Removal of motion blur in images of human faces using deep learning

Franov, Marcelo

UNIVERSITY OF SPLIT
FACULTY OF SCIENCE

GRADUATE THESIS

REMOVAL OF MOTION BLUR IN IMAGES OF HUMAN FACES USING DEEP LEARNING

Marcelo Franov

Split, September 2024.

Basic documentation card

Graduate thesis

University of Split

Faculty of Science

Department of Computer Science

Ruđera Boškovića 33, 21000 Split, Croatia

Removal of motion blur in images of human faces using deep learning

Marcelo Franov

Abstract

Facial image degradation by motion blur is a problem that continues to occur with the increased usage and presence of cameras in our daily lives. Whether the purpose of image capturing is personal use or further higher level computer vision tasks, they both require deblurred facial image. This thesis covers the latest literature and describes every aspect of facial deblurring problem. The importance of training dataset is tested by training different models on different facial datasets. After the trained models are tested and compared in detail the discovered conclusions were used for further development and testing of deblurring models with the emphasis on common problems and limitations.

Keywords: motion deblurring, facial images, deep learning

Graduate thesis deposited in library of Faculty of science, University of Split

Thesis consists of: 59 pages, 23 figures, 19 tables and 118 references

Original language: English

Mentor: **Saša Mladenović, Ph.D.** *Full Professor, Faculty of Science, University of Split*

Reviewers: **Saša Mladenović, Ph.D.** *Full Professor, Faculty of Science, University of Split*
Goran Zaharija, Ph.D. *Assistant Professor, Faculty of Science, University of Split*
Nika Jerković *Instructor, Faculty of Science, University of Split*

Thesis accepted: September 2024

Temeljna dokumentacijska kartica

Diplomski rad

Sveučilište u Splitu

Prirodoslovno-matematički fakultet

Odjel za informatiku

Ruđera Boškovića 33, 21000 Split, Croatia

Uklanjanje zamućenja kretanja na slikama ljudskog lica korištenjem dubokog učenja

Marcelo Franov

Sažetak

Degradacija slike lica zbog zamućenja kretanja problem je koji se i dalje javlja s povećanom upotrebom i prisutnošću kamera u našem svakodnevnom životu. Bilo da je svrha snimanja slike osobna upotreba ili daljnji zadaci više razine računalnog vida, oboje zahtijevaju uklanjanje zamućenja slike lica. Ovaj rad pokriva najnoviju literaturu i opisuje svaki aspekt problema uklanjanja zamućenja lica. Važnost skupa podataka za treniranje testirana je treniranjem različitih modela na različitim skupovima podataka lica. Nakon što su uvježbani modeli testirani i detaljno uspoređeni, otkriveni zaključci korišteni su za daljnji razvoj i testiranje modela uklanjanja zamućenja s naglaskom na uobičajene probleme i ograničenja.

Ključne riječi: uklanjanje zamućenja kretanja, slike lica, duboko učenje

Rad je pohranjen u knjižnici Prirodoslovno-matematičkog fakulteta, Sveučilišta u Splitu

Rad sadrži: 59 stranica, 23 grafičkih prikaza, 19 tablica i 118 literaturnih navoda. Izvornik je na engleskom jeziku.

Mentor: **Dr. sc. Saša Mladenović,** *redoviti profesor Prirodoslovno-matematičkog fakulteta u Splitu, Sveučilišta u Splitu*

Ocjenjivači: D**r. sc. Saša Mladenović,** *redoviti profesor Prirodoslovno-matematičkog fakulteta u Splitu, Sveučilišta u Splitu*

Dr. sc. Goran Zaharija, *docent Prirodoslovno-matematičkog fakulteta u Splitu, Sveučilišta u Splitu*

Nika Jerković, asistent *Prirodoslovno-matematičkog fakulteta u Splitu, Sveučilišta u Splitu*

Rad je prihvaćen: **Rujan 2024**

Acknowledgments

First and foremost, I would like to extend my deepest gratitude to my supervisor Full professor Saša Mladenović, Ph.D. for his boundless patience and invaluable advice. His support and guidance throughout this work made the process enjoyable and purposeful. I am thankful for his supervision and his efforts to not only show the right way but to show why we should strive for it.

I would also like to thank my family for their unyielding support. They were my foundation that I could always rely upon.

And finally, I would like to thank my friends who brought me joy and words of encouragement on all the steps of this journey.

Content

DECLARATION

of the independent preparation of the graduate thesis

I declare under full material and moral responsibility that this work was created independently by me and that there are no copied or duplicated parts of the text of other people's works in it, without being properly marked as quotations with the specified source from which they were taken.

In Split, 19.09.2024.

Marcelo Franov

(student)

1. Introduction

Facial image deblurring is a process of recovering sharp facial images from a blurred input image. It is a low-level computer vision task and it is necessary as a prerequisite for high-level computer vision tasks such as face detection, face recognition, face verification [1]. As complex computer vision tasks become increasingly popular, the need for robust face motion deblurring also increases [2]. Motion blur, which occurs when either the object or the camera moves during and exposure duration, impairs image quality especially when taking the pictures of human faces. Older mobile phones with less capable cameras still suffer from motion blur in their images if human subjects move during the image taking process. Facial motion deblurring can also be useful in sports where athletes are moving at high speeds which could be hard to capture in high quality. With the rise of autonomous systems such as cars and humanoid robots, some level of interaction with human beings is expected [3, 4]. That could range from detecting a fast moving human being which is about to cross the road or reading the facial emotions for a better interaction by a robot during a conversation with humans. Both of those tasks require undistorted image of human face which is used for further processing. With all those things considered facial motion deblurring presents a specific task which has potential for research and improvement [5].

1.1. Motivation

Motion blur happens when the object moves while the background remains stationary or the camera moves while the scene does not. In specific case of facial motion blur, the human is the object that moves relatively to the background and because of that the region of his face suffers from motion blur. That is the usual origin of real world motion blur. When the deep learning deblurring models are trained, they are usually trained in supervised manner on pairs of sharp image and motion blurred image of the same face. Motion blur which is in one of the image pairs is achieved synthetically by convolving the sharp image with the blur kernel. Most of the time the convolution is applied to the entire image of sharp face including background. In addition to that, the blur kernel is uniform in contrast to real life blur which is often non-uniform [6, 7]. This leads to poor performance of models trained on synthetic data when they are tested on real life images containing facial motion blur.

The main motivation of this thesis is to test whether segmenting faces or facial regions and blurring only them with motion blur kernels could lead to better and more robust models which could then be used in real life scenarios such as emotion recognition or animation capture [8, 9].

In order to insure that, the thesis will cover key issues of facial deblurring domain which are needed to completely understand the problem in order to build the appropriate solution. To better understand how to achieve this it is important to understand the context and the definition of the problem we are dealing with.

1.2. Problem Context

Face motion deblurring represents one subset of tasks which belong to the face restoration domain which itself falls under general image restoration [10]. General image restoration is the all-encompassing term covering restoration of various kinds of image which could contain buildings, natural environments, text and face [11]. The task depends on the type of degradation the image suffered. There are various kinds of image degradations such as blur, noise, haze, resolution issues and artifacts. Restoration tasks aim to recover lost details or correct artifacts in images to enhance their visual quality and utility. Face image restoration is a subtask of general image restoration and consists of face deblurring, face denoising, face artifact removal, face super-resolution also known as face hallucination [12, 13] and old photo restoration [14]. Furthermore, face motion deblurring represents a specific form of deblurring because there are several types of image blur such as motion blur, defocus blur and Gaussian blur. Deblurring is particularly challenging because it requires estimating the blur kernel and reversing the complex distortions that result from motion.

Each of these tasks has its own characteristics starting from image type which is being restored. General images contain various objects and backgrounds, often with sharp edges which can aid in image deblurring because they show where one object ends and other begins. On the other hand, human faces lack sharp edges and because of that are hard for motion deblurring tasks. List of face image restoration tasks and their causes can be found in Table 1.1.

Table 1.1 List of face image restoration tasks and their causes

Tasks	Deblurring	Denoising	Super-resolution (face hallucination)	Artifact removal	Old photo restoration
Causes of degradation	Motion blur	Gaussian noise	Low resolution	JPEG compression artifacts	Scratches
	Defocus blur	Salt-and-pepper noise			Film noise fading
	Gaussian blur	Poisson noise			Color fading

Another new problem worth mentioning are Adversarial perturbations which are intentionally added noises to fool face recognition systems. [15, 16].

1.3. Problem Definition

Blurry images are still widespread in modern times despite the modern camera equipment and image taking technology. The blur in the image can be cause by various reasons such as optical aberrations, camera shake and object movement. Degradation process which could be used to describe the blur of the image is shown through Eq. (1):

$$B = D(C, X, n) \quad (1)$$

Here, B refers to the blurred image which is achieved by applying degradation process D using the blur operator X to the clear image C, while n refers to the noise which could also appear in the degradation process.

There are several kinds of blur, most notably motion blur, defocus blur and Gaussian blur but in this thesis we will focus solely on motion blur.

Motion blur is caused by the motion of the camera or the movement of the object. The natural motion blur is very complicated to model because it is rarely uniform. For example, the direction and degree of the camera movement can vary and in dynamic scenes some objects like people and cars move while the other objects and backgrounds are stationary. In order to simplify the deblurring process we often assume that the motion blur is uniform. Then the formula would look like in Eq. (2):

$$B = k * C + n \quad (2)$$

Where B is the blurry image obtained by convolution operation * between C which is the original clear image and k which is the blur kernel. Once again, n is the random noise. Blur kernel can be modeled using Eq. (3):

$$k(i,j) = \begin{cases} \frac{1}{\pi r^2}, & if\ i^2 + j^2 \leq r^2 \\ 0, & otherwise \end{cases} \quad (3)$$

where r is the radius of the blur and (i, j) refers to the pixel coordinates.
Now that we understand how we define and model motion blur, we can continue with further understanding of the intentions of this thesis.

1.4. Thesis Goals

Emerging from previous discussion and motivation, the subsequent research questions, goals and objectives of this thesis can be found in Table 1.2.

Table 1.2 Research questions, goals and objectives

Research Questions	Research Objectives	Research Goals
RQ1: What are the previous related literature reviews about deblurring (facial) images?	-Search the article databases for appropriate literature reviews -Describe the process -Analyze the reviews and show what they cover	Gain a thorough understanding of the literature reviews conducted and areas covered.
RQ2: What is the newest literature about deblurring facial images?	-Search the article databases for appropriate articles - Describe the process - Analyze the articles and show what they cover	Gain a thorough understanding of the newest approaches to the problem.
RQ3: What algorithms are used for deblurring face images?	-Find the algorithms used in the literature for this task -List their characteristics -Describe their advantages and disadvantages	Gain a thorough understanding of the deep learning algorithms for face image deblurring including their limitations and flaws

RQ4: What loss functions are used in facial deblurring?	-Define the loss functions used in the literature -List specific uses of loss functions based on which specific face characteristic they preserve	Gain a thorough understanding of the loss functions and which face element they preserve the best
RQ5: What are the ways for creating the training and testing datasets for face image deblurring models?	-List and describe face deblurring datasets used in the articles from literature review -Analyze the mentioned methods for their creation -Evaluate the strengths and limitations of dataset creation methods	Identify the methods for dataset creation. Learn which foundational face image dataset is suitable for the task.
RQ6: What are the evaluation metrics for evaluating the effectiveness of different algorithms for face deblurring task?	-Describe and compare evaluation metrics used in the literature -Investigate alternative methods for evaluating face deblurring model which are more suited for face image related tasks	Describing the model evaluation framework depending on what aspect of the deblurred image we prefer.
RQ7: What training dataset yields the best deblurring results?	-Create different training datasets with previously described methods -Train different facial deblurring models on created datasets -Compare their performance on described evaluation metrics	Obtaining objective test results supported by evaluation metrics which will tell which training dataset is best for this kind of problem.

1.5. Methodology

This thesis uses a combination of methodologies to achieve the stated goals. In order to establish a theoretical framework in this complex domain, a thorough literature review will be conducted. The review is divided into two parts: (1) a synthesis of existing literature reviews, essential for building the domain context, and (2) an in-depth review of articles

published after the most recent relevant literature review, specifically focusing on advancements in face motion deblurring.

In processing the information from these literature reviews and articles various different quantitative and qualitative methodologies will be used. Among qualitative methodologies we can mention case studies of existing motion deblurring techniques, in order to understand their strengths and weaknesses and content analysis of academic publications. This will also include an evaluation of the characteristics of human face images, as well as the datasets used for deblurring and the techniques applied in creating them. For quantitative methodologies this thesis will use data collection and analysis to identify key metrics for evaluating the performance of face deblurring models. Additionally, it involves creating and training face deblurring models on custom-made datasets, which are segmented to analyze facial regions distinct from backgrounds, mimicking real-world blurred face conditions. Two sets of blurred datasets will be created to test model robustness across different deblurring scenarios. These datasets will then be used in experiments, where face deblurring models are developed and evaluated based on quantitative metrics, such as PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index). In the model development phase, attention will be given to understanding the basic layers and blocks that make up these models and their effectiveness in face deblurring. Furthermore, this research will carefully select and apply appropriate loss functions during training, as these functions significantly influence model performance and convergence. All of those elements and processes will be described in detailed and analyzed in the context of facial motion deblurring. The thesis structure and process itself is shown in the Figure 1.1.

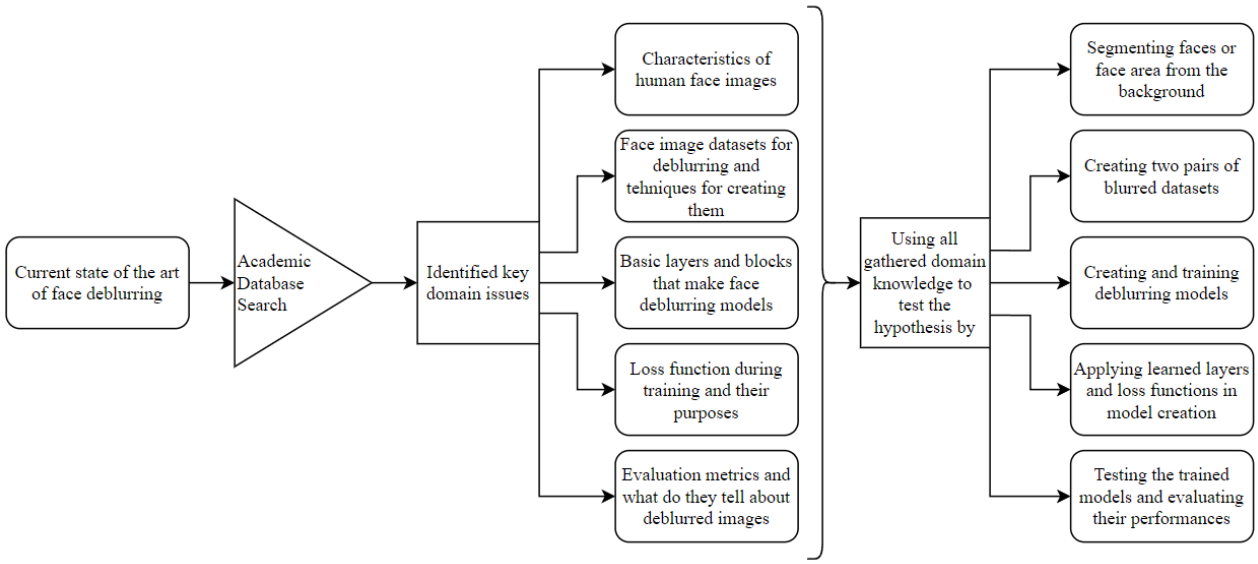

Figure 1.1 Thesis structure and process

1.6. Thesis Outline

In Chapter 2 previous literature reviews are described and their most important characteristics are shown. Following that, a literature review is conducted with the purpose of showing newest solution to facial motion deblurring problem. This chapter also presents the most popular datasets, basic layers and blocks which make the deblurring models, loss functions and relevant evaluation metrics.

Chapter 3 presents and overview of dataset creation methods. It gives descriptions of segmentation models which can be used and ways to artificially add motion blur to images or parts of images.

In Chapter 4 the experimental setup is described along the chosen models described in previous chapters. Evaluation metrics used for assessing the trained models are also described.

