School

of

Electronics and Communication Engineering

Minor Project Report

on

Self-supervised video De-blurring

By:

1. **Aditya krishna vamsy Mudragada** USN:01fe18bec006
2. **Kuljeet singh** USN:01fe18bcs100

Semester: VI, 2020-2021

Under the Guidance of

Ramesh Tabib, Uma Mudenagudi and Sandeep

K.L.E SOCIETY'S
KLE Technological University,
HUBBALLI-580031
2020-2021

SCHOOL OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that project entitled " **Self-supervised video De-blurring** " is a bonafide work carried out by the student team of "**Aditya krishna vamsy Mudragada(01fe18bec006) , Kuljeet singh(01fe18bcs100)** ". The project report has been approved as it satisfies the requirements with respect to the minor project work prescribed by the university curriculum for BE (VI Semester) in School of Electronics and Communication Engineering of KLE Technological University for the academic year 2020-2021.

Ramesh Tabib	**Nalini C. Iyer**	**N. H. Ayachit**
Guide	**Head of School**	**Registrar**

External Viva:

Name of Examiners **Signature with date**

1.

2.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Ramesh Tabib , Uma mudenagudi and Sandeep for assisting us in the completion of this project, without whose help it could not have been possible.All their contributions are deeply appreciated and acknowledged.It was a great privilege and honour to work and study under their guidance. We are highly indebted to Dr. Ashok S Shettar , Vice Chancellor of KLE Technological University, Hubli , Head of School of Electronics and communication department for giving us an opportunity to implement a solution for such a noble cause.

- Aditya krishna vamsy Mudragada(01fe18bec006)
Kuljeet singh(01fe18bcs100)

ABSTRACT

Motion blurry that occurs in images challenge many computer vision algorithms, for example feature detection, motion estimation, or object recognition and many more .To solve this problem existing techniques can be classified into two categories as in the first approach the problem is is considered as optimization problem the blur kernel or the latent sharp image will be optimized with the help of gradient descent an advantage of this approach is that there is no need to ground truth images but this approach take more time and the complexity increases due to this drawback its use is limited and not useful when situation is time constrained, the second approach considers the problem to be learning problem where convolution neural network are considered as state-of-the-art for image deblurring but the problem is that obtaining training data having both blurr and corresponding sharp image is hard to collect so to overcome this challenge self-supervised approach is used where it helps the network to learn from real-world blurry image sequences that yield sufficient information to inform the deblurring task.

Contents

List of Tables

List of Figures

Chapter 1

Introduction to self-supervised video de-blurring

Motion is blur is a major factor that degrades the image quality .It arises due to rapid change in image content hence necessitating longer exposure times. The existing techniques to address the challenge are divided into 2 categories: first approach is optimization based approach which suffer from high complexity of computation which limits their application in time-restricted environment. Second approach considers the deblurring task as learning problem with the help of convolution neural networks(CNN) state-of-the-art results are obtained but obtaining both blur and sharp image pairs are not easy as not every camera has high fps that can be used to train the data and another reason is the obtaining quality in real scenarios is difficult where the real motion blur occurs , so a novel self-supervised approach is needed which only relies on real-world blurry image frame sequences for training which helps in improving the networks generalization performance.

1.1 Motivation for the study

1. Motion blur images is a major factor that degrades the image quality as image does not represent a single instant of time during long exposure times, the moving objects will tend to look sharper while the background will become more blurred a solution to this problem will surely help.

2. Motion blur caused due to due to rapid movement or long exposure challenges many computer vision algorithms like depth prediction,feature detection, motion estimation and object recognition.

3. A proper approach to solve the problem will allow in improving the performance of many computer vision algorithms.

1.2 Objectives of the problem statement

1.Deblur input blurry videos using a self-supervised approach for deblurring the videos that don't require and ground truth images.

2. Train the model proposed in base paper in a self-supervised fashion as collecting real-world blurry and sharp is hard.

3.Test model on blurry videos after training the model and using the saved model to test upon the input blur images.

