T.C.

ALTINBAS UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

MULTI-EXPOSURE IMAGE FUSION WITH EDGE PRESERVING SMOOTHING FILTER

Ibrahem Bayan yaseen ALABDULLAH

Master of Science

Supervisor

Asst. Prof. Dr. Sefer KURNAZ

Istanbul, 2021

MULTI-EXPOSURE IMAGE FUSION WITH EDGE PRESERVING SMOOTHING FILTER

by

Ibrahem Bayan yaseen ALABDULLAH

Electrical and Computer Engineering

Submitted to the Institute of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2021

The thesis titled "MULTI-EXPOSURE IMAGE FUSION WITH EDGE PRESERVING SMOOTHING FILTER" prepared and presented by "Ibrahem Bayan yaseen ALABDULLAH" was accepted as a Master of Science Thesis in Electrical and Computer Engineering.

____

Asst. Prof. Dr. Sefer KURNAZ

Supervisor

Thesis Defense Jury Members:

Asst. Prof. Dr. Sefer KURNAZ	School of Engineering and Natural Sciences, Altinbas University	______________
Prof. Dr. Osman Nuri UÇAN .	School of Engineering and Natural Sciences, Altinbas University	______________
Prof. Dr. Mesut RAZBONYALI	Engineering and Architecture Faculty, Maltepe University	______________

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Approval Date of Institute of Graduate Studies: ____/____/____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Ibrahem Bayan yaseen ALABDULLAH

DEDICATION

I devote and pledge this research work to my supervisor who is salient for guiding me through whole research work as well as my family for always assisting me in my hard time.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Asst. Prof. Dr. Sefer KURNAZ for guiding and helping me along the way in writing this dissertation. Discussing my progress, problems, and ideas with my supervisor Asst. Prof. Dr. Sefer KURNAZ a couple of times every week helped me tremendously in understanding the logic behind the research. It made me better realize the technical need for this research work.

ABSTRACT

MULTI-EXPOSURE IMAGE FUSION WITH EDGE PRESERVING SMOOTHING FILTER

ALABDULLAH, Ibrahem Bayan yaseen,

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Asst. Prof. Dr. Sefer KURNAZ

Date: April /2021

Pages: 63

This master's thesis focuses on multi-exposure image fusion using image processing and deep learning techniques. The development of image edge smoothening system using CNN based on artificial bee colony optimization task is in hot pursuit, with special attention being given to the smoothening of all the edges of image. Given its high propensity to meta-size, going hand in hand with severe decreases in preservation rates, and the high inter-edge variability in image appearance, as well as a strong requirement on the training of the physician properly de-noising an image can be considered a daunting task. The purpose of this research thesis is to use a deep learning and image processing pipeline for multi-exposure image fusion for the segmentation of edges in an image using with hybrid techniques in deep learning and imaging. The literature review of different papers was conducted with different imaging model architectures. The CNN custom model was created for the task, and deep learning technique (CNN) was used with different levels of fine tuning of hybrid image processing techniques. Screening for high edge filter to identify edges at high accuracy has been under debate. In current discussion, the suggested smoothening procedure is bone density based including possible prescreening techniques using the Artificial Bee Colony (ABC) optimization. Development of new prediction models and automated smoothening could contribute to general screening programs in the future.

The custom deep learning model architectures were designed to represent different depths. The idea behind this is to analyze the effect of increasing representational capacity to the results and visualizations for all edges in an image. Additionally, deep learning CNN model was created to represent traditional automated image processing approach. Image processing has been used in some edging of images, producing good results for example in segmentation of image area structure and segmentation of image smoothening areas under consideration. The study also attempts to find solutions to practical deep learning challenges such as low training speed and lack of transparency with an accuracy of 97.59% absolutely. The imaging and deep learning pipelines are optimized in order to exploit the available parallelism using the MATLAB programming language with multiple tools under consideration.

Keywords: Smoothening, Filter, Deep Learning, Segmentation, Image, CNN, Mulit-Exposure, ABC, Model, Optimization.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

LIST OF ABBREVIATIONS

IPS	:	Image Processing Systems
IPC	:	Image-Based Control
AIAS	:	Advanced Imaging Assistance Systems
ISP	:	Image Signal Processor
DSP	:	Digital Signal Processors
TM	:	Tone Mapping
HDR	:	High Dynamic Range
CNN	:	Conventional Neural Network
ABS	:	Artificial Bee Colony

1. INTRODUCTION

1.1 INTRODUCTION

Image fusion and edge smoothening is a trend that has been driven by the fact that human beings are known to be susceptible to filtering mistakes [1], whether it is distractions from the direct environment or the fact that driving is done under fatigue, alcohol or drugs. Although having a fully automated vehicle is still a challenge [2], significant progress has been achieved by ongoing research [3]; from the [4], the very first radio-controlled car, to the Mercedes-Benz Van in 2018, which incorporated computer vision, and to the cars of today from most major car manufacturers that possess numerous electronic aids such as collision avoidance, advanced imaging assistance systems (AIAS), and image processing systems (IPS). Those aids, which can be considered as a substantial step towards fully automated vehicles, use at least one camera to enhance safety. As part of a Sensing-Perception-Decision architecture [5] of autonomous driving systems, cameras are a fundamental sensing device, the data of which can be utilized in object recognition and tracking tasks that allow for proper path planning. With cameras being largely integrated into both mid and luxury-class vehicles [6], their annual growth rate is expected to be around 20% per vehicle by 2023 [7], playing a vital role in computer vision and, therefore, in image-based control (IBC).

As cameras are being increasingly adopted by the auto industry, the urge for diverse functionality becomes highly relevant. Camera sensors need to be able to provide pleasant images for human display, as well as predictable and reliable images to be used by computer vision [7]. Those camera capabilities allow AIAS to be employed for vision-based perception tasks, such as object recognition, localization, and mapping, and path or motion planning [8], which are, in principle, IBC tasks and may enable a safer driving experience and bridge the gap towards fully autonomous driving.

As can be seen in Figure 1.1, an image colors may consist of an image signal processor (ISP), the image processing algorithm and the controller. A traditional image signal processor processes the RAW output of an image sensor and produces a compressed image that can be used by a vision application. Typical ISP stages include smoothening, de-noising, white balancing and color mapping, gamut mapping, tone mapping, and compression. Those stages, albeit standard, require excessive computational intensity to produce a high-quality output image, which in the case of computer vision applications is not necessarily essential [9] and can be approximated.

Approximate computing is a concept that relies on building systems with acceptable behavior from inexact hardware or software components. It allows to trade-off application accuracy in order to achieve considerable performance and energy gains [10] at design time. A key challenge in approximate computing is the identification of those sections either in hardware or software that can actually be approximated, as there is always a risk of crashing an application if a critical component is being approximated.

Figure 1.1: The basic example of image fusion with multiple exposures [10].

1.2 OVERVIEW AND MOTIVATION

An image fusion uses multiple cameras for additional safety during navigation. A use case of a vision-based lateral control example using a single camera is studied in this project, where the camera output is being processed by an image processing application and the autonomous functionality is maintained by a controller that actuates based on the input from the camera. The camera produces 60 frames per second and the rest of the application needs to achieve real time performance in order to process all those frames as given in [10].

The camera is the sensor and is attached to the system. Each frame from the camera sensor goes through a series of processing stages. Initially, the frames need to be processed by an image signal processor (ISP), which converts the RAW output of the camera sensor to a format that is useful for the human and computer vision. The resulting image is, then, ready to be processed by the image processing stage, which performs feature extraction and provides information about the image processing environment to the image-based controller as mentioned in [11]. The controller uses the visual feedback from the camera and actuates the required steering angle in order to allow the car to drive autonomously. In principle, a fundamental parameter in the control design is the sampling period. It depends on the duration the application needs in order to finish the required computations, from sensing to actuating. Typically, shorter sampling periods are necessary to achieve real time performance and maintain high quality-of-control.

The autonomous car application with the pipeline of stages can be seen in Figure 1.1. The ISP stage can be rather time consuming and is usually mapped onto specialized hardware, such as digital signal processors (DSP). Given its high computational complexity, we attempt to approximate the ISP, by coarsely skipping some of the ISP stages. Approximate computing is a technique that allows to improve the runtime performance of an error-resilient application, which is desired by the controller as mentioned in [12]. The error resilience of the application can be achieved in the image processing stage. When modified accordingly for the use case of lane detection, the image processing can handle the approximation, by successfully extracting features from the inaccurate images.

For the purpose of improving the quality-of-control of this image-based control application, seven different approximated ISP pipelines are developed. This allows to analyze the trade-off between runtime improvement and quality degradation with respect to quality-of-control for different degrees of approximation. The pipelines are optimized in order to exploit the available parallelism using the Matlab programming language.

1.3 MULTI-EXPOSURE IMAGE FUSION

The scenes in the physical world include brusque lightening circumstances leading to highlights (over-exposed areas) or shadows (under-exposed areas) in images taken digitally. Conventional digital cameras usually fail capturing details in over as well as under exposed areas in the case when high or low exposure configurations are set in camera. To illustrate the scene as realistically as possible, some modifications are made on the hardware side, however, to reduce the costs, some are made on software side, where most of the problems were solved from this perspective. This enabled creating images from scenes in analogy to one captured via the eyes of human. Such software solutions, which are applied software-based high dynamic range (HDR) approach which has been a major focus for various studies, in which a lot of efficient methods were created [13]. The major task of such approaches is that well-exposed images could be implemented through taking images' sequence with various levels of exposure taken via digital camera as input. Furthermore, synthesize output image which is extremely useful and perceptually attractive in comparison to input image.

Present HDR imaging methods could be divided to two major categories: Image Fusion (IF)-based and Tone Mapping (TM)-based approaches. With regard to TM-based approaches, HDR images are initially re-constructed by using particular HDR re-construction (HDR-R) approach applied over multiple low dynamic ranges (LDR) images regarding the same scene [14]. Therefore, TM-based approaches have two important stages, which are, HDR re-construction as well as tone mapping (HDR-R+TM). A lot of efficient TM algorithms were developed recently [15].

Distinguishing from TM-based approaches require generating intermediate HDR images, IF-based approaches have the task of directly obtaining everywhere well-exposed images through a process of merging complementary information related to the input LDR image and have the ability of skipping the process of HDR image construction. Therefore, IF-based approaches are typically more effective compared to TM-based approaches [16]. Such IF-based HDR approach is referred to as the multi-exposure fusion (MEF) as well, that was developed as a major subject of study in HDR imaging and image fusion. In comparison to the general image fusion task [17], MEF has certain properties due to the fact that over and under exposed regions in LDR images have certain distinguishing features like extreme luminance and saturation. Even though that certain some universal fusion approaches could be used to the MEF [18], it has considerable meaning for developing extra detailed algorithms for achieving optimum performance.

For instance, in [19] suggested block-based MEF approach through selecting block that contain maximal information. At the same time, in [20] developed influential exposure fusion approach on the basis of weight maps. With regard to their approach, weight map related to each one of the source images has been estimated through combining three quality measures that are contrast, saturation and well-exposedness at pixel level, which in addition to other quality measures, were selected in the comparison conducted in this work.

MEF fill a gap between the HDR in natural scene and low LDR images captured via traditional digital camera [21]. MEF offer inexpensive alternative to bridge a gap between LDR display and HDR imaging [22]. The main aim of exposure fusion is obtaining scene's full dynamic range through a process of blending a number of exposure images in one high-quality composite image and keeping texture information and details as much as possible [23]. Since it has been initially developed in 2017 [22], MEF was a major field of study and its importance has elevated recently. Although there are several types of approaches proposed to solve the multi-exposure image fusion problems. Most of them lack the efficiency of their outcomes, due to the inaccuracy of their parameters. Finding the optimal value of parameters is a critical issue in order to provide highly accurate results. With the existence of various MEF algorithms, it is very important comparing their performance of quality measures, for the purpose of finding the optimal best in addition to directions for more development. The quality measures are important factors that influence the performance of multi-exposure image fusion as given in [24]. However, the problem of this work can be defined by identifying the quality measures families and feature groups that are highly relevant to the MEF.

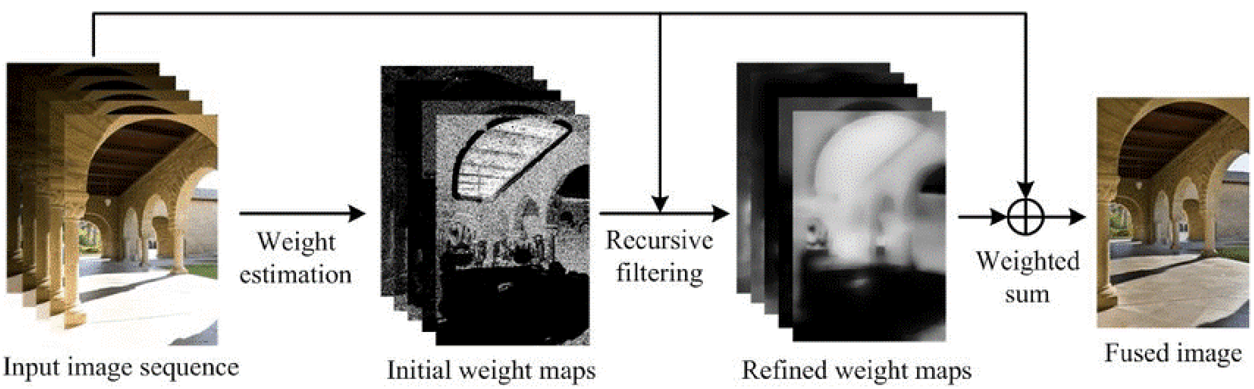


Figure 1.2: The pathology of multi-exposure image fusion keeping texture information and details [24].

1.4 PROBLEM STATEMENT

The compute-intensive image processing stage of an image fusion system result in longer sampling periods, which are considered during design time and negatively affect the QoC of the underlying system. Usually the control design of data intensive feedback control applications relies on the worst case approach [25]. This includes image processing computations, the duration of which might vary, but is not taken into account. Researchers in [26] proposed to deal with workload variations that are caused by image processing using a scenario-aware approach. Instead of designing the controller for the worst-case (WC) scenario, continuously adapt the sampling period of the controller on the actual case at run-time. In that way, the average sampling period is improved with respect to a WC design. However, this approach treats the image preprocessing pipeline as a black box and assumes a constant image preprocessing delay. The image processing step is required in order to process the output of the image sensor and extract useful information that can be used by the controller for decision making. We identified the impact that excessive computations in the image processing stage can have for real-time applications and proposed a feature-based detection algorithm, which only performs basic operations in order to identify the lane markers. In the preprocessing step, the RGB image is converted to gray-scale in order to select the region-of-interest (ROI) and extract two binary images, out of which one is used for global thresholding and the other for edge detection using the Sobel filter. Those two binaries are, then, combined using neighborhood-AND to generate the bird's eye view (BEV) of this image, and fusion identification is performed based on the maximum likelihood estimation on histogram plots.

