IMPACT OF LOSS FUNCTION ON MULTI-FRAME SUPER-RESOLUTION

Anzhelika Mezina
Doctoral Degree Programme 1st year, FEEC BUT
E-mail: xmezin00@vutbr.cz

Supervised by: Radim Burget
E-mail: burgetrm@feec.vutbr.cz

Abstract: Nowadays, one of the most popular topics in image processing is super-resolution. This problem is getting more actual even in security, since monitoring cameras are everywhere and in the case of an incident, it is necessary to recognize a person from records. A lot of approaches exist, which are able to reconstruct image, and the most of them are based on deep learning. The main focus of this work is to analyze, which loss function for neural networks is more effective for real-world image reconstruction. For this experiment chosen architecture and dataset are used for multi-frame super-resolution for $\times 8$ scaling.

Keywords: super-resolution, image processing, loss function, deep learning

1 INTRODUCTION

Last time, the image processing, especially, improvement of image quality, takes a big place in research area. In spite of the fact, that there are a lot of techniques for image processing, and it is not difficult to get cameras with good optics, there is still a need to reconstruct images from low quality. Especially, this problem is critical in the security cameras. These cameras usually have low quality of records and in the case of incidents the police is not able to recognize a person in such videos.

Recent years, many approaches have been introduced and a lot of them are based on neural networks, which require huge datasets for training. Currently, the most works use the datasets, which were created from Youtube videos or films, and these images are really nice, because of light and color correction. The security cameras do not have such good quality and recognition of a person from such records is getting more complicated. For that reason, it is necessary to have algorithms and methods, which would be able to improve quality of image.

The main focus in research area is on the methods for single frame super-resolution, however, there are not so many methods, which perform reconstruction from a couple of images. Additionally, there are not so many works, which pay attention on loss functions for deep learning methods. For these reasons, this paper provides the comparison study of loss functions which can be applied for multi-frame super-resolution task. In this work the experiment is performed with dataset of images from real-world for 8 scale factor. The results introduced in this paper can be useful for future study of problem of image reconstruction from low quality images.

2 RELATED WORK

The often used methods in the problem of super-resolution are based on convolutional neural networks (CNN) and generative adversarial networks (GAN). The well known methods are based on single frame super-resolution, for example, SRCNN [1], which is based on CNN and has only three layers. The authors applied Mean Squared Error (MSE) as a loss function. It achieved better results than

methods without application of deep learning. Another method introduced in 2018 is ESRGAN [2]. It is based on GAN and has been developed as enhanced version of SRGAN [3]. The authors used perceptual loss function, which is better for reconstruction tasks.

Multi-frame super-resolution is not so popular area and the most methods are based on mathematical approaches, however, there are some methods, which utilize the deep leaning. One of the latest works [4] introduced the method based on CNN to perform motion-based multi-frame super-resolution. Also, some approaches use residual learning, for example, in the work [5] authors combined sub-pixel registration method for mapping into the high-resolution grid and deep residual learning approach for restoring features without noise. Both mentioned methods utilized MSE as a loss function.

3 METHODOLOGY

This work is focused on impact of loss functions used in neural network on quality of a reconstructed image from a sequence of images. There are some kind of loss functions, which are frequently used in image reconstruction, such as Mean Square Error (MSE), Feature reconstruction loss, Charbonnier loss and, relatively new one, is Learned Perceptual Image Patch Similarity (LPIPS). In this section, the experiment with these loss functions is described.

3.1 DATASET

Nowadays, there are not so many datasets for multi-frame super-resolution. In this work the new dataset Multi-Frame Labeled Faces Database (MLFDB) [6] is used for experiments. This dataset contains sequences of 7 images, extracted from real-life videos. Moreover, this dataset allows to do experiments with different scale factors: 2, 4, 8. For this experiment the chosen scale factor is 8 and training, validation, testing sets have 2,504, 837 and 753 images respectively. The example of input and output (label) images are shown in Fig. 1. The input size of images is 32×32 px and output is 256×256 px.

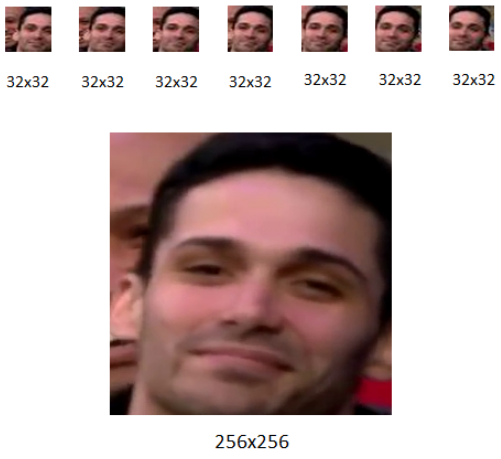


Figure 1: Example from dataset.

