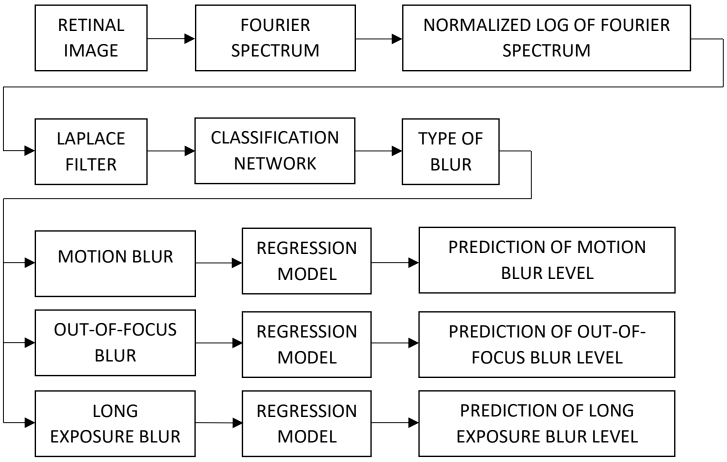


Fig. 4.11: Block diagram of the process of image blur identification and evaluation.

4.3.1 InceptionResNetV2

Similarly to section 4.2.1 a Inception-ResNet-v2 was chosen for the task of evaluating the image blur distortion. Three separate networks are trained in this case, each for the certain type of blur, and the network is based on the results of the training obtained in the section 4.2.1. As the network with pre-trained imagenet weights, Adam optimizer with learning rate 0.0001, β_1 0.9, β_2 0.99, and ϵ $1x10^{(-8)}$ outperformed the other networks trained with different parameters, these were used also for the training of the separate networks.

Fig. 4.12-4.14 suggest, that removal of the blur identification problem resulted into only slightly better training behaviour of the network, but over-fitting in all three networks is still very much present. It is obvious that these networks are not well suited for the evaluation of image blur.

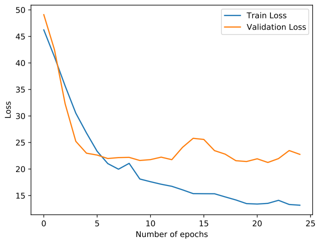


Fig. 4.12: Training of Inception-ResNet-v2 network for motion blur estimation.

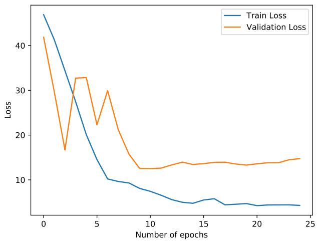


Fig. 4.13: Training of Inception-ResNet-v2 network for out-of-focus blur estimation.

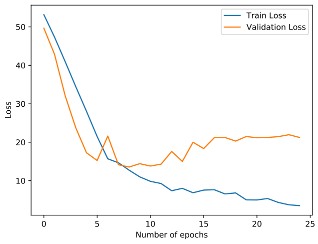


Fig. 4.14: Training of Inception-ResNet-v2 network for long exposure blur estimation.

4.3.2 ResNet50 Feature Extraction

As it seems that the amount of data is not sufficient to provide satisfactory results, training deep neural network is not the optimal method of evaluating the image blur. On the other hand deep neural networks offer a lot of high-level semantic features that might be used in further analysis. Based on an idea from Li et al. [42] were these features extracted using ResNet50 network. [42]

ResNet50 is a 50 layer deep network using residual connections, making the network easier to optimize and offering better accuracy. Depth of the network is a very important parameter affecting the overall performance of the network but it might be the cause to other issues, such as vanishing/exploding gradient. Residual learning, as shown in Fig. 4.15 is supposed to eliminate another problem known as degradation of training accuracy, meaning that deeper neural networks lead to greater training error than shallow neural networks trained for the same problem. [42], [43]

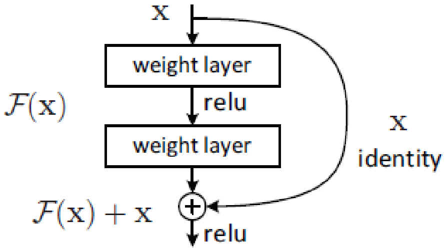


Fig. 4.15: Building block of residual learning. [43]

According to [42], residual learning is the reason why features extracted from last layer of ResNet50 offer better results compared to e.g. AlexNet or GoogleNet. Based on this apprehension, the next method is based on the feature extraction using ResNet50.

Every image was normalized by mean substraction and adjusted into range 0 to 1 in the first step and divided into patches of the input size of ResNet50 pre-trained on imagenet dataset. The used pre-trained ResNet50 is available in PyTorch open source library. The size of the patches was 224x224 pixels, with stride of 112 pixels. These patches are then advanced to the ResNet50 model to extract the features. Features of all patches from one image are extracted from last layer and aggregated into three vectors representing the image - mean and std aggregated, moment aggregated, and quantile aggregated. Vector X1 is is a vector of 4096 elements, 2048 elements resulting from mean of the columns of the output, and the other 2048 elements resulting from std of the columns of the output. Other vectors are computed in similar manner. X2 feature vector is consisting of three different aggregated moments - mean of the original network output, 3rd root of mean of the output that was raised to the 3rd power, and 4th root of mean of the output that was raised to the 4th power. The last feature vector X3 is consisting of 5 quantiles - 0th, 1st, 2nd, 3rd, and 4th. Three feature vectors X1, X2, and X3 are the output of this algorithm. These features are then advanced into another model, each feature to a separate model. The final prediction are acquired by averaging the separate predictions. [42]

Two different types of regression models were tested, general regression neural networks (GRNN) and partial least sqaures regression (PLSR).

GRNN

GRNN is a memory base neural network that is able to provide continuous estimates. It is a one-pass learning algorithm based on statistical principles, with no need of back-propagation. The network can be used for a regression problem, even if linearity is not justified. The output of GRNN is computed as:

$$\hat{Y}(X) = \frac{\sum\limits_{i=1}^{n} Y_i exp(-D_i^2/2\sigma^2)}{\sum\limits_{i=1}^{n} exp(-D_i^2/2\sigma^2)}, \tag{4.6}$$

where n is the number of sample observations, $D_i^2 = (X - X_i)^T(X - X_i)$, Y_i are sample values, and σ is the spread parameter. The spread parameter σ determines the smoothness of the approximation, ideally it should be smaller than the average distance between input vectors. [44], [45]

The GRNN algorithm was implemented using Python library *Neupy* and its function *GRNN*, where the only parameter to determine is σ, which was set to 1.8 for all three networks.