1.3 Literature survey related to self-supervised video deblurring

As per the base paper we have referred we have the following method useful to perform self-supervised video deblurring. Paper[1] Self-Supervised Linear Motion Deblurring Motion blur is challenging many computer vision algorithms.Deep convolution neural networks are perform the deblurring task very well but obtaining real blurry and sharp image pairs are not easy in the paper a self-supervised approach is proposed for deblurring linear motion blur which does not depend the ground truth sharp images .The netowrk will be able to learn from real-world consecutive blur image sequences which is sufficient to inform deblurring task.

The method proposed has four main components Deblur ,flownet, reblur block and image warping using deblur net we can perform single image deblurring where the network accepts single image as input and gives corresponding sharp image using 2 of these networks and taking consecutive blur images as input and it is given as input to the flownet to estimate a bi-directional optical flow which is used to calculate a spatially blur kernel for the blur images using this kernel we reblur the image and compare the input image with the reblurred image using L1 loss function and the sharpe estimate is compared with warped deblurred image using L1 loss to constraint optical flow.

$$\mathcal{L}_{\text{self}} = \|\mathbf{B}'_a - \mathbf{B}_a\|_1 + \|\mathbf{B}'_b - \mathbf{B}_b\|_1$$

Ba' and Bb' is the reblurred image Ba and Bb is the input blur image

$$\mathcal{L}_{\text{fw/bw}} = \|\mathbf{I}'_a - \mathbf{I}_a\|_1 + \|\mathbf{I}'_b - \mathbf{I}_b\|_1$$

Ia' and Ib' are warped deblurred images and Ia and Ib are sharp estimate of the deblur net

$$\mathcal{L} = \mathcal{L}_{\text{self}} + \lambda \mathcal{L}_{\text{fw/bw}},$$

L is the weighted sum of both the loss function where lambda is set to 0.2 empirically. The results of this approach are competitive with the supervised approach even though there is no usage of ground truth images.

Single Image Deblurring on Synthetic Dataset. Supervised-Retrained Denotes We Retrained the Networks With Our Training Data. Supervised-Pretrained Denotes We Use the Official Pretrained Models to Evaluate on Our Test Data Directly

	Method	PSNR↑	SSIM↑	Time ↓
Opt.-based	Xu et al. [39]	26.04 dB	0.78	377.8 s
Supervised -retrained	DeepDeblur [21]	33.55 dB	0.92	3.45 s
	DeblurGAN [17]	33.23 dB	0.91	0.06 s
	SRN-Deblur [32]	34.64 dB	0.93	0.13 s
Supervised -pretrained	DeepDeblur [21]	29.91 dB	0.87	3.45 s
	DeblurGAN [17]	28.70 dB	0.88	0.06 s
	SRN-Deblur [32]	30.71 dB	0.88	0.13 s
self-supervised	Madam et al. [19]	21.69 dB	0.75	0.25 s
	Ours	32.24 dB	0.91	0.05 s

Figure 1.1: Comparison between different approaches for deblurring

In short, by considering deblurring as inverse rendering problem first 2 consecutive blur images are given into the network which predicts the sharp estimate of the image using deblur net from which the corresponding optical flow is estimated using this prediction we again blur the images in order to compare with the original blur images .Both synthetic as well as real dataset is used for evaluation of the model.

In order to improve the results of the network we have concentrated on hyperparameter tuning and loss function improvement.

1.3.1 Hyperparameter tuning

The Hyperparameter tuning is a process of choosing a set of optimal hyperparameters for a neural network . A hyperparameter is a model argument that is set before training model before the model begins to learn . Hyperparameter tuning is a key process that needs to be followed in order to obtain the best out of the neural network model.

The hyperparameters choosen:

1.learning rate:It is one of the important hyperparameter used in the training of neural network. Learning rate controls how quickly the model will adapt to the data.Here learning rate 0.0001 is set.