In Chapter 5 we will be discussing the results of the experiment.

And finally in Chapter 6 conclusion and future work will be presented.

2. Background

For better understanding of the topic it is important to start with the literature reviews. The literature reviews were made using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology [17].

2.1. Previous Literature Reviews

Databases are the discovery mechanisms used for searching and browsing scholarly information [18]. After consulting with the literature on scientific databases [19, 20], four databases were chosen for literature search. Chosen databases are IEEE Xplore, Web of Science (WOS), Scopus and ScienceDirect. The first three literature databases were searched for review and survey papers covering the topic. Due to its long existence, popularity and frequent advancement in the field, there is a large number of literature reviews covering image deblurring and face restoration. The search term was intentionally broad with the goals of encompassing all possible reviews of this complex topic. There was no publishing year filter in order not to limit any reviews.

Publications were retrieved from IEEE Xplore digital database on June 13. 2024. Database was searched with the following term: " image deblurring survey" in the title. No further filters were applied. The search gave 20 results. After reading the titles and abstracts 17 papers were excluded because they were not on topic of image deblurring and three papers were taken for full further reading and examination.

Web of Science database was searched with " image deblurring literature review " in the all fields category. The search gave 13 results. After reading the titles and abstracts 12 papers were excluded because they were not on topic. One was on topic and it was taken for further reading and examination.

Scopus database was searched with the term "image AND deblurring AND survey" within: Article title, Abstract, Keywords. The search gave 52 results. After reading the titles and abstracts 50 papers were excluded because they were not on topic. Two were on topic and they were taken for further reading and examination. One of those two was specifically about facial image deblurring.

Additionally, from these articles found in databases, three more literature reviews were collected from other sources and were added to the final number of articles, totaling nine, for further reading and examination. The purpose of finding and analyzing these surveys is to get familiar with the topic and to the various approaches of covering it. Important information shown in Table 2.1 will be extracted from these literature reviews and shown separately.

Table 2.1 Extracted Data Items

Number	Data Item
1	Published year
2	Title
3	Main content
4	Information about which domain topics are covered

Basic characteristics of each survey can be found in Table 2.2.

Table 2.2 Basic characteristics of found surveys

Number	Reference	Year	Survey Title	Main Content
1	Wang et al. [21]	2023.	A survey on facial image deblurring	Covers entire deblurring process of facial images
2	Biyouki and Hwangbo [22]	2023.	A COMPREHENSIVE SURVEY ON DEEP NEURAL IMAGE DEBLURRING	Review of recent progress of deep neural architectures in image deblurring
3	Wang et al. [23]	2022.	A Survey of Deep Face Restoration: Denoise, Super-Resolution, Deblur, Artifact Removal	Survey focuses on methods for face restoration
4	Su et al. [24]	2022.	A Survey of Deep Learning Approaches to Image Restoration	Review on deep learning methods for image restoration tasks
5	Mahalakshmi and Shanthini [25]	2016.	A Survey on Image Deblurring	Review of techniques for removing blur in images
6	Huixin Zheng [26]	2021.	A Survey on Single Image Deblurring	Summarizes the recently published image dehazing methods

7	Xiang et al. [27]	2024.	Application of Deep Learning in Blind Motion Deblurring: Current Status and Future Prospects	Provides an exhaustive overview of the role of deep learning in blind motion deblurring
8	Zhang et al. [28]	2022.	Deep Image Deblurring: A Survey	Comprehensive survey of recently published deep learning based image deblurring approaches
9	Ranjan and Ravinder [29]	2022.	Deep Learning based Image Deblurring: A Comparative Survey	Comprehensive survey of all deblurring techniques

Table 2.3 shows which domain topics are adequately covered.

Table 2.3 Domain topics covered in literature surveys

	Important themes to cover for domain understanding						
Surveys	Detailed Problem description	Description of models for deblurring	Advantages and disadvantages of models	Detailed comparison of numerical and visual results of previous models	Description of datasets	Description of Loss functions	Description of Evaluation methods
Wang et al. [21]	✓	✓	✓	✓	✓	✓	✓
Biyouki and Hwangbo [22]	✓	✓	✓	✓	✓	✓	✓
Wang et al. [23]	✓	✓	✓	✓	✓	✓	✓
Su et al. [24]	✓	✓	✓	✓	✓	✓	✓
Mahalakshmi and Shanthini [25]	✗	✓	✓	✗	✗	✗	✗
Huixin Zheng [26]	✗	✓	✓	✗	✓	✗	✓
Xiang et al. [27]	✓	✓	✓	✓	✓	✗	✓
Zhang et al. [28]	✓	✓	✓	✓	✓	✓	✓

Ranjan and Ravinder [29]	✓	✓	✓	✘	✘	✘	✘

These literature reviews give us good cover of the domain of image restoration. They also show us how methods which are good for general image deblurring are not necessarily the best or adequate for facial image deblurring.

2.2. Literature Review

Four online databases were searched for publications which could be useful in the writing process of this thesis. The most recent literature review about facial motion deblurring is from 2024. And because of that, only more recently published papers were searched for by filter the search years to 2023. and 2024.

The papers were collected for literature review according to PRISMA statement in accordance with inclusion and exclusion criteria which can be seen in Table 2.4.

Table 2.4 Inclusion and exclusion criteria

Criterion	
Exclusion Criteria	Inclusion Criteria
EC1 Papers not written in the English language	IC1 Papers written in the English language
EC2 Papers which are not strictly on facial deblurring	IC2 Papers which are about facial deblurring
EC3 Papers which are not available to the researcher	IC3 Papers which are using deep learning methods
EC4 Papers which are using non deep learning methods for deblurring	IC4 Papers from 2023. And 2024.
EC4 Duplicate records	/

Publications were retrieved from IEEE Xplore digital database on June 16. 2024. Papers were selected in the database search based on the following two conditions: (1) "face image deblurring" should appear in the title; (2) years of publication are 2023. and 2024. The search gave 31 results. After reading the titles and abstracts 14 papers were excluded based on exclusion criteria and 17 papers which satisfied inclusion criteria were taken

for full further reading and examination.

Publications were retrieved from Web of Science digital database on June 16. 2024. Papers were selected in the database search based on the following two conditions: (1) "face image deblurring" should appear in „All Fields“; (2) years of publication are 2023. and 2024. The search gave 21 results. After reading the titles and abstracts 8 papers were excluded based on exclusion criteria and 13 papers which satisfied inclusion criteria were taken for full further reading and examination.

Publications were retrieved from Scopus digital database on June 16. 2024. Papers were selected in the database search based on the following search query within Article title, Abstract, Keywords: "face AND image AND deblurring". Years of publication are 2023. and 2024. The search gave 45 results. After reading the titles and abstracts 40 papers were excluded based on exclusion criteria and 5 papers which satisfied inclusion criteria were taken for full further reading and examination.

Publications were retrieved from ScienceDirect digital database on June 16. 2024. Papers were selected in the database search based on the following two conditions: (1) "face deblurring" should appear in title; (2) years of publication are 2023. and 2024. The search gave 7 results. After reading the titles and abstracts 4 papers were excluded based on exclusion criteria and 3 papers which satisfied inclusion criteria were taken for full further reading and examination.

The simpler search term “face image deblurring” or “face deblurring” was chosen because it produced less false positives in accordance with assumptions given by Kitchenham et al. [117]. All papers, which were able to be found, were taken for further reading and examined. Those not strictly on topic of facial motion deblurring were excluded. PRISMA statement describing the process can be seen on Figure 2.1.

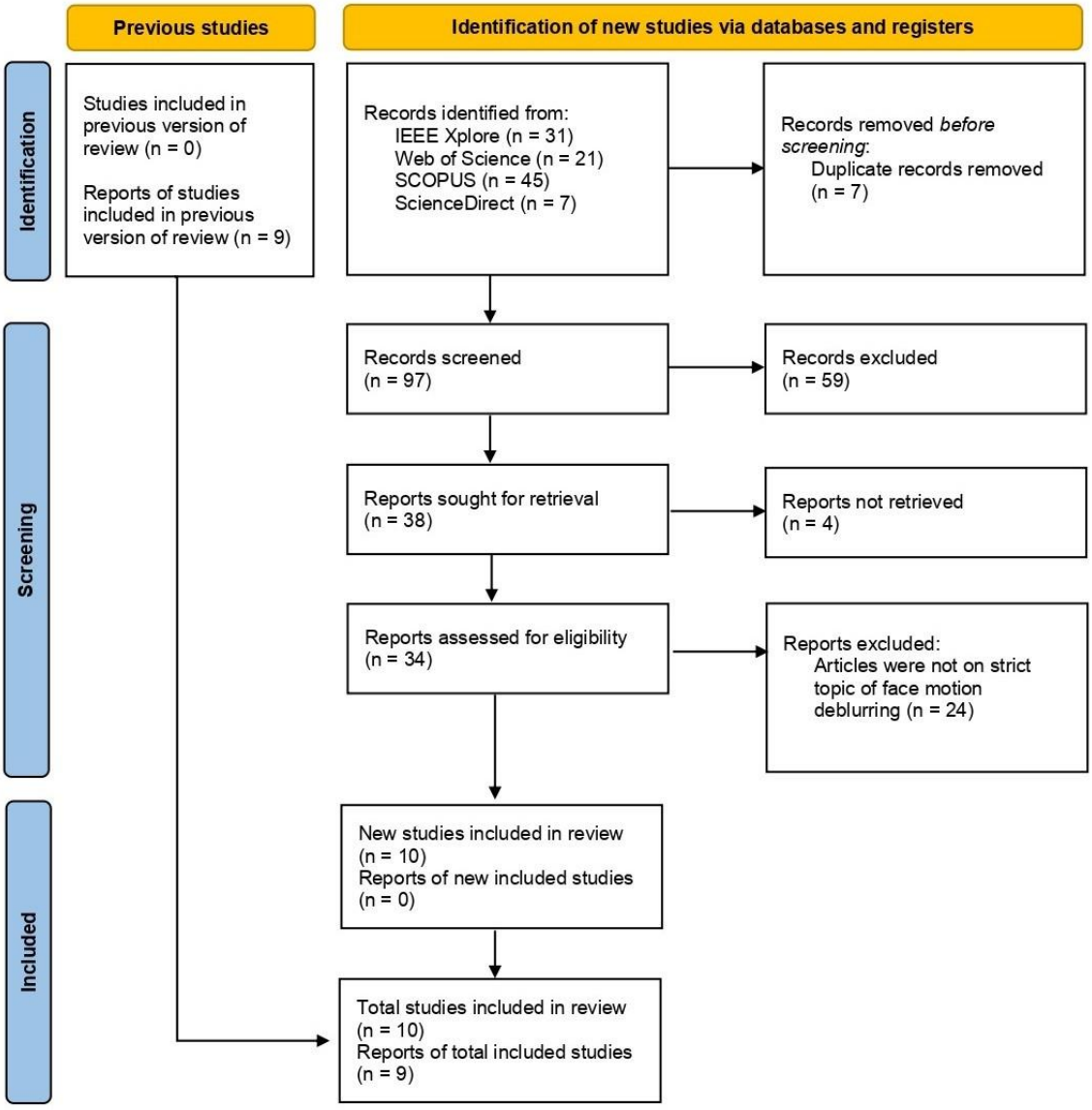

Figure 2.1 Prisma Statement

After the articles were collected data shown in Table 2.5 was extracted from the articles.

Table 2.5 Data items extracted in correspondence to which Research question (RQ)

Number	Related RQ	Data Item
1	RQ2	Published year
2	RQ2	Title
3	RQ2	Main goals
4	RQ5	Datasets used in the article
5	RQ3	Models used in the deblurring process

6	RQ3	Custom improvements made to the model or anywhere in the deblurring process
7	RQ6, RQ7	Reported measures
8	RQ5	Dataset creation description
9	RQ4	Loss functions used in model training procedure
10	RQ3	Information relevant to the model training (optimizer, learning rate, batch size, number of epochs etc.)
11	RQ3	Training environment description

The extracted data is presented to the reader in several tables. The purpose of this is easy comparison of articles based on their various characteristics. Tables are often more effective for comparing multiple data points or attributes across different categories. They also present a large amount of information in a compact, structured format. The grid structure of tables allows readers to quickly scan both horizontally and vertically to locate specific information that could be relevant for their own research.

Table 2.6 Table of extracted articles and their main characteristics

Number	Reference	Year	Article Title	Main Goals	Target images
1	Xiao and Pan [30]	2023.	A facial motion deblurring algorithm based on semantics and GAN	Paper proposes new face deblurring network based on GAN network called PSFD-GAN	Only deblurring of facial images
2	Gao et al. [31]	2023.	Blind deblurring of single image based on kernel estimation of texture image	Paper proposes blind image deblurring method based on image texture autocorrelation	General image deblurring with application to facial images
3	Zhang et al. [32]	2024.	Blind Face Restoration: Benchmark Datasets and a Baseline Model	Creation of two blind face restoration benchmark datasets and development of an model called Swin Transformer U-Net	Only deblurring of facial images
4	Yang et al. [33]	2023.	Blind Image Deblurring with Extreme Gradient and Dark Channel Priors	Paper presents a novel algorithm for blind image deblurring based on extreme gradient and dark channel priors	General image deblurring with application to facial images

5	Yu and Xie [34]	2023.	Blurry Facial-Image Deconvolution via Model-Guided Deep Neural Network Inspired From Edge Regularization	Proposes a new deep neural network for facial blur deconvolution	Only deblurring of facial images
6	Ahalya et al. [35]	2024.	Deep Learning for Single Image Deblurring	Paper gives an evaluation of newly released deep learning deblurring techniques	General image deblurring with application to facial images
7	Shi et al. [36]	2023.	Face deblurring based on regularized structure and enhanced texture information	Present a new network for face deblurring utilizing more regularized structure and enhanced texture information	Only deblurring of facial images
8	Cui et al. [37]	2023.	Joint Face Super-Resolution and Deblurring Using Multi-Task Feature Fusion Network	Proposes a new multi-task feature fusion network based on double branches	Only deblurring of facial images
9	Han et al. [38]	2023.	Semantic-Aware Face Deblurring With Pixel-Wise Projection Discriminator	Proposes a new type of discriminator and introduces a prediction-weighted loss	Only deblurring of facial images
10	Wang et al. [39]	2023.	Take a Prior from Other Tasks for Severe Blur Removal	Proposes a semantic prior embedding layer with multi-level aggregation and semantic attention transformation to integrate the priors effectively	General image deblurring with application to facial images

Table 2.7 Table of articles with extracted model information

Reference	Dataset used	Model used	Custom improvements	Reported measures
Xiao and Pan [30]	CelebAMask-HQ dataset for training and FFHQ for testing	Proposed PSFD-GAN	progressive training strategy	On proposed method PSNR: 27.52 SSIM: 0.901