The separate predictions are averaged afterwards, resulting into the final prediction. The final predictions of motion blur, out-of-focus blur, and long exposure blur level on test data are shown in Fig. 4.16-4.18. It is clear that the most accurate prediction is motion blur level prediction, while prediction of long exposure blur level is the worst one.

PLSR

When the number of predictors is much higher than the number of observations, it is usually necessary to modify the input vector. One of the possibilities is to eliminate some predictors, or perform principal component analysis (PCA), acquiring principal components serving as input into the regression method. This method has problem of choosing the optimum subset of predictors, which can be solved by choosing number of first principal components. The disadvantage is that these components rather describe the input, not the output. Contrarily, the PLSR aims to select set of components that explain as much as possible of the covariance between input and output. This step is followed by regression, where the decomposition of input is used to predict output. [46]

This method was implemented in Python using function *PLSRegression* from library *sklearn*. The number of components to keep was set to 15 after considering other numers too.

The final predictions were acquired similarly to GRNN method. Three separate predictions were averaged into one final prediction belonging to one type of blur. The final predictions for motion blur, out-of-focus blur, and long exposure blur on test data are shown in Fig. 4.19-4.21. It is clear that the PLSR predictions are more accurate than the GRNN predictions. The most accurate is the predictions of motion blur level and the least accurate is the prediction of long exposure blur level, similarly to GRNN predictions.

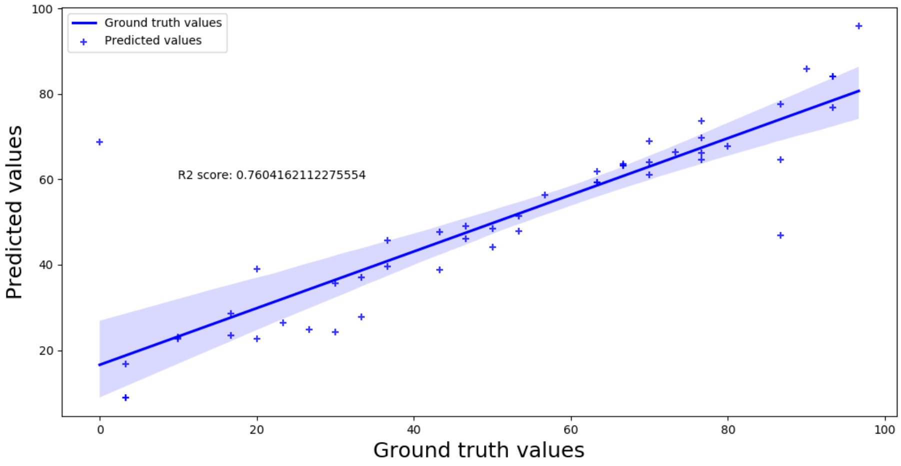


Fig. 4.16: Prediction of motion blur extent by GRNN method.

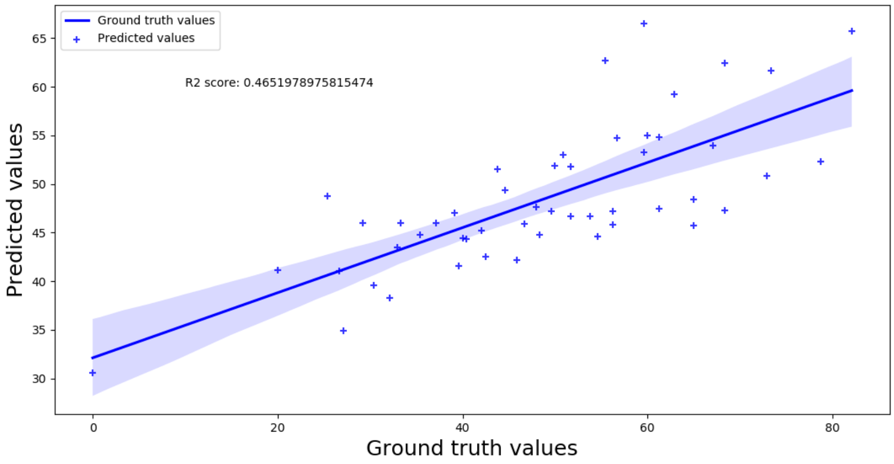


Fig. 4.17: Prediction of out-of-focus blur extent by GRNN method.

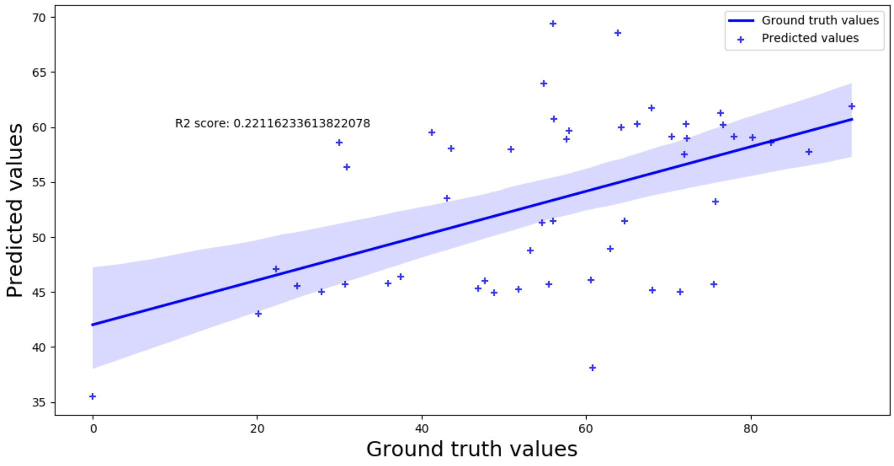


Fig. 4.18: Prediction of long exposure blur extent by GRNN method.

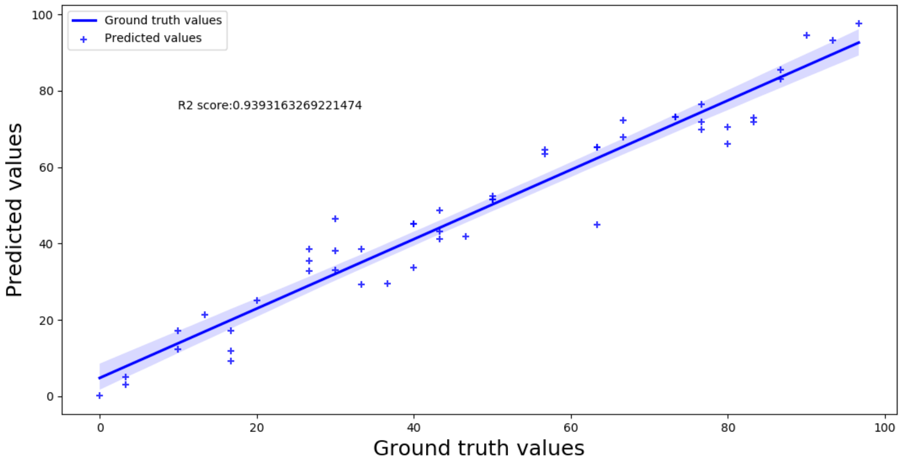


Fig. 4.19: Prediction of motion blur extent by PLSR method.