2.Loss function:It is a component of Neural Networks that predicts error of Neural Net.Initially the default and the virtual choice of the loss function for training neural network on images was L2 but actually the L1 loss function works well on images than L2 as it does not over-penalize large errors so it is clear that L1 loss is better than L2 loss for images but L1 loss does not provide good contrast.[2]

On other hand SSIM helps the network to learn and produce visually pleasing images and as the choice of standard deviation values impact the quality of the images of a network that is trained with SSIM that is when standard deviation values are small then network looses the ability to preserve local structure and irregularly shaped spot,stain artifacts appear in flat regions of image and when the value is large then the neural network tends to conserve noise in nearby edges so rather than finetuning the standard deviation value we can use multi-scale version of SSIM that is MSSIM.

$$\mathcal{L}^{\text{Mix}} = \alpha \cdot \mathcal{L}^{\text{MS-SSIM}} + (1-\alpha) \cdot G_{\sigma_G^M} \cdot \mathcal{L}^{\ell_1},$$

Figure 1.2: Mixed loss function

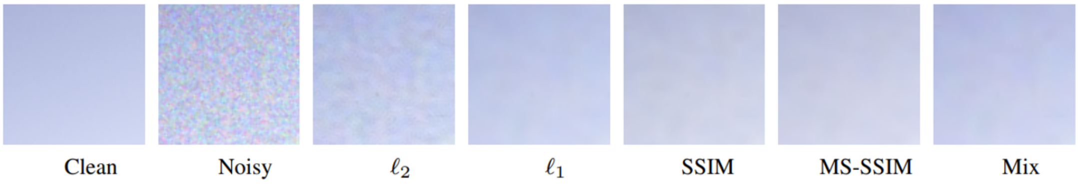


Figure 1.3: Comparison of loss functions

So in order to preserve contrast we can combine both L1 loss and MSSIM it helps in taking advantage of both preserving contrast and colour and luminance.In order to combine both the loss functions we use a hyperparameter alhpa that is empherically set to 0.84 and add both loss functions added as weighted sum in the following way.

1.4 Problem Statement

Train a self-supervised convolution neural network for video deblurring.

1.5 Applications in Societal Context

1. Motion blurr in images question many computer vision algorithms like depth prediction, feature detection, motion estimation, or object recognition as these algorithms depend on the visual input.

2.Collecting training data with corresponding sharp and blurry image pairs might not be easy and when the environment is illuminated poorly this aggregates the problem.

3. At a time only blur or clear frames can be captured so a self-supervised way of training is needed that does not depend on the ground truth sharp images for deblurring task.