Although a feature-based lane detection approach such as the allows for increased flexibility for each one of the intermediate steps of the algorithm, it uses excessive steps, which may hinder the real-time performance. We followed a slightly different approach; using a front-mounted camera view instead of BEV for selecting the ROI in their algorithm, they attempted to minimize the steps needed by the algorithm as mentioned in [27]. The preprocessing stage includes conversion from RGB to intensity image followed by an averaging filter to reduce noise. Then, the intensity image is thresholded in order to obtain the lane marks. Finally, for lane identification, the Hough transformation is used in combination with a horizontal differencing filter, which has the role of the edge detector. This approach achieves good performance, but is susceptible to the performance of the edge detection, since the Hough transformation requires accurately detected edges.

1.5 AIM OF CONTRIBUTION

The research in this works aims to contribute to the following key aspects of the use of deep learning and image processing for smoothening of edges in an image. The main contributions of this research include the following:

- Integration of an image preprocessing pipeline and approximate computing in a baseline ABC algorithm and CNN framework.
- Characterization of accurate and approximate algorithms to identify the regions-of-interest for approximation.
- Study and characterization of approximation choices of the imaging pipeline & computer vision algorithms with respect to edge smoothening with filters.
- Approximation of image fusion without taking into account the timing impact.
- Approximation of image fusion taking into account the timing impact on sampling period.
- Application profiling to obtain execution times and to compute optimized sampling periods.
- Trade-off analysis between approximation and image fusion for optimized sampling period designs.
- Development of a tool-chain written in a high performance language (MATLAB) that allows the exploration of different approximations and their impact on quality-of-control.
- Identify the highly performance quality measure approach.
- Identify the optimal combination between the quality measure approaches that provide the best outcomes.

2. LITERATURE REVIEW

Approximate computing is a technique that leverages from the tolerance of applications to errors or inexact computations that reduce the quality in a controllable and acceptable manner. A large number of modern applications can tolerate those inexact computations and boost performance and energy efficiency. Software techniques such as loop perforation, memorization, precision scaling, task dropping, and data sampling or hardware techniques such as over scaling, clock over-gating, body-biasing and refreshing rate, may yield benefits of possibly up to 50% in execution time and analogous improvements in energy efficiency as mentioned in [28]. In the case of image fusion systems that are used in the context of image fusion and smoothening, the obvious bottom line is to avoid crashing the controlled vehicle. Thus, the output of an approximate computing algorithm needs to be carefully monitored. There are different approximation strategies available, either for software-based approximation, such as precision scaling and loop perforation, or hardware-based approximation, such as using inexact hardware. A common denominator of those approaches is that they are application-specific.

2.1 IMAGE FUSION PIPELINE

Image fusion microscopic pipeline can combine the advantages of image-throughput and image-content analysis with several processing steps for producing high quality image fusion with smoothened edges. In the following, these concepts are explained.

2.1.1. Image-Content Analysis. In image content analysis, rich information, such as high-resolution images, is obtained from each sample. Most recent advancements of high-content analysis include super-resolution microscopy and imaging mass spectrometry, which increase the amount of spatial information or number of labels measured in a given sample, respectively [29].

2.1.2. Image-Throughput Analysis. In image throughput analysis, a larger number of samples are analyzed quickly, though not necessarily at a single cell resolution as given in [30]. Examples for high-throughput methods are traditional flow cytometry or fluorescence intensity measurement using plate readers that yield a fluorescence response for each well in a plate. Thereby, thousands of cells are contained in each well.

2.1.3. Multi-well Plates for Image Flow

Some techniques combine the multi-well plates for image fusion as the advantages of high-content and high-throughput analysis raise this issue for smoothening the image. Capturing biological image fusion is now highly automated, which enables methods that deliver rich information for a large number of cells at the same time as mentioned in [31].

2.1.4. Independent Pixel Processing

An independent pixel processing in imaging flow captures high resolution images of each single cell passing through a flow channel and thus using dyes and CNN feature extraction, one can obtain spatial information about substances of interest sample images for image fusion. The robots are used to automate the independent pixel processing and handling of multi-well plates. Imaging flow cytometry and fluorescence microscopy are examples for methods that allow high-throughput and high-content analysis at the same time.

2.1.5. High Resolution Fusion

In high resolution fusion, the elements of microscopes capture the fluorescence response of multiple cells in one well at once. Images typically depict hundreds of cells. This method's spatial resolution is very high and thus morphological details are visible for each cell. This is why this method is used in morphological profiling as provided in [32]. Compared to imaging flow cytometers, fusion microscopes provide higher resolution images, are more sensitive and allow higher magnifications.

2.2 SEGMENTATION OF FUSION CHANNEL

The research also provides the improve the segmentation of fusion in the DNA channel. The research compares the performance of segmentation algorithms against the current segmentations obtainable with proposed CNN which is the standard approach. The CNN features does a combination of semantic and instance segmentation for fusion given in [33]. It classifies pixels into foreground and background pixels and then assigns the foreground pixels to an object instance. The CNN current implementation for segmentation is implemented in the module Identify Primary Objects and consists of three steps:

- Using thresholding, foreground regions are identified. Thus, the CNN filer distinguishes between pixels that belong to fusion and background pixels.

- Clumped fusion are identified and segmented. For the segmentation, a seeded watershed algorithm is used. The user can choose to either apply it on the distance transform of the thresholded image or on pixel intensity values. Other traditional computer vision methods such as image smoothing are used to improve the segmentation results.

- Fusion images whose features do not fall into user-specified ranges are discarded. For example, CNN offers to discard nuclei that are close to the image's boundary, or nuclei below a certain size.

The segmentation quality of this approach is highly affected by the choice of parameters for the used algorithms. Thus, the user's input is required for choosing the parameters. The interface for the parameter selection of the module as described in [34]. For parameter optimization, users usually start off with a given set of parameters and adapt these empirically based on the obtained segmentation. The user can then optimize the parameters in an iterative process. As ground truth annotations are usually lacking, the user has no quantitative segmentation quality measure and is relying on visual inspection only. These methods:

- They typically require an initialization or starting point which can be set manually or estimated automatically for instance as local minima of some global criterion.

- The optimization is usually performed iteratively. Example of sample image and a region growing isolation of multi-exposures.

2.2.1. Isolation of Multi-Exposure Images

The contrast in an image is used for isolation of multi-exposure images because of general problem that occurs in computer vision. Its task is to identify and split an image into regions with different context-dependent meanings, such as the four classes (each marked by a different color) that corresponds to various biological tissues in the input image as mentioned in [35]. Formally, a segmentation function assigns label to each pixel, the application-specific set of classes (labels). In most segmentation tasks, one can assume that the regions of equally labeled pixels should be somehow compact, i.e., neighboring pixels share a label. In this sense, image fusion can be seen as an image segmentation but with a difference that the number of a label is much larger and there is just local meaning for each label [36]. Following the previous definition, image fusion segmentation is then a function assigning a label to each image fusion. Note, that the image fusion label can be simply propagated (distributed) to all pixels belonging to the particular image fusion as given in [37]. Through the use of Laplace kernel, the summation of the differences over nearest neighbors of central pixel in selected block are estimated. In the case where there is an edge, or sudden changes in that block, then the resulting block will contain higher values, that can be clearly noticed in comparison with the other blocks as predicted in [38]. This convolution process is repeated on all the blocks of the image in order to produce a new block. The combination of all these blocks represents the resultant image, which is the results after conducting the contrast measure.

2.2.2. Pixel-Wise Segmentation

Let us define the pixel-wise segmentation commonly used pipeline for multi-exposure image based segmentation which is very similar to much pixel-wise segmentation. Existing image fusion-based segmentation methods usually consist of 3 steps:

- Computing multi-exposures, which preserve required details such as object boundary and reasonable image fusion size
- Extracting multi-exposures appearance features based on image intensity or multi-exposures shape
- Using an estimated model or classifier to assign a label to each multi-exposure.

This classification can be performed by a standard classifier in a supervised or unsupervised manner with a fundamental assumption about the input images such as a number of classes or interclass differences as presented in [39]. Especially in multi-exposure imaging, many image segmentations based on image fusion have been presented mainly as a preprocessing phase in further image analyses, segmentation, registration, etc. The advantage of using image fusion except reducing the degree of freedom (number of variables) comparable image information is also higher robustness to noise as can occur for multiple modalities, typically CNN with MRI as given in [40]. It has been shown that image fusion lead to more accurate segmentation than direct pixel-level segmentations while having lower demands on processing time and memory resources. This measurement is a Gauss filter, which is why, is performed the weighing of every one of the pixels according to the closeness of that pixel to 0.5. This measurement is important in the selection of a pixel via viewing its exposed value. In this approach, the map of intensity is utilized for getting a pixel's exposed value.

2.2.3. Watershed Edge Smoothening

The watershed edge smoothening segmentation on distances to boundaries is a typical example of region growing method starting from local minima, and stopping at touching the neighboring region as mentioned in [41]. In level-set segmentation, the object of interest is described by a function and the resulting segmented region corresponds to the zero-level set for entropy. Multiple objects can also evolve at the same time, but policy definition on their intersections is much harder.

2.2.4. Independent Pixel Saturation

The independent pixel saturation classification and employing spatial relations between neighboring pixels through some other graph-based regularization, e.g. graph cuts improved the segmentation results regarding accuracy and robustness to noise, but at the costs of increased computational demands. The most similar previous work was done by Researchers in [37], the authors also use super pixels and graph cuts as regularize, but their work focuses on a single application — single object binary segmentation. Researchers in [38] uses a simpler image fusion extraction method and only basic edge terms in the graph regularization. Mean shift clustering, and Researchers in [39] worked on image fusion with a region-adjacency-graph regularization, while researchers in [40] considers long-range similarity-based interactions instead of interactions based on neighborhoods. Finally, there seems to be a great promise in deep learning methods using convolutional neural networks, such as CNN requiring a small number of training examples but it assumes reasonable homogeneous objects and produces only binary segmentation where proper object separation relays on exact boundaries prediction. In which CNN represent saturation, μ represent the average value of the R, G, and B. Saturated colors are required and make images appear vivid. Saturation measure is of high importance in keeping as much color information as possible in final through taking just the pixels of the highest saturation. The major aim of such measurement is to select optimum pixel's color in final image.

2.3 NEURAL NETWORK EXPLANATION

A short introduction to convolutional neural networks. The research start with a brief explanation of Feed Forward Neural Networks before covering the single-layer perceptron and how it relates to logistic regression. We then generalize to the multi-layer perceptron and finally introduce convolutional neural networks.

The term neural network describes systems of parametrized mathematical operations that are applied on some input data in order to obtain out- put values. These mathematical operations can be seen as synapses between intermediate results which in turn are called neurons due to the similarity to the function of the brain as presented in [41]. It is to be noted that biological neural networks indeed inspired the design of neural networks, but the function is still different. One central difference between both models is that biological nerve cells have spikes in their action potential at discrete time points whereas neural networks use continuous values to describe the excitement of a neuron. The CNN model presented here are all feed-forward neural networks which is an artificial neural network whose operations do not form loops.

2.3.1 Feed-Forward Neural Network

A basic task in supervised learning is the binary classification of data. Given two classes A and B and data points $x \in R^d$ that belong to exactly one of these classes, the goal is to learn a model that classifies new data points correctly.

The simplest form of classification is finding a hyperplane $w^T x + b = 0$ in the d-dimensional space that separates the data. Thus, one wants to find weights W and a bias b such that $w^T x + b > 0$ for $x \in A$ and $w^T x + b < 0$ for $x \in B$. Thus, the model summarizes as follows:

$$f(x) = sgn\{w^T + b\} \tag{2.1}$$

Without loss of generality, if $f(x) = 1$, x is classified as belonging to class A and belonging to class B if $f(x) = -1$. This model is called a single-layer perceptron.

2.3.2 Single-Layer Perceptron

In contrast to logistic regression, a single-layer perceptron does a hard assignment of the data points to one of the two classes given in [42]. Logistic regression performs a soft assignment in the sense that a probability $p \in (0, 1)$ for the data point belonging to class A is predicted.

$$P = (Y = c|x, \theta) = Ber(c|\sigma(x, \theta)) \tag{2.2}$$

Ber hereby denotes the Bernoulli distribution, Θ summarizes the model parameters w and b. The logistic function σ is defined as follows:

$$\sigma(x, \theta) = \frac{1}{1+e^{-(w^T+b)}} \tag{2.3}$$

However, the model of logistic regression is similar to a single-layer perceptron. Assuming a maximum, a posteriori estimator, a linear decision boundary $w^T\ x + b = 0$ can be obtained at e. As the logistic function is applied on the distance to this hyperplane, the model's predicted probability of the data point belonging to class A becomes more degenerate with increasing distance to the hyperplane.

2.3.3 Multi-Layer Perceptron

A multi-layer perceptron (MLP) is a parallel and sequential concatenation of single-layer perceptron's. Two consecutive layers of an MLP are connected by a series of parallel single-layer perceptron's as described in [43]. The layers are stacked sequentially and instead of applying the sign function to the output of each layer as in a single-layer perceptron, non-linear activation functions are used. The transformed outputs of a given layer serve as the input data for the subsequent layer.

2.4 CNN FEATURES EXTRACTION

The research presents a feature-based CNN technique for feature isolation of multi-exposures using CNN model and random forest for some instances using an innovative combination of image fusion and Graph Cut regularization to impose spatial compactness in images. This makes the segmentation/isolation both fast and robust. Researchers in [43] have also introduced a new edge weight factor. Finally, we have tested our method extensively on five real applications, and have compared it with previously used methods. We found not only that image fusion methods work better than non-image fusion methods, but also that graph cut regularization yields further improvement. Also, an advantage is that the supervised segmentation can be trained from just a few images or even partially annotated examples given in [44]. Interestingly, an unsupervised variant of our method can generate completely acceptable results for some applications, without the need for manual segmentation.

In comparison to alternative decomposition methods, our method yielded higher resistance to internal deformations and input noise on both synthetic and real datasets. These methods are developed as part of the large project in development with MATLAB R2019b. The complete image processing pipeline and deep learning based convolutional neural network was developed in the beginning of this work, an automatic image analysis pipeline for mining gene patterns from a vast set of microscopic based multi-exposure images. Possible continuation of this thesis is extending these methods to work in 3D entirely and release it as close form software/plugin (MATLAB) which can be simply used by a biologist. Even though being developed for the particular application of multi-exposure images, the proposed method can be easily generalized and adapted to other application domains, not only in the multi-exposure context and on multi-exposure images as presented in [45]. To ease the access to the presented methods and also to encourage the community to possibly improve the proposed methods, the implementations, as well as the source code, were released as open-source for public usage.