Gao et al. [31]	Köhler et al. dataset	Custom algorithm	Autocorrelation of texture images	CDV:56.125 CPBDM: 0.275 MDM: 0.905 WEW: 2.812
Zhang et al. [32]	EDFace-Celeb dataset EDFace-Celeb-1M (BFR128) EDFace-Celeb-150K (BFR512)	STUNet which adopts the strong Transformer architecture	Design two task-driven metrics, AFLD and AFICS Created new face image datasets with degradations	STUNet for Face Deblurring: PSNR: 27.3912 SSIM: 0.8080
Yang et al. [33]	Levin dataset Köhler et al. dataset	Custom algorithm	They combine local minimum and maximum gradient prior information to better constrain the solution space	PSNR: 20.89 SSIM: 0.75
Yu and Xie [34]	CelebAMask-HQ, CelebA-short, pubFig-40	MG-DNN	MG-DNN is proposed by a multichannel deep network with several KI-blocks extending from the optimization framework	On the CelebA-short: PSNR: 24.7753 SSIM: 0.7779 MSSSIM: 0.8575 BIQA: -16.3483
Ahalya et al. [35]	No datasets mentioned	Some models described	General vague description of the image deblurring domain	No numerical results mentiones
Shi et al. [36]	CelebA, CelebA-HQ	RSETNet	Face parsing network with fine-tuning, feature adaptive denormalization (FAD)	On CelebA-HQ: PSNR: 29.21 SSIM: 0.878
Cui et al. [37]	CelebA	Multi-task feature fusion network	Adopt a dual-branch network structure	PSNR: 26.3697 SSIM: 0.7651 LPIPS: 0.1859 NME: 0.0417
Han et al. [38]	MSPL	SAPPGAN	Propose a SAPP discriminator that considers face component information when it makes the pixel-wise real/fake decision.	MSPL-Center: CelebA: PSNR: 28.32 SSIM: 0.926 LPIPS: 0.071
Wang et al. [39]	GoPro, HIDE, RealBlur	Pre made models with	Propose a semantic prior	Improvements in measurements: +0.77dB in

		their proposed semantic prior		PSNR +0.98dB in MPRNet

Table 2.8 Table of articles with extracted technical information

Reference	Dataset creation description	Loss functions used	Other relevant information	Training environment descriptions
Xiao and Pan [30]	They convolve 20 000 images with 10 000 blur kernels and add Gaussian noise $\sigma = 0.03$. The sizes of these blur kernels are divided into five types: 13×13, 17×17, 21×21, 24×24, and 27×27.	Content loss for generator G_1 Content loss, face structure loss, perceptual loss and the adversarial loss for generator G_2	Adam optimizer, $\partial 1 = 0.99$, $\partial 2 = 0.999$ learning rate is set to 1×10^{-4}	Single NVIDIA 3060-Ti GPU.
Gao et al. [31]	There are 4 images and 12 blur kernels in dataset which are used for deblurring experiments in the paper.	None	Algorithm is implemented in MATLAB R2022a	Windows 10 operating system with an Intel(R) Core(TM) i5-8250U CPU @1.60GHz 1.80GHz.
Zhang et al. [32]	To obtain the Blur setting, they convolved HQ images with a Gaussian or Motion blur kernel. The images in the noise setting are generated by adding one of Gaussian, Laplace, and Poisson noise.	L1 pixel loss as the loss function	training epochs: 3 learning rate: 0.001	3090 GPU

Yang et al. [33]	Levin dataset: 32 blurry images generated by 4 clear images and 8 different types of blur kernels. Köhler dataset: 48 blurred images, generated from 4 high-quality images and 12 different types of blur kernels	None	MATLAB 2016b	PC with an Intel i5 processor, a 64-bit Windows 10 operating system, CPU is 4GHz
Yu and Xie [34]	They randomly chose 10 000 images from CelebAMask-HQ dataset and used pre-trained MTCNN to extract the clear face images from those images which they then convolved with the blur kernels and added 1%–5% random Gaussian noise. Kernels were randomly generated ranging from 11 × 11 to 39 × 39 by using the trajectory constructed method.	Loss function is a combination of: reconstruction loss, perceptual loss, facial identification loss, edge loss and adversarial loss	Optimization function: Adam, $\beta 1 = 0.9$ and $\beta 2 = 0.999$ Learning rate is initially set as 0.0001 and after 100 epochs it is linearly decreased to zero over the next 300 epochs batch size: 20 Each model took about 57 h to complete	Training on: Four Titan-X GPUs Some performance measures conducted on a computer with an i7 3.6-GHz CPU and a Titan-X GPU
Shi et al. [36]	For CelebA-HQ they used 25 blur kernels to blur the motion of the original image, with a 45 motion angle	Multi-region reconstruction loss and adversarial loss	Stage-I: Optimizer: SGD (stochastic gradient descent) Initial learning rate of 0.02,	Training on: NVIDIA RTX 2080Ti GPU

	For CelebA they use 3D camera trajectories to produce 25,000 blur kernels with sizes ranging from 13×13 to 29 × 29. They generate millions of pairs of clean-blurry data using both 160 × 160 image patches and convolution with 2500 blur kernels. They also apply Gaussian noise with σ = 0.03 to blurry images		momentum = 0.9, weight decay = 0.0005 Stage-II: Optimizer: Adam Initial learning rate of 0.0002, beta1 = 0.5, and beta2 = 0.999.	
Cui et al. [37]	CelebA dataset as the training set. They use MTCNN to detect face region and then adjust the region's size to 128x128 as Ground Truth. To simulate motion blur, they synthesize 20 000 motion blur kernels by random camera trajectories.	Super- resolution loss, deblurring loss, local loss and pixel-based	Optimizer: Adam beta1 = 0.9, and beta2 = 0.999. two-step training strategy: 25 epochs + 25 epochs	No information
Han et al. [38]	MSPL dataset consists of training set and a test set for face deblurring. 18000 motion blur kernels are synthesized from random 3D	Discriminator loss: Prediction-weighted loss Generator loss: reconstruction	Optimizer: Adam β1 = 0.9, β2 = 0.999. Learning rates of the generator and	Two NVIDIA Titan Xp GPUs

	trajectories, where the size of blur kernel ranges from 13 × 13 to 27 × 27 including {13 × 13, 15 × 15, 17 × 17, 19 × 19, 21 × 21, 23 × 23, 25 × 25, 27 × 27}. Each blurred image is obtained by convolving the sharp image with one of blur kernels with the addition of Gaussian noise with standard deviation of 0.015.	loss, prior feature loss, adversarial loss	discriminator are initialized to 1×10^{-4} and decayed exponentially by 0.99 every epoch Batch size: 16 Epochs: 300	
Wang et al. [39]	Pre made datasets. The blur in the HIDE dataset is more severe due to the aggregating of 11 consecutive frames, leading to huge movements in every blurry image	Cross-level feature distillation loss, Hierarchical Context Loss (HCL)	Batch size: 10 Optimizer: AdamW $\beta1 = 0.9$, $\beta2 = 0.999$ learning rate: 2×10^{-4} , which is steadily decreased to $1 \times 10{-}6$ using the cosine annealing strategy Epochs: 700	Single NVIDIA Geforce RTX 3090 GPU

By conducting the literature review and gathering this information into a clear table we can see some commonalities between the papers and approaches to facial image deblurring. We see that a lot of the authors use the same loss functions and same

optimizer Adam and that the most commonly used metrics for model evaluation are PSNR and SSIM. All this information can be used to make the decisions in the moment of making our experiment. Throughout the thesis, other relevant scientific papers were cited but were not included in the PRISMA statement itself.

2.3. Characteristics of Face Images

Human face contains a large amount of information and characteristics which could be further used for various tasks. For example, human face geometry information also known as facial priors can be used for face restoration, face deblurring and face super-resolution [40, 41]. Information contained in an image of a human face can be divided into three categories: human attribute information, human identity information and other prior information.

Human attribute information denotes to us whether face contains special attributes such as gender, age, emotion and others. These attributes are used for other high-level computer vision tasks such as gender detection, age detection, face recognition and emotion detection. The process of creating a dataset containing these attributes and deep learning models for predicting them is described in detail by Liu et al. [46]

Each face also has human identity information which are unique for each face. This identity information can be used to generate faces as close as possible to the real face identity. This type of information is always used for keeping the identity consistency between the deblurred result and the ground truth through the loss function which calculates the difference. It offers high-level constraints to the deblurring task because it would not be good should the person change visually after the deblurring process.

Among other prior information we can count facial landmarks [44], facial heatmaps, facial parsing maps [45] and 3D face prior. Facial landmarks are and important reference points of facial components. For example, they are eye centers, nose tips, mouth corners... Their number depends on the dataset. CelebA [46] dataset contains 5 landmarks, FFHQ [47] dataset has 68 landmarks while Helen [48] dataset contains 194 landmarks.

Facial heatmaps describe the probability that reference points are facial landmarks [49]. Facial parsing maps are semantic feature maps of face images which are separated out face components. These could be nose, skin, eyes and hair [45]. 3D face prior provides 3D knowledge based on the fusion of different face attributes. Incorporating these 3D face priors helps models grasp sharp facial structure and decrease the occurrence of face artifacts in

super-resolution tasks [51]. They can also be used in general face restoration tasks [52]. In dealing with face restoration Li et at. matched and selected the most similar facial component features from their corresponding facial dictionaries and transfered the high-quality details to the degraded images using dictionary feature transfer block [53]. There are also reference based priors such as high-quality guided images of the same identity [54] and generative priors such as rich and diverse priors encapsulated in a pre-trained face GAN [55].

2.4. Datasets

Face image deblurring datasets can be divided into two categories based on way they were made. Real-shot datasets contain blurry images which are made by averaging the video frames or moving the camera through a specific trajectory. Synthetic datasets are made by artificially adding blur to the sharp images. The blur could be added through convolution operation or by using deep neural networks for artificial blurring. The majority of datasets used in facial image deblurring are synthetic blurry-sharp image pairs, usually custom made for each experiment by the paper writers because of a severe lack of benchmark datasets for facial image deblurring. In this chapter we will cover general face image datasets which are used as a foundation for synthetic datasets. It is important to know their characteristics because of various image conditions such as image illumination or availability of image priors can affect the choice of starting dataset and at a later point, model performances depending on which dataset it was trained on.

2.4.1. Face image datasets

CelebA [46] dataset has more than 200 000 images. The number of images comes from a large number of identities in the dataset. There are 10 000 identities and each is represented with 20 images. The dataset is annotated by a professional labeling company and each face has 40 face attributes and five key points. The original images were selected and labeled from the CelebFaces dataset [56].

Helen [48] dataset contains 2 330 high-resolution face images for which it also provides 194 landmarks per face.

FFHQ [47] dataset was created by crawling images from Flickr and it consist of 70 000 high-quality human face images. Additional benefit of this this dataset is a good variety in age and ethnicity of human individuals and image backgrounds. It also covers some human accessories such as hats, glasses and sunglasses.

Further information about various datasets can be found in a survey by Wang et al. [23] where where he descibes facial image datasets and their characteristics.

2.4.2. Synthetic datasets

In the making of MSPL dataset Lee et al. [57] used 30 000 high resolution facial images from CelebA-HQ dataset [58] which the convolved with 18 000 motion blur kernels while also adding Gaussian noise.

Dataset made by Shen et al. [59] is made of images collected from several other facial image datasets. 2 000 from Helen dataset [48], 2 164 from CMU PIE dataset [60] and 2 300 from CelebA dataset [46]. Once collected, these images were convolved with 20 000 artificially created motion blur kernels with the addition of Gaussian noise. The final number is 130 million blurry-sharp image training pairs. They also created test set in a similar way which contains 16 000 images.

By cropping 110 000 images of size 320 × 320 from FaceScrub [62] dataset and convolving them with 10 000 random generated motion blur kernels while adding white Gaussian noise with standard deviation 2.55, Jin et al. [61] made their own dataset which they used to avoid overfitting.

In order to make their dataset, Lin et al. [63] collected images from various face databases. 2 184, 2 000, 2 000 and 2 400 clear facial images were collected from the CMU PIE, Helen, CelebA and PubFig [64] databases. After that, they were convolved with 20 000 motion blur kernels to generate blurry images. By utilizing data augmentation such as random rotation, cropping and scaling they increased the diversity of the data in the dataset.

HIDE dataset by Shen et al. [105] is not exclusively face image dataset but it contains images of pedestrians and city streets. It is composed of 8 422 clear-blurred image pairs. 6 397 pairs for training and 2 025 pairs for testing.

2.4.3.Real-shot datasets

These datasets contain images suffering from realistic blur which is hard to synthetically copy but the main issue is that often, there is no ground truth image to the corresponding blurred image.

Lai et al. dataset [65] presents a collection of real-world blurry images obtained in the wild and captured by different users using different cameras and settings. There are 100 real

blurry images of various kinds such as facial and text images. Since the images are captured in real life scenarios where blur occurred they do not contain corresponding ground truth images.

2MF2 dataset [66] was created by processing and extracting faces from videos. Motion blur is generated by averaging multiple frames of the same face. The middle frame of the averaging is considered as the ground truth image. Dataset consists of 1150 videos containing 2.1 million frames of facial images. For each image there are also facial landmarks.

2.5. Foundational terms and information

In facial image deblurring, the input of that goes into the model is typically a blurred facial image that needs to be deblurred into a sharp, clear image which is called output. This input image is a three dimensional tensor because it consists of height, width and color channels which represent its dimensions. Input images are colored images and they have red, green and blue channels, while if the input images would be grayscale, they would have only one channel

Convolution is a mathematical operation in deep learning architectures used for facial image deblurring. It involves sliding a small matrix (also called a kernel) across the input image computing dot products between the kernel and the local regions of the input image to extract relevant features like edges, textures, and fine details of the face. Once the feature maps are extracted, they need to be reduced.

Pooling is a downsampling operation used to reduce the spatial dimensions of feature maps in deep neural networks which helps to reduce computational complexity.

In order to learn complex patterns of facial images the deep learning network needs to learn to model non-linearity. That is achieved by activation functions. Two of the most used activation functions are ReLu and LeakyRelu. ReLU (Rectified Linear Unit) outputs zero for negative values in the input and the value itself for zero and positive inputs. LeakyReLU modifies the ReLU function to allow a small, non-zero gradient when the input is negative.

Batch normalization [42] is a method for normalizing layer inputs throughout model architecture for each training mini-batch. By doing this input normalization process it enables us to use higher learning rates while also being less strict about initial weight initialization because it ensures that the distribution of activations remains consistent during training. For different levels and types of blurring in facial images which cause high variability in the training data this regularization technique helps the model by reducing

internal covariate shifts.

Optimization algorithms such as Stochastic Gradient Descent (SGD), Adam [116], or RMSprop are used to adjust the weights of the model during training to minimize the loss function.

Upsampling techniques like interpolation, transposed convolutions, or sub-pixel convolutions increase the spatial resolution of feature maps. They are used for reconstructing high-resolution images from compressed representations, ensuring that the deblurred facial images retain important details.

Latent space representation is the compressed feature representation of the input data in the bottleneck layer of an autoencoder.

An epoch is one complete pass through the entire training dataset, while batch size is the number of samples processed before the model is updated.

Data Augmentation techniques are used to artificially expand the training dataset by applying transformations like rotation, scaling, flipping, and adding noise. It enhances the ability of the model to generalize by exposing it to a wider variety of images.

Hyperparameters are configurable parameters external to the model (e.g., learning rate, number of layers, filter sizes).

2.6. Basic Layers and Building Blocks

This section will go over some of the most common deep learning network layers and blocks used for image deblurring.

Convolutional layers form the foundation of many facial deblurring networks. These layers are essential for extracting features from blurred images by applying convolution filters and can be trained to directly recover sharp images without kernel estimation steps. Multi-scale convolutions can be employed to capture both coarse and fine image details, improving the model's ability to handle complex motion blur patterns.

Recurrent layer can extract features across images at multiple scales in a coarse-to-fine manner. This ability to model temporal dependencies allows recurrent layers to progressively refine the deblurred image by retaining useful information from previous iterations, leading to sharper and more accurate outputs. For example, Ren et al. use a combination of recurrent and convolutional layers for better deblurring [67].

Residual layer is used to directly connect low-level and high-level layers in the area of image deblurring in order to avoid vanishing or exploding gradients during training by

introducing skip connections. ResBlock uses local residual layers, similar to residual layers in ResNet [68]. These blocks ensure that information passes through the network more effectively, reducing the degradation problem that can occur in deeper networks. Global residual connections are also used to learn global feature information in an overall manner.

Dense layer is also used to address gradient vanishing problem but also to improve feature propagation and reduce the number of parameters. DenseBlocks are used to replace CNN layers or ResBlocks. Dense connections in residual dense networks can connect all layers from the previous state to the current state, allowing the network to adaptively learn more effective features [69]. These connections between layers allow the model to better capture complex blur patterns and improve the efficiency of deblurring, as each layer receives inputs from all preceding layers.

Attention layer helps deep networks to focus on the most important regions for deblurring. It can also be employed to extract better feature maps. In facial motion deblurring, attention layers allow networks to prioritize facial features that are critical to the restoration of sharpness, such as eyes and facial contours. These layers help networks extract more informative feature maps by dynamically weighting the importance of different spatial regions, improving the overall performance of the deblurring model. Various types of attention mechanisms have been developed: (1) Channel Attention which focuses on informative channels closely related to blur artifacts [70], (2) Spatial Attention which helps perceive blurry spatial position information [71] and Feature Attention which is composed of both channel and pixel attention mechanisms. This approach focuses on blurred pixels and important channel information, effectively solving the problem of uneven blurred distribution in images [72].