Chapter 2

System design of self-supervised video de-blurring

2.1 Functional block diagram

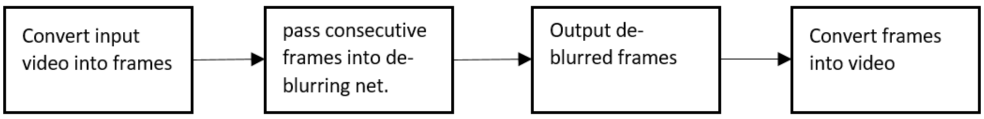


Figure 2.1: functional block diagram

2.2 Design alternatives

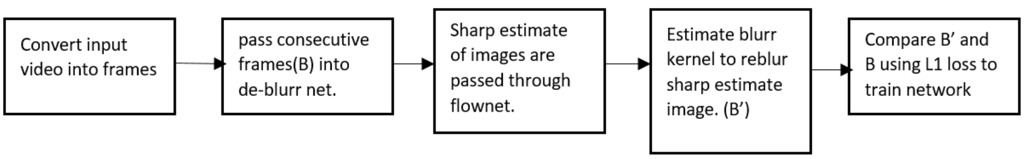


Figure 2.2: Design alternative

2.3 Final design

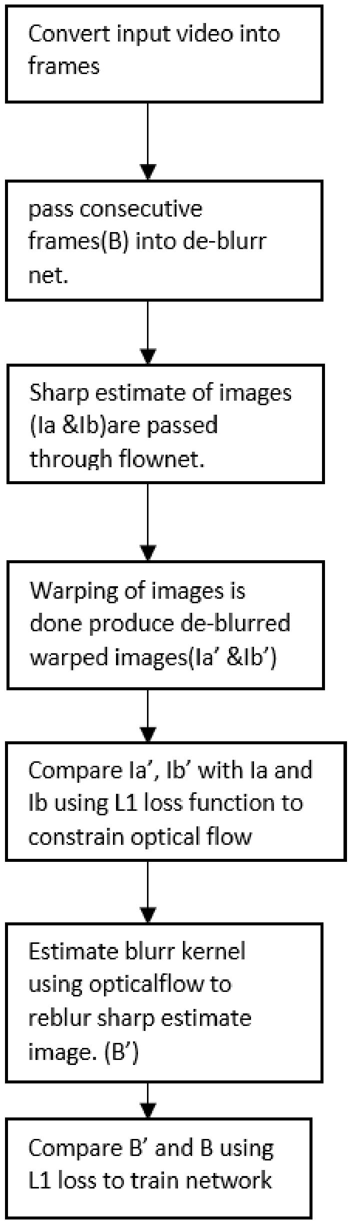


Figure 2.3: Final design

Chapter 3

Implementation details of self-supervised video de-blurring

3.1 Specifications and final system architecture

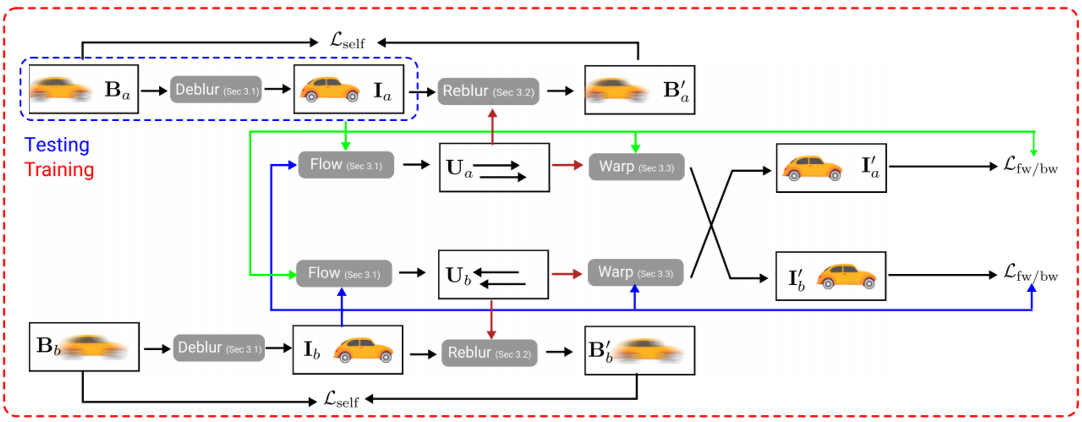
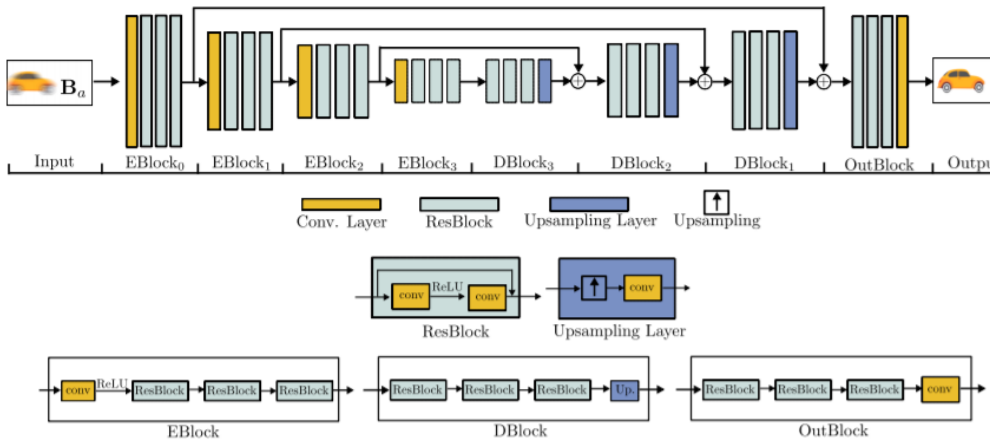


Figure 3.1: System architecture for self-supervised video deblurring



3.2 Algorithm

1. With two consecutive blurry frames sequence of a video , we first estimate the deblurred images of corresponding images using a deblur net.