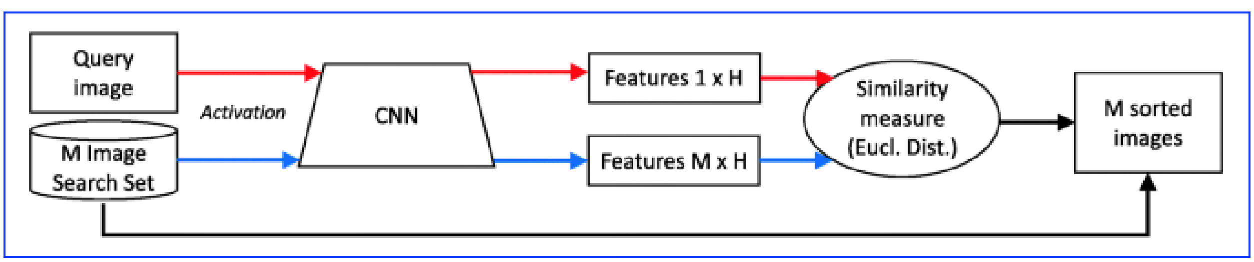


Figure 2.1: A robust convolutional neural network based system for feature extraction [45].

2.5 QUALITY OF EXPERIENCE (QOE)-BASED MULTI-EXPOSURE FUSION

Researchers in [46] have developed a deep learning and image processing based multi-exposure detection and isolation application with the growing amounts of available images together with their increasing resolution, a manual analysis by an expert becomes almost infeasible. The semantic segmentation of ovary is difficult to die to high similarity between two pairs of classes; the estimation of multi-exposures center has not been proposed yet; performing an instance segmentation for such highly structured object is however very challenging, and standard techniques for a pattern extractions tend to be sensitive to noise and internal deformations using MATLAB R2019a. We introduced used datasets with a variety of user annotation depending on the image processing task as given in [47]. Later, we proposed four image analysis methods and presented performances on real microscopy images and we compared them with standard methods.

The central objective of this work was to improve a fully automatic processing and analysis pipeline for microscope images of the Multi-exposures. In particular, the proposed methods are — image segmentation, object center detection, region growing for sensed individual images and followed by binary pattern extraction as atlas estimation from a broad set of sensed images. Our proposed enhancement to the image fusion extraction regarding processing time and smarter post-processing as given in [48]. Research has presented enhancement to multi-exposure isolation extraction led to improved performance compared to standard pixel-wise methods on real microscopy images using Convolutional neural network with comparatively better performance than previously performed techniques on MATLAB R2019b.

2.5.1 Perceived Local Contrast

address the first step of the image processing pipeline - segmenting individual object in the input image and marking the rest as background. This image segmentation is commonly called ‘instance segmentation’. The cell segmentation is a crucial step in Multi-exposures ovary analysis such as edges stage classification which assumes only single cell in an image and further creating atlases of locations of possible gene activations. This task became very popular in computer vision, and several methods based on Convolutional Neuronal networks (CNN) have been recently proposed. Unfortunately, all these deep approaches require a large amount of annotated training data which is always an issue in multi-exposure imaging where the annotated images are counted in tens not in thousands or millions as it is common in other applications as mentioned in [49].

This problem has several challenging aspects. First, the objects (edges) are highly structured and cannot be easily distinguished by texture or intensity due to high similarity between pairs of tissue classes (background—cytoplasm and nurse—follicular edges). Second, there are several edges in the image, often touching, with unclear boundaries between them as mentioned in [50]. Third, the algorithm should be fast, as there is a high number of images to be processed. Note also that identifying individual edges with mutually similar appearance is more challenging than a standard binary or multi-class segmentation.

The most straightforward way of segmenting Multi-exposures would be to apply region growing from a manually or automatically marked seed point in the approximate center of each cell until it reaches the boundary or touches a neighboring cell. However, the real boundary cannot be quickly recovered because of the various structures inside each cell. Template matching methods would suffer from the different appearance of the edges' inside, for example, the high number of various positions.

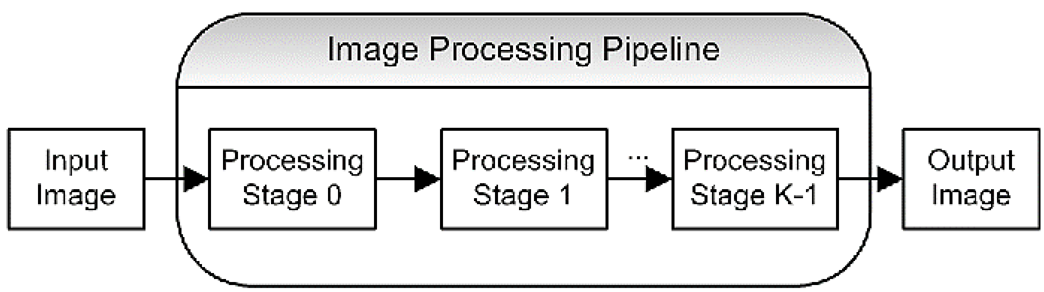


Figure 2.2: The general image processing pipeline for processing the image from input to output [50].

A texture-based structure segmentation method for cell chamber images was described in [51] but there is no simple way how to obtain a segmentation of individual edges directly. Unfortunately, the semantic segmentation is not sufficient for identifying individual cell chambers, for example, using watershed segmentation. Another standard approach is to post-process the foreground/background segmentation using mathematical morphology and connected component analysis, but in this case, it turned out not to be sufficiently robust due to the touching objects. Nevertheless, such tissue segmentation can be used as the initial step of our approach.

2.5.2 Transducer and Psychometric Functions

The image analysis with transducer is a critical phase in many multi-exposure and image processing applications and treatments. This thesis is focusing on multi-exposure imaging, namely in images processing to detect and isolate multi-exposures in edging of the functions. Compared to other applications the main characteristic of multi-exposure imaging is an appearance of noise, image distortion and a small number of annotated data. On the other hand, the importance of using some image processing methods and tools is increasing with the rapid growth of image resolutions captured with new technologies and amount of sensed images compare to stagnating number of human experts performing the analyses manually. It shows the high demand for (semi)automatic image analyses which would solve some tasks entirely independently or at least significantly speed up the process and assist to a human expert. In this thesis, we aim at automatic image analysis of microscopy images of Multi-exposures, which is an important study subject in genetic biology. The used processing pipeline starts with sensed images and ends by extracting an atlas of gene activations. The methods presented in this work are image segmentation on multi-exposures, estimation object centers, region growing on multi-exposures and final decomposition of a stack of segmented images into a small number or non-overlapping patterns. Note that similar pipeline can be used in other research and notably proposed methods can also be applied in other fields as given in [52].

2.5.3 Color Saturation

To define the multi-exposure color saturation, multi-exposures have different variant as they have characteristics to develop into various types of edges for the purpose of detection of disease as malignant or non-malignant and can generate completely acceptable results for some applications, without the need for manual segmentation because these multi-exposures are progenitor edges and have the ability of regeneration.

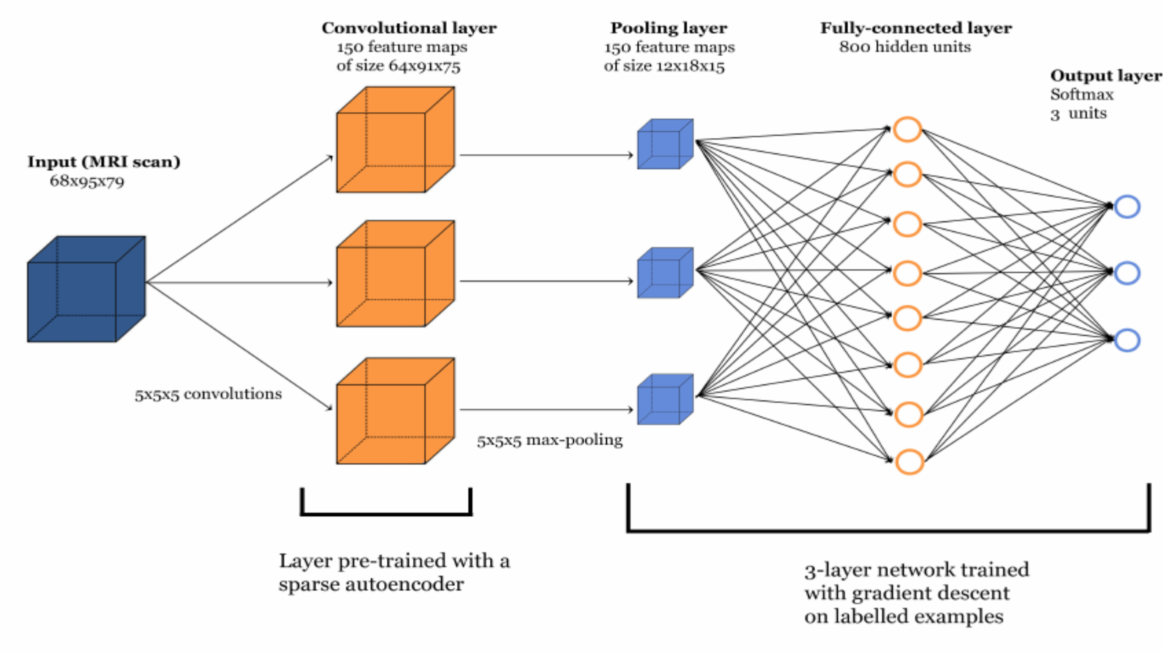


Figure 2.3:A robust pattern recognition using the convolutional neural network [53].

Researchers in [53] proposed a method for precise extraction of edge region which is based on deep learning methodologies. To lessen noisy artifacts, the input image, after pre-processing, was applied to CNN. CNN produced a segmentation mask which detected the area of the image edge. Using some post processing operations, the quality image of the mask is being further improved. Their input images were produced by customary cameras; henceforth, those were pre-processed in order to manage noisy artifacts. They used a filter to decrease the noise of the images as mentioned in [54]. They created two patches, one is local patch and another is global patch. The local patch is illustrated as a window around each pixel. It shows the local texture around the middle pixel as mentioned in [55]. Whereas, the global patch is used to show the global structure of the area. After that, both the local and global textures are sent into the proposed CNN architecture as the training input. The experimental outcomes showed that their suggested method can outpace other architectures to detect lesions.

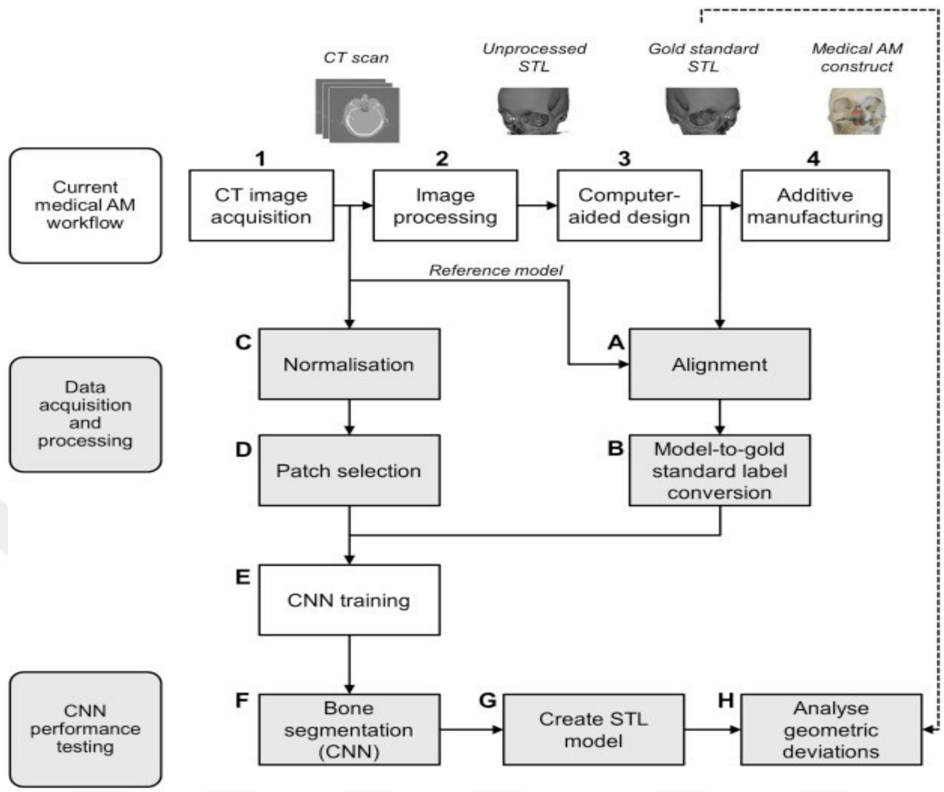


Figure 2.4: The disturbed intracellular signaling process for detection and segmentation sample with CNN [55].

Researchers in [56] proposed an innovative automated methodology to detect and recognize image edge using CNN. They performed three steps. The first step was to enhance the contrast of the images using first local Laplacian filtering with HSV color conversion. The second step was to extract lesion boundary using XOR operation of color CNN methodology. The final step was to extract the features by applying transfer learning by using ABC algorithm in order to feature fusion using hamming distance technique as mentioned in [57]. They introduced a feature selection technique by controlling entropy which was presented for the collection of the most discriminant types. Their proposed system was verified by different datasets. They achieved excellent accuracy using their proposed system. We need multi-exposures to detect and isolate because the multi-exposures are potential treatment for the diseases as well as they are Embryonic in nature for the treatment of diseases or any multi-exposure treatment. The research explains the need for new processing methods and formulate the related problems we face in this work. In the end, we summarize the main contributions of this work and present the structure of the thesis by referring to my publications.

2.6 IMAGE SMOOTHENING TECHNIQUES WITH HIGH-THROUGHPUT

2.6.1 Fluorescence Microscopy in Smoothening

Fluorescence microscopy can combine the advantages of high-throughput and high-content analysis. In the following, these two concepts are explained. In high-content analysis, rich information, such as high-resolution images, is obtained from each sample. Most recent advancements of high-content analysis include super-resolution microscopy and imaging mass spectrometry, which increase the amount of spatial information or number of labels measured in a given sample, respectively as given in [58]. In high throughput analysis, a larger number of samples are analyzed quickly, though not necessarily at a single cell resolution. Examples for high-throughput methods are traditional flow cytometry or fluorescence intensity measurement using plate readers that yield a fluorescence response for each well in a plate. Thereby, thousands of edges are contained in each well. Some techniques combine the advantages of high-content and high-throughput analysis. Capturing biological images is now highly automated, which enables methods that deliver rich information for a large number of edges at the same time. Figure 2.5 shows how robots are used to automate the handling of multi-well plates.

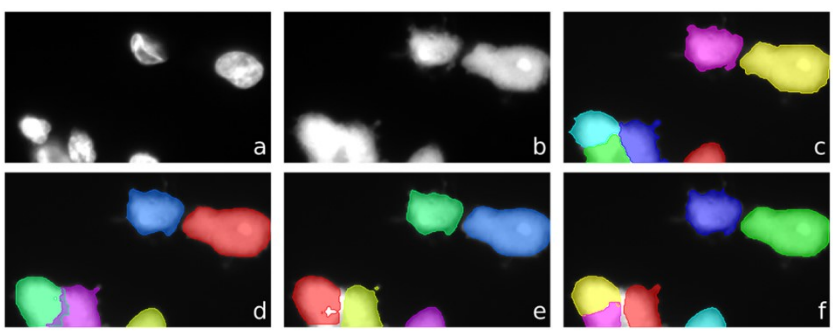


Figure 2.5: The disturbed fluorescence microscopy used in high-content analysis of images [58].

