UDC 004.93

LAMA-WAVELET: IMAGE IMPAINTING WITH HIGH QUALITY OF FINE DETAILS AND OBJECT EDGES

Kolodochka D. O. – Student of the Institute of Computer Systems, National University "Odessa Polytechnic", Odessa, Ukraine.

Polyakova M. V. – Dr. Sc., Associate Professor, Professor of the Department of Applied Mathematics and Information Technologies, National University "Odessa Polytechnic", Odessa, Ukraine.

ABSTRACT

Context. The problem of the image impainting in computer graphic and computer vision systems is considered. The subject of the research is deep learning convolutional neural networks for image inpainting.

Objective. The objective of the research is to improve the image inpainting performance in computer vision and computer graphics systems by applying wavelet transform in the LaMa-Fourier network architecture.

Method. The basic LaMa-Fourier network decomposes the image into global and local texture. Then it is proposed to improve the network block, processing the global context of the image, namely, the spectral transform block. To improve the block of spectral transform, instead of Fourier Unit Structure the Simple Wavelet Convolution Block elaborated by the authors is used. In this block, 3D wavelet transform of the image on two levels was initially performed using the Daubechies wavelet db4. The obtained coefficients of 3D wavelet transform are splitted so that each subband represents a separate feature of the image. Convolutional layer, batch normalization and ReLU activation function are sequentially applied to the results of splitting of coefficients on each level of wavelet transform. The obtained subbands of wavelet coefficients are concatenated and the inverse wavelet transform is applied to them, the result of which is the output of the block. Note that the wavelet coefficients at different levels were processed separately. This reduces the computational complexity of calculating the network outputs while preserving the influence of the context of each level on image inpainting. The obtained neural network is named LaMa-Wavelet. The FID, PSNR, SSIM indexes and visual analysis were used to estimate the quality of images inpainted with LaMa-Wavelet network.

Results. The proposed LaMa-Wavelet network has been implemented in software and researched for solving the problem of image inpainting. The PSNR of images inpainted using the LaMa-Wavelet exceeds the results obtained using the LaMa-Fourier network for narrow and medium masks in average by 4.5%, for large masks in average by 6%. The LaMa-Wavelet applying can enhance SSIM by 2–4% depending on a mask size. But it takes 3 times longer to inpaint one image with LaMa-Wavelet than with LaMa-Fourier network. Analysis of specific images demonstrates that both networks show similar results of inpainting of a homogeneous background. On complex backgrounds with repeating elements the LaMa-Wavelet is often more effective in restoring textures.

Conclusions. The obtained LaMa-Wavelet network allows to improve the image inpainting with large masks due to applying wavelet transform in the LaMa network architecture. Namely, the quality of reconstruction of image edges and fine details is increased.

KEYWORDS: image inpainting, wavelet transform, LaMa network, Daubechies wavelet, Fréchet inception distance, wavelet convolution.

ABBREVIATIONS

CNN is a convolutional neural network;

MSNPS is Muli-Scale Neural Patch Synthesis;

GLCIC is Globally and Locally Consistent Image Completion;

CoModGAN is co-modulated generative adversarial network;

LaMa-Fourier is Large Mask Inpainting with Fourier Convolutions;

LaMa-Wavelet is Large Mask Inpainting with wavelet transform;

FFC is a fast Fourier convolution;

FFT is a fast Fourier transform;

iFFT is an inverse fast Fourier transform;

LLL, LLH, LHL, HLL, LHH, HLH, HHL, HHH are wavelet coefficient subbands;

DWT is a discrete wavelet transform;

iDWT is an inverse discrete wavelet transform;

BN is a batch normalization layer;

ReLU is a rectified linear unit;

LPIPS is Learned Perceptual Image Patch Similarity;

FID is Fréchet inception distance;

PSNR is a peak signal-to-noise ratio;

MSE is a mean square error;

SSIM is a structural similarity index measure.

NOMENCLATURE

n, m is the number of image rows and columns;

(x,y) are coordinates of the image pixel;

$I(x,y)$ is a vector function representing an image by color channels;

$I_R(x,y)$, $I_G(x,y)$, $I_B(x,y)$ are the red, green, blue color channels of an image;

$\circ$ is an element-by-element product of matrixes;

f_θ is an inpainting network;

$struct_{f\theta}$ is the architecture of the f_θ network;

$param_{f\theta}$ is the set of parameters of the f_θ network;

$I_{in}(x,y)$ is an inpainted three-channel color image;

L_2 is a pixel loss;

L_P is a perceptual loss;

L_D is a competitive loss;

k, a, b are the coefficients controling the impact of each of the losses;

m_r and m_g are vectors of mean feature values for real and generated image sets, respectively;

R_r and R_g are the covariance matrices of the features of the real and generated sets of images;

DOI 10.15588/1607-3274-2024-1-19

tr is the trace of the matrix;

L is the number of intensity levels on the image;

$I(v, w)$, $c(v, w)$, $s(v, w)$ are the luminance, contrast, and structure between images v and w ;

m_v, m_w are the local means of images v and w ;

σ_v, σ_w are the standard deviations of images;

σ_{vw} is the cross-covariance for images v and w.