2. The other flownet(a neural net) takes both deblurred images as input and computes the optical flow of corresponding images.

3. With the help of this prediction, and assuming a locally linear blur kernel, the model again renders the blurred images and compares the results to the original blurry inputs using a photometric loss function.

4. Moreover, we constrain the optical flow network using a photo-consistency loss function(comparing the warped deblurred images with input sharp images).

5.The entire model can be trained end-to-end using pairs of successive blurry images sequences captured with a consumer video camera.

6. During testing , the network takes a single blurred image and deblurs it in real time on with the help of the learned parameters.

7. The deblurred frames are again made into video by combining the images at the FPS of input video.

3.3 Flowchart

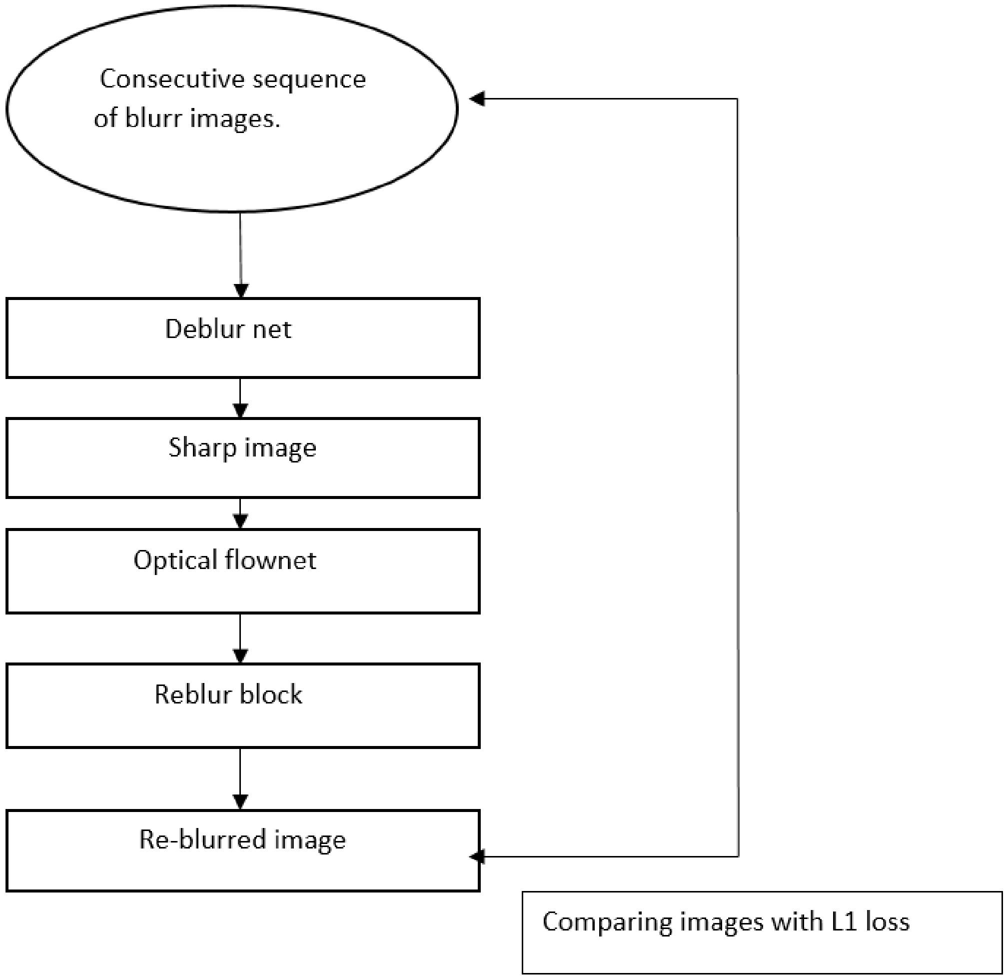


Figure 3.2: Flowchart

Chapter 4

Optimization of self-supervised video de-blurring

4.1 Introduction to optimization

In making decisions and assessing physical and logical processes, optimization is a critical stage. It improving a system's performance in time(speed) or space(size). The motive behind optimization is to attain the best functional design. This involves optimizing characteristics of the system in use (range, energy consumption).