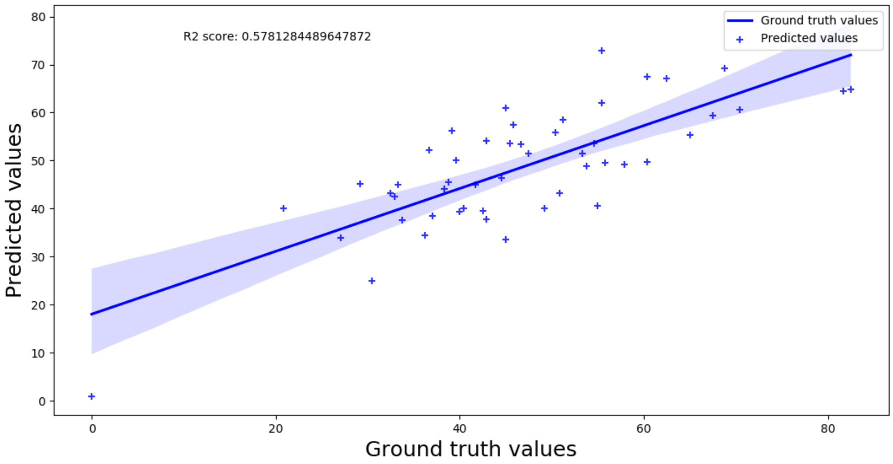


Fig. 4.20: Prediction of out-of-focus blur extent by PLSR method.

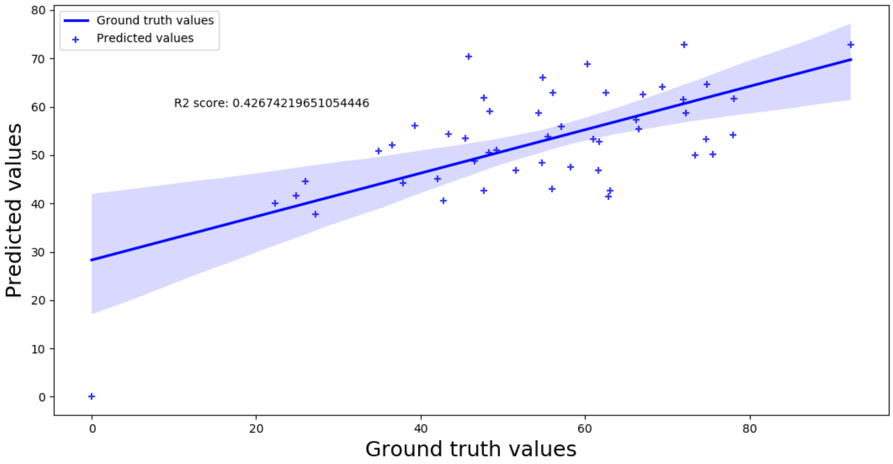


Fig. 4.21: Prediction of long exposure blur extent by PLSR method.

5 Evaluation of Retinal Image Sequences

One task of this master's thesis is to evaluate the quality of the whole retinal image sequence and also to evaluate individual images from the given retinal sequence. In the previous chapter, the implemented methods were firstly tested on simulated blurred images and on blurred retinal images. This section provides the evaluation on real retinal sequences.

5.1 Performance Criteria

To evaluate the performance of the blur classification model two criteria were used. The first is accuracy, computed according to the equation (5.1). The other metric is confusion matrix, which provides general overview of results. The columns of this matrix represent the actual class and the rows usually represent the predicted class. The confusion matrix was in this case computed using *confusion_matrix* from *sklearn* library and accuracy was computed using function *accuracy score* from the same library.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \tag{5.1}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. [47]

The performance of trained regression models for each type of blur is evaluated by SROCC, PLCC and root mean square error RMSE. These coefficients are supposed to describe the strength of association between predicted and measured values.[47]

SROCC is a nonparametric measure, evaluating the monocity of the relation between the measured and the predicted dataset. It does not expect that the datasets are normally distributed. SROCC values lie between -1 and 1, 1 means exactly linear relation and both measured and predicted values, both increasing, while -1 means exactly opposite relation. SROCC value was in this case computed with function *stats.spearmanr* from Python library *Scipy.* [47]

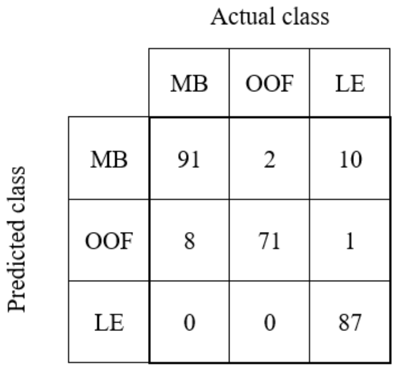
Another computed measure is PLCC. It measures the linear relationship between two datasets, and unlike SROCC it does expect normally distributed datasets. The values again extend in range from -1 to 1, 0 implying no correlation between two datasets and equivalently as in SROCC, 1 and -1 mean linear relation between two datasets. The PLCC values computed using function *stats.pearsonr* from Python library *Scipy.* [47]

The last metric is the R2 score, or coefficient of determination, is the measure of regression models. 1 is the best possible score, 0 means that the prediction is constant, regardless of the input. It can also reach negative values. The R2 score was computed using the function *r2_score* from *sklearn* library. [48]

5.2 IQA of Retinal Images

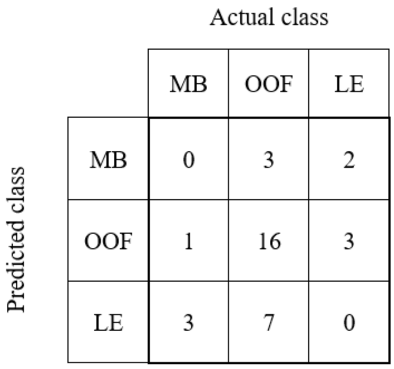
Image Blur Identification

The blur classification network was trained on 630 simulated blurred retinal images, originated from three sharp retinal images. All types of blur in the train dataset were approximately equally represented. The network was afterwards tested firstly on 270 simulated blurred images and also 35 blurred images from 10 retinal sequences. Real blurred images were selected and labeled subjectively as motion blurred, out-of-focus blurred, or long exposure blurred. Considering the bad quality of retinal sequences the subjective selection of blurred images was a problematic task. The accuracy of the classification of simulated blurred images was 0.92, while the accuracy of blurred images from retinal sequences was only 0.46. The classification results from both tests are also represented in confusion matrices in Fig. 5.1.