2.7. Models for facial image deblurring

Before describing the facial image deblurring models it is important to understand the division of approaches to the deblurring problem. As we previously stated, image blur occurs after the image is degraded by the blur kernel. If that blur kernel is known to us before the deblurring process, then we are dealing with non-blind facial image deblurring. Since the purpose of our work is to deblurr real life facial motion blur it will always be the case that we simply don’t know the blur kernel, especially since that blur kernel is usually non uniform. In that case we are dealing with blind facial image deblurring in which we have to estimate the blur kernel and negate or fix the image degradation it has caused. Furthermore,

we can divide the facial deblurring methods into model-based and deep learning based methods. Most model-based deblurring methods try to restore edges implicitly or to estimate the blur kernel explicitly and thus deblurr the image but they are dealing with facial images who lack a large number of edges needed for blur kernel estimation. After the arrival of convolutional neural networks, model-based methods no longer represent the state-of-the-art. In this thesis we will be focusing on deep learning models to achieve facial deblurring and because of that only them will be descibed.

Deep learning based methods for facial image deblurring can be further divided based on the learning strategy into supervised and unsupervised learning. Emerging from the difficulty of creating large datasets of paired sharp and blurred images, unsupervised learning should have enabled the models to learn only from the blurred images and in such way enabled them to effectively deblurr images. Since the datasets lacked sharp images as a reference, the images generated by unsupervised deblurring models were of low quality [21]. In this thesis we will focus on supervised learning methods in which the models have access to the ground truth images and their blurred counterparts.

Depending on the architecture of the models we can differentiate deep autoencoders, Generative Adversarial Networks (GAN), cascade networks and multi-scale networks. In order to further improve their models, some authors use various facial priors in a combination with some architectures, so we can talk about prior guided deblurring or no prior guided facial image deblurring. Out of these networks, cascade networks and multi-scale networks are usually used in restoring heavier damage to the picture or simply trying to upscale the image using face hallucination. Zhu et al. [43] solved the upscaling problem of faces with significant pose and illumination variations by using deep cascaded network. Their work falls in the face hallucination category of face image restoration since their foundational input images are of very low quality. Similarly, Chen et al. [45] solve restoration of low quality facial images by using multi-scale network. The severely degraded images go through different scales which restore them in a coarse-to-fine manner. Because in this thesis we are dealing only with motion blur which usually presents the biggest and most common cause of distortion in taking pictures, we will be focusing more on similar approaches to the problem which mostly use autoencoders and generative adversarial networks.

Deep autoencoders are neural networks designed to learn efficient data representations. They consist of two main parts: (1) Encoder which takes an input image and compresses it into a

lower-dimensional latent representation or "code". The purpose is to extract the most important features from the input while discarding irrelevant noise or detail. (2) Decoder which takes the latent representation from the encoder and attempts to reconstruct the original image from it. In the case of image restoration tasks like deblurring, the goal is to produce a clearer, sharper version of the input. Yu and Porikli [50] demonstrated that autoencoders could even be used for more demanding tasks such as face hallucination from very low resolution input images.

Generative adversarial networks consist of two competing neural networks: (1) Generator which creates synthetic data, in case of facial image deblurring that data are deblurred facial images. (2) Discriminator which distinguishes between real and generated data [77]. The purpose of these two networks is for them to work in competition in order to achieve good deblurring results and realistic output images. The generator learns how to produce sharp, realistic facial images from blurred inputs while the discriminator learns to differentiate between real sharp images and generated deblurred images. Throughout the literature we can see improvements and experimentation with both modified generators and with modified discriminators.

Xiao and Pan [30] propose a progressive GAN network which consists two generators which incrementally improve the image. Both generators are convolutional U-net networks [112] which consist of encoder and decoder.

Han et al. [38] focus their modifications of discriminator whose task is to decide whether the picture given to him is real or fake (synthetically made). Their proposed discriminator considers face component information when it makes a decision.

Generator in the network by Shi et al. [36] consist of several components. A series of ResBlocks forms the encoder, extracting shallow and deep features. Each ResBlock consists of two residual modules, each with convolution layers, spectral normalization , and ReLU activation. Decoder block consists of three modules (convolution, batch normalization (BN), and ReLU activation) for upsampling. Laplace depth-wise separable convolution enhances texture information in shallow feature maps during decoding. Feature adaptive denormalization block is used for adaptive normalization of facial regions, based on face parsing information. Their network has two discriminators who influence each other. Detailed image information are obtained by the multi-patch discriminator while the realism of the restored image is obtained by global discriminator.

2.8. Loss Functions

To optimize face deblurring network, numerous loss functions have been proposed in the literature. The choice of loss function during network training is critical to the final deblurring performance. [73] Loss functions measure difference between the deblurred outcome and the original target image. Many studies will combine several loss functions in the form of a weighted sum for better deblurring effects because the model learns to consider several different aspects of good image quality. Below are descriptions, impacts and flaws of several popular loss functions.

Content loss also known as reconstruction loss can be formulated in two types:
L2-norm content loss also known as mean squared error (MSE) and L1-norm or mean absolute error (MAE). This loss computes the discrepancy of pixel values between the ground truth image and an output image of a network according to the corresponding norm. Minimizing this loss helps the deep learning network in restoring overall content and structure of the image and makes sure that the pixel values of the deblurred output image should be as close to the sharp ground truth image as possible. However, this method of pixel difference comparing does consider human subjective visual perception and tends to lead to over smoothed output results.
Perceptual loss [74] compares the ground truth and output images in their CNN feature representations instead of pixel-wise differences as would content loss. The purpose of this loss function is to make an output image perceptually indistinguishable from the ground truth image. The authors in [75] extracted pool4 layer from pre-trained VGGFace network [76] to compute the perceptual loss. VGGFace network is pre-trained on a large face recognition dataset and its intermediate features capture useful facial characteristics that can guide the face deblurring process.
Adversarial loss [77] is employed to generate realistic images in a GAN structure. The loss aims to restore more texture details of the output image by encouraging the generation of more realistic and visually appealing results.
Because of the special characteristics of the human face it is possible to use face-related losses. This kind of loss aims at incorporating information related to the human face structure in face deblurring process. Among these we count heatmap loss and identity preserving loss.

Each loss function helps achieve different results in deblurred images. It is necessary to carefully select a variety of loss functions according to the actual needs of the model. More information about some loss functions and their characteristics can be found in Table 2.9.

Table 2.9 Loss functions and their characteristics

Loss Function	Description	Strengths	Weaknesses	Computational Cost	Facial Detail Preservation
L2 Content Loss (MSE)	Measures the pixel-wise differences between the ground truth image and output using squared differences.	Encourages smooth, globally consistent images.	Leads to over-smoothed results; doesn't consider human perception.	Low	Low (treats each pixel equally, it doesn't emphasize critical facial regions like eyes, mouth, or nose.)
L1 Content Loss (MAE)	Computes the absolute differences between pixel values of ground truth and deblurred output.	Encourages sharpness and preserves structure better than MSE.	Can lead to poor gradient learning in certain cases.	Low	Low (It treats all pixels equally)
Perceptual Loss [74]	Compares features extracted from a pre-trained CNN (VGGFace) to measure high-level differences.	Produces perceptually better results, preserves texture and structure.	Requires pre-trained networks.	High (feature space comparisons, significantly increasing GPU memory usage and computational time)	Medium (not explicitly tuned to preserve facial details like landmarks or identity)
Adversarial Loss [77]	Encourages the model to produce outputs that are indistinguishable from real images by training a GAN.	Produces more realistic images with finer details and texture.	Can lead to unstable training and mode collapse if not tuned properly.	High (GAN structure demands additional forward and backward passes through both networks)	Medium (does not specifically target facial landmarks or features)

Relativistic Loss [78]	A variation of adversarial loss that considers the relative realism between generated and real images.	Improves visual realism by comparing the realism of pairs of images (generated vs real).	Can require careful tuning.	High (comparing real and generated images relative to each other adds complexity)	Medium (effectiveness in preserving facial details is limited unless used with complementary losses)
Local Structure Loss	Focuses on preserving local structural information (e.g., edges and contours) in the deblurred image.	Helps in preserving fine local details such as facial features, edges, etc.	Sensitive to noise, can be difficult to optimize when combined with other global losses.	Medium (gradient-based or edge-based computations)	High (eyes, nose, and mouth often contain strong local gradients and edges)
Optical Flow Loss [79]	Measures the consistency of motion between frames in a sequence of images, important for video deblurring.	Helps to preserve temporal consistency and smoothness in dynamic scenes.	More computationally demanding, especially in real-time video deblurring.	High (especially when high-resolution videos or long sequences are involved)	Medium (it doesn't explicitly focus on preserving facial details)
Heatmap Loss [40]	Utilizes heatmaps that highlight important facial landmarks to guide the deblurring process.	Helps preserve important facial regions like eyes, mouth, etc.	Can overlook general image quality if not well combined with other loss functions.	Medium (requires generating heatmaps, but heatmaps are typically low-dimensional)	High (designed to preserve facial details, focusing on critical regions of the face by guiding the network to pay attention to key features)
Identity Preserving Loss [80]	Ensures that the deblurred image retains the identity of the subject, often using a face recognition model.	Important for applications where facial recognition or verification is required.	Adds complexity and computation, relies heavily on pre-trained face recognition models.	Medium (adds some computational cost due to identity-related feature extraction)	High (it focuses on preserving the features most critical to facial recognition)
Unsupervised Losses [81]	Used when ground truth images are not available,	Can be applied to real-world datasets	Generally less accurate than supervised losses, can	Medium (need for additional operations	Medium (may struggle to perfectly

	combining techniques like reblurring loss and self-measurement.	without ground truth data, increasing versatility.	require intricate model designs.	like self-supervised learning tasks)	preserve facial details)
Landmark Loss [82]	Preserves the spatial alignment of facial landmarks such as eyes, nose, and mouth during deblurring.	Useful for preserving facial symmetry and alignment.	Overly focused on landmarks, can miss global quality.	Medium (there is cost of detecting and aligning landmarks)	High (excels at preserving facial details related to the position and structure of facial landmarks)
Facial structure Loss [83]	Focus on spatial relationships between facial components and their visibility	Ensures specific facial component s retain their sharpness and structure.	Adds complexity, may not generalize well across all images.	Medium	High (directly focuses on key facial components)

2.9. Evaluation Metrics

In order to judge the motion deblurring model quality it is important to evaluate the output image quality. Since we are dealing with images of human face it is difficult to construct a sufficiently objective metric that conforms to human perception. There are three types of evaluation metrics in facial image deblurring. The first type of metric evaluates the images on the pixel level of the image. The second type evaluates the quality of the images in terms of visual perception by extracting the deep features of the image. The third type of metrics are task oriented, which evaluates the quality of images by comparing their accuracy in advanced visual task. Furthermore, we can divide evaluation metrics on whether they need ground truth images for calculating the metrics. Full-reference metrics assess the image quality by comparing the restored image with the ground-truth and no-reference metrics use only the deblurred images to measure the quality.

2.9.1. Image-level evaluation metrics

These metrics do not require any additional inputs and can be obtained through basic calculations. In their calculation they use ground truth images corresponding to the deblurred outputs of the models.

Peak Signal Noise Ratio (PSNR) is acquired by calculating the pixel-level mean squared error of two images. The larger the value the smaller the difference between the two images. Using PSNR tends to lead to over smoothed results.

Structural Similarity Index Measure (SSIM) [84] is a metric modeled after the visual system of humans. It measures the difference between two images in terms of contrast, brightness and structure. The larger the value, the more similar are the images. The metric can be deceiving because even the images with lower SSIM can have good visual experience to the human visual perception.

2.9.2. Perceptual evaluation metrics

LPIPS [85] calculates the distance between two images in the high dimensional feature space. These high dimensional feature spaces are extracted using a pre-trained classification network. The smaller the value, the more similar two images are.

NIQE (Natural Image Quality Evaluator) [86] is a no-reference image quality assessment metric which makes it suitable for evaluating deblurred facial images when the ground truth is unavailable. NIQE uses a natural scene statistic (NSS) model to capture deviations from statistical regularities observed in natural images. For deblurred facial images, a lower NIQE score would suggest more natural-looking results

2.9.3. Advanced visual task evaluation metrics

The whole purpose of facial image deblurring is to prepare the blurred image for an advanced visual task. It can improve accuracy of high-level visual tasks such as face detection or face recognition. For example, face verification rate measures the ability to correctly match deblurred facial images to their corresponding high-quality counterparts. Face identification rate assesses how well the deblurred facial images can be identified in a larger database of face images.

Beyond these mentioned metrics there are many more. Many of them are described and compared by Ding et al. [87] It is also important to have in mind various advantages and disadvantages of these metrics depending on the deep learning network that we are using for deblurring. Evaluation metrics for GAN are described and analyzed with examples by Borji [88, 89]. Further information about various evaluation metrics can be found in table 2.10.

Table 2.10 Evaluation metrics and their characteristics

Metric	Range	Higher result means	Strengths	Weaknesses	Best used for	Noise Sensitivity	Requires Pre-trained Network
MSE / RMSE	[0,∞)	Worse	Easy to compute, widely used	Poor correlation with human perception	Simple comparisons when perception is not critical	Low	No
PSNR [90]	[0,∞) dB	Better	Easy to compute, widely used	Poor correlation with human perception	Simple comparisons, used when perception is not critical	Low	No
SSIM [84] / MS-SSIM [91]	[−1,1] [0,1]	Better	Better correlation with human perception, incorporates structure	Computationally expensive, less suitable for deep learning	Comparing methods with different structural properties	Medium	No
LPIPS [85]	[0,1]	Worse	Good correlation with human perception, robust to noise	Requires a pre-trained deep network, expensive computationally	Evaluating deep learning methods when human perception is key	High	Yes
FID [92]	[0,∞)	Worse	Compares feature distributions , correlates well with perception	Requires pre-trained network, sensitive to noise	Evaluating generated images with deep learning	High	Yes
IS [93]	[1,∞)	Better	Measures quality and diversity, common for GAN evaluation	Sensitive to noise, not always correlated with perception	Evaluating generated images and diversity	High	Yes
NIQE [86]	[0,∞)	Worse	No-reference, doesn’t need pre-trained data	Can underperform compared to perceptual-based metrics	No-reference evaluations of real-world deblurred images	Low	No

BRISQUE [94]	[0,100]	Worse	No-reference, non-deep-learning based	Can be sensitive to non-structural changes	No-reference evaluations, perceptual quality	Medium	No
VIF [95]	[0,1]	Better	Measures information preserved in blurred vs. original image	More computationally intensive	Evaluating structural similarity and deblurring performance	Medium	No
PI [96]	[0,∞)	Worse	Combines NR-Metrics (NIQE, MA) and perceptual FID for balance	Computationally expensive, may overestimate noise effects	Comprehensive evaluation of perceptual quality	High	Yes
GMSD [97]	[0,∞)	Worse	Fast, incorporates global structure similarity	Less robust to image artifacts	Comparing global image structures and deblurring effectiveness	Medium	No

3. Dataset Creation Methods

As described previously, the majority of training datasets are made using some facial image database and convolution operation with motion blur kernels with some additional Gaussian noise added to the blurred image. These synthetic datasets with uniform blur kernels are not good representations of realistic motion blur. It is difficult to simultaneously capture sharp ground truth image of the human face and image from the same angle of the same face blurred with motion blur. One way of creating realistic motion blur is to use specially mounted cameras. Camera motion has six degrees of freedom in two categories, translational and rotational motions. Translational motion relates to depth variation [98,99], while rotational camera motion and object motion are independent factors that also lead to non-uniform blurs in the image. By shaking the camera in specific directions, realistic motion blur can be achieved. Another example mention is averaging frames containing faces from a video sequence and achieving motion blur that way. One such example is GoPro dataset by Nah et al. [104]. By averaging varying number of successive latent frames (7 - 13) they managed to produce blurs of different strengths. This dataset is used for general scene deblurring but it descibes in great detail its synthesizing. Another maybe more useful example is HIDE

dataset by Shen et al. [105]. Its focus on motion blur in pedestrian and city streets, which also includes camera shake and object movement enables us to test some of our models on blurred image patches containing human faces. The blur was synthetically made by averaging frames from high frame rate videos captured by GoPro Hero camera. The ground truth image is the center frame in the average. The problem with all these methods is that the resulting motion blur covers the entire image which is in contrast to some real life scenarios of photo taking where only the person, and subsequently their face, moves while the background is stationary and sharp in comparison.