4.2 Types of Optimization

1)Data optimization
2)Data processing optimization
3)Deterministic and Stochastic optimization

4.3 Selection with justification of optimization method

For video deblurring we have taken 2 blur image sequences for processing and calulated the loss using L1 loss function. For better results we have hyperparameter tuning is done with improving loss function.

1)Hyperparamters set:After trying combinations of hyperparameters we have set learning rate to be 0.0001 , batch size is 2,image of height and width 256 X 256 ,learning rate decay is also set all of the following helps in proper learning of the model as per the experiments made.

2)Loss function :To improve loss function a mixed loss function is used which preserves contrast as well as luminance of image when deblurred the loss function is a combination of L1 and MSSIM loss function.

Chapter 5

Results and discussions of self-supervised video de-blurring

5.1 Result Analysis

5.1.1 using L1 loss function

Pre-Trained model trained for 500+ epochs.

The blur input image

Sharp estimate of input blur image

Figure 5.1: Deblurring using l1 loss on colour images of a video taken in low lit scenes

The video is taken at low lit environment during evening before sunset from the figure it is clear that the model is able to deblur objects like water tank and window ,edges of building.

Blur input image

Sharp estimate of input blur image

Figure 5.2: Deblurring using l1 loss on colour images of a video taken in artificial lighting

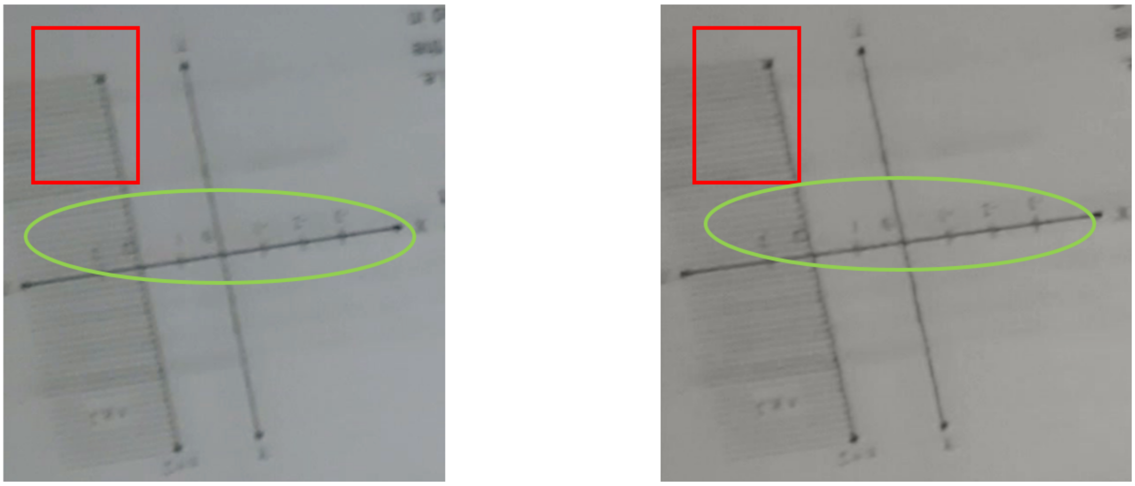
The video is taken at artificial light(room lighting) environment during evening after sunset from the figure it is clear that the model is able to deblur texts,edges of design ,geometry of the figure.

The blur input image

Sharp estimate of input blur image

Figure 5.3: Deblurring using l1 loss on grayscale images

The image is taken form base paper dataset which is taken at low lit scenes from the figure it is clear that the model is able to deblur texts,edges of building,contents of poster.