		Actual class		
		MB	OOF	LE
Predicted class	MB	91	2	10
	OOF	8	71	1
	LE	0	0	87

Fig. 5.1: Confusion matrix of classification of simulated test blurred images (MB = motion blur, OOF = out-of-focus blur, LE= long exposure blur).



		Actual class		
		MB	OOF	LE
Predicted class	MB	0	3	2
	OOF	1	16	3
	LE	3	7	0

Fig. 5.2: Confusion matrix of classification of real retinal images (MB = motion blur, OOF = out-of-focus blur, LE= long exposure blur).

The reason for the substantial difference between the results of simulated blurred images and blurred images from retinal sequences is caused by the generally bad quality of retinal sequences. Most of the sequences are considerably distorted by combination of different types of blur and it is difficult to subjectively assess which frames are blurred and what type of blur is the prevalent one. As seen in Fig. 5.1 motion blur was mostly confused with long exposure blur and the other way around, what is understandable as they often appear quite similar in real retinal images.

Image Blur Evaluation

As no exact labels describing the amount of blur distortion are available for blurred images from retinal sequences, the performance of PLSR and GRNN models is further evaluated only on simulated blurred data. The results are summarized in Tab. 5.1-5.3.

Tab. 5.1: Comparison of performance of GRNN and PLSR models evaluating motion blur in images.

	SROCC	**PLCC**	**RMSE**	**R2 SCORE**
GRNN	0.87	0.88	13.84	0.76
PLSR	0.96	0.97	6.61	0.94

Tab. 5.2: Comparison of performance of GRNN and PLSR models evaluating out-of-focus blur in images.

	SROCC	**PLCC**	**RMSE**	**R2 SCORE**
GRNN	0.74	0.74	11.86	0.47
PLSR	0.72	0.77	9.57	0.58

Tab. 5.3: Comparison of performance of GRNN and PLSR models evaluating long exposure blur in images.

	SROCC	**PLCC**	**RMSE**	**R2 SCORE**
GRNN	0.45	0.49	16.93	0.22
PLSR	0.50	0.66	12.96	0.43

Tab. 5.1-5.3 interpret that overall better performance has the PLSR model. Even though GRNN also offers decent results, PLSR outperforms this method, perhaps because PLSR selects a smaller number of significant features, rather than only memorizing the test dataset. This might be the reason why PLSR reacts better to differently blurred test images, specifically their features.

5.3 IQA of Retinal Sequences

Dataset

The dataset consists of 15 retinal image sequences captured by experimental VO described in section 1.4, with length 10 s and frequency 25 fps. All the images were 1000x770 pixels in size, but they were cropped to size 600x470 pixels in order to exclude the time tag and light artefacts, in the same manner as the data described and employed in chapter 4. Each image was subjectively labeled as sharp/distorted (0/1), and if possible, the type of blur was estimated too. Examples of the images are in Fig 5.3-5.6.

Fig. 5.3: Example of sharp retinal image .

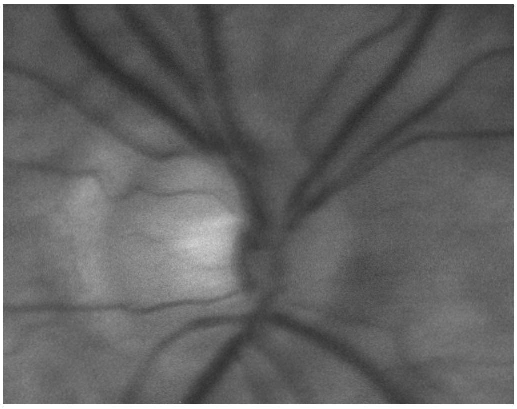


Fig. 5.4: Example of motion blurred image.

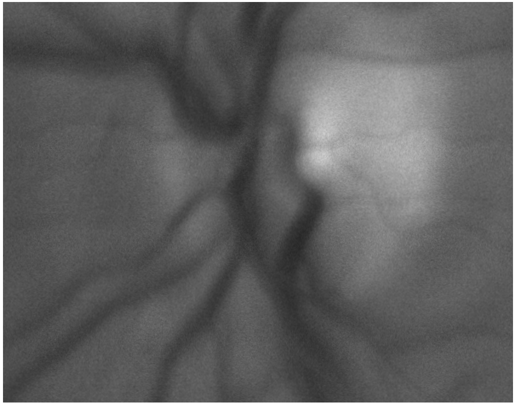


Fig. 5.5: Example of out-of-focus blurred image.

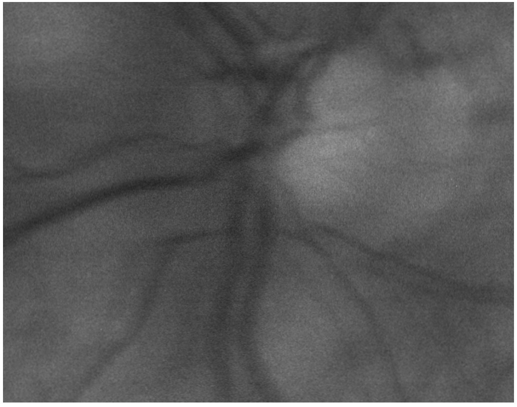


Fig. 5.6: Example of long exposure blurred image.

Detection of Blurred Images

The previously proposed methods all operated only with blurred images. Detection of blurred images is crucial to be able to asses the quality of retinal sequences. Therefore a train dataset consisting o 35 real blurred retinal images and 35 real sharp retinal images from 10 train videos, and a test dataset consisting of 26 retinal images from 5 test videos were prepared. These images were also divided into patches and features were extracted from the patches by ResNet50. These aggregated feautures, concatenated into one vector served as an input for the classification model.