3.1. Traditional Motion Blur synthesis methods

In order to successfully train image deblurring model it is important to simulate realistic motion blur in the training dataset image pairs. Several methods of generating simulated motion blur kernels have been proposed.

Boracchi and Foi [106] model motion blur through Point Spread Function (PSF) trajectories that represent the motion of the camera or object during exposure time. These trajectories are treated as random processes, and the PSFs are generated as a function of time, exposure, and motion characteristics. The PSF in this case defines how a single point of light spreads over the image sensor due to motion during exposure, causing the blur. The blurred image is modeled as a convolution of the original image with the PSF and corrupted by noise. The convolution process simulates how the image would look if the camera or object moved during the exposure.

An article by Chakrabarti [107] presents an older description of synthetic dataset creation for training neural network for deblurring. Training dataset was constructed by blurring extracted sharp image patches with synthetically generated kernels, and adding Gaussian noise. The motion kernels are generated randomly by selecting six points on a grid and fitting a spline through them. The intensity of the kernel at each point is sampled from a Gaussian distribution. These values are clipped to be non-negative and normalized to have a unit sum. This process simulates a variety of real-world motion blurs caused by camera shake or object motion.

In the paper by Sun et al. [108] motion blur is modeled using motion vectors that represent the length and direction of the blur at each pixel. Each vector corresponds to a local motion blur kernel, which defines how the motion affects the image during exposure. This approach

captures the non-uniform nature of blur, where different regions of the image may experience different blur magnitudes and directions. The lenghts and orientations are sampled from different sets of values and blur kernel are created. These motion kernels are then convolved with sharp images to create synthetic blurred images.

3.2. Realistic Motion Blur synthesis methods

Some newer articles offer more realistic motion blur simulation methods.

Li et al. [109] introduce Blur Space Disentangled Network (BSDNet) which is designed to extract latent blur features from blurry images and use these features to generate synthetic blurry images. This helps in simulating realistic blur styles and augmenting training datasets. Motion blur in this framework is modeled by treating blurry images as a combination of two spaces: the blur space (which encodes the characteristics of the blur) and the sharp image content. The BSDNet network is trained to disentangle these spaces from paired sharp-blurry images to synthesize blurry images with the same blur styles.

Rim et al. [110] created RSBlur dataset and used it to analyze the difference between the generation process of real and synthetic blur. Based on their analysis, the authors develop a new synthesis pipeline to create realistic blurred images. Motion blur is modeled using sharp video frames, the number of original frames is expanded using interpolated and averaged to simulate the accumulation of light which closely mimics the real blur process. The pipeline synthesizes saturated pixels by generating a mask of saturated regions in each sharp frame. It then calculates a final mask for saturated pixels and uses this to adjust the synthetic blurred image, adding saturated effects to make the image more realistic. To simulate real-world conditions, the pipeline adds noise, modeling it as a combination of Gaussian and Poisson noise. Parameters for the noise model are estimated from real camera data to enhance realism. Finally, after adding noise, the image is processed through the camera's image signal processing, including steps like white balance, demosaicing, color correction, and gamma adjustment. This results in a blurred image that closely mimics the appearance of real-world blurred photos.

In the paper by Bahat et al. [111] Motion blur is modeled as a non-uniform field, where each pixel has its own local blur kernel, making the problem more complex than usual uniform blur deblurring. The blur at each pixel is assumed to be linear (1D), but the direction and magnitude of the blur vary spatially. The blur matrix K models this non-uniformity, with each row of the matrix representing a blur kernel specific to a pixel. Because they do not

have the sharp images, the authors use the concept of Re-blurring to model the blur field. This means that the input blurry image is blurred again using different 1D blur kernels (varying in length and direction), and the kernel that minimizes the difference between the re-blurred image and the input blurry image is selected as the estimated local blur kernel. This allows the algorithm to estimate the non-uniform blur field pixel by pixel.

3.3. Face Segmentation Model

The purpose of segmentation model is to segment the face from the surrounding background for precise motion blur application just to the face area. This would more accurately simulate real world scenario in which the person moved while the photo was being taken. The entire human head can be segmented or just the face area depending on the final purpose of image deblurring. More about face segmentation and image segmentation can be found in reviews by Khan et al. [100] and Minae et al. [101]. In this thesis, for face segmentation we will use DeepLabV3 model and MediaPipe framework. Chen et al. [103] use atrous (dilated) convolution which enables the segmentation model to have a larger receptive field without increasing the number of parameters and to retain spatial resolution in the feature maps and capture multi-scale information more efficiently. DeepLabv3 achieves improved segmentation performance in comparison to previous DeepLab versions. MediaPipe [102] is a framework developed by Google for creating pipelines that process sensory data in real time.

4. Experiment Setup

The purpose of the experiment is to see whether face segmentation before synthetic image blurring will improve deblurring results. Human face dataset was chosen, then two different ways of segmenting faces were selected. In each case two different dataset were made. One of those two had image pairs of sharp images and blurred images where the convolution operation was applied to the entire image while the other dataset had image pairs of sharp images and blurred images where the convolution operation was applied only to the segmented face area while the surroundings remained sharp. It is important to say that the blur kernels were the same for both version of the dataset and the only difference was whether the blur kernel was applied to the entire image or the segmented face area.

After the dataset creation, two different deep learning models were created for image deblurring. Created models were trained on image pairs and they were tested on different test datasets.

4.1. Hypothesis

H0: There is no significant difference in PSNR and SSIM values between the same models trained on different datasets.

H1: Models trained on a dataset where motion blur is applied only to a segmented facial area achieve significantly different results of PSNR and SSIM values compared to the same models, trained on a dataset where motion blur is applied to the entire image, because they more realistically simulate the real case of the appearance of motion blur where the person is moving and the background is static.

4.2. Environment description

The initial experimentation was conducted on a local laptop computer with these specifications: Windows 10 64-bit operating system, AMD Ryzen 5 5600H with Radeon Graphics 3.30 GHz, NVIDIA GeForce RTX 3050 Ti Laptop GPU, 16 GB RAM. The input data of every model are colored face images, each of 3 channels (RGB) which makes the training computationally expensive. Because of that, training runs on previously stated hardware were simply too long. For one of the models, one epoch or one pass through the training image set took eight hours. In order to deal with that, another environment was selected. All experimentation, including dataset creation, model training and testing was done using Google Colab platform. Dataset images were loaded from Google Drive. Because experimentation was run on their side and depended on their hardware, there was no point in measuring training time or similar performance metrics.

The code was written using the Python programming language and PyTorch [118] library for machine learning. All training and testing was done on A100 40GB inside Google Colab.

4.3. Dataset

The foundational facial image dataset is the CelebA-HQ train image dataset consisting of 24 183 images. Only the train set was selected as a starting point for further processing because working with images is computationally demanding. The train set was also chosen instead of random subsampling from the entire dataset because it is easier to take

the same train set of images for experiment replication. This selected dataset was subsequently divided into 3 parts: the train dataset for training, val dataset for validation during training and test dataset for model testing. 90% of the original images went into train, 5% into val and 5% into test datasets.

After that division, two dataset pairs were created. Each pair contains two more datasets. One of those two has blur kernels applied to the entire image and the other has blur kernels applied only to the segmented face area. In one pair MediaPipe [102] was used for face segmentation and in the other pair, DeepLabV3 model [103] was used for face segmentation. The point was to train the same models on two different datasets and see does the selective blur kernel application improve model evaluation metrics. The purpose of the fully blurred dataset in each pair is to serve as ground truth method that is representative of general approach in the literature. Each pair is created using their own unique blur kernels to ensure diversity but it is important to notice that within the pair, each original sharp image is convolved with the same blur kernel whether the blur was applied to the entire image or the segmented part of the image. Motion blur was simulated in similar way as examples in Table 2.8. The length which controls the size (strength of the blur effect) of the blur kernel and the angle at which the blur is applied, were created randomly and applied to the entire image or just face region. Diagram of dataset creation can be seen in Figure 4.1.

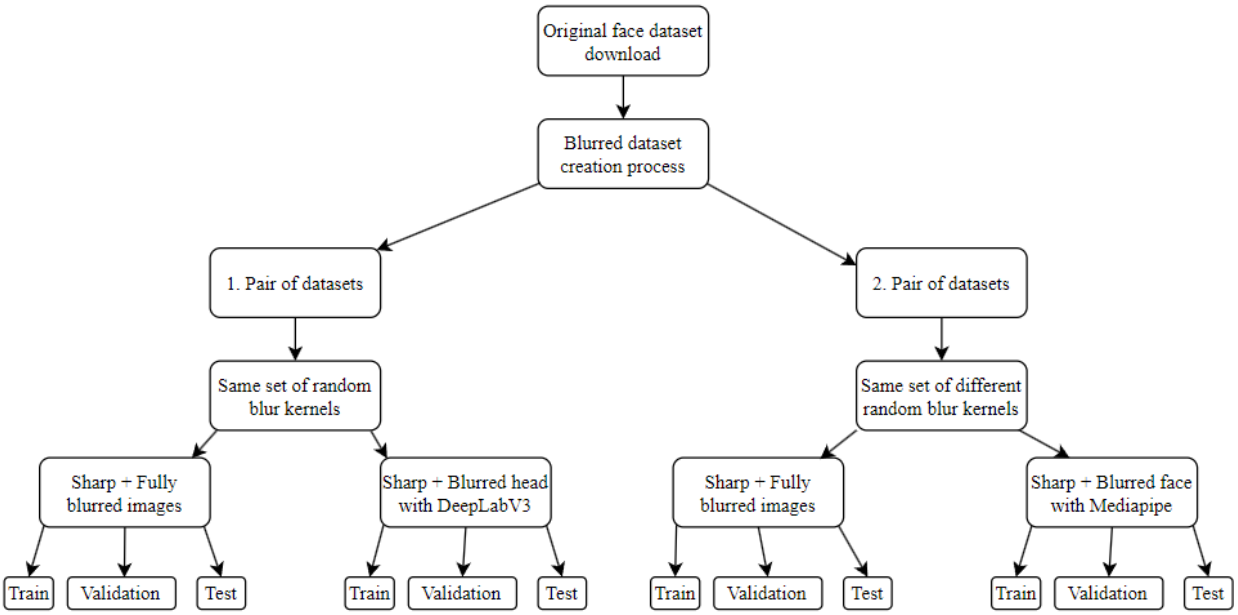

Figure 4.1 Dataset creation diagram

Examples of motion blur kernels and how they affect the image can be seen in Figures 4.2 and 4.3.

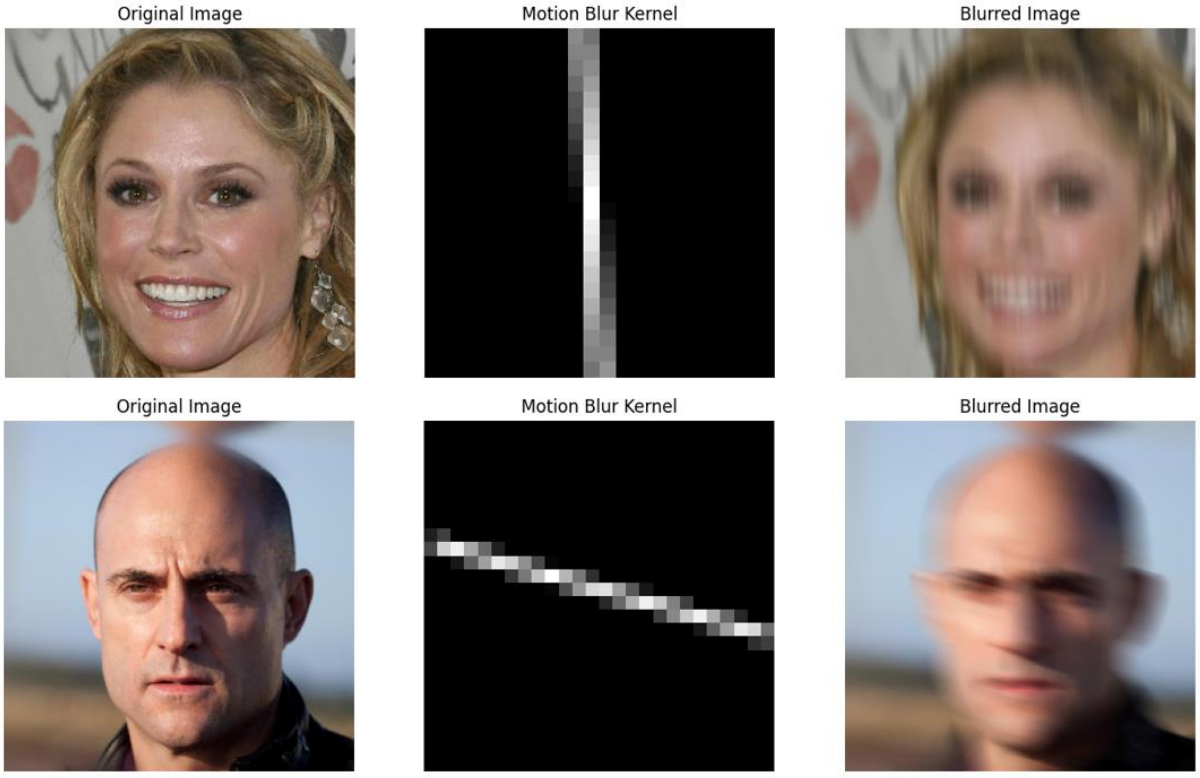

Figure 4.2 Different angles of motion blur kernels

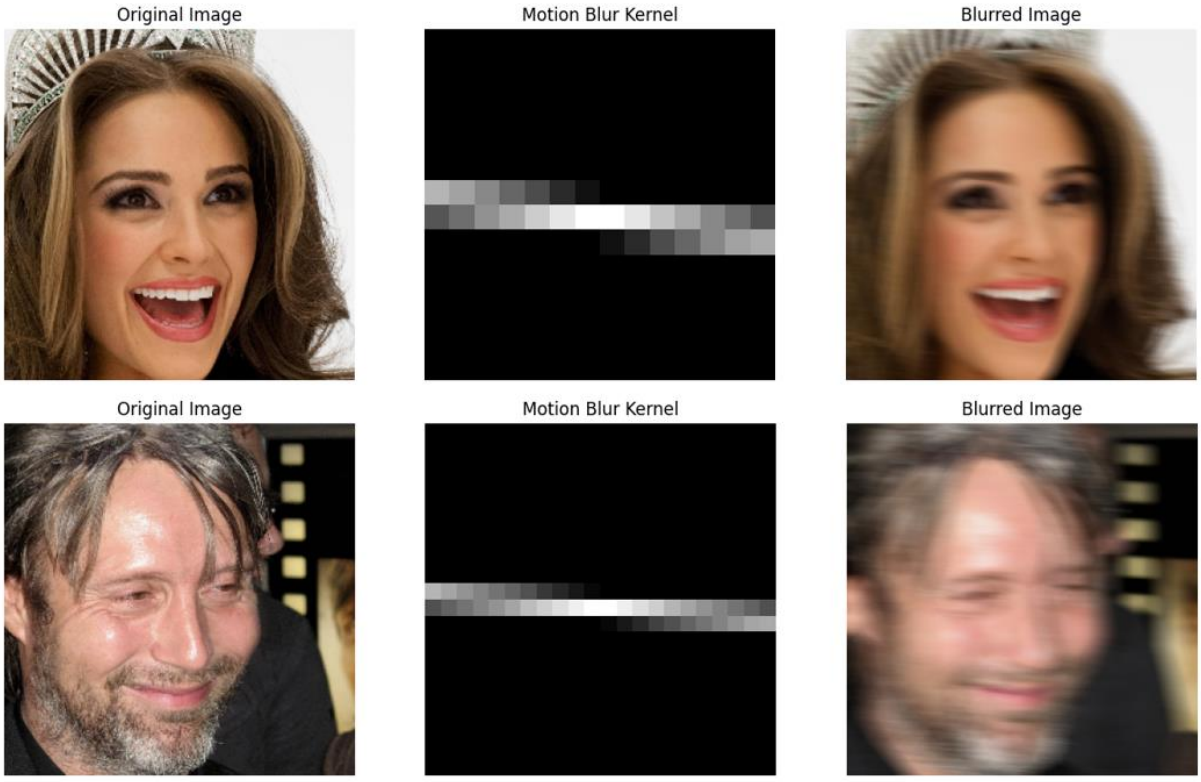

Figure 4.3 Different lengths of motion blur kernels

4.4. Models for Face Deblurring

Two models are created for testing the hypothesis. Throughout this thesis we have seen the immense complexity of the facial image deblurring problem. There are various elements which contribute to the performances of the model such as training dataset selection, model architecture, loss functions and criteria for evaluation. The models created for testing this hypothesis are relatively simple. There are no complicated loss functions or priors used because the emphasis is on testing just the different datasets, made by different approaches. Previously, through the thesis, we have described important building elements of these models and other important terms mentioned in this chapter of model descriptions.