Imaging flow cytometry and fluorescence microscopy are examples for methods that allow high-throughput and high-content analysis at the same time. An imaging flow cytometer captures high resolution images of each single cell passing through a flow channel and thus using dyes, one can obtain spatial information about substances of interest. In fluorescence microscopy, high resolution microscopes capture the fluorescence response of multiple edges in one well at once. Images typically depict hundreds of edges. This method's spatial resolution is very high and thus morphological details are visible for each cell. This is why this method is used in morphological profiling. Compared to imaging flow cytometers, fluorescence microscopes provide higher resolution images, are more sensitive and allow higher magnifications. Cell profiling experiments based on fluorescence microscopy and the Cell Painting protocol which is a standard staining strategy for fluorescence microscopy experiments

2.6.2 Morphological Profiling

In this research [59], the author gives a more detailed introduction to morphological cell profiling with fluorescence microscopy images, which is the key application relevant to this research. The goal of each profiling experiment is to obtain a set of features for each cell that describes its phenotype. This set of features is called the cell's profile. Cell profiles can be analyzed and compared to each other. In a drug screening experiment for example, profiles of edges that were treated can be compared against profiles of edges in the control set to quantify important cellular changes. For example, cell profiles can reveal a cell's disease state or can be used for classification in cell states such as phases of the cell cycle or hematopoietic differentiation.

Profiles can be of very different kinds. Examples include expression profiles that quantify the transcription of genes and morphological profiles that quantify the shape of the cell and its compartments. Fluorescence microscopy is an important tool in morphological profiling and captures the images used to obtain morphological cell profiles.

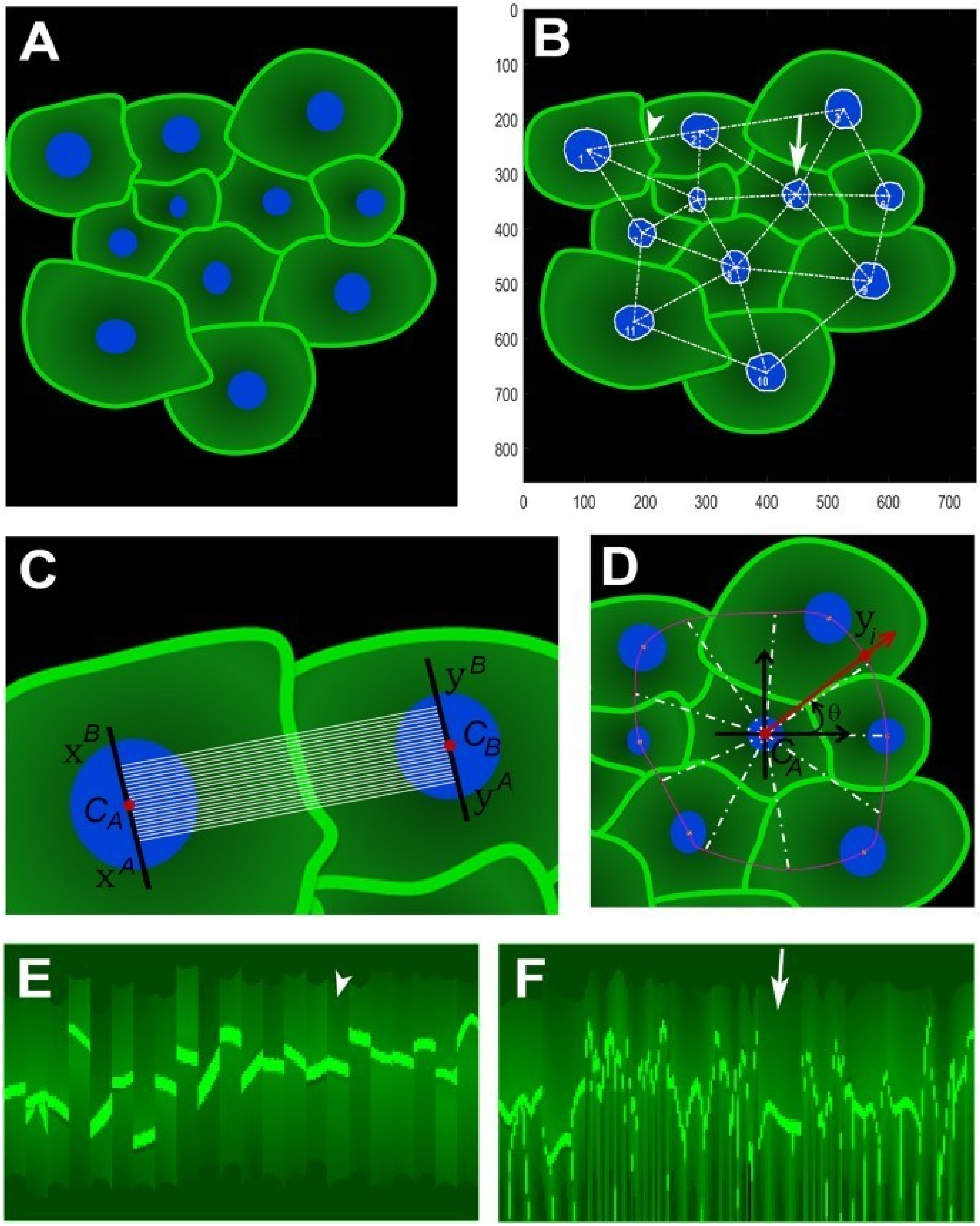


Figure 2.6: Types of morphological profiling that separates the foreground and background in series of steps [59].

As mentioned earlier, the goal of this project is to improve the segmentation of nuclei in the DNA channel. We compare the performance of our segmentation algorithms against the current segmentations obtainable with image fusion which is the standard approach. Image fusion does a combination of semantic and instance segmentation for nuclei as mentioned in [50]. It classifies pixels into foreground and background pixels and then assigns the foreground pixels to an object instance. Image fusion current implementation for segmentation is implemented in the module Identify Primary Objects and consists of three steps:

- Using thresholding, foreground regions are identified. Thus, image fusion distinguishes between pixels that belong to nuclei and background pixels.
- Clumped nuclei are identified and segmented. For the segmentation, a seeded watershed algorithm is used. The user can choose to either apply it on the distance transform of the thresholded image or on pixel intensity values. Other traditional computer vision methods such as image smoothing are used to improve the segmentation results.
- Nuclei whose features do not fall into user-specified ranges are discarded. For example, image fusion offers to discard nuclei that are close to the image's boundary, or nuclei below a certain size.
- The segmentation quality of this approach is highly affected by the choice of parameters for the used algorithms. Thus, the user's input is required for choosing the parameters. The interface for the parameter selection of the module.

For parameter optimization, users usually start off with a given set of parameters and adapt these empirically based on the obtained segmentation. The user can then optimize the parameters in an iterative process as mentioned in [60]. As ground truth annotations are usually lacking, the user has no quantitative segmentation quality measure and is relying on visual inspection only.

3. METHODOLOGY

It is notable that the image fusion has successively used deeper networks but advanced deep learning techniques will provide great efficiency due to its architectural flexibility with artificial bee colony algorithm. Having more data models possess greater representational power making it possible to approximate more complex functions. Also, fast development in the field continuously introduces new techniques and architectures expanding the toolbox for design. However, science has always embraced simplicity, and it is reasonable to argue that the network should be kept as simple as possible. A study has been using the CNN technique even questioned the common practice of using pooling and sophisticated activation functions in machine learning. Instead, they suggest using just recurrent layers with properly selected stride and filter size parameters.

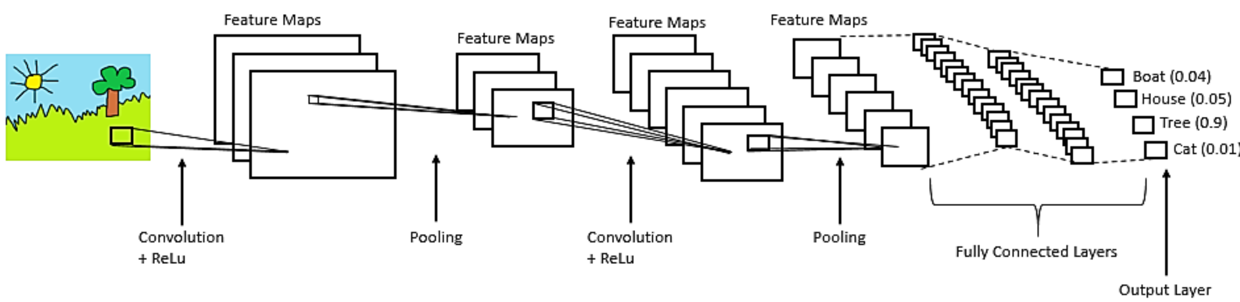


Figure 3.1: Proposed system depicted by flowchart using CNN.

3.1 DATASET DESCRIPTION

The dataset has been acquired from an open-source repository known as the TNO image fusion dataset [61]. The physical appearance (shape, size, texture, color, etc.) of objects in images can vary depending on many physical and human geographical factors, both from the ground perspective as well as in satellite imagery. Possible sources of variation include the cultivated species, plant status, topography, soil properties, weather conditions, cultivation methods and other human influences. Due to such local variations, the task of accurately differentiating between several field objects can become ambiguous. This makes it difficult to create robust algorithms that recognize all kinds of field instances without overfitting to the specific scene. In addition to the variation within a single scene, satellite images can show even stronger intra-class variations. This applies to images from different agricultural regions, from different dates or growing seasons, or under different atmospheric and illumination conditions. Also, in view of increasing public availability of high quality geo datasets, the use of additional and more heterogeneous ground truth and satellite image data (e.g. from multiple study areas in parallel) could improve the general model accuracy, robustness and transferability. Additional data augmentation (e.g. random scaling, distortions, color offset or jitter, introduction of image noise) might also help mitigate overfitting and increase the robustness to different geographic settings as well as atmospheric and lightning conditions. Figures above shows images under various exposures in all sequence. More details concerning every one of the sequences, like name, number of source images, and spatial resolution have been stated in the caption of figures.

3.2 CONVOLUTIONAL NEURAL NETWORK

In this architecture, the CNN are responsible for feature extraction, where the features to which they will respond are determined through a learning process of the variable input connections. Simple features, such as specific orientations of lines and edges, will be extracted by lower-level CNN models, while higher-level CNN models extract more global features (e.g. parts of learned patterns). The CNN model receive their inputs through fixed connections from CNN models in the previous layers and are responsible for a robust pattern recognition (i.e. decreasing sensitivity to a deformation or location shift of a pattern). Each CNN model receives inputs from a number of CNN models (’the input window’) in one cell plane, i.e. from CNN models that extract the same feature, but at different spatial locations. The models will fire, if it receives an input from at least one of the models. Consequently, the feature will be detected even under a shift in the location of the input features, rendering the system less sensitive to the exact feature locations. The behavior of the models can however also be interpreted from an alternative point of view. The input windows for the different models strongly overlap thus models can be considered as performing a spatial blur on the excitatory signals they receive from the models. This spatial pooling is obtained by averaging these signals from the input window, which are models that perceive the same feature at slightly different locations. Furthermore, the excitatory cell input window is often framed by a small inhibitory region. This enables the models to distinguish a continuous line from the end-point of the line and allows the network to still recognize features after blurring. A consequence of the fixed input windows of the models is a reduction in size of the CNN-planes, thus allowing the network to compress data as it progresses to higher stages.

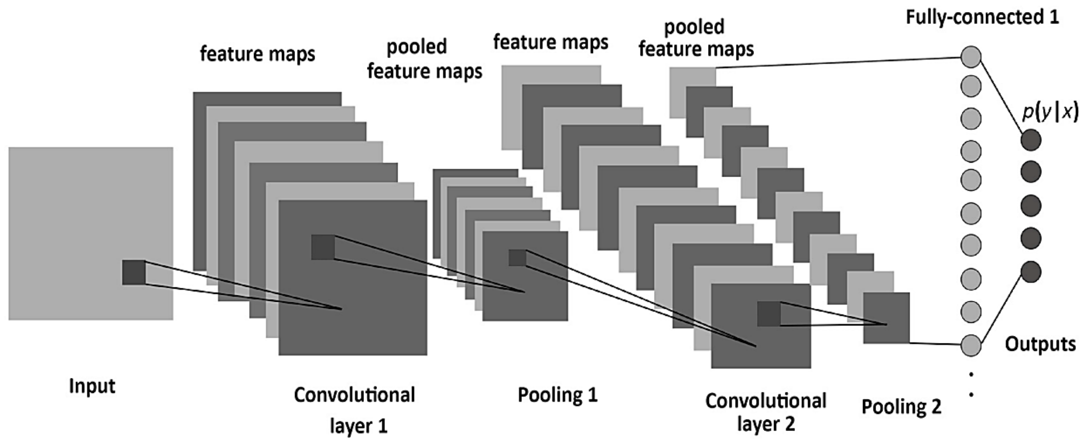


Figure 3.2: Illustration of a deep convolutional neural network.

CNN stands for convolutional neural networks originated from this work whose network has a similar architecture cognition. It translates the behavior of the models in mathematical formulations, respectively as convolutions and subsampling. The network primarily distinguishes itself however from the cognition in its training process. Whereas the neo-cognition relies on a layer-wise, unsupervised training process of the cell layers and supervised training of the output layer, CNN is trained to find a global minimum over all the parameters through back-propagation. The key notion equation that mimic and render CNN's feature fusion for image smoothening is given by (1) from [62].

$$s[t] = (x * w)[t] = \sum_{a=-\infty}^{a=\infty} x[a]w[a+t] \quad (3.1)$$

Where the 's' represents the fusion feature maps, 'x', 'w' represents the input parameters, however the remaining equation represents the CNN kernel for smoothening. Due to the high density of research in image illumination estimation, multi-exposure images often appears clumped. Thus, the main challenge for this research is to find the patch-wise solution for illumination estimation with human interference with CNN technique by separating different instances of multi-exposure fusion images. Traditional segmentation methods such as the seeded watershed algorithm, which is implemented in fusion images, are slow and sometimes make errors. In addition, the user needs to be experienced in order to use these algorithms as parameter tuning is required. The goal of this research is to overcome these dominant problems in the patch-wise illumination estimation for multi-exposure image fusion. We will use a convolutional neural network for multi-exposure image fusion images. We will train the neural network on hand-annotated images and evaluate its performance with object-based metrics.

3.3 POWERFUL CNN'S FOR IMAGE EDGE SMOOTHENING

3.3.1 Key Concepts of CNN's

Convolutional layer as images contain highly correlated 2D information (generally in three RGB-channels), the idea of local connections has been exploited by creating units in a convolutional layer which receive weighted input from local patches in the previous layer, referred to as their receptive fields (RF's). By relying on local RF's, this introduces sparse connections, rather than the whole-scale image statistics passed on to units and causing high-dimensional problems in fully-connected feed-forward networks. The weight vector of a unit is also known as filter bank and through training the weights will be updated to filter for specific features, giving the units the behavior of filter detectors. The weighted inputs and additive bias are then passed through the activation function of said unit, where ReLU's have portrayed much faster learning in deep networks compared to other activation functions.