A logistic regression model was trained on the train dataset and further tested on the test dataset. Logistic regression attempts to model posterior distribution $P(y|x)$, therefore providing not only the labels but also the probabilities of class membership. This model generally calculates the class membership probability for one of the two categories, and $P(0|x, \alpha) = 1 - P(1|x, \alpha)$. The equation states that the posterior probability is dependent on model parameters α. The decision boundary between two categories is formed by a hyperplane satisfying condition $\alpha.x = 0$. This method was chosen particularly because of the small size of dataset and easy implementation. [49]

The accuracy of logistic regression was only 0.46. On the other hand it is also necessary to notice the probabilities predicted for each test image as shown in Fig. 5.7-5.10. The images suggest that even though the accuracy of the classification is not very satisfactory, the model is able to detect more distorted images.

Proposed method of IQA on real retinal image sequences

The video quality assessment in this master's thesis is based on the previous methods built on simulated images. The algorithm is briefly described in the Fig. 5.11.

Firstly, the retinal sequence is loaded and divided into 250 respective frames. Each frame is divided into patches. Pre-trained ResNet50 extracts the features from these patches, and next, the features from all patches of each image are aggregated, resulting into three feature vectors X1, X2, and X3 as was already described in section 4.3.2. These feature vectors concatenated into one vector are advanced to blur detection model. The frames classified into class 1 (blurred images) are further subjected to classification of the image blur type. A spectrum is then computed using Fourier transform, and the logarithmic spectrum is normalized, and filtered by Laplace operator. The modified spectrum is flatten into a vector and advanced to a model trained on the simulated data. The position of maximum output value of this model is the present type of blur. Feature vectors X1, X2, and X3 are advanced to the corresponding regression models, which evaluate the blur level of the predicted type of blur. This predicted value is assigned to the analyzed frame into the field of

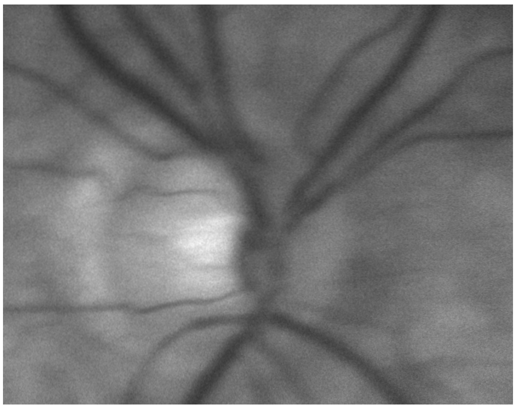


Fig. 5.7: Blurred image misclassified as sharp with probability of class 1 (blurred images) 0.49.

Fig. 5.8: Sharp image misclassified as blurred with probability of class 0 (sharp images) 0.20.

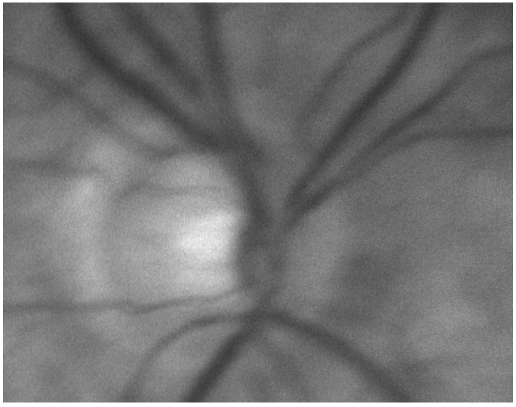


Fig. 5.9: Blurred image correctly classified as blurred with probability of class 1 (blurred images) 0.97.

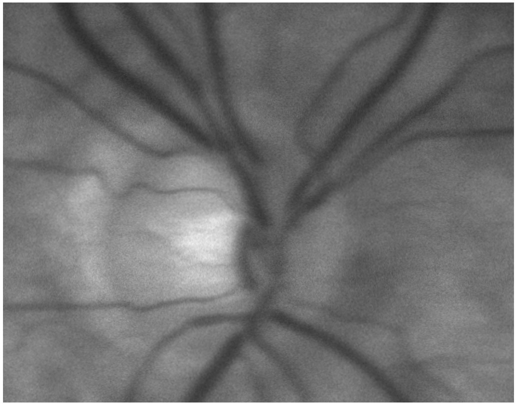


Fig. 5.10: Sharp image correctly classified as sharp with probability of class 0 (sharp images) 0.75.

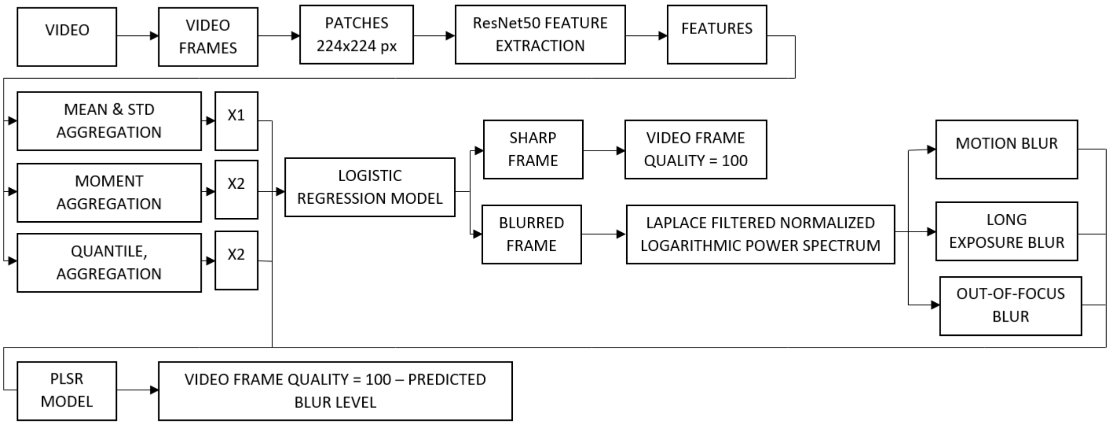


Fig. 5.11: Simplified block diagram of video quality assessment.

the analyzed type of blur. The other levels of other types of blur for this frame are set to 0. Each frame is therefore assigned three values, each representing blur level of certain type. The predicted values represent the blur level value of each frame. Video frames that were in the first step classified as sharp are assigned with blur level value 0 for all types of blur. In the end, only one value represents the quality of the frame and that is 100 minus the sum of the three predicted blur levels. If no blur was detected in the given frame, its quality is therefore set to 100. If some blur was detected, the blur levels are summed and the resulting quality of this frame is this sum subtracted from 100. For simplification purposes it can be also interpreted that the mean of the maximum blur level of all the frames is subtracted from 100 (100 is the highest possible predicted blur level value) and this value represents the overall quality of the video. The outputs of this algorithm is the overall quality of the video, percentage of the blurred frames, and a detailed table where each row represents one frame, and each column represents one out of the three types of blur, example in Tab 5.4. The values in the table are the predicted blur level values within the range 0-100.