The first model is autoencoder model. It consists of two main components: an encoder and a decoder, with additional skip connections between corresponding layers in the encoder and decoder to improve learning by providing fine-grained information during reconstruction.

The encoder performs downsampling and feature extraction through a series of convolutional layers. ReLU (Rectified Linear Unit) activation function is applied after each convolution, introducing non-linearity. Each layer progressively reduces the spatial dimensions while increasing the depth of the features, capturing more complex information at each stage.

The decoder reconstructs the deblurred image by upsampling the feature maps back to the original resolution using transpose convolutions. Two skip connections help the network retain detailed information from earlier stages in the encoder.

In the forward pass the image goes through the encoder, producing four sets of feature maps. Then the decoder reconstructs the image, starting from the deepest feature map and progressively upsamples it. Skip connections are added at intermediate stages.

The model uses a combined loss function that incorporates both Mean Squared Error (MSE) and a perceptual loss based on VGG19 features.

A VGGFeatureExtractor is used to extract features from both the deblurred output and the ground truth image. Depending on the features we want to extract, we can cut off a pre-trained network on different layers. This network uses a pre-trained VGG19 network up to layer 21 (Conv4_2).

The perceptual loss is calculated as the MSE between the feature maps of the output and the target image. This ensures that the deblurred image not only matches the pixel values of the target but also has perceptually similar features.

The total loss is a combination of the MSE loss at the pixel level and the perceptual loss and can be described by Eq (4).

$$\text{combined\_loss} = \text{MSE (output, target)} + 0.001 \times \text{Perceptual Loss} \quad (4)$$

The perceptual loss is given a smaller weight (0.001) to balance the importance of pixel accuracy and perceptual similarity.

The optimizer used is Adam [116] due to its adaptive learning rate capabilities and momentum properties. We have previously seen that many models mentioned in the literature review also used Adam as an optimizer.

The model is trained for 50 epochs on A100 inside Google Colab. Other models details can be seen in Table 4.1.

Table 4.1 Autoencoder model training details

Parameter	**Value**
Learning rate	0.001
Batch size	16
Optimizer	Adam
Criterion	Combined Loss (MSE + Perceptual Loss)
Maximum epoch	50

The second model is a Generative Adversarial Network (GAN) for facial motion deblurring, consisting of two primary components: Generator whose purpose is to attempt to generate deblurred images from blurred ones and Discriminator which aims to classify images as either real (original) or fake (generated by the generator). The generator architecture is based on a convolutional neural network (CNN) with residual blocks, which is a powerful way to preserve feature representations while progressively refining them. The purpose of the CNNs is to progressively reducing the spatial resolution of the input image. The core of the generator consists of 9 residual blocks, which help the network retain important low-level information and stabilize training. The residual connection (skip connection) helps in preserving features by adding the input of the block to the output, allowing gradients to flow more easily during backpropagation.

After the residual blocks, upsampling is performed using transposed convolution layers (also called deconvolution), which increase the spatial resolution. The ReLU activations are used for non-linearity, except for the final layer where Tanh is used to map the output image to a specific range suitable for image representation. The discriminator is designed to classify whether an input image is real (original) or fake (generated by the generator). It is structured as a deep convolutional neural network (CNN). The architecture consists of multiple convolutional layers that progressively reduce the spatial resolution of the image while increasing the feature depth. Each convolution is followed by LeakyReLU activations. Batch normalization is applied after each layer (except the first) to stabilize and speed up training.

The GAN is trained using a combination of adversarial loss, content loss, and a custom SSIM loss. The adversarial loss is based on binary cross-entropy loss (BCELoss). It measures how well the generator fools the discriminator and how well the discriminator distinguishes between real and fake images. For the generator, this is the loss that encourages it to produce images that look like real images. For the discriminator, it encourages it to correctly classify real and fake images. The content loss is based on Mean Squared Error (MSELoss) between the generated (deblurred) image and the original image. It ensures that the deblurred image is similar to the original in pixel space, emphasizing per-pixel similarity. Structural Similarity Index (SSIM) is used as an additional perceptual loss that focuses on image structure and luminance similarity. SSIM computes the similarity between two images based on the degradation of structural information. The final loss for the generator is a weighted combination of the adversarial loss, content loss, and SSIM loss. It can be described by Eq. (5).

$$g\_loss = g\_loss\_adv + 100 \times g\_loss\_content + g\_loss\_ssim \quad (5)$$

The high weight for content loss ensures that the generator produces high-quality images close to the original in pixel space, while the adversarial and SSIM losses help in maintaining realism and structural similarity.

Adam optimizer is used for both the generator and discriminator. The model is trained for 50 epochs on A100 inside Google Colab. This was necessary because during the initial testing phase, which was conducted with the same number of images on a local system, one training epoch ran for eight hours. Other model details can be found in Table 4.2.

Table 4.2 GAN model training details

Parameter	**Value**
Learning rate	0.0002
Batch size	16
Optimizer	Adam (betas = (0.5, 0.999))
Generator Loss	Adversarial + MSE (Content) + SSIM
Discriminator Loss	Binary Cross Entropy (BCE)
Maximum epoch	50

Each type of these described models was trained on all four previously described datasets, thus creating eight different models.

4.5. Evaluation Metrics

In order to evaluate the trained models, evaluation metrics of PSNR and SSIM were selected and calculated on the test dataset. Looking at the literature review we can clearly see that these two metrics are most commonly used to measure performance of facial image deblurring models.

5. Results and Discussion

Each one of eight previously mentioned models was tested on the test set which was created along their train sets. After that, the models were also tested on other test sets to see the performance of models on facial deblurring tasks for which they have not been specifically trained on.

5.1. Results

Results of PSNR and SSIM evaluation metrics will be shown in the tables and comparison images will also be shown because the objective evaluation metrics are still not the best guarantee that the characteristics of a human face are faithfully restored in the deblurred image.

Results on the first dataset pair where DeepLabV3 model was used for face segmentation in the second dataset are presented in Table 5.1.

Table 5.1 First dataset pair where DeepLabV3 model was used for face segmentation

Model name	No segmentation		DeepLabV3 segmentation	
	PSNR	SSIM	PSNR	SSIM
Autoencoder	26.2243	0.7491	25.3114	0.7225
GAN	23.0206	0.8117	24.2852	0.8635

Below are resulting images of autoencoder and GAN model on the first dataset pair.

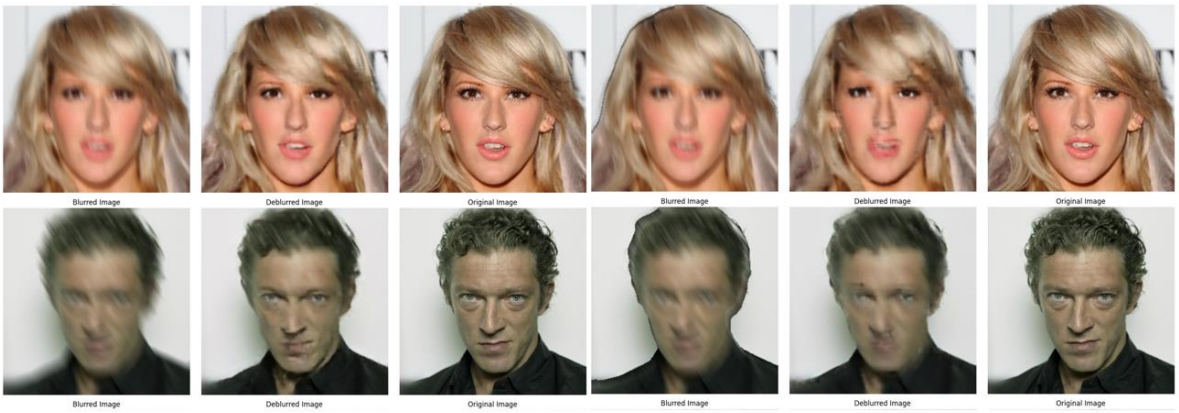

Figure 5.1 Autoencoder model, no segmentation on the left, DeepLabV3 segmentation on the right

Figure 5.2 GAN model, no segmentation on the left, DeepLabV3 segmentation on the right

Results on the second dataset pair where MediaPipe was used for face segmentation in the second dataset are presented in Table 5.2.

Table 5.2 Second dataset pair where MediaPipe was used for face segmentation

Model name	No segmentation		MediaPipe segmentation	
	PSNR	SSIM	PSNR	SSIM
Autoencoder	26.2784	0.7513	29.3614	0.8834
GAN	27.1409	0.8860	28.5285	0.9396

Below are resulting images of autoencoder and GAN model on the second dataset pair.

Figure 5.3 Autoencoder model, no segmentation on the left, MediaPipe segmentation on the right

Figure 5.4 GAN model, no segmentation on the left, MediaPipe segmentation on the right

In order to further judge the performance and robustness of these trained models, we have decided to test them on datasets which they have not been trained on. The final result can be seen in Table 5.4. The letter A or G denotes whether the model is autoencoder or GAN and the following letters differentiate the dataset which was used for the model training. Explanations of dataset names can be found in Table 5.3.

Table 5.3 Explanation of dataset names

SFB1	Sharp + Fully blurred images blurred by random blur 1
SB1SD	Sharp + random blur 1 applied on head area segmented with DeepLabV3
SFB2	Sharp + Fully blurred images blurred by random blur 2
SB2SM	Sharp + random blur 1 applied on face area segmented with MediaPipe

Table 5.4 Results for all models on all datasets. Best results for that dataset are in bold

Autoencoder (A) GAN (G)	SFB1		SB1SD		SFB2		SB2SM	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ASFB1	26.223	0.7491	24.4925	0.7269	26.2395	0.7495	26.4802	0.8024
ASB1SD	24.6737	0.6945	**25.3114**	0.7225	24.7044	0.6954	27.1715	0.8313
ASFB2	26.252	0.7507	24.3855	0.7255	26.2784	0.7513	25.9232	0.7817
ASB2SM	25.7062	0.7367	24.7136	0.7404	25.7311	0.7372	**29.3614**	0.8834
GSFB1	23.0206	0.8117	21.0733	0.7870	23.0092	0.8115	20.7365	0.7708
GSB1SD	23.8784	0.8544	24.2852	0.8635	23.9055	0.8546	25.0381	0.8822
GSFB2	**27.1548**	**0.8860**	23.8647	0.8503	**27.1409**	**0.8860**	23.0679	0.8312
GSB2SM	25.4178	0.8703	24.2205	**0.8675**	25.4670	0.8714	28.5285	**0.9396**

Bellow we can see results of some different models on different test sets.