Furthermore, images tend to portray similar features at different spatial locations. Accordingly, units with receptive fields at different locations will have the same weight vectors to extract similar features, which is known as parameter sharing and further reduces the dimensionality of the problem.

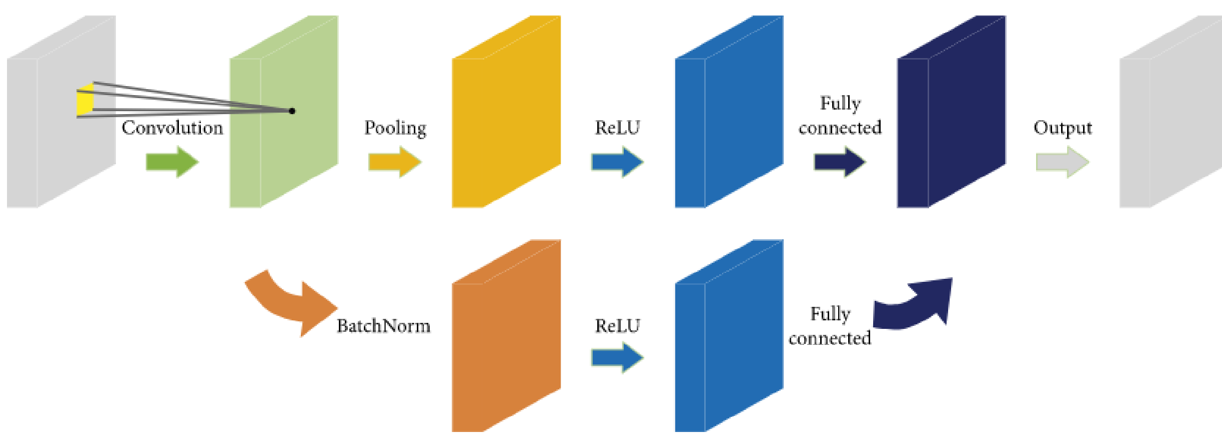


Figure 3.3: Basic building blocks of convolutional neural network.

On a structural level, the outputs of units with the same weight vector are organized in feature maps. Multiple feature maps (with different weight vectors) in each convolutional layer enable scanning for different features in each local image patch. The optimal number of feature maps needed to capture the image dynamics tends to increase with the complexity of the input images.

3.3.2 PoolingLayer

The feature maps in the convolutional layer serve as input for the next stage, being a pooling layer. Given that the relative location of features with respect to each other contains significantly more information than their precise pixel-location, further stressed by the notion that small shifts in the precise feature locations are to be expected in different images, the idea of local pooling of features to reduce spatial resolution is implemented at this stage. This results in a network invariant to small shifts in the images and the stages of convolutional/pooling layers equip the network to handle variable size inputs.

Pooling units will scan the feature maps with (non-)overlapping p1 x p2 receptive fields and output a single value. Typical choices for pooling units are subsampling, where the output of a unit in the jth layer is the average of units in the receptive field, supplemented with additive bias b and passed through an activation function and max pooling.

3.3.2 Deep Hierarchies

Finally, the use of many layers represents the different stages in the simplified model of the ventral stream. Features at higher-level stages results from combinations of lower-level features, thus enabling deep structures to capture the compositional hierarchy of natural images.

3.4 TRAINING AND VALIDATION

An implementation is based on hybrid models of CNN and ABC algorithm from various approaches. The visualization method proved to be helpful in interpreting the predictions, although it produced different results than expected. The expectation was that the CNN with ABC model would learn to identify small signs of skin disease risk. Instead, with custom CNN models they learned wider patterns that possibly predict resistance to fractures. This can of course be seen as the other side of the same thing. The experiments show that deep learning techniques can be used with imaging techniques to produce predictive ability different from the standard clinical methods currently in use. However, due to limited amount of data and wide range of architectural possibilities, their true potential in this context remains to be unfolded.

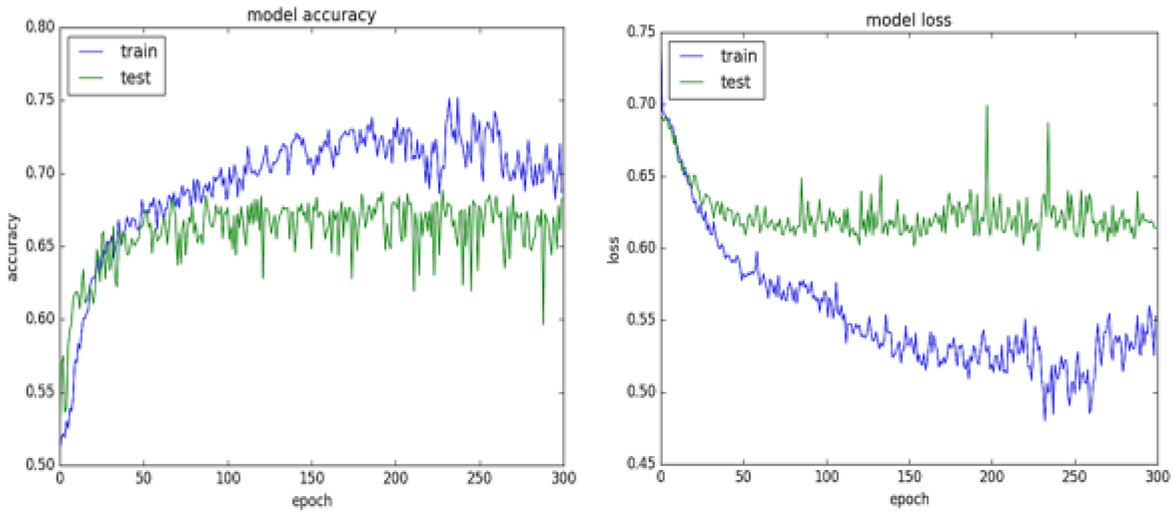


Figure 3.4: The ABC+CNN Model Accuracy (accuracy vs epochs) Model Loss (loss vs epochs).

Given the current hype of deep learning, optimizations of CNN architectures and designs is in hot pursuit. This section will therefore be devoted to some of the powerful CNN models that have been developed in the past few years and will be used throughout this work. Specific, revolutionary design-elements that have assisted these improvements are marked throughout this discourse. It is also likely that the data analysis alone cannot provide the best fracture risk prediction, regardless of the model. It could contribute to more comprehensive prediction method that utilizes a wider set of input data to cover different aspects of the image edge smoothening. Using hybrid deep learning techniques produced the best image edge smoothening prediction results in the experiments. The results support the advertised benefits of hybrid deep learning, such as faster training speed and suitability for larger datasets.

3.4.1 Cross Validation

The cross validation is used to validate the data in artificial intelligence based systems, interpretable machine may even teach humans about how to make better decisions. To address the transparency problem, visualization techniques can be used to identify the image regions most important for the model's prediction. It is also possible to highlight the shapes and textures the network sees by visualizing pixel-wise impact on prediction score. Another visualization technique is called guided backpropagation.

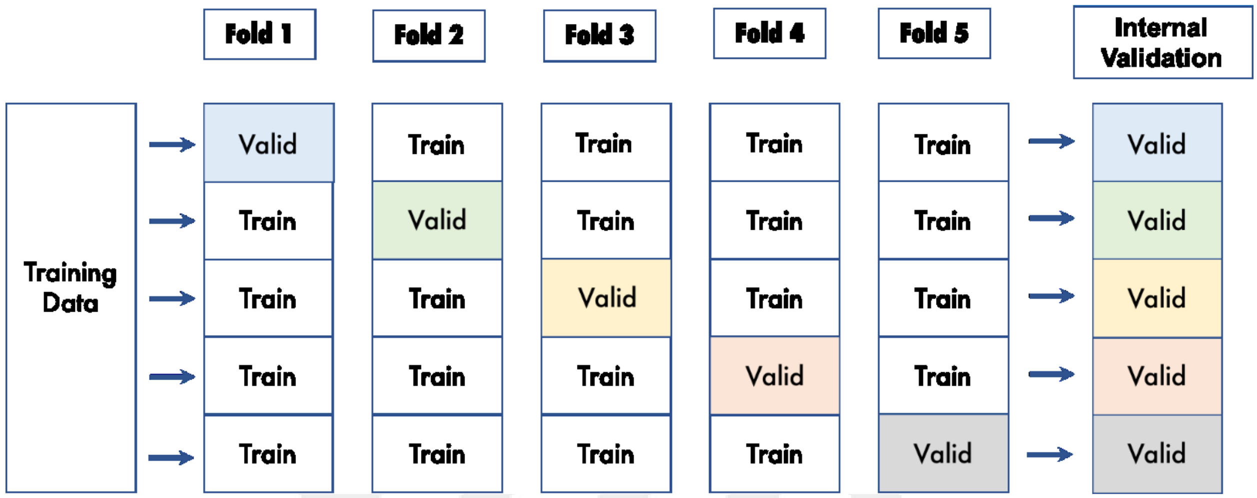


Figure 3.5: An example of dividing dataset for cross validation.

It produces an illustration of the detected features that contributed to the image edge smoothening prediction. This is done by taking account only the neurons that had a positive effect on the output, while others are set to zero. Combining different visualization techniques can produce effective visual explanations for the CNN model's predictions. To gain better coverage of the problem space, the number of training samples can be artificially expanded. This method called cross validation means modifying samples with random transformations such as rotation and flipping, varying properties like brightness and contrast or by adding random noise.

3.4.2 Training and Validation

As deep learning techniques take numerical data as input, image analysis might seem like a simple case of feeding the network with raw image data. After all, images are nothing but grids of numbers representing pixel intensities. Actually, this approach can work well for simple recognition tasks with small, normalized images. The problem is poor scalability. When the input size and number of layers' increase, the amount of required computation explodes.

Table 3.1: The number of epochs per training instances

Number of Epochs	Loss	Accuracy	Validation Loss	Validation Accuracy
25	0.3301	0.8492	0.6911	0.5360
50	0.2418	0.9047	0.6942	0.5360
75	0.1891	0.9350	0.7985	0.5360
100	0.1702	0.9426	0.9698	0.5379
125	0.1384	0.9606	0.4709	0.7803
150	0.1246	0.9625	0.3569	0.8220
175	0.1020	0.9768	0.3543	0.8295
200	0.0809	0.9829	0.3598	0.8371
250	0.0747	0.9853	0.3662	0.8295
300	0.0678	0.9867	0.3709	0.8314

3.5 IMAGE PRE-PROCESSING IN MATLAB

Possible sources of variation include the smoothening techniques, filter status, topography, filter properties, smoothening conditions, segmentation methods and other classification influences. Due to such local variations, the task of accurately differentiating between several field image filter can become ambiguous. This makes it difficult to create robust image processing algorithms that recognize all kinds of field instances without overfitting to the specific scene. In addition to the variation within a single scene, satellite images can show even stronger intra-class variations. This applies to images from different filter regions, from different dates or growing seasons, or under different atmospheric and illumination conditions. Also, in view of increasing public availability of high quality fusion image datasets, the use of additional and more heterogeneous ground truth and fusion image data (e.g. from multiple study areas in parallel) could improve the general model accuracy, robustness and transferability. Additional data augmentation (e.g. random scaling, distortions, color offset or jitter, introduction of image noise) might also help mitigate overfitting and increase the robustness to different fusion settings as well as image lightning conditions.

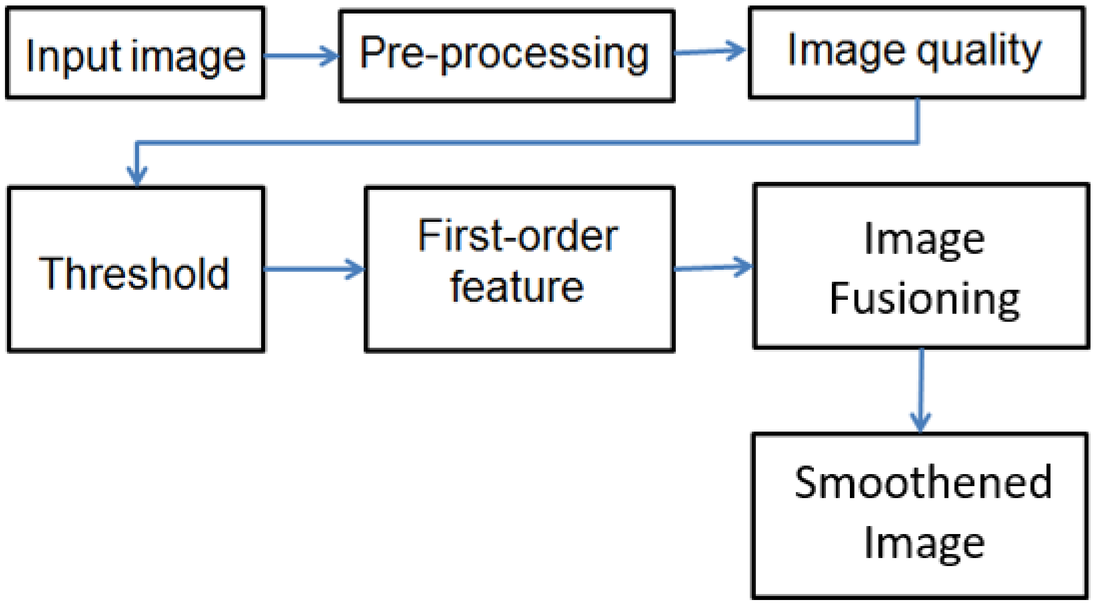


Figure 3.6: The flowchart approach for smoothened image by series of steps being followed.

In the proceeding discussion, the image fusion will serve as a guideline to discuss the most poignant results obtained with CNN's with indication of some remarkable new elements. To give an indication of the performance of these models, the top-5 error rate on the ImageNet dataset of the aforementioned nets is presented. It has to be noted that the submitted nets, training and testing schemes offer a myriad of other hyper-parameters, which were tuned to achieve these results (e.g. a 7-network ensemble of CNN-architectures).

3.6 ARTIFICIAL BEE COLONY (ABC) ALGORITHM

The research focuses on compiling the delineate details and classifies image fusion with edge smoothing using artificial bee colony algorithm and convolutional neural network while distinguishing between image instances of the same class. Instance segmentation lies at the intersection of edge detection (predicting the bounding box and class of a variable number of images, but not segmenting them) and semantic segmentation (labeling each image pixel with a semantic category label, but not distinguishing between edges of the same category). With semantic segmentation it would be possible to find single instances of ABC algorithm on fields if they were completely surrounded by areas of different image classes.