3.2 LOSS FUNCTIONS

For this work used such loss functions, as MSE, Feature reconstruction loss, Charbonnier loss and LPIPS.

MSE loss function is the most frequently used in image processing. This loss function calculates squares between original image and predicted one. It can be formulated as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i)^2, \tag{1}$$

where X_i – original high-resolution image, Y_i – predicted super-resolution image, n – number of training samples.

Charbonnier loss was also applied for super-resolution task [7]. This loss function might better handle outliers and improve the reconstruction accuracy.

$$L = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(Y_i - X_i)^2 + \varepsilon^2}, \tag{2}$$

where X_i – original high-resolution image, Y_i – predicted super-resolution image, ε – constant, 0.001, n – number of training samples.

In spite of high Peak Signal-to-Noise Ratio (PSNR) value of the image, which is reconstructed with MSE loss, there is still a problem, that quality of image do not correlate well with human perception [7]. For that reason, the perceptual loss function is very popular in image super-resolution task. Feature Reconstruction Loss [8], which is a kind of perceptual loss, is used in this work. In contrast to MSE loss function, it encourages the similar of feature representation, instead of matching of pixels. This loss function can be defined as:

$$\ell_{feat}^{\theta,j}(\hat{I}, I) = \frac{1}{C_j H_j W_j}||\phi_j(\hat{I}) - \phi_j(I)||^2, \tag{3}$$

where I – original high-resolution image, $\hat{I}$ – predicted super-resolution image, $\phi_j(\hat{I})$ – activations of jth layer of the VGG19 network ϕ for processing the image I, $C_j H_j W_j$ – the shape of the feature map.

LPIPS was applied for extreme super-resolution with 16 scale factor [9]. The authors state, that for extreme super-resolution the VGG-based loss function is not the best choice, since VGG network is used for classification task. But the authors proposed to apply the AlexNet network as a perceptual loss function. This loss function can be described as:

$$\ell_{lpips} = \tau(\phi_j(\hat{I}) - \phi_j(I)), \tag{4}$$

where I – original high-resolution image, $\hat{I}$ – predicted super-resolution image, ϕ – feature extractor, τ – transformation of deep embedding to scalar LPIPS score.

3.3 ARCHITECTURE

For this experiment used neural network architecture, which was proposed and applied for multi-frame super-resolution task, is U-Net based network with GEU blocks [6]. This model processes the sequence of images. Originally, in the paper it was used for 2 scale factor, however, for this experiment it was adapted for 8 scale factor: some layers with subpixel convolution operation for upsampling were added. The scheme of architecture is shown in Fig. 2. The model was trained with hyperparameters: optimizer is Adam with learning rate 0.0001, the number of epochs is 300 with 500 steps for epochs, batch size is 4.