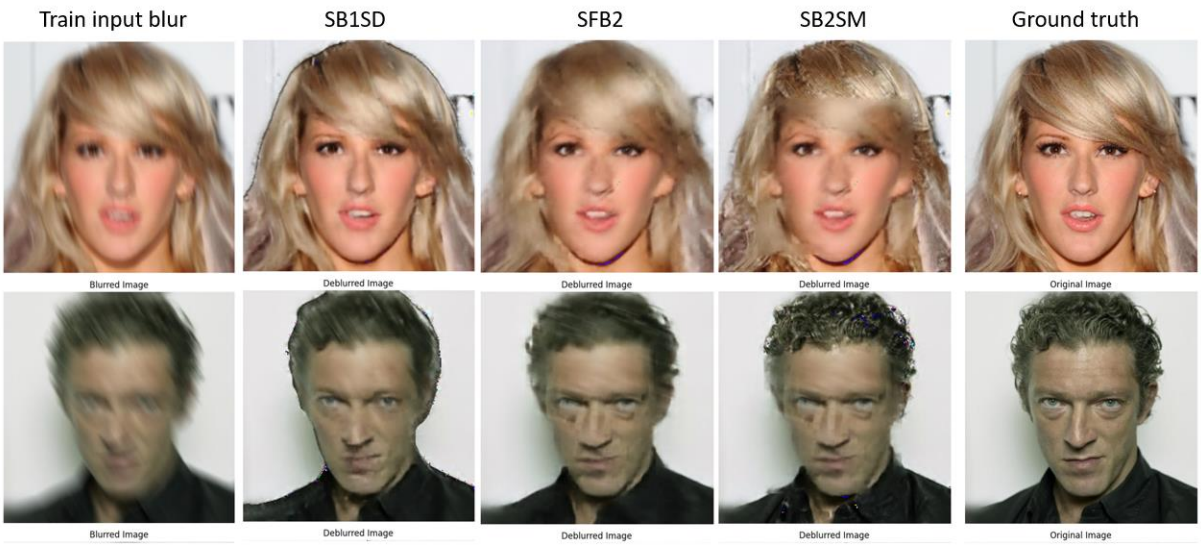

Figure 5.5 ASFB1 model deblurring visual performance on other test datasets

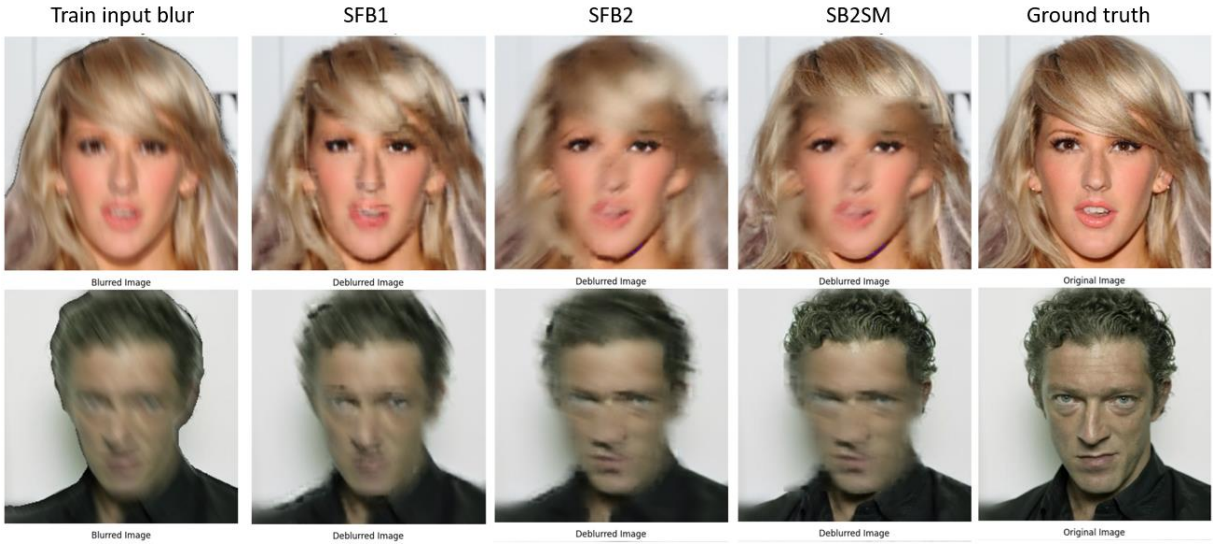

Figure 5.6 ASB1SD model deblurring visual performance on other test datasets

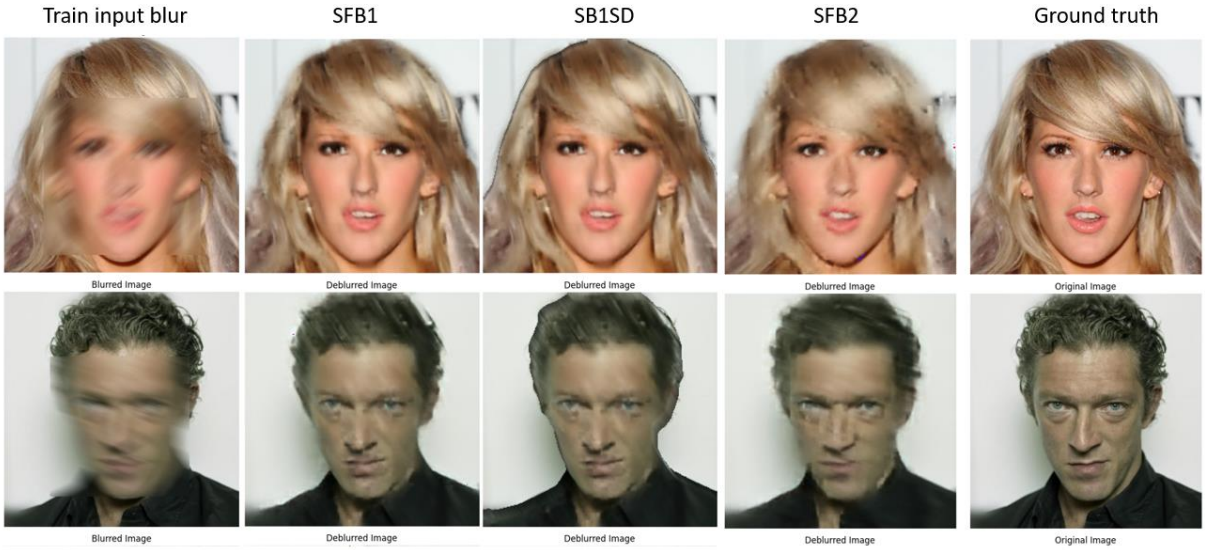

Figure 5.7 ASB2SM model deblurring visual performance on other test datasets

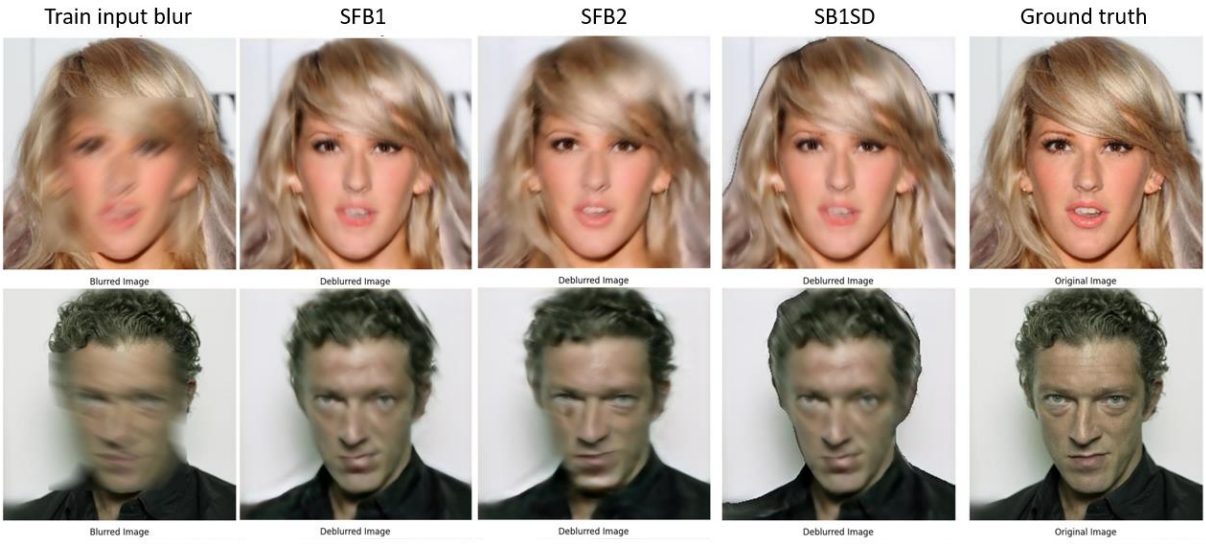

Figure 5.8 GSB2SM model deblurring visual performance on other test datasets

Looking at the Table 5.4 we can see that the GSFB2 model achieves good metrics on two datasets in which the images are completely blurred without segmentation. To further examine visual performance of that model, we will show more images for easier comparison.

Figure 5.9 GSFB2 model on SFB2 test dataset

Figure 5.10 GSFB2 model on SFB1 test dataset

Figure 5.11 GSFB2 model on SB1SD test dataset

Figure 5.12 GSFB2 model on SB2SM test dataset

5.2. Discussion

Looking at the final result in the Table 5.4 we can see that most models achieve best results on test set of image which has the same blur type as the images they were trained on. One exception would be GSFB2 model. Judging by the final table where models have been tested on various databases we can see that GSFB2 model achieves the best results on fully blurred images. Because of that we further tested that model. We tested it on all datasets after 25 epochs instead of 50 to see what difference in results does 25 more epochs make. The results can be seen in Table 5.5. We see that training the model for 50 epochs actually decreases the metrics on all databases and we can assume that the model is overfitted. The initial idea was to train the GSFB2 on SFB2 dataset for 100 epochs and compare that to the same model but trained on an even split of 4 x 25 epochs on all train datasets. Since the model was already overfitted on 50 epochs we gave up on the first half of the test. We still decided on testing the second idea and we trained GSFB2 model first for 25 epochs on SFB2 dataset then for 25 more epochs on SB1SD dataset, then for 25 more on SB2SM dataset and for the final 25 epochs we trained it on SFB1 dataset. The training was conducted in Google Colab with the same parameters previously described for the GAN model. The final result of evaluation metrics can be seen in the table 5.5.

Table 5.5 Further testing of GSFB2 model

Models	SFB1		SB1SD		SFB2		SB2SM	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
GSFB2_50	27.1548	0.8860	23.8647	0.8503	27.1409	0.8860	23.0679	0.8312
GSFB2_25	27.2817	0.8872	24.3905	0.8615	27.3047	0.8873	24.7057	0.8623
GSFB2_100	24.6067	0.7991	22.8405	0.7902	24.6562	0.7998	22.2620	0.7750
GSFB2_20	25.4839	0.8256	25.4993	0.8266	23.8199	0.8144	24.8443	0.8401

Figure 5.13 Comparison of GSFB2 model trained for 25 and 50 epochs

Judging by the Figure 5.13, we can see that even if the evaluation metrics are higher for the model version which is trained for 25 epochs, the visual performances and clarity are lacking, especially in the eye area which is often critical for GANs. The final skin textures are also over smoothed for both model. Final deblurred faces lack pores and sharpness.

Looking at the Table 5.5 we can see that the performances of the model trained for 100 epochs on four different datasets are worse on every test set by a large margin. Even the difference in blur type in those four datasets did not train a better model. Because of those poor results we decided to train another GAN model on the same four datasets in the same training order, just for 20 epochs, 5 epochs on each of four datasets. The test results of that model are in the final row of Table 5.5. We can see that they are worse than the results of the model trained on just one dataset, but they are still better than the results of the model trained on four different datasets for 100 epochs.

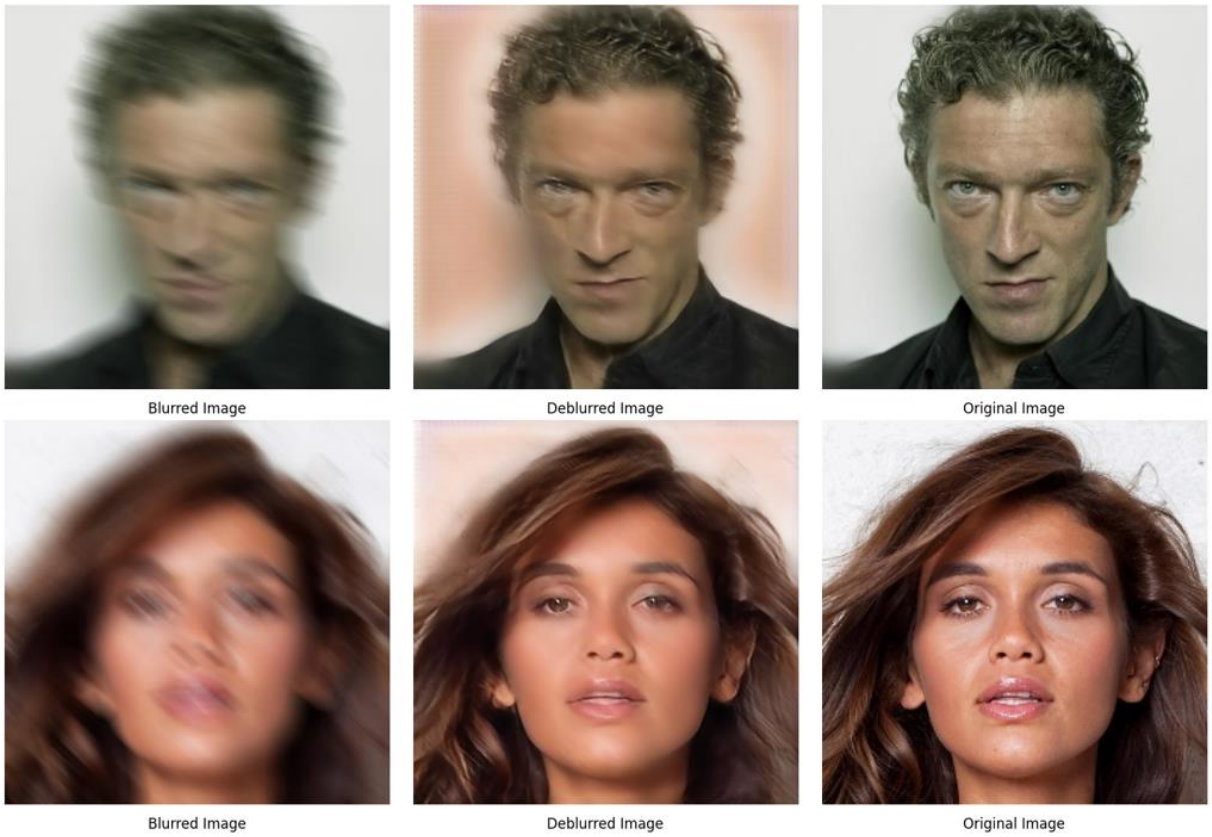

Figure 5.14 GSFB2_100 model on SFB2 test dataset

Figure 5.15 GSFB2_100 model on SFB1 test dataset

Figure 5.16 GSFB2_100 model on SB2SM test dataset

Even though the performance of GSFB2_100 may seem fine visually on Figures 5.14 and 5.15, we can clearly see artifacts in the Figure 5.16 which are degrading the clear part of the image. The deblurring of just the blurred face area with the preservation of clearness and sharpness in the non-blurred area is not achieved by GSFB2_100 model.

Figure 5.17 GSFB2_20 model on SFB2 test dataset

Figure 5.18 GSFB2_20 model on SB2SM test dataset

With the GSFB2_20 model we can see that evaluation numbers are better, but in the deblurred images we can see that there are less artifacts, but also that the face area deblurring is worse than in GSFB2_100 model.

6. Conclusion and Future Work

6.1. Conclusion

By conducting a thorough literature review we have made a good foundation for better understanding the complexity of facial motion deblurring problem. In mentioning previous attempts of solving this problem with the focus on their approaches and modifications we have shown diversity of possible approaches and in which aspects they are similar. We have explained the decision making process in the conduct of the experiment which is based on the collected literature and previous attempts. After looking at the evaluation metrics of models in our experiment and having in mind visual perception of the results in deblurred facial images we can say that the models specifically trained on segmented and then blurred datasets performed worse than the models trained on non-segmented, fully blurred images. Further testing revealed that using a combination of training datasets resulted in poorer performance, leaving more artifacts in the output. This experiment highlights the critical importance of the training dataset and the specific methods used in its creation for optimizing the performance of facial motion deblurring models.

6.2. Future Work

The performance of facial motion deblurring models depends on many things, but the main thing it depends on is the training dataset and the type of motion blur that is simulated in it. For future work the focus should be put on more realistic blur modeling. Generative adversarial network could be trained to artificially apply realistic motion blur during the dataset creation. Issue with that idea comes again from the fact that even that model needs to have a good representation of realistic motion blur in its dataset in order to learn its modeling. There are various ways of modeling motion blur described in the literature and in order to see which type performs the best, a large ablation study should be conducted. By training this blurring generative adversarial network on different kinds of blur and maybe even on a combination of them, realistic and diverse motion blur

modeling could be achieved which could in turn lead to creation of robust models capable of handling diverse types of motion blur which degrade facial images.

Literature

[1] Dutta, A., Veldhuis, R., & Spreeuwers, L. (2012, May). The impact of image quality on the performance of face recognition. In 33rd WIC Symposium on Information Theory in the Benelux and the 2nd Joint WIC/IEEE Symposium on Information Theory and Signal Processing in the Benelux 2012 (pp. 141-148). Werkgemeenschap voor Informatie-en Communicatietheorie (WIC).

[2] Nahli, A., Cao, Y., & Xu, S. (2020, April). Face Image Deblurring: A Data-Driven Strategy. In GCAI (pp. 59-69).

[3] Wang, R., Zhang, C., Zheng, X., Lv, Y., & Zhao, Y. (2023). Joint Defocus Deblurring and Superresolution Learning Network for Autonomous Driving. IEEE Intelligent Transportation Systems Magazine.

[4] Park, K., Shin, S., Jeon, H. G., Lee, J. Y., & Kweon, I. S. (2014, November). Motion deblurring using coded exposure for a wheeled mobile robot. In 2014 11th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI) (pp. 665-671). IEEE.

[5] Joshi, A. R., Cuadros, X. S., Sivakumar, N., Zappella, L., & Apostoloff, N. (2022, March). Fair SA: Sensitivity analysis for fairness in face recognition. In Algorithmic fairness through the lens of causality and robustness workshop (pp. 40-58). PMLR.

[6] Wei, H., Ge, C., Qiao, X., & Deng, P. (2022). Rethinking blur synthesis for deep real-world image deblurring. arXiv preprint arXiv:2209.13866.

[7] Cho, H., Jeong, Y., Kim, T., & Yoon, K. J. (2023). Non-coaxial event-guided motion deblurring with spatial alignment. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 12492-12503).

[8] Madnani, M. (2023, December). Real-Time Emotion Recognition System for Dynamic Real-World Environments. In 2023 16th International Conference on Developments in eSystems Engineering (DeSE) (pp. 731-736). IEEE.

[9] Tian, Z., Weng, D., Fang, H., & Bao, Y. (2023, October). Deep Detector and Optical Flow-based Tracking Approach of Facial Markers for Animation Capture. In 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) (pp. 625-630). IEEE.

[10] Fan, S., & Luo, Y. (2021, February). Deblurring processor for motion-blurred faces based on generative adversarial networks. In Proceedings of the 2021 5th International Conference on Digital Signal Processing (pp. 272-277).

[11] Zhai, L., Wang, Y., Cui, S., & Zhou, Y. (2023). A comprehensive review of deep learning-based real-world image restoration. IEEE Access, 11, 21049-21067.

[12] Li, P., Prieto, L., Mery, D., & Flynn, P. (2018). Face recognition in low quality images: A survey. arXiv preprint arXiv:1805.11519.

[13] Jiang, J., Wang, C., Liu, X., & Ma, J. (2021). Deep learning-based face super-resolution: A survey. ACM Computing Surveys (CSUR), 55(1), 1-36.

[14] Chen, R., Guo, T., Mu, Y., & Shen, L. (2024). Learning Compact Hyperbolic Representations of Latent Space for Old Photo Restoration. IEEE Transactions on Image Processing.

[15] Xue, H., Liu, B., Yuan, X., Ding, M., & Zhu, T. (2023). Face image de-identification by feature space adversarial perturbation. Concurrency and Computation: Practice and Experience, 35(5), e7554.

[16] Goswami, G., Agarwal, A., Ratha, N., Singh, R., & Vatsa, M. (2019). Detecting and mitigating adversarial perturbations for robust face recognition. International Journal of Computer Vision, 127, 719-742.

[17] Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*, T. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine, 151(4), 264-269.

[18] Tomaszewski, R. (2021). A study of citations to STEM databases: ACM Digital Library, Engineering Village, IEEE Xplore, and MathSciNet. Scientometrics, 126(2), 1797-1811.