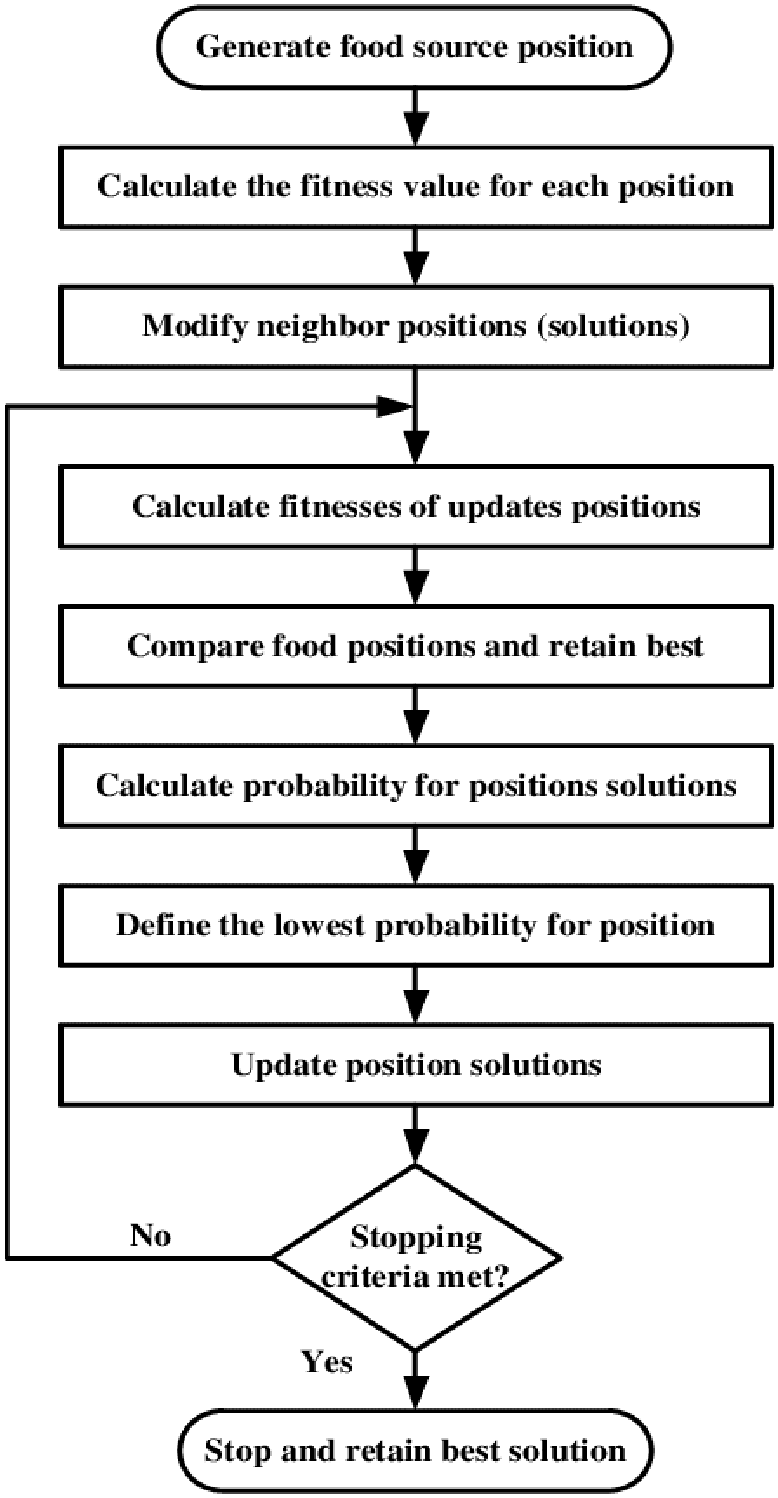


Figure 3.7: Artificial bee colony flowchart that is being followed in this work.

These datasets usually image fusion from manual tracing of field boundaries. However, manual tracing of image edge smoothening is heavily dependent on the used imagery and supplementary data. It also is a highly subjective task, inevitably leading to inaccuracies and ambiguities depending on the priorities of the operator. The ground truth dataset can be complemented by existing property parcel information that could be indistinguishable based on the imagery alone. Furthermore, a single field image edge smoothening can potentially show several subfields due to unreported land use changes within the growing season. Given these constraints, even a very powerful automated image processing algorithm can only become as good as the ground truth data but will never be in perfect alignment with reality. Although the model can generalize within certain limits (local generalization), its transferability is mainly limited by the variability of the training data. However, filter field are often directly adjacent to each other and have touching boundaries. In this case, edge segmentation would yield the image edge smoothening. Same filter instance segmentation is able to distinguish between these connected or overlapping instances of the same class. Deep learning has not reached mainstream adoption in the remote sensing community yet.

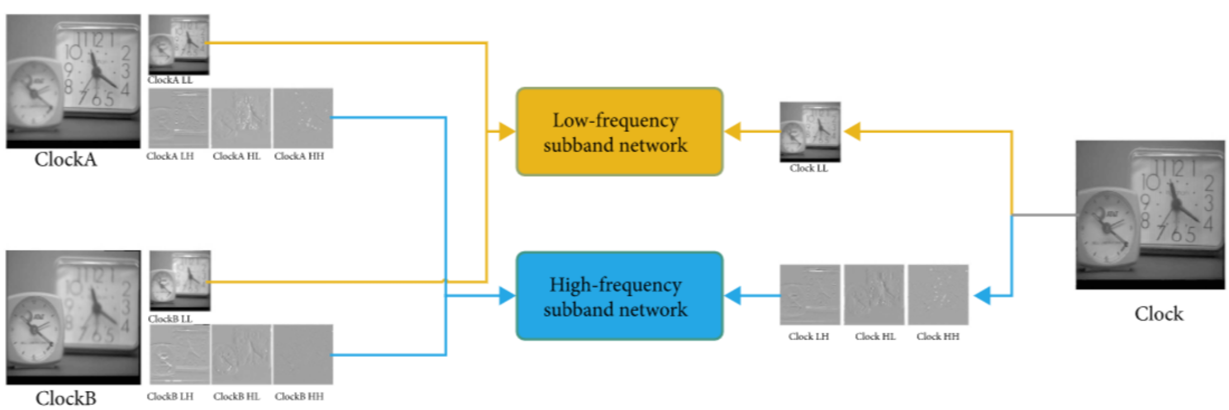


Figure 3.8: Working performed with low frequency sub-bands vs high frequency sub-bands.

3.6.1 Edge smoothening on RAW Image

The RAW image is basically a buffer of intensities per pixel. For this reason, the image at this stage is very dark and not suitable for a computer vision application. When the image is fusion, each pixel is mapped to a three-channel RGB representation with red, green and blue channels.

The ABC and CNN algorithm determines the intensities of the other channels, based on information from neighboring pixels. This is a loss stage, as information in terms of bits per pixel is being lost. The image fusion has a huge tone, which is due to the Bayer mosaic pattern. In the Bayer mosaic, green is the dominant color as the color arrays in the image sensor are placed in rows of green-blue and red-green. As can be seen, color transformation corrects for that hue tone. The effect of gamma correction is not directly visible when compared to the output of the color transformation stage, but adds to the highlights and black tone appearance of the final result. The effect of the tone mapping stage is clearly visible, as it makes the image pleasant to the human vision, by fixing the brightness of different parts of the image. Furthermore, the training of the aforementioned deep CNNs generally required multiple GPU's and was a time-consuming process (up to 3 weeks). Thus, it is clear that training the same, high-performance CNN architectures on real-life data is in most cases unfeasible. The CNNs trained to recognize the ImageNet dataset have nevertheless a strong capability to extract very distinctive features from natural images, rendering them useful for datasets of other images. This is supported by the idea that the lower levels of these architectures extract general features from the images and become increasingly more task-specific deeper in the network. From this the idea of transfer learning sprung forth, which aims to kick start the process of the random weight initialization by using the pre-trained weights of deep CNN's to facilitate the training process.

3.6.2 Key Notions Concerning Image Fusion and Smoothening

The ABC and CNN algorithm learning entails training a source network on a source dataset for a source task, transferring the learned knowledge (e.g. features or weights) to a new network for a different learning task and/or on a different target domain.

Transfer learning is used with the objective of increasing performance on the target task by exploiting the trained knowledge of the source task. It can either serve as a 'Kick-starter' for the process improving initial performance, increase final performance or speed up the training of the target task. Three main types of transfer learning can be distinguished with respect to similarities in the source/target domains/tasks: (1) inductive transfer learning with the goal of inducing a generalizable, predictive model for the target task from a different source task; (2) transfer learning where a model should be derived in the target domain for the target task lacking labeled data, based on the source task containing a substantial amount of labeled data; and (3) unsupervised transfer learning which targets unsupervised learning in the target domain based on a different source task.

For purposes of image classification, the approach of inductive transfer learning has gained momentum in the past years. Given the large training time and requirements on labeled data, architectures are pre-trained on the ImageNet (source domain) for 1000-class classification purposes (source task) and knowledge is transferred to a labeled dataset of choice (target domain) for e.g. classification as target task. Thus, it primarily exploits the learned feature-representations in the source domain and aims to transfer these to the target domain with the aim of improving classification performance whilst making use of deep, data-greedy CNN architectures.

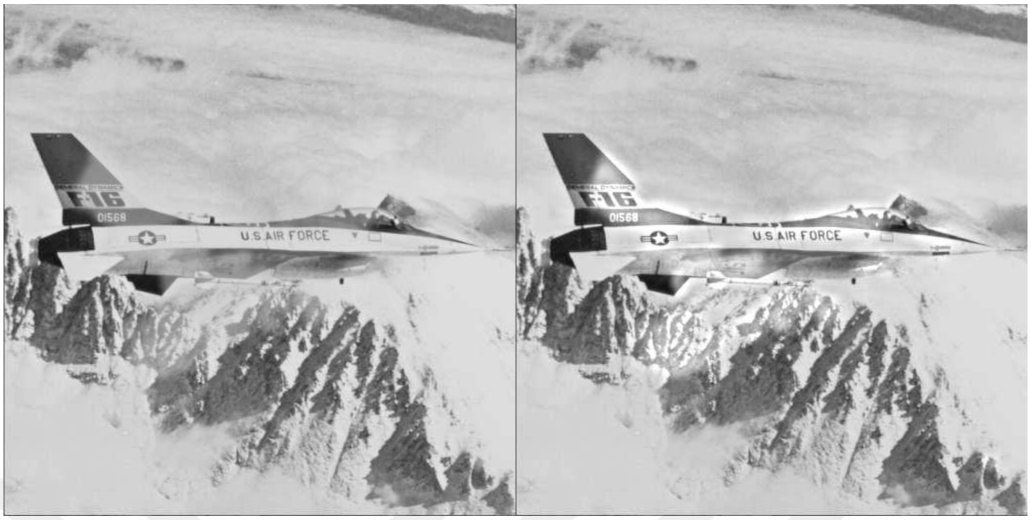
The transferability of different levels of features was touched upon in this work with results confirming that the deepest layers are most task-specific and better results can be obtained by using features from levels upwards in the model. This was confirmed by the work, in which furthermore was suggested that pooling of features as well as combinations of features may improve performance. We added to the evidence by documenting higher performances for CNN's initiate with pre-trained weights and fine-tuned for the task at hand, rather than through random weight initialization and training from scratch. Furthermore, they investigated the transferability of specific feature layers and reported co-adaptation of intermediate feature layers, where splitting between frozen and trainable layers at this level might result in a drop in performance.

4. RESULTS

The goal of each image fusion patch-wise illumination estimation experiment is to obtain a set of features for each image that describes its sequence. This set of features is called the image profile. Image profiles can be analyzed and compared to each other. In a patch-wise illumination estimation experiment for example, illumination estimation of images that were treated can be compared against profiles of images in the control set to quantify important matrix changes using CNN. For example, image patches can reveal an image sequence state or can be used for classification in image states such as phases of the image cycle or hematopoietic differentiation. Patch-wise illumination estimation using CNN can be of very different kinds. Examples include expression profiles that quantify the transcription of genes and morphological profiles that quantify the shape of the image and its compartments. Patch-wise illumination estimation is an important tool in morphological profiling and captures the images used to obtain morphological image profiles.

(a) Input Image

(b) Gaussian Noise Image



(c) Background Focused (d) Foreground Focused

(e) Image Fusion

Figure 4.1: Image fusion with smooth edges in an example of occluded jet and cloud.

The behavior that noticed in this figure, was the local contrast which is rapidly reduced as soon as the under-exposed image has been started to capture with foreground and background focused. Secondly, slight variation and robustness with entropy measure detected. furthermore, at the over-exposed image the CNN feature start to increase and gradually decreased. the other features ranging between (no affected to light effect) with the light condition. The main challenge for this research is to find the patch-wise solution for illumination estimation with human interference with CNN technique by separating different instances of multi-exposure fusion images.

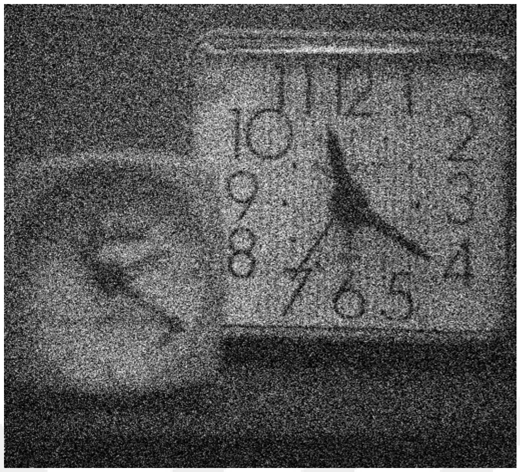
Traditional segmentation methods such as the seeded watershed algorithm, which is implemented in fusion images, are slow and sometimes make errors. In addition, the user needs to be experienced in order to use these algorithms as parameter tuning is required.

(a)Background focused

(b) Foreground focused

(c) Image Fusion

(d) Distorted Image



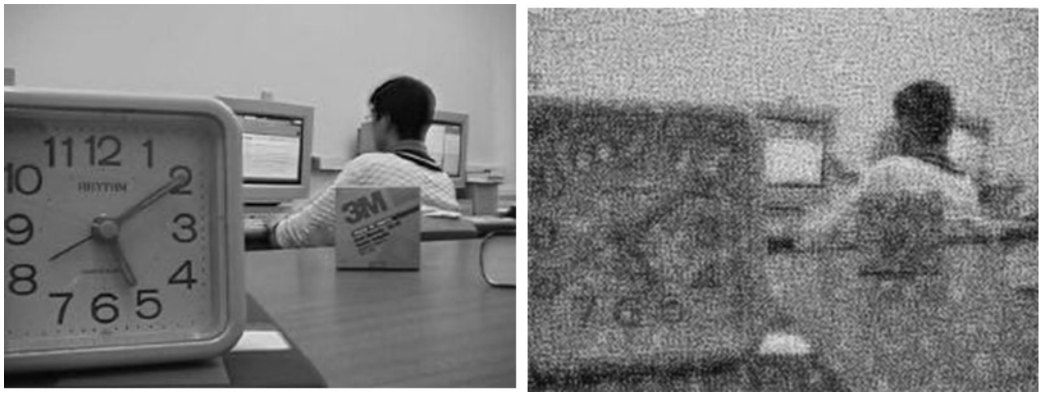
(e) Gaussian Noise Image

Figure 4.2: Image fusion with smooth edges in an example of two parallel placed table clocks.