The blur input image

Sharp estimate of input blur image

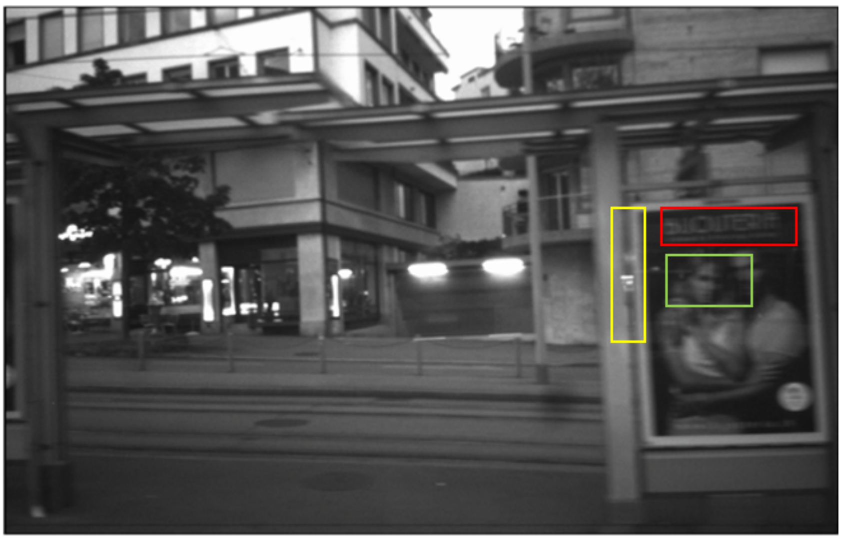
Figure 5.4: Deblurring using l1 loss on images with text and lines

The image is taken form our dataset wich is taken at low lit scene from the figure it is clear that the model is able to deblur texts,digits,lines of a book.

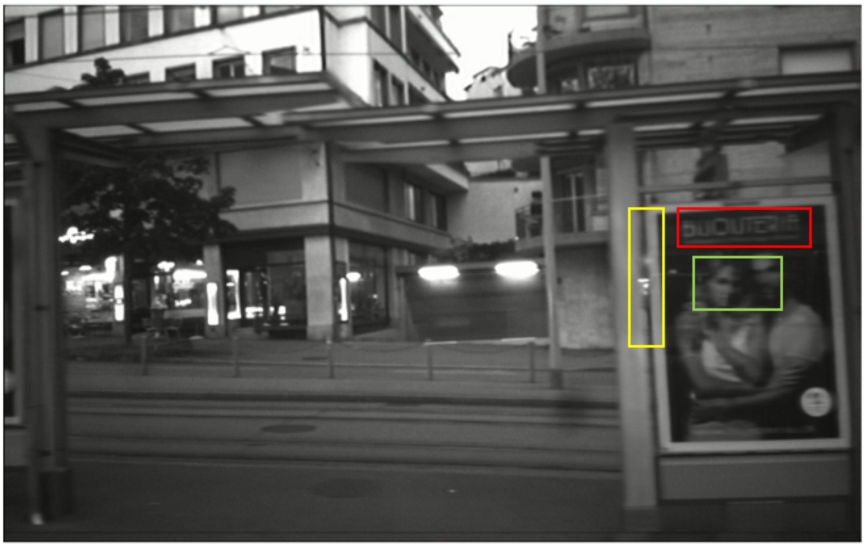
From the following figures where there is an input blur image that is made to pass through our model our model is able to recover lines ,edges , geometry.

5.1.2 Using L1 + MSSIM loss function

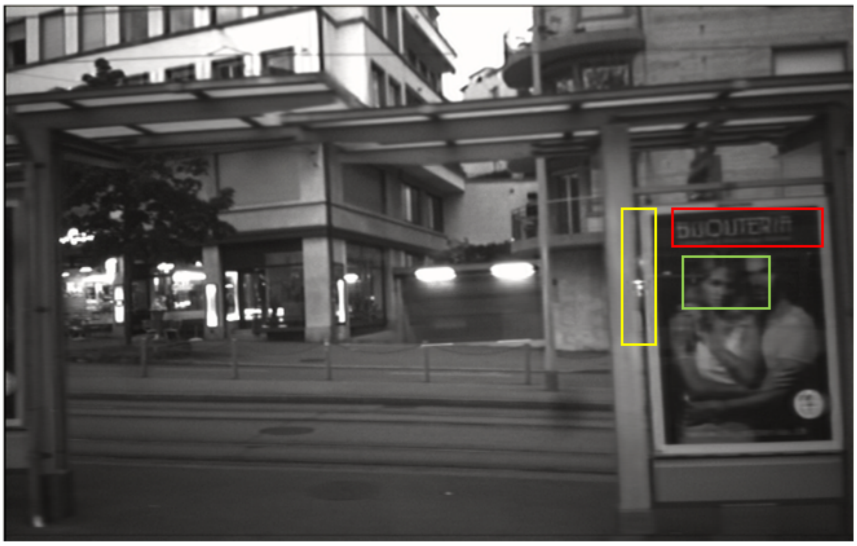
On grayscale images.