Tab. 5.4: Example of output table of video quality assessment (frames 20-25, test video 1).

	MOTION PREDICTED BLUR LEVEL	**OUT-OF-FOCUS PREDICTE BLUR LEVEL**	**LONG EXPOSURE PREDICTED BLUR LEVEL**
FRAME 20	0	0	0
FRAME 21	0	36.91669	0
FRAME 22	0	0	0
FRAME 23	0	0	0
FRAME 24	0	0	65.52092
FRAME 25	0	0	0

The output quality of this quality assessment is shown in Tab. 5.5, along with the percentage of the detected blurred frames and the percentage of the frames labeled as blurred. The predicted quality is similar to the truth - the better is the overall quality of video, the higher is the output quality value and the other way around. The percentage of detected blurred frames is much higher than in labeled data but it must be mentioned that the boundary between blurred and sharp image is in this case very difficult to assess, as even the images labeled as sharp can be already perceived as blurred.

Tab. 5.5: Results of IQA of real retinal sequences.

	QUALITY	**% OF DETECTED BLURRED FRAMES**	**% OF LABELED BLURRED FRAMES**
VIDEO 1	86.76	31.60%	4.00%
VIDEO 2	90.16	23.30%	2.00%
VIDEO 3	90.63	24.00%	2.00%
VIDEO 4	91.85	19.60%	0.00%
VIDEO 5	93.07	21.60%	0.40%

6 Conclusion

The focus of this master's thesis is the assessment of retinal image quality.

First of all it was necessary to present theoretical overview of the acquisition of the retinal sequences along with their characteristics.

As the main focus of this master's thesis is to assess the quality of the retinal images, the second chapter is a summary of different approaches to retinal image quality, comparing their advantages and disadvantages, in order to modify described methods to design a optimal method.

The third chapter discusses image blur, and the process of image degradation. In addition to this discussion are presented different generally recognised types of image blur, their sources and characteristics. Further are proposed image blur identification methods that serve as IQA algorithms with focus on blur.

Chapter 4 describes two main methods of IQA. The first being a method for direct blur identification and evaluation. This method is based on the fact that in real retinal images different types of image blur often occur together. The dataset contained 300 images, all distorted by three types of image blur at the same time and as the method for evaluation of each image blur level was proposed a deep neural network Inception-ResNet-v2. Due to the size and quality of the dataset this network could not converge to an optimal solution, even after many modifications. Other deep neural networks were tested too, but they were not successful. Thus another dataset was provided, this time was each image distorted only by one type of blur at once. A neural network with normalized logarithmic Fourier spectrum as input was trained to classify the blur type. Subsequently three separate Inception-ResNet-v2 were trained on the new dataset, each for the particular type of blur. The assumption was that such simplification will lead to better results. Based on the training results Inception-ResNet-v2 was not appropriate even for this problem. As it was not possible to successfully train a deep neural network model, a pre-trained deep neural network model, particularly ResNet50, was used to only extract features. These extracted features were advanced to two simple regression models, GRNN and PLSR. Both of these models learn and predict faster than previously used Inception-ResNet-v2 and they are more accurate.

The results of the proposed method are discussed in chapter 5. In the first section of this chapter are described metrics used to evaluate performance of the proposed classifier and regression models. The second section contains the results of image blur classification on both simulated blurred images and real blurred images. The accuracy of classification of simulated images is 0.92, while the accuracy of classification of real images is only 0.46. These results are caused by the difference between simulated and real blurred images. The blur in real images is more subtle and different types of blur may affect one image. The second section also contains comparison of two regression models, GRNN and PLSR. From the regression performance measures is clear that PLSR outperformed GRNN. The performance measure values of PLSR in evaluating the motion blur level are SROCC 0.96, PLCC 0.97, RMSE 6.61, and R2 score 0.94. For PLSR model evaluating the out-of-focus blur are the performace measure values SROCC 0.72, PLCC 0.77, RMSE 9.57, and R2 score 0.58. The worst PLSR model is the model evaluating the long exposure blur, with performance measure values SROCC 0.50, PLCC 0.66, RMSE 12.96, and R2 score 0.43. It is not possible to statistically evaluate the video quality assessment but the predicted quality meets the subjective perception of the quality of the tested videos. The percentage of detected blurred frames was much higher than the labeled blurred frames, meaning that the model is more sensitive to blur.

Bibliography

[1] LECUN, Y. et al. *Deep learning.* Nature, 521 (7553), pp. 436–444, 2015.

[2] TORNOW, Ralf-Peter; ODSTRCILIK, Jan; KOLAR, Radim. *Time resolved quantitative inter-eye comparison of cardiac cycle induced blood volume changes in the human retina.* Biomedical optics express, 2018, 9.12: 6237-6254.

[3] FATT, Irving; WEISSMAN, Barry A. *Physiology of the eye: an introduction to the vegetative functions..* Butterworth-Heinemann, 2013.

[4] SENIOR, Kara Rogers, et al. (ed.). *The eye: the physiology of human perception.* The Rosen Publishing Group, Inc, 2010.

[5] MARTINEZ-CONDE, Susana; MACKNIK, Stephen L.; HUBEL, David H. *The role of fixational eye movements in visual perception.* Nature reviews neuroscience, 2004, 5.3: 229.

[6] DAVSON, Hugh. *Physiology of the Eye.* Macmillan International Higher Education, 1990.

[7] RYAN, Stephen J. *Retina.* 5th ed. London: Saunders/Elsevier, 2013.

[8] KOLB, Helga; FERNANDEZ, Eduardo; NELSON, Ralph. *Webvision: The Organization of the Retina and Visual System.* [Internet]. Salt Lake City (UT): University of Utah Health Sciences Center. 1995.

[9] BESHARSE, Joseph; BOK, Dean (ed.). *The retina and its disorders.* Academic Press, 2011.

[10] KOLAR, Radim; TORNOW, Ralf P.; ODSTRCILIK, Jan. *Retinal image registration for eye movement estimation.* In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015. p. 5247-5250.

[11] HOOVER, A. *Structured analysis of the retina.* [2000-04-11]. http://www. ces. clemson. edu/-ahoover/stare, 2000.

[12] LAGENDIJK, Reginald L.; BIEMOND, Jan. *The essential guide to image processing.* Academic Press, 2009.