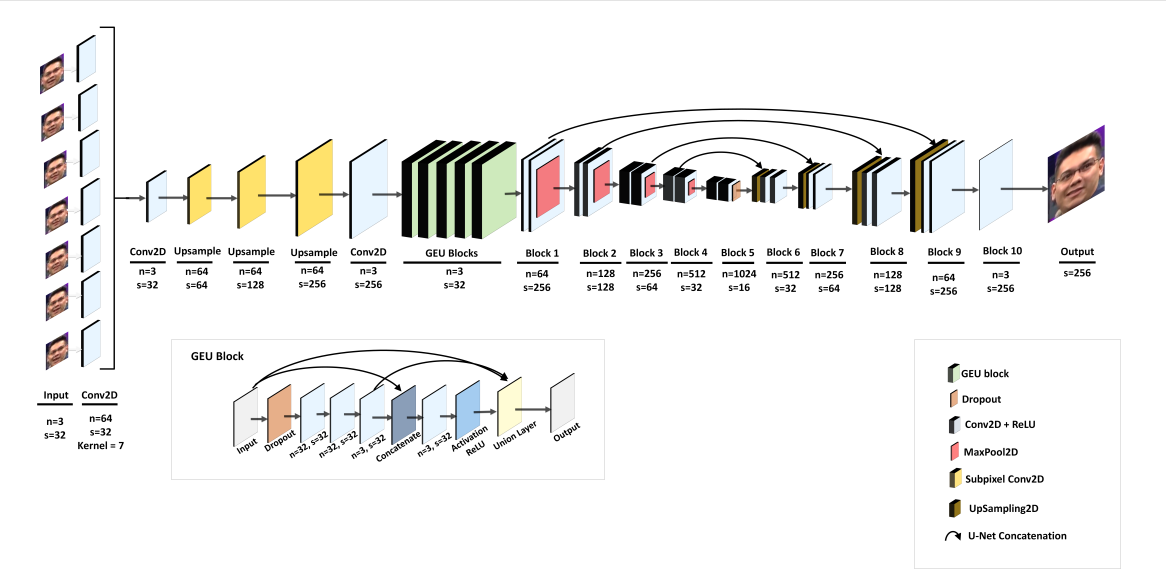


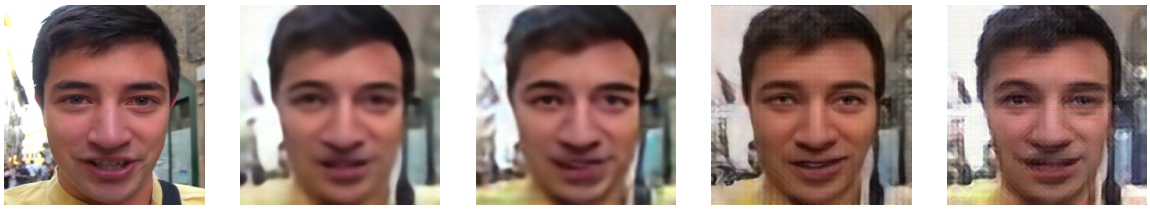
Figure 2: Adapted for 8 scale U-Net+GEU 3 from [6].

4 RESULTS

The U-Net+GEU 3 model was trained on mentioned dataset with different loss functions which were described in this work. The results of evaluation of methods are shown in Table 1 and Fig. 3, where the last column, FFR, is the number of failed face recognition. According to the metrics the best objective results – SSIM, MSE, PSNR, achieved model with Charbonnier loss function. It achieved SSIM – 0.7251, MSE – 302.5284, PSNR – 24.1890 dB. However, it can be seen from Table 1, that the number of failed face recognition is the least for LPIPS loss function – 435, that can be a useful point in case of person identification. On the other hand, from subjective side it seems that image reconstructed with feature reconstruction loss looks better and more details are seen. Moreover, the sharpness difference between original and reconstructed images is the least for feature reconstruction loss.

Table 1: Results for 8 scale factor.

Loss function	SSIM	MSE	PSNR, dB	Sharp. difference	FFR
MSE	0.7083	352.6861	23.4785	0.2071	473
Charbonnier	**0.7251**	**302.5284**	**24.1890**	0.2073	466
Feature reconst.loss	0.6490	616.5493	20.8621	0.0270	450
LPIPS	0.5240	435.4415	22.3790	-0.4884	**435**



(a) Ground truth **(b)** Charbonnier **(c)** MSE **(d)** Feat. reconst. **(e)** LPIPS

Figure 3: Results of evaluation

5 CONCLUSION

The aim of this work was to investigate the impact of loss functions on image reconstruction from a sequence of frames. For experiment the new dataset and the method for multi-frame super-resolution were used, and applied loss functions were Charbonnier, MSE, feature reconstruction loss and LPIPS. According to results, the best metrics such as SSIM, MSE, PSNR, achieved images reconstructed with Charbonnier loss. However, the resulting image seems very smooth and for human it is difficult to recognize a person there. The number of failed face recognition is the least for LPIPS loss. Subjectively, more details are seen in the image reconstructed with feature reconstruction loss, but the face has some deformations. In this way, it is obviously that loss function takes a big role in image reconstruction.

As the conclusion, it is necessary to prepare more experiments with architectures and loss functions. There is still a lack in research field for image super-resolution for real-world application. Nowadays, there are a lot of approaches with nice reconstructed images, which were tested on datasets with good quality images. However, these methods can fail on images from monitoring cameras. For real-world application, it is better to use loss functions in neural networks, which are focused on improvement quality for human perception. If the reconstructed image is used by police for person identification, it does not matter, which objective metrics it is achieved, it is more important to recognize a culprit.

REFERENCES

[1] Dong, Chao, et al. "Image super-resolution using deep convolutional networks." IEEE transactions on pattern analysis and machine intelligence 38.2 (2015): 295-307.

[2] Wang, Xintao, et al. "Esrgan: Enhanced super-resolution generative adversarial networks." Proceedings of the European Conference on Computer Vision (ECCV) Workshops. 2018.

[3] Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[4] Elwarfalli, Hamed, and Russell C. Hardie. "FIFNET: A convolutional neural network for motion-based multiframe super-resolution using fusion of interpolated frames." Computer Vision and Image Understanding 202 (2021): 103097.

[5] Noor, Dewan Fahim, et al. "Multi-frame super resolution with deep residual learning on flow registered non-integer pixel images." 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019.

[6] Rajnoha, Martin, Anzhelika Mezina, and Radim Burget. "Multi-frame labeled faces database: Towards face super-resolution from realistic video sequences." Applied Sciences 10.20 (2020): 7213.

[7] Lai, Wei-Sheng, et al. "Deep laplacian pyramid networks for fast and accurate super-resolution." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[8] Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." European conference on computer vision. Springer, Cham, 2016.

[9] Jo, Younghyun, Sejong Yang, and Seon Joo Kim. "Investigating loss functions for extreme super-resolution." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.