Input



L1 loss



L1+MSSIM

Figure 5.5: Deblurring using L1+MSSIM loss on grayscale image

observing the above images and comparing the l1 losss and l1+mssim loss function results we can say that with l1+mssim loss function we are able to see the letters in the more properly that is l1+mssim loss function is able to increase contrast of images so we are able to differentiate between the parts of image as it is a grayscale image colour and luminance preservation is not seen.

On Colored images.

Input

L1 loss

Figure 5.6: Deblurring using L1 loss on coloured image

L1+MSSIM

Figure 5.7: Deblurring using L1+MSSIM loss on coloured image

In the following images we can see that both colour and contrast of images are preserved by the mixed loss function as seen in in the bounding boxes deblurring is happening ,the colour shift comparing both the loss function results we can say in mixed loss function's results there is less color shift with the help of mssim loss combined with l1 loss function we are able to preserve contrast in high frequency region(part represents the rough parts (such as contours, lines and so on) of the image).

Chapter 6

Conclusions and future scope of self-supervised video de-blurring

6.1 Conclusion

De-blurring is very much needed upon motion blurred images as it challenges many computer vision algorithms that rely on visual input for further analysis of the image and in order to train a neural network in a supervised way needs both input blur and sharp images which are hard to collect which leads to need for a self-supervised way of training the neural network that improves the networks generalization performance and we have improved the results by setting proper hyperparamters and improved loss function by making the loss function to preserve contrast and colour of image when deblurred.

6.2 Future scope

We can train the model for sufficient number of epochs nearly for 500 epochs and decide the hyperparameters automatically using optuna,hyperopt hyperparamter optimization framework and we can improve any neural network performance and make it lighter using depthwise separable convolution layer in place of convolution layer in the future depthwise separable convolution are going to replace normal convolution due to their high representational efficiency.

Chapter 7

References

[1] "Self-Supervised Linear Motion Deblurring",IEEE ROBOTICS AND AUTOMATION LETTERS, VOL. 5, NO. 2, APRIL 2020.

[2]"Loss Functions for Image Restoration with Neural Networks",IEEE Transactions on Computational Imaging.

2_Self-supervised video De-blurring

ORIGINALITY REPORT

18%	12%	8%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to B.V. B College of Engineering and Technology, Hubli Student Paper	7%
2	arxiv.org Internet Source	5%
3	Peidong Liu, Joel Janai, Marc Pollefeys, Torsten Sattler, Andreas Geiger. "Self-Supervised Linear Motion Deblurring", IEEE Robotics and Automation Letters, 2020 Publication	2%
4	en.wikipedia.org Internet Source	1%
5	dokumen.pub Internet Source	1%
6	Submitted to Indian Institute of Technology, Bombay Student Paper	<1%
7	M. Hermann, B. Ruf, M. Weinmann, S. Hinz. "SELF-SUPERVISED LEARNING FOR MONOCULAR DEPTH ESTIMATION FROM	<1%

AERIAL IMAGERY", ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2020
Publication

8	Submitted to Universiteit van Amsterdam Student Paper	<1 %
9	export.arxiv.org Internet Source	<1 %
10	Submitted to Visvesvaraya Technological University Student Paper	<1 %
11	docplayer.net Internet Source	<1 %

Exclude quotes	Off	Exclude matches	< 5 words
Exclude bibliography	Off		