[13] TRUCCO, Emanuele; MACGILLIVRAY, Tom; XU, Yanwu (ed.). *Computational Retinal Image Analysis: Tools, Applications and Perspectives.* Academic Press, 2019.

[14] WANG, Shuigen, et al. *Gradient-based no-reference image blur assessment using extreme learning machine.* Neurocomputing, 2016, 174: 310-321.

[15] HUNTER, Andrew, et al. *An automated retinal image quality grading algorithm.* In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2011. p. 5955-5958.

[16] LEE, Samuel C.; WANG, Yiming. *Automatic retinal image quality assessment and enhancement.* In: Medical Imaging 1999: Image Processing. International Society for Optics and Photonics, 1999. p. 1581-1590.

[17] LALONDE, Marc, et al. *Automatic visual quality assessment in optical fundus images.* In: Proceedings of vision interface. Ottawa, 2001. p. 259-264.

[18] USHER, D. B., et al. *Automated assessment of digital fundus image quality using detected vessel area.* In: Proceedings of Medical Image Understanding and Analysis. 2003. p. 81-84.

[19] NIEMEIJER, Meindert; ABRAMOFF, Michael D.; VAN GINNEKEN, Bram. *Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening.* Medical image analysis, 2006, 10.6: 888-898.

[20] YU, FengLi, et al. *Image quality classification for DR screening using deep learning.* In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017. p. 664-667.

[21] SAHA, Sajib Kumar, et al. *Deep learning for automated quality assessment of color fundus images in diabetic retinopathy screening.* arXiv preprint arXiv:1703.02511, 2017.

[22] MAHAPATRA, Dwarikanath, et al. *Retinal image quality classification using saliency maps and CNNs.* In: International Workshop on Machine Learning in Medical Imaging. Springer, Cham, 2016. p. 172-179

[23] SUN, Jing, et al. *Retinal image quality classification using fine-tuned CNN.* In: Fetal, Infant and Ophthalmic Medical Image Analysis. Springer, Cham, 2017. p. 126-133.

[24] TENNAKOON, Ruwan, et al. *Image quality classification for DR screening using convolutional neural networks.* 2016.

[25] ZAGO, Gabriel Tozatto, et al. *Retinal image quality assessment using deep learning.* Computers in biology and medicine, 2018, 103: 64-70.

[26] YAN, R., Shao, L., Member, S. *Blind image blur estimation via deep learning.* IEEE Transactions on Image Processing 25.4 (2016): 1910-1921.

[27] ZHANG, S., Shen, X., Lin, Z., Costeira, P. (n.d.). *Learning to understand image blur.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[28] BANHAM, M. R. and Katsaggelos, A. K. *Digital image restoration.* in IEEE Signal Processing Magazine, vol. 14, no. 2, pp. 24-41, March 1997. doi: 10.1109/79.581363

[29] BOVIK, Al. *The Essential Guide to Image Processing.* Academic Press, 2009. ISBN 9780123744579, https://doi.org/10.1016/B978-0-12-374457-9.00033-0.

[30] PIRES, Ramon, et al. *Retinal image quality analysis for automatic diabetic retinopathy detection.*In: 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images. IEEE, 2012. p. 229-236.

[31] CHERNOMORETS, Alexandra A.; KRYLOV, Andrey S. *Blur detection in fundus images.* In: 2012 5th International Conference on BioMedical Engineering and Informatics. IEEE, 2012. p. 243-246.

[32] KANG, Le, et al. *Convolutional neural networks for no-reference image quality assessment.* In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. p. 1733-1740.

[33] WANG, Rui, et al. *Blur image classification based on deep learning.*In: 2017 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, 2017. p. 1-6.

[34] WILLIAMS, Bryan M., et al. *Fast blur detection and parametric deconvolution of retinal fundus images.*In: Fetal, Infant and Ophthalmic Medical Image Analysis. Springer, Cham, 2017. p. 194-201.

[35] ZHANG, Shanghang, et al. *Learning to understand image blur.* In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. p. 6586-6595.

[36] YU, Shaode, et al. *A shallow convolutional neural network for blind image sharpness assessment.* PloS one, 2017, 12.5: e0176632.

[37] SZEGEDY, Christian, et al. *Inception-v4, inception-resnet and the impact of residual connections on learning.* In: Thirty-first AAAI conference on artificial intelligence. 2017.

[38] Keras: the Python deep learning API. *Keras: the Python deep learning API* [online]. Retreived from https://keras.io/

[39] [online]. Retrieved from https://ai.googleblog.com/2016/08/improving-inception-and-image.html

[40] KINGMA, D. P.; BA, J. L. *Adam optimizer.* arXiv preprint arXiv:1412.6980, 2014, 1-15.

[41] YAN, Ruomei; SHAO, Ling. *Blind image blur estimation via deep learning.* IEEE Transactions on Image Processing, 2016, 25.4: 1910-1921.

[42] LI, Dingquan; JIANG, Tingting; JIANG, Ming. *Exploiting high-level semantics for no-reference image quality assessment of realistic blur images.* In: Proceedings of the 25th ACM international conference on Multimedia. 2017. p. 378-386.

[43] HE, Kaiming, et al. *Deep residual learning for image recognition.*In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.

[44] SPECHT, Donald F., et al. *A general regression neural network.* IEEE transactions on neural networks, 1991, 2.6: 568-576.

[45] LI, Chaofeng; BOVIK, Alan Conrad; WU, Xiaojun. *Blind image quality assessment using a general regression neural network.*IEEE Transactions on neural networks, 2011, 22.5: 793-799.

[46] ABDI, Hervé. *Partial least square regression (PLS regression).*Encyclopedia for research methods for the social sciences, 2003, 6.4: 792-795.

[47] SciPy.org — SciPy.org. *SciPy.org — SciPy.org* [online]. Copyright © Copyright 2020 SciPy developers. [cit. 10.05.2020]. Retrieved from https://www.scipy.org/

[48] sklearn.metrics.r2_score — scikit-learn 0.23.1 documentation. *scikit-learn: machine learning in Python — scikit-learn 0.16.1 documentation* [online]. Copyright © 2007 [cit. 10.05.2020]. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html

[49] DREISEITL, Stephan; OHNO-MACHADO, Lucila. *Logistic regression and artificial neural network classification models: a methodology review.* Journal of biomedical informatics, 2002, 35.5-6: 352-359.