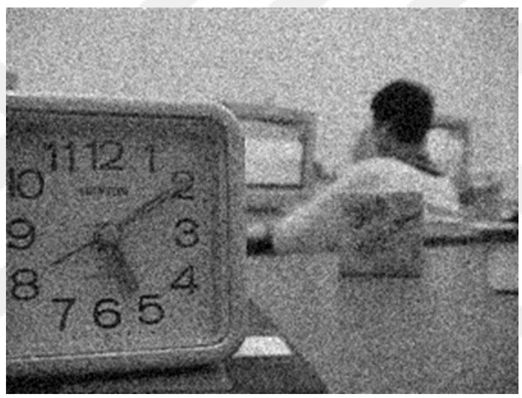
The goal of this research is to overcome these dominant problems in the patch-wise illumination estimation for multi-exposure image fusion. We will use a convolutional neural network for multi-exposure image fusion images. We will train the neural network on hand-annotated images and evaluate its performance with object-based metrics.

(a) Foreground focused

(b) Background focused

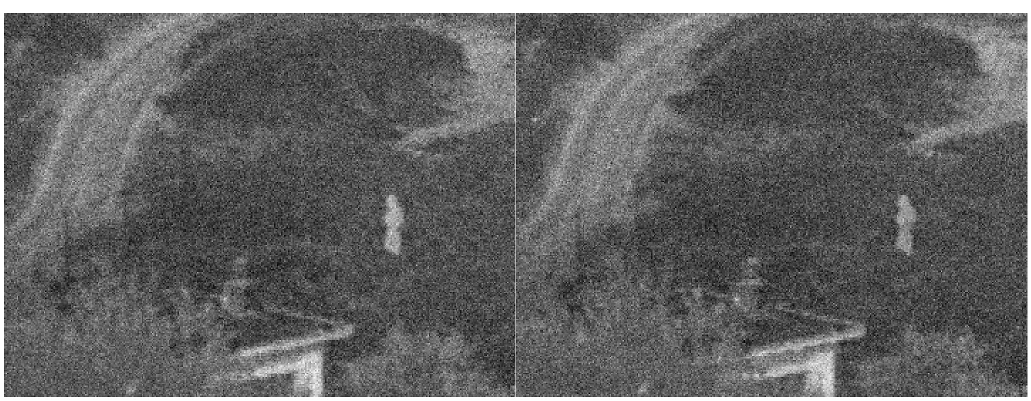


(c) Image Fusion (d) Distorted Image

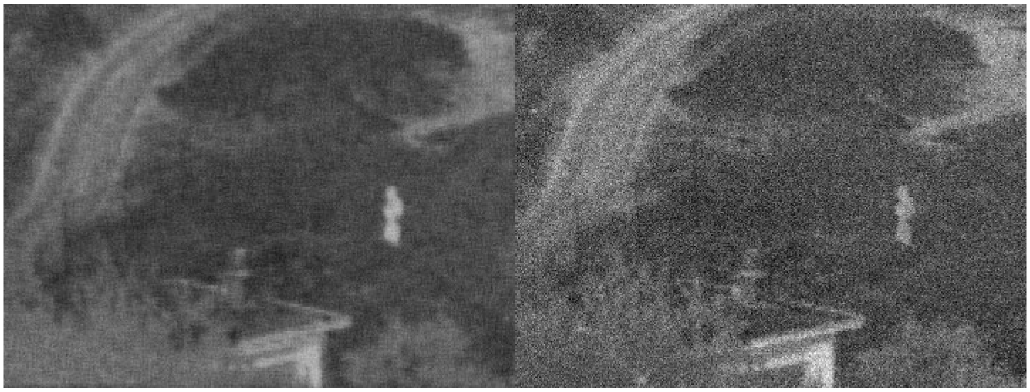


(e) Gaussian Noise Image

Figure 4.3: Image fusion with smooth edges in an example of table clock in foreground and person on a computer in the background.



(a) Foreground Focused (b) Background Focused



(c) Image Fusion (d) Distorted Image

(e) Gaussian Noise Image

Figure 4.4: Image fusion with smooth edges in an example of off-road path sidelined by ground where person is standing.

Research has refine and finish the recursive filtering to incorporate multi-exposure fusions, all things considered, and impediments for order utilizing Convolutional Neural Network (CNN), which the classifier had the option to perceive just by taking a classifier at a solitary image. This was significant so as to do address patch-wise illumination for multi-exposure image fusion.

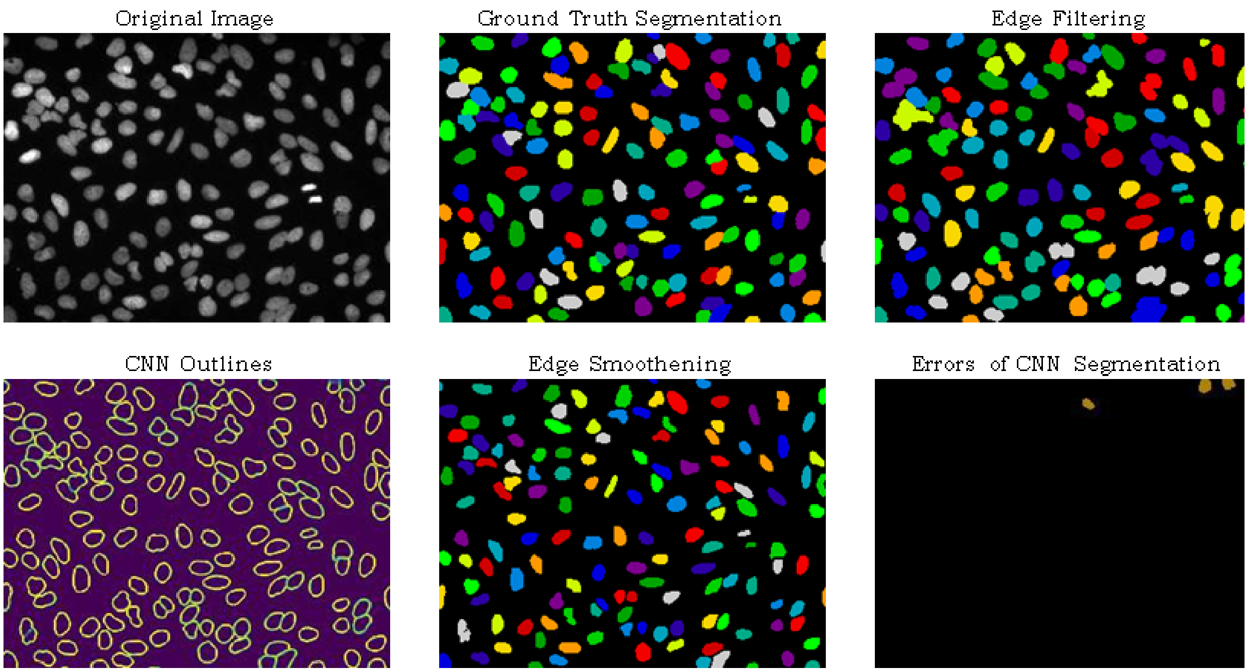


Figure 4.5: The multi-exposure image fusion for edge smoothening comprises of different steps on samples.

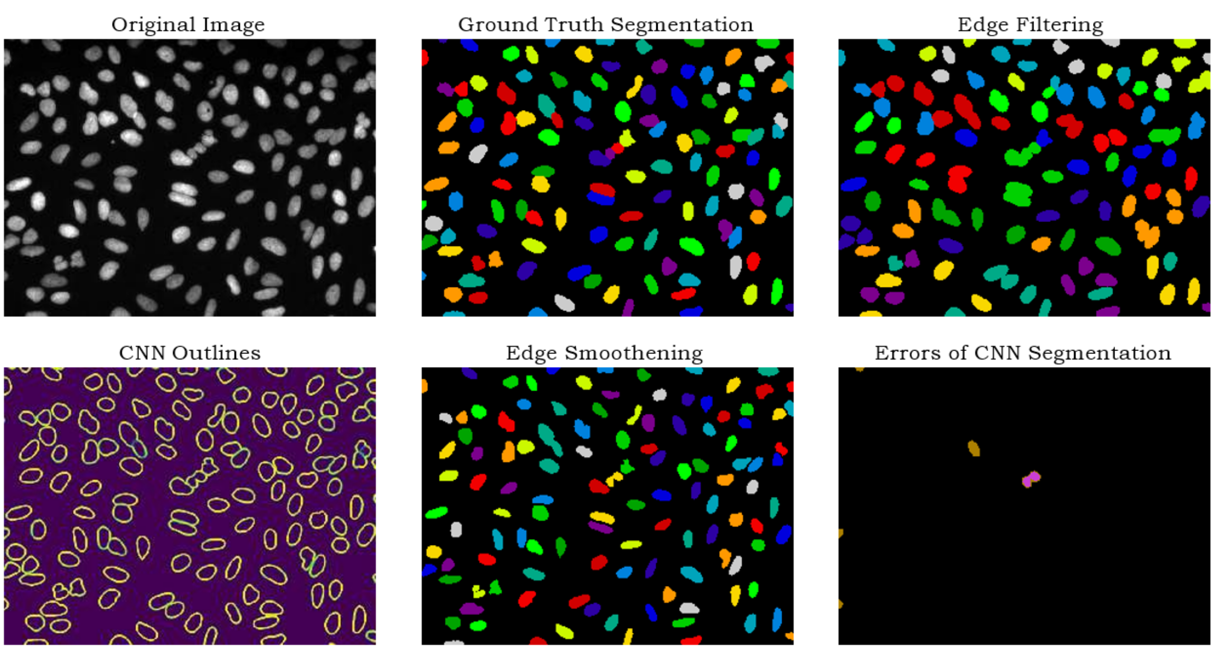


Figure 4.6: The multi-exposure image fusion for edge smoothening comprises of different steps on samples.

5. DISCUSSION

This work has shown the impact of image fusion algorithmic approximation to the quality-of-control for image-based control systems for smoothening. For the analysis, we used the use-case of a lateral controller that performs lane keeping for an autonomous vehicle. The application that performs the image signal processing, lane detection and control computations, was developed in the high performance MATLAB programming language, with the programmed in the domain specific language Halide. The simulations were run on the CNN simulator, using a vehicle and two separate tracks, namely a straight and a curved track. The application was made error-resilient by adapting the lane detection algorithm to be able to process the degree of approximation that the different approximate ISP versions required. We ran this application on an Intel i7 processor, utilizing the available parallelism of the application through MATLAB on the eight available threads distributed in four cores. We, initially, benchmarked the application by conducting careful and detailed profiling on a total of 8 different pipeline versions, each of which on a dataset of 200 images that were obtained in CNN. Then, we evaluated the degradation in quality-of-control that is caused by approximation, without taking into account the impact of approximation on runtime performance. Finally, the runtime performance gains due to approximation was taken into account and the improvement in quality-of-control was quantified. We used two separate metrics, namely the settling time and sum of squared errors, to evaluate the performance of the controller. Additionally, we used the energy, memory footprint and sensor-to-actuator delay to measure the improvement in the application performance.

Using this approach, we shed light on how real-time ABC algorithm can benefit from approximate computing. The approximated CNN pipelines achieved a maximum speedup of factor 3.5 for the sensor-to-actuator delay and a 40% improvement in settling time. The overall performance that was achieved by the multi-dimensional application was 50% better than the baseline. We showed the improvement of the application in metrics such as energy and memory footprint, which suggests that approximate computing is an efficient approach, when there are more constraints for the application, such as in an embedded environment.

We suggest a new, compact representation of patch-wise illumination estimation for multi-exposure image fusion estimation, which is independent of the camera (image projection matrix).

It utilizes the projections of the 3D image sequencing to the input image as opposed by putting away image boundaries utilizing the CNN model. This makes the indicator equipped for arranging patch-wise illumination estimation in images from input image applying the weights to the pre-processed images through feature extraction. The combination image position estimation would then be able to be reproduced when the right image projection framework and the ground plane condition.

Table 5.1: The comparison of proposed work with literature that is already available

Article	**Technique**	**Accuracy/Precision**
[63]	Convolutional Neural Network	96.10%
[64]	Illumination Estimation	89.76%
Proposed	ABC + CNN	98.17%

The examination will make commitments to the patch-wise estimation remuneration and multi-exposure imaging, which comprises of testing images of light in different stances from arrangement. We will refine and finish the recursive filtering to incorporate multi-exposure fusions, all things considered, and impediments for order utilizing Convolutional Neural Network (CNN), which the classifier had the option to perceive just by taking a classifier at a solitary image. This was significant so as to do address patch-wise illumination for multi-exposure image fusion. In addition to this classification problem, we also framed the outline prediction formulation as a regression model. In this problem, we allowed the annotations to be continuous instead of categorical and thus represent a probability with which a pixel belongs to a nucleus' outline. The closer a given pixel is to the boundary, the higher we set its probability to belong to it. This boundary was created in a preprocessing step as a linear combination of the boundaries with two, four, six and eight pixels. The coefficients of the linear combinations were chosen in such a way that the probability of a pixel belonging to the outline follows a Gaussian bell in the distance of the pixel to the boundary. It is thereby assumed that the boundary is in between the two pixels of the boundary of width two. The Gaussian bell was scaled such that the two pixels adjoin to the boundary had a probability of 1 to belong to the boundary. For this model, we keep the setup as previously defined for the classification problem. The only difference is that the mean squared error is used as a loss function.

- An intuitive approach is to threshold the original image and subtract the thresholded probability map and use the remaining foreground pixels as segmented nuclei. This would separate clumped nuclei if the pixels in between them are correctly classified as boundary. However, this would also classify debris and other artifacts that are assigned as fore- ground by the thresholding method and correctly undetected as nuclei by the deep learning method.

- A second method we considered is finding connected components among pixels that are not classified as boundary and filtering the largest connected component as background. This method was unsatisfactory because it failed for cases where the background was not connected, which is often the case at the image's boundaries.

- We implemented a method that is very similar but uses a more involved filtering strategy. Among all connected components in the pixel's not classified as boundary, those whose mean image intensity is above a threshold are defined as nucleus. The threshold can be learned using the training data. We show the graphs for the ABC algorithm on fluorescence microscopy images from which we learned a threshold of 50 for this purpose using an 8-bit encoding that yields values ranging from 0 to 255.

In a last optimization step, we change the training procedure in such a way that for each training step, we sample a random crop from all training images which is in addition rotated by a multiple of 90◦ and flipped randomly. This approach seemed promising to us, as we achieved very high pixel-wise accuracy with the previous models. However, a common error on the test set for this models is the misclassification of boundary pixels between clumped nuclei as pixels belonging to a nucleus. On the training set, these pixels were correctly classified as boundary, which is an indicator for a model over-fit on the training data. Thus, by using data augmentation, we avoid this over-fit, and show that boundaries between nuclei get correctly classified as boundaries with this technique.