[19] Zhao, P., Zhang, X., Cheng, M. M., Yang, J., & Li, X. (2024). A Literature Review of Literature Reviews in Pattern Analysis and Machine Intelligence. arXiv preprint arXiv:2402.12928.

[20] Valente, A., Holanda, M., Mariano, A. M., Furuta, R., & Da Silva, D. (2022, October). Analysis of academic databases for literature review in the computer science education field. In 2022 ieee frontiers in education conference (fie) (pp. 1-7). IEEE.

[21] Wang, B., Xu, F., & Zheng, Q. (2024). A survey on facial image deblurring. Computational Visual Media, 10(1), 3-25.

[22] Biyouki, S. A., & Hwangbo, H. (2023). A comprehensive survey on deep neural image deblurring. arXiv preprint arXiv:2310.04719.

[23] Wang, T., Zhang, K., Chen, X., Luo, W., Deng, J., Lu, T., ... & Zafeiriou, S. (2022). A survey of deep face restoration: Denoise, super-resolution, deblur, artifact removal. arXiv preprint arXiv:2211.02831.

[24] Su, J., Xu, B., & Yin, H. (2022). A survey of deep learning approaches to image restoration. Neurocomputing, 487, 46-65.

[25] Mahalakshmi, A., & Shanthini, B. (2016, January). A survey on image deblurring. In 2016 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-5). IEEE.

[26] Zheng, H. (2021, January). A Survey on Single Image Deblurring. In 2021 2nd International Conference on Computing and Data Science (CDS) (pp. 448-452). IEEE.

[27] Xiang, Y., Zhou, H., Li, C., Sun, F., Li, Z., & Xie, Y. (2024). Application of Deep Learning in Blind Motion Deblurring: Current Status and Future Prospects. arXiv preprint arXiv:2401.05055.

[28] Zhang, K., Ren, W., Luo, W., Lai, W. S., Stenger, B., Yang, M. H., & Li, H. (2022). Deep image deblurring: A survey. International Journal of Computer Vision, 130(9), 2103-2130.

[29] Ranjan, A., & Ravinder, M. (2022, December). Deep Learning based Image Deblurring: A Comparative Survey. In 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 996-1002). IEEE.

[30] Xiao, K., & Pan, Z. (2023, December). A facial motion deblurring algorithm based on semantics and GAN. In 2023 3rd International Conference on Information Technology and Contemporary Sports (TCS) (pp. 26-29). IEEE.
[31] Gao, Y., Hou, Q., Yan, C., Zhou, W., Xie, F., Wang, M., & Li, J. (2023, December). Blind deblurring of single image based on kernel estimation of texture image. In 2023 9th International Conference on Computer and Communications (ICCC) (pp. 1891-1895). IEEE.
[32] Zhang, P., Zhang, K., Luo, W., Li, C., & Wang, G. (2024). Blind face restoration: Benchmark datasets and a baseline model. Neurocomputing, 574, 127271.
[33] Yang, C., Li, Q., Li, C., & Zheng, Y. (2023, July). Blind Image Deblurring with Extreme Gradient and Dark Channel Priors. In 2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS) (pp. 577-584). IEEE.
[34] Yu, X., & Xie, W. (2022). Blurry Facial-Image Deconvolution via Model-Guided Deep Neural Network Inspired From Edge Regularization. IEEE Transactions on Cognitive and Developmental Systems, 15(1), 285-297.
[35] Ahalya, C., Boya, R., Munisetty, H., Damam, C., & Telugu, R. (2024, March). Deep Learning for Single Image Deblurring. In 2024 3rd International Conference for Innovation in Technology (INOCON) (pp. 1-6). IEEE.
[36] Shi, C., Zhang, X., Li, X., Mumtaz, I., & Lv, J. (2024). Face deblurring based on regularized structure and enhanced texture information. Complex & Intelligent Systems, 10(2), 1769-1786.
[37] Cui, Y., Tang, C., & Huang, Q. (2023, November). Joint face super-resolution and deblurring using multi-task feature fusion network. In 7th International Conference on Vision, Image and Signal Processing (ICVISP 2023) (Vol. 2023, pp. 57-61). IET.
[38] Han, S., Lee, T. B., & Heo, Y. S. (2023). Semantic-Aware Face Deblurring with Pixel-Wise Projection Discriminator. IEEE Access, 11, 11587-11600.
[39] Wang, P., Zhu, Y., Xue, D., Yan, Q., Sun, J., Yoon, S. E., & Zhang, Y. (2024). Take a prior from other tasks for severe blur removal. Computer Vision and Image Understanding, 245, 104027.
[40] Kim, D., Kim, M., Kwon, G., & Kim, D. S. (2019). Progressive face super-resolution via attention to facial landmark. arXiv preprint arXiv:1908.08239.
[41] Chen, Y., Tai, Y., Liu, X., Shen, C., & Yang, J. (2018). Fsrnet: End-to-end learning face super-resolution with facial priors. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2492-2501).
[42] Ioffe, S. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.
[43] Zhu, S., Liu, S., Loy, C. C., & Tang, X. (2016). Deep cascaded bi-network for face hallucination. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14 (pp. 614-630). Springer International Publishing.
[44] Wang, H., Teng, Z., Wu, C., & Coleman, S. (2022, July). Facial Landmarks and Generative Priors Guided Blind Face Restoration. In 2022 IEEE 20th International Conference on Industrial Informatics (INDIN) (pp. 101-106). IEEE.
[45] Chen, C., Li, X., Yang, L., Lin, X., Zhang, L., & Wong, K. Y. K. (2021). Progressive semantic-aware style transformation for blind face restoration. In

Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11896-11905).
[46] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision (pp. 3730-3738).
[47] Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4401-4410).
[48] Le, V., Brandt, J., Lin, Z., Bourdev, L., & Huang, T. S. (2012). Interactive facial feature localization. In Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III 12 (pp. 679-692). Springer Berlin Heidelberg.
[49] Yu, X., Fernando, B., Ghanem, B., Porikli, F., & Hartley, R. (2018). Face super-resolution guided by facial component heatmaps. In Proceedings of the European conference on computer vision (ECCV) (pp. 217-233).
[50] Yu, X., & Porikli, F. (2017). Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3760-3768).
[51] Hu, X., Ren, W., LaMaster, J., Cao, X., Li, X., Li, Z., ... & Liu, W. (2020). Face super-resolution guided by 3d facial priors. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16 (pp. 763-780). Springer International Publishing.
[52] Hu, X., Ren, W., Yang, J., Cao, X., Wipf, D., Menze, B., ... & Zha, H. (2021). Face restoration via plug-and-play 3D facial priors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12), 8910-8926.
[53] Li, X., Chen, C., Zhou, S., Lin, X., Zuo, W., & Zhang, L. (2020, August). Blind face restoration via deep multi-scale component dictionaries. In European conference on computer vision (pp. 399-415). Cham: Springer International Publishing.
[54] Li, X., Liu, M., Ye, Y., Zuo, W., Lin, L., & Yang, R. (2018). Learning warped guidance for blind face restoration. In Proceedings of the European conference on computer vision (ECCV) (pp. 272-289).
[55] Wang, X., Li, Y., Zhang, H., & Shan, Y. (2021). Towards real-world blind face restoration with generative facial prior. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9168-9178).
[56] Sun, Y., Wang, X., & Tang, X. (2013). Hybrid deep learning for face verification. In Proceedings of the IEEE international conference on computer vision (pp. 1489-1496).
[57] Lee, T. B., Jung, S. H., & Heo, Y. S. (2020). Progressive semantic face deblurring. IEEE Access, 8, 223548-223561.
[58] Lee, C. H., Liu, Z., Wu, L., & Luo, P. (2020). Maskgan: Towards diverse and interactive facial image manipulation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5549-5558).
[59] Shen, Z., Lai, W. S., Xu, T., Kautz, J., & Yang, M. H. (2018). Deep semantic face deblurring. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8260-8269).

[60] Sim, T. (2003). BS, and M. Bsat,"The cmu pose, illumination, and expression database,". IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(12), 1615-1618.
[61] Jin, M., Hirsch, M., & Favaro, P. (2018). Learning face deblurring fast and wide. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 745-753).
[62] Ng, H. W., & Winkler, S. (2014, October). A data-driven approach to cleaning large face datasets. In 2014 IEEE international conference on image processing (ICIP) (pp. 343-347). IEEE.
[63] Lin, S., Zhang, J., Pan, J., Liu, Y., Wang, Y., Chen, J., & Ren, J. (2020, April). Learning to deblur face images via sketch synthesis. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11523-11530).
[64] Kumar, N., Berg, A. C., Belhumeur, P. N., & Nayar, S. K. (2009, September). Attribute and simile classifiers for face verification. In 2009 IEEE 12th international conference on computer vision (pp. 365-372). IEEE.
[65] Lai, W. S., Huang, J. B., Hu, Z., Ahuja, N., & Yang, M. H. (2016). A comparative study for single image blind deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1701-1709).
[66] Chrysos, G. G., & Zafeiriou, S. (2017). Deep face deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 69-78).
[67] Ren, W., Zhang, J., Pan, J., Liu, S., Ren, J. S., Du, J., ... & Yang, M. H. (2021). Deblurring dynamic scenes via spatially varying recurrent neural networks. IEEE transactions on pattern analysis and machine intelligence, 44(8), 3974-3987.
[68] Jiang, W., & Liu, A. (2022). Image motion deblurring based on deep residual shrinkage and generative adversarial networks. Computational Intelligence and Neuroscience, 2022(1), 5605846.
[69] Qian, P., Wu, Y., & Zhang, X. (2021, March). Dense connected residual generative adversarial network for single image deblurring. In 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (Vol. 5, pp. 461-466). IEEE.
[70] Lee, H. S., & Cho, S. I. (2023). Locally adaptive channel attention-based spatial–spectral neural network for image deblurring. IEEE Transactions on Circuits and Systems for Video Technology, 33(10), 5375-5390.
[71] Zhao, Y., Cui, H., Zhao, B., & Ma, J. (2021, July). Edge prior and spatial attention fusion enhanced hierarchical multi-patch network for image deblurring. In 2021 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
[72] Chen, B., Jiang, L., Wu, P., Zheng, F., Li, K., Chen, T., & Li, R. (2022, October). A feature attention based multi-stage network for image deblurring. In 5th International Conference on Computer Information Science and Application Technology (CISAT 2022) (Vol. 12451, pp. 432-437). SPIE.
[73] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. IEEE Transactions on Computational Imaging, 3(1):47–57, 2016.
[74] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision, pages 694–711. Springer, 2016.

[75] Song, Y. B.; Zhang, J. W.; Gong, L. J.; He, S. F.; Bao, L. C.; Pan, J. S.; Yang, Q. X.; Yang, M. H. Joint face hallucination and deblurring via structure generation and detail enhancement. International Journal of Computer Vision Vol. 127, Nos. 6–7, 785–800, 2019.
[76] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. In Proceedings of the British Machine Vision Conference (BMVC)
[77] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.
[78] Jolicoeur-Martineau, A. (2018). The relativistic discriminator: a key element missing from standard GAN. arXiv preprint arXiv:1807.00734.
[79] Gong, D., Yang, J., Liu, L., Zhang, Y., Reid, I., Shen, C., ... & Shi, Q. (2017). From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2319-2328).
[80] Wang, X., Li, Y., Zhang, H., & Shan, Y. (2021). Towards real-world blind face restoration with generative facial prior. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9168-9178).
[81] Lu, B., Chen, J. C., & Chellappa, R. (2019). UID-GAN: Unsupervised image deblurring via disentangled representations. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2(1), 26-39.
[82] Otto, C., Chandran, P., Zoss, G., Gross, M., Gotardo, P., & Bradley, D. (2023, October). A perceptual shape loss for monocular 3D face reconstruction. In Computer Graphics Forum (Vol. 42, No. 7, p. e14945).
[83] Yu, X., Fernando, B., Ghanem, B., Porikli, F., & Hartley, R. (2018). Face super-resolution guided by facial component heatmaps. In Proceedings of the European conference on computer vision (ECCV) (pp. 217-233).
[84] Wang, Z.; Bovik, A. C.; Sheikh, H. R.; Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing Vol. 13, No. 4, 600–612, 2004.
[85] Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 586–595, 2018.
[86] Mittal, A., Soundararajan, R., & Bovik, A. C. (2012). Making a “completely blind” image quality analyzer. IEEE Signal processing letters, 20(3), 209-212.
[87] Ding, K., Ma, K., Wang, S., & Simoncelli, E. P. (2021). Comparison of full-reference image quality models for optimization of image processing systems. International Journal of Computer Vision, 129(4), 1258-1281.
[88] Borji, A. (2019). Pros and cons of GAN evaluation measures. Computer vision and image understanding, 179, 41-65.
[89] Borji, A. (2022). Pros and cons of GAN evaluation measures: New developments. Computer Vision and Image Understanding, 215, 103329.
[90] Hore, A., & Ziou, D. (2010, August). Image quality metrics: PSNR vs. SSIM. In 2010 20th international conference on pattern recognition (pp. 2366-2369). IEEE.
[91] Wang, Z., Simoncelli, E. P., & Bovik, A. C. (2003, November). Multiscale structural similarity for image quality assessment. In The Thrity-Seventh

Asilomar Conference on Signals, Systems & Computers, 2003 (Vol. 2, pp. 1398-1402). Ieee.
[92] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems, 30.
[93] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans. Advances in neural information processing systems, 29.
[94] Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain. IEEE Transactions on image processing, 21(12), 4695-4708.
[95] Sheikh, H. R., & Bovik, A. C. (2006). Image information and visual quality. IEEE Transactions on image processing, 15(2), 430-444.
[96] Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., & Zelnik-Manor, L. (2018). The 2018 PIRM challenge on perceptual image super-resolution. In Proceedings of the European conference on computer vision (ECCV) workshops (pp. 0-0).
[97] Xue, W., Zhang, L., Mou, X., & Bovik, A. C. (2013). Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE transactions on image processing, 23(2), 684-695.
[98] Z. Hu, L. Xu, M.-H. Yang, Joint depth estimation and camera shake removal from single blurry image, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2014, pp. 2893–2900.
[99] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, B. Curless, Single image deblurring using motion density functions, in: Proc. IEEE Eur. Conf. Comput. Vis. (ECCV), Springer, 2010, pp. 171–184.
[100] Khan, K., Khan, R. U., Ahmad, K., Ali, F., & Kwak, K. S. (2020). Face segmentation: A journey from classical to deep learning paradigm, approaches, trends, and directions. IEEE Access, 8, 58683-58699.
[101] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence, 44(7), 3523-3542.
[102] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.
[103] Chen, L. C. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.
[104] Nah, S., Hyun Kim, T., & Mu Lee, K. (2017). Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3883-3891).
[105] Shen, Z., Wang, W., Lu, X., Shen, J., Ling, H., Xu, T., & Shao, L. (2019). Human-aware motion deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 5572-5581).
[106] Boracchi, G., & Foi, A. (2012). Modeling the performance of image restoration from motion blur. IEEE Transactions on Image Processing, 21(8), 3502-3517.
[107] Chakrabarti, A. (2016). A neural approach to blind motion deblurring. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14 (pp. 221-235). Springer International Publishing.

[108] Sun, J., Cao, W., Xu, Z., & Ponce, J. (2015). Learning a convolutional neural network for non-uniform motion blur removal. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 769-777).
[109] Li, Z., Gao, Z., Yi, H., Fu, Y., & Chen, B. (2023). Image deblurring with image blurring. IEEE Transactions on Image Processing.
[110] Rim, J., Kim, G., Kim, J., Lee, J., Lee, S., & Cho, S. (2022, October). Realistic blur synthesis for learning image deblurring. In European conference on computer vision (pp. 487-503). Cham: Springer Nature Switzerland.
[111] Bahat, Y., Efrat, N., & Irani, M. (2017). Non-uniform blind deblurring by reblurring. In Proceedings of the IEEE international conference on computer vision (pp. 3286-3294).
[112] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.
[113] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022).
[114] Z. Shen, W.-S. Lai, T. Xu, M.-H. Yang, and G. Cloud, “Deep semantic face deblurring,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 8260–8269.
[115] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE signal processing letters, 23(10), 1499-1503.
[116] Kingma, D. P. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
[117] Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. Information and software technology, 51(1), 7-15.
[118] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 32.

Figure Index

Table Index