C_1, C_2, C_3, α, β and γ are the positive constants.

INTRODUCTION

In computer vision and computer graphics systems, there is a need to inpaint missing areas of the image. This problem arises in the case of faded colors or physical damage of the surface on which the image was located. In another case, there is a need for removing unwanted objects from the image in such a way that the resulted image looks realistic and fits the context. For example, when photographing, distracting objects in a scene such as strangers and objects that get in the way are usually unavoidable but at the same time undesirable for users. Before sharing a photo, users may want to make some changes, such as removing distracting elements from the scene or adjusting the position of objects in the image for a better composition. Image inpainting techniques provide automatic filling of missing regions of an image with a plausible hypothesis. These techniques are used in many real-world tasks, such as removing distracting objects, restoring damaged parts, and filling the missing areas of images [1, 2].

The object of research is inpainting of real scene images in computer vision and computer graphics systems.

Due to the development of modern technologies and the increase in computing power, a number of image inpainting methods based on deep learning CNNs have been elaborated, which can generate missing regions of the image with good global consistency and local fine textures. Thus, CNNs Context Encoder, MSNPS, GLCIC, DeepFill v1–2 differ in such characteristics as speed, the size of the processed image, and the quality of filling of image regions [3]. The disadvantage of the listed methods is that when using large masks, the result becomes unsatisfactory in terms of generating both image context and texture.

The subject of the research is methods of image inpainting using CNNs of deep learning.

Unlike the Context Encoder, MSNPS, GLCIC, DeepFill v1–2 methods, the LaMa-Fourier [4] is able to obtain a good result even when the missing areas occupy most of the image. In general, CNNs for image inpainting usually achieve better results by complicating the network architecture or by dividing it into sub-networks with separate tasks. LaMa-Fourier, on the contrary, uses a single network and a fewer variables [4].

The advantages of the LaMa-Fourier method are a higher speed of image processing and network training; better quality than other neural network methods when using narrow masks; better quality of large mask inpainting of spectral textures. The disadvantage of the LaMa-Fourier network is the insufficient quality of inpainting of fine details of images and edges of objects. To eliminate this shortcoming, it is appropriate to use a wavelet transform representing both global and local features of images.

The aim of the paper is to improve the quality of image inpainting in computer vision and computer graphics systems with applying wavelet transform in the LaMa-Fourier network architecture.

1 PROBLEM STATEMENT

The color natural image is represented as $I(x,y)=(I_R(x,y), I_G(x,y), I_B(x,y)\}$, where $x=1, \ldots, n$; $y=1, \ldots, m$. Then each pixel of the image is described by three features $I_R(x,y)$, $I_G(x,y)$, $I_B(x,y)$ which take values from the interval [0, 255]. A mask is introduced to represent the missing areas of the image. This is a binary image $M(x, y)$ of the same size as each channel of the original image. The mask is element-by-element producted by image features. Then, using a mask, the image with missing areas is represented as $I_M(x,y)=(I_R(x,y)\circ M(x, y), I_G(x,y)\circ M(x, y), I_B(x,y)\circ M(x, y))$. It is necessary to transform the image $I_M(x,y)$ so as to fill missing areas. In this case, the resulting image should approximate the original one in the sense of some criterion [4].

Let the CNN $f_\theta=\{struct_{f\theta}, param_{f\theta}\}$ was preliminarily designed to inpaint the images. The set $struct_{f\theta}$ includes blocks with layers of the designed network. Taking $I_M(x,y)$ the inpainting network processes the input in a fully-convolutional manner, and produces $I_{in}(x,y) = f_\theta(I_M(x,y))$ which approximates the original image $I(x,y)$.

The problem of the refining CNN architecture is as follows. It is necessary to make structural changes to the existing architecture $struct_{f\theta}$ of the network $f_\theta(\bullet)$. These changes should improve the image inpainting performance compared to the initial $f_\theta(\bullet)$ network after training the parameters of the resulting network [5, 6]. The training is performed on a dataset of (image $I(x,y)$, mask $M(x, y)$) pairs obtained from natural images and synthetically generated masks.

2 REVIEW OF THE LITERATURE

The analysis of CNNs for image inpainting showed that such methods are primarily focused on the properties of processed images. Also, the architectures of the used CNNs are determined by the computing power available to the researcher and the quality requirements for the inpainted images.

Thus, the methods [7, 8] are recommended for the filling localized missing areas of small-sized images with not very high quality of the result. These methods use one CNN, which is quite easy to train on another class of images. It is not require relatively significant time and computing power. For example, in [7] the Context Encoder was designed on the basis of a generative-competitive network. This CNN includes a fully connected layer. Due to convolutional layers, all locations of spatial objects on the previous layer contribute to the location of spatial objects on the current layer. Thus, the

DOI 10.15588/1607-3274-2024-1-19

network can learn the relationship between the locations of objects. Also, the Context Encoder can be trained to understand the overall context of the entire image.