List of symbols, physical constants and abbreviations

IQA	image quality assessment
RPE	retinal pigment epithelium
OD	Optic disc
CRA	central retinal artery
RPCs	radial peripapillary capillaries
CRV	central retinal vein
VO	videoophtalmoscope
SNR	Signal-to-noise ratio
CNN	convolutional neural network
ReLU	Rectified Linear Unit
SVM	support vector machine
PSF	point spread function
GRNN	general regression neural network
PLSR	partial least sqaures regression
PCA	principal component analysis
SROCC	Spearman's rank order correlation coefficient
PLCC	Perason correlation coefficient
RMSE	root mean squaree error

7 Appendices

Attachments contain source codes written during the practical part of the master's thesis.

- `Appendices_Zuzana_Vasickova`
  - `00_Blurred Images` Source code for simulation of blurred images
    - `Image_Blurring_Mix.py`
  - `01_Inception-ResNet-v2 train and test` Folder containing source codes of different trainings of Inception-ResNet-v2
    - `InceptionResNetV2_ALL_epsilon.py` Inception-ResNet-v2 with added dropout layer
    - `InceptionResNetV2_ALL_frozen-layers.py` Inception-ResNet-v2 with frozen layers
    - `InceptionResNetV2_ALL_no-dropout.py` Inception-ResNet-v2 without dropout layer
    - `InceptionResNetV2_LE_no-dropout.py` Inception-ResNet-v2 for long exposure blur estimation
    - `InceptionResNetV2_MB_no-dropout.py` Inception-ResNet-v2 for motion blur estimation
    - `InceptionResNetV2_OOF_no-dropout.py` Inception-ResNet-v2 for out-of-focus blur estimatio
    - `TEST_MODEL.py` Load and test model on test dataset
  - `02_Blur Type Classification` Folder containing blur type classification source codes
    - `Blur_classification.py` Training of neural network
    - `Blur_classification_TEST.py` Load and test model on test dataset
  - `03_Feature extraction ResNet50` Image feature extraction ResNet50
    - `Feature Extraction.py`
  - `04_Blur level estimation` Training of regression models for blur level estimation
    - `GRNN.py` GRNN model
    - `PLS_regression.py` PLSR model
  - `05_Evaluation of retinal sequences`
    - `LogReg_VID.py` Training logistic regression model for detection of blurred frames
    - `RESULTS_VIDIQA.py` Source code for evaluation blur level prediction from RETINA_SEQ_IQA.py
    - `RETINAL_SEQ_IQA.py` Source code for loading and evaluating retinal sequences
  - `06_Data` Folder containing sample data.
    - `Features_separate_ResNet50`
      - `LE`
        - `X1`
          - `X1_Features_Blurred_imageStudy_02_00008_02_R.pngLE0.png_[1].npy`
          - `X1_Features_Blurred_imageStudy_02_00008_02_R.pngLE1.png_[2].npy`

```
        X2
            X2_Features_Blurred_imageStudy_02_00008_02_R.pngLE0.png_[1].npy
            X2_Features_Blurred_imageStudy_02_00008_02_R.pngLE1.png_[2].npy
        X3
            X3_Features_Blurred_imageStudy_02_00008_02_R.pngLE0.png_[1].npy
            X3_Features_Blurred_imageStudy_02_00008_02_R.pngLE1.png_[2].npy
    MB
        X1
            X1_Features_Blurred_imageStudy_02_00008_02_R.pngMB0.png_[1].npy
            X1_Features_Blurred_imageStudy_02_00008_02_R.pngMB1.png_[2].npy
        X2
            X2_Features_Blurred_imageStudy_02_00008_02_R.pngMB0.png_[1].npy
            X2_Features_Blurred_imageStudy_02_00008_02_R.pngMB1.png_[2].npy
        X3
            X3_Features_Blurred_imageStudy_02_00008_02_R.pngMB0.png_[1].npy
            X3_Features_Blurred_imageStudy_02_00008_02_R.pngMB1.png_[2].npy
    OOF
        X1
            X1_Features_Blurred_imageStudy_02_00008_02_R.pngOOF0.png_[1].npy
            X1_Features_Blurred_imageStudy_02_00008_02_R.pngOOF1.png_[2].npy
        X2
            X2_Features_Blurred_imageStudy_02_00008_02_R.pngOOF0.png_[1].npy
            X2_Features_Blurred_imageStudy_02_00008_02_R.pngOOF1.png_[2].npy
        X3
            X3_Features_Blurred_imageStudy_02_00008_02_R.pngOOF0.png_[1].npy
            X3_Features_Blurred_imageStudy_02_00008_02_R.pngOOF1.png_[2].npy
IMG
    IMG_ALL
```

```
         Blurred_imageStudy_02_00008_02_R.pngMB+OOF+LE0.png
         Blurred_imageStudy_02_00008_02_R.pngMB+OOF+LE1.png
      IMG_LE
         Blurred_imageStudy_02_00008_02_R.pngLE99.png
         Blurred_imageStudy_02_00014_02_R.pngLE99.png
      IMG_MB
         Blurred_imageStudy_02_00014_02_R.pngMB99.png
         Blurred_imageStudy_02_00016_01_L.pngMB99.png
      IMG_OOF
         Blurred_imageStudy_02_00014_02_R.pngOOF99.png
         Blurred_imageStudy_02_00016_01_L.pngOOF99.png
   PAR
      PAR_ALL
         Blurred_imageStudy_02_00008_02_R.pngMB+OOF+LE0.npy
         Blurred_imageStudy_02_00008_02_R.pngMB+OOF+LE1.npy
      PAR_LE
         Blurred_imageStudy_02_00008_02_R.pngLE99.npy
         Blurred_imageStudy_02_00014_02_R.pngLE99.npy
      PAR_MB
         Blurred_imageStudy_02_00014_02_R.pngMB99.npy
         Blurred_imageStudy_02_00016_01_L.pngMB99.npy
      PAR_OOF
         Blurred_imageStudy_02_00014_02_R.pngOOF99.npy
         Blurred_imageStudy_02_00016_01_L.pngOOF99.npy
   SHARP_IMG
      imageStudy_02_00008_02_R.png
   VID_test
      imageStudy_02_00113_01_R_registered.avi_2.png
      imageStudy_02_00113_01_R_registered.avi_21.png
   VIDEOS_train_frames_blurred
      imageStudy_02_00008_02_L_registered.avi_14.png
      imageStudy_02_00008_02_L_registered.avi_106.png
   VIDEOS_train_frames_sharp
      imageStudy_02_00008_02_L_registered.avi_13.png
      imageStudy_02_00008_02_L_registered.avi_104.png
05_AVI to PNG.py.Source code for loading a video and saving the frames as .png
```