6. CONCLUSION

6.1 CONCLUSION

This master's thesis focuses on multi-exposure image fusion using image processing and deep learning techniques. The deep learning challenge originates from the accuracy of the ground truth data sets that are used as validation and/or training data for automated image processing algorithms. These datasets usually stem from manual tracing of field boundaries. However, manual tracing of image fusion is heavily dependent on the used imagery and supplementary data. It also is a highly subjective task, inevitably leading to inaccuracies and ambiguities depending on the priorities of the operator. The ground truth dataset can be complemented by existing property parcel information that could be indistinguishable based on the imagery alone. Furthermore, a single field parcel polygon can potentially show several subfields due to unreported land use changes within the growing season. Given these constraints, even a very powerful automated image processing algorithm can only become as good as the ground truth data but will never be in perfect alignment with reality. Although the model can generalize within certain limits (local generalization), its transferability is mainly limited by the variability of the training data. Furthermore, higher spatial resolution would certainly improve the exact delineation of the field boundaries, especially for small objects. At the time of writing, presents the highest resolution of freely available satellite imagery. Instead, the methodology focuses on a simple input-output mapping. Additional pre-processing or post-processing techniques could certainly add value later on, but would distract from the current goal of establishing baseline result. To exploit the pre-trained CNN for image edge smoothening of the different datasets, a top model is to stacked on the deep learning with an accuracy of 98.17%. We explore three different top models in this thesis to maximally make use of the labeled, supervised data. A first and obvious choice is a mere fully-connected layer, applied on the flattened CNN-output with added dropout. In light of the high parametric demand this top-model places on the learning process, we opted to also investigate performance with residual blocks as top model. Not only are these residual blocks hypothesized to facilitate training by learning the residual mapping, they also decrease the number of parameters while allowing a multi-layered top model.

6.2 FUTURE RECOMMENDATIONS

- In future, we want to add up the research line of deep learning with big data, training on RNN with Apache Spark engine for combinations of different feature from different pre-trained hybrid models leads to an enhanced classification performance compared to stand-alone features for image fusion and edge smoothening.
- A more advanced use of fine-tuning however proved to be adapting the pre-trained model, to the task at hand and fine-tuning up to a certain depth in the RNN model.
- This scheme will give rise to the optimal results achieved in this study for edge smoothening and image filtering. Finally, these results will be compared developed using RNN framework in future for pre-training on any image dataset.
- The best performance results will be achieved with high precision entered the same ballpark as the higher-scoring RNN classifiers on multiple pre-trained model framework, but could compete with the any other edge detection and smoothening work in previous studies.
- We note however that in light of the trends observed in the RNN framework, a more thorough exploration (of architecture, training time and hyper-parameter settings) could potentially improve these results and open a window of opportunity to thoroughly using three labeled datasets for kick-starting the supervised training phase with limited three labeled datasets.

.

REFERENCES

[1] Wang, Z.; Liu, Q.; Ikenaga, T. Visual salience and stack extension based ghost removal for high-dynamic-range imaging. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 2244–2248.

[2] Zhang, W.; Hu, S.; Liu, K. Patch-based correlation for deghosting in exposure fusion. Inf. Sci. 2017, 415, 19–27.

[3] Li, Z.; Zheng, J.; Zhu, Z.; Wu, S. Selectively Detail-Enhanced Fusion of Differently Exposed Images With Moving Objects. IEEE Trans. Image Process. 2014, 23, 4372–4382.

[4] Nejati, M.; Karimi, M.; Soroushmehr, S.M.R.; Karimi, N.; Samavi, S.; Najarian, K. Fast exposure fusion using exposedness function. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 2234–2238.

[5] Li, H.; Wang, Y.; Yang, Z.; Wang, R.; Li, X.; Tao, D. Discriminative dictionary learning-based multiple component decomposition for detail-preserving noisy image fusion. IEEE Trans. Instrum. Meas. 2019, 69, 1082–1102.

[6] Zhu, Z.; Yin, H.; Chai, Y.; Li, Y.; Qi, G. A novel multi-modality image fusion method based on image decomposition and sparse representation. Inf. Sci. 2018, 432, 516–529.

[7] Zhao, Q.; Sbert, M.; Feixas, M.; Xu, Q. Multi-Exposure Image Fusion Based on Information-Theoretic Channel. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 1872–1876.

[8] Shen, J.; Zhao, Y.; Yan, S.; Li, X. Exposure Fusion Using Boosting Laplacian Pyramid. IEEE Trans. Cybern. 2014, 44, 1579–1590.

[9] Zhu, Z.; Zheng, M.; Qi, G.; Wang, D.; Xiang, Y. A Phase Congruency and Local Laplacian Energy Based Multi-Modality Multi-Exposure Image Fusion Method in NSCT Domain. IEEE Access 2019, 7, 20811–20824.

[10] Li, Y.; Sun, Y.; Huang, X.; Qi, G.; Zheng, M.; Zhu, Z. An Image Fusion Method Based on Sparse Representation and Sum Modified-Laplacian in NSCT Domain. Entropy 2018, 20, 522.

[11] Li, H.; He, X.; Tao, D.; Tang, Y.; Wang, R. Joint multi-exposure image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning. Pattern Recognit. 2018, 79, 130–146.

[12] Kinoshita, Y.; Shiota, S.; Kiya, H.; Yoshida, T. Multi-Exposure Image Fusion Based on Exposure Compensation. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 1388–1392.

[13] Prabhakar, K.R.; Babu, R.V. Ghosting-free multi-exposure image fusion in gradient domain. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 1766–1770.

[14] Liu, Y.; Wang, Z. Dense SIFT for ghost-free multi-exposure fusion. J. Visual Commun. Image Represent. 2015, 31, 208–224.

[15] Kou, F.; Wei, Z.; Chen, W.; Wu, X.; Wen, C.; Li, Z. Intelligent Detail Enhancement for Exposure Fusion. IEEE Trans. Multimedia 2018, 20, 484–495.

[16] Ma, K.; Li, H.; Yong, H.; Wang, Z.; Meng, D.; Zhang, L. Robust multi-exposure image fusion: a structural patch decomposition approach. IEEE Trans. Image Process. 2017, 26, 2519–2532.

[17] Ma, K.; Duanmu, Z.; Yeganeh, H.; Wang, Z. Multi-Exposure Image Fusion by Optimizing A Structural Similarity Index. IEEE Trans. Comput. Imag. 2018, 4, 60–72.

[18] Li, Y.; Sun, Y.; Zheng, M.; Huang, X.; Qi, G.; Hu, H.; Zhu, Z. A Novel Multi-Exposure Image Fusion Method Based on Adaptive Patch Structure. Entropy 2018, 20, 935.

[19] Qin, Z.; Fan, J.; Liu, Y.; Gao, Y.; Li, G.Y. Sparse Representation for Wireless Communications: A Compressive Sensing Approach. IEEE Signal Process Mag. 2018, 35, 40–58.

[20] Liu, S.; Shi, M.; Zhu, Z.; Zhao, J. Image fusion based on complex-shearlet domain with guided filtering. Multidimension. Syst. Signal Process. 2017, 28, 207–224.

[21] Ma, K.; Duanmu, Z.; Zhu, H.; Fang, Y.; Wang, Z. Deep Guided Learning for Fast Multi-Exposure Image Fusion. IEEE Trans. Image Process. 2020, 29, 2808–2819.

[22] Wang, K.; Qi, G.; Zhu, Z.; Chai, Y. A Novel Geometric Dictionary Construction Approach for Sparse Representation Based Image Fusion. Entropy 2017, 19, 306.

[23] Ma, K.; Zeng, K.; Wang, Z. Perceptual Quality Assessment for Multi-Exposure Image Fusion. IEEE Trans. Image Process. 2015, 24, 3345–3356.

[24] Petrovic, V. Subjective tests for image fusion evaluation and objective metric validation. Inform. Fusion 2007, 8, 208 – 216.

[25] Zhu, Z.; Qi, G.; Chai, Y.; Chen, Y. A Novel Multi-Focus Image Fusion Method Based on Stochastic Coordinate Coding and Local Density Peaks Clustering. Future Internet 2016, 8, 53.

[26] A. Saleem, A. Beghdadi, and B. Boashash, "Image fusionbased contrast enhancement," EURASIP Journal on Image and Video Processing, vol. 2012, no. 1, p. 10, 2012.

[27] J. Wang, G. Xu, and H. Lou, "Exposure fusion based on sparse coding in pyramid transform domain," in Proceedings of the 7th International Conference on Internet Multimedia Computing and Service, ser. ICIMCS '15. New York, NY, USA: ACM, 2015, pp. 4:1–4:4.

[28] Z. Li, J. Zheng, Z. Zhu, and S. Wu, "Selectively detailenhanced fusion of differently exposed images with moving objects," IEEE Transactions on Image Processing, vol. 23, no. 10, pp. 4372–4382, 2014.

[29] T. Sakai, D. Kimura, T. Yoshida, and M. Iwahashi, "Hybrid method for multi-exposure image fusion based on weighted mean and sparse representation," in 2015 23rd European Signal Processing Conference (EUSIPCO). EURASIP, 2015, pp. 809–813.

[30] M. Nejati, M. Karimi, S. M. R. Soroushmehr, N. Karimi, S. Samavi, and K. Najarian, "Fast exposure fusion using exposedness function," in 2017 International Conference on Image Processing (ICIP). IEEE, 2017, pp. 2234–2238.

[31] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in ACM SIGGRAPH. ACM, 1997, pp. 369–378.

[32] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," ACM Transactions on Graphics (TOG), vol. 21, no. 3, pp. 267–276, 2017.

[33] T.-H. Oh, J.-Y. Lee, Y.-W. Tai, and I. S. Kweon, "Robust high dynamic range imaging by rank minimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 6, pp. 1219–1232, 2015.

[34] Y. Kinoshita, S. Shiota, M. Iwahashi, and H. Kiya, "An remapping operation without tone mapping parameters for hdr images," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. 99, no. 11, pp. 1955–1961, 2016.

[35] Y. Kinoshita, S. Shiota, and H. Kiya, “Fast inverse tone mapping with reinhard’s grobal operator,” in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017, pp. 1972–1976.

[36] “Fast inverse tone mapping based on reinhard’s global operator with estimated parameters,” IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. 100, no. 11, pp. 2248–2255, 2017.

[37] Y. Q. Huo and X. D. Zhang, “Single image-based hdr imaging with crf estimation,” in 2016 International Conference On Communication Problem-Solving (ICCP). IEEE, 2016, pp. 1–3.

[38] T. Murofushi, M. Iwahashi, and H. Kiya, “An integer tone mapping operation for hdr images expressed in floating point data,” in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2013, pp. 2479– 2483.

[39] T. Murofushi, T. Dobashi, M. Iwahashi, and H. Kiya, “An integer tone mapping operation for hdr images in openexr with denormalized numbers,” in 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014, pp. 4497–4501.

[40] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, “Robust multiexposure image fusion: A structural patch decomposition approach,” IEEE Trans. Image Process., vol. 26, no. 5, pp. 2519–2532, May 2017.

[41] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, “DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs,” in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 4724–4732.

[42] F. Kou, Z. Li, C. Wen, and W. Chen, “Multi-scale exposure fusion via gradient domain guided image filtering,” in Proc. IEEE Int. Conf. Multimedia Expo, Jul. 2017, pp. 1105–1110.

[43] K. Ma, Z. Duanmu, H. Yeganeh, and Z. Wang, “Multi-exposure image fusion by optimizing a structural similarity index,” IEEE Trans. Comput. Imag., vol. 4, no. 1, pp. 60–72, Mar. 2018.

[44] K. Ma, K. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” IEEE Trans. Image Process., vol. 24, no. 11, pp. 3345–3356, Nov. 2015.

[45] Q. Yang, K.-H. Tan, and N. Ahuja, “Real-time O (1) bilateral filtering,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2019, pp. 557–564.

[46] J. Ragan-Kelley, A. Adams, S. Paris, M. Levoy, S. Amarasinghe, and F. Durand, "Decoupling algorithms from schedules for easy optimization of image processing pipelines," ACM Trans. Graph., vol. 31, no. 4, pp. 32:1–32:12, Jul. 2012.

[47] S. W. Hasinoff et al., "Burst photography for high dynamic range and low-light imaging on mobile cameras," ACM Trans. Graph., vol. 35, no. 6, pp. 192:1–192:12, Nov. 2016.

[48] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," ACM Trans. Graph., vol. 36, no. 4, pp. 118:1–118:12, Jul. 2017.

[49] Q. Chen, J. Xu, and V. Koltun, "Fast image processing with fully convolutional networks," in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 2516–2525.

[50] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 3431–3440.

[51] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," ACM Trans. Graph., vol. 36, no. 4, pp. 144:1–144:12, Jul. 2017.

[52] Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5835–5843.

[53] Zhao, Y.; Li, G.; Xie, W.; Jia, W.; Min, H.; Liu, X. GUN: Gradual Upsampling Network for Single Image Super-Resolution. IEEE Access 2018, 6, 39363–39374.

[54] Wang, Y.; Wang, L.; Wang, H.; Li, P. End-to-End Image Super-Resolution via Deep and Shallow Convolutional Networks. IEEE Access 2019, 7, 31959–31970.

[55] Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. IEEE Trans. Pattern Anal. Mach. Intell. 2016, 38, 295–307.

[56] Shen, L.; Yue, Z.; Feng, F.; Chen, Q.; Liu, S.; Ma, J. MSR-net:Low-light Image Enhancement Using Deep Convolutional Network. arXiv 2017, arXiv:1711.02488.

[57] Liu, C.; Cheng, I.; Zhang, Y.; Basu, A. Enhancement of low visibility aerial images using histogram truncation and an explicit Retinex representation for balancing contrast and color consistency. ISPRS J. Photogramm. Remote Sens. 2017, 128, 16–26.

[58] Zollini, S.; Alicandro, M.; Cuevas-González, M.; Baiocchi, V.; Dominici, D.; Buscema, P.M. Shoreline extraction based on an active connection matrix (ACM) image enhancement strategy. J. Mar. Sci. Eng. 2020, 8, 9.

[59] Dominici, D.; Zollini, S.; Alicandro, M.; della Torre, F.; Buscema, P.M.; Baiocchi, V. High resolution satellite images for instantaneous shoreline extraction using new enhancement algorithms. Geosciences 2019, 9, 123.

[60] Shen, L.; Yue, Z.; Feng, F.; Chen, Q.; Liu, S.; Ma, J. MSR-net:Low-light Image Enhancement Using Deep Convolutional Network. arXiv 2017, arXiv:1711.02488.

[61] Toet, Alexander (2014): TNO Image Fusion Dataset. figshare. Dataset. https://doi.org/10.6084/m9.figshare.1008029.v1

[62] Available Online: https://medium.com/datadriveninvestor/convolutional-neural-network-cnn-simplified-ecafd4ee52c5

[63] S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang, "Deep Guided Learning for Fast Multi-Exposure Image Fusion," in Proc. Eur. Conf. Comput. Vis., 2020, pp. 1–16.

[64] Vonikakis, Vassilios & Bouzos, & Andreadis, (2011). Multi-Exposure Image Fusion Based on Illumination Estimation. 10.2316/P.2011.738-051.