A significant number of image inpainting methods [7, 9–11] require the correct shape of the missing area. However, the PartialConv algorithm [8] is able to fill several areas of arbitrary shape at once. PartialConv is a Unet type network which differs by the applying of partial convolutions in the convolutional layers. When partial convolution is used, the processing image fragment is first multipled by a binary matrix element by element, and then only a filter mask is applied. The disadvantage of the PartialConv is the reducing of the quality of filling missing regions with few details lagging behind each other.

In contrast to CNNs that use one subnet, the MSNPS [9] fills the missing regions of images with the help of two CNNs applied sequentially. The first CNN is used to generate the global context, the second CNN is used to further add local texture. This approach improves the quality of the inpainted images, but significantly increases the training time. MSNPS is a further developing of [7], but differs in higher details of local textures due to the use of a separate CNN for their generation.

Methods [10–14] use ensembles of three or more CNN to achieve high-quality image inpainting. But they require the significant computing power and much training time.

In [10] the GLCIC consist of three CNNs to inpaint the images. One CNN is applied to fill missing regions of the image and two auxiliary networks is used as local and global discriminators. The latter networks are used only while training. Their role is to assess the realism of the resulting image by comparing the original image areas with the inpainted ones. At the same time, the generative network is learned to deceive the discriminators, and the discriminators are learned to better identify unrealistic images.

The GLCIC fills missing regions not only based on the current image, but also based on the images used while training. A fully connected layer is not used, which significantly reduces the time to inpaint an image as compared with [7, 9] and practically removes the limitation on the size of the input image. The disadvantage of GLCIC is that, in addition to the generative network, discriminators must also be trained.

In [11], the DeepFill v1 network was proposed as a sequential combination of the networks from [9, 10]. The peculiarity of the DeepFill v1 is that when searching for suitable areas for copying, not only similar areas are determined, but also the contribution of all visible objects to the missing area is estimated. As a result, a combination of the most significant visible objects is used to fill the missing region. This enhances the quality of the obtained result. However, for correct inpainting the missing area of the image must be square.

Instead of the filling of missing regions by global context generation and local texture generation EdgeConnect [12] is proposed to generate edge map and to inpaint an image based on the obtained edges. A discriminator is used while training of each of the two subnets of the EdgeConnect [12]. The EdgeConnect can fill missing regions of arbitrary shape and shows better results than previous methods when generating objects of complex shape. But the EdgeConnect is needed to train discriminators in addition to generative networks.

The DeepFill v2 network [13] is based on the DeepFill v1, EdgeConnect, and PartialConv networks. DeepFill v2 sequentially applies three CNNs. First network is designed to generate the global context. Second network generates a local texture based on the global context of the image. Third network is a discriminator assessing the realism of the resulting image. After training the DeepFill v2 network is able to fill missing regions not only on the basis of other parts of the image, but also to evaluate the contribution of surrounding objects to the content of the missing area. This network can process regions of arbitrary shape, and use an edge map when generating objects. The DeepFill v2 shows better performance in terms of inpainted image quality and processing time compared to [7–12]. However, the training of an ensemble of three CNNs with more than 4 million parameters requires significant time and computer resources.

The CoModGAN network [14] architecture is similar to DeepFill v2, but greatly enhanced. The connectivity of filled regions to the context is better compared to DeepFill v2, but visible artifacts are possible in the center of the inpainted region. The CoModGAN network on average shows better results than DeepFill v2, but due to increasing the number of parameters to 108 million. Then, the time of network training and image inpainting increases several times.

The considered CNNs improves of the quality of image inpainting by complicating the network architecture and/or by processing taking into account, in addition to the color components, other features of the images. In particular, the texture or the edges of objects are processed. The main feature of the LaMa-Fourier network compared to considered networks is the use of a new type of a convolutional layer which is FFC [4]. It allows to significantly increase the logical connectivity of the filling missing regions with known image regions and at the same time to reduce the number of network parameters several times.

The LaMa-Fourier uses the FFC, learning mask generator, and loss function different from previously proposed methods. Its architecture is simpler compared to DeepFill v2 and CoModGAN. LaMa-Fourier has 27 million parameters, and is faster than CoModGAN due to fewer layers. At the same time, it shows better results, the absence of visible artifacts and variations of the texture structure, especially when restoring large areas and spectral textures. However, the LaMa-Fourier requires more computational resources for the implementation of FFC compared to convolution. In addition, there is

DOI 10.15588/1607-3274-2024-1-19

sometimes a slight blurring of the areas inpainted by the LaMa-Fourier, which is a side effect of applying the Fourier transform [15].

Therefore, in the paper it is proposed to enhance the LaMa-Fourier network by applying of wavelet transform which can be considered as a generalization of the Fourier transform [16]. Wavelet decomposition is performed with the help of functions with a limited extension to get information about image details.

3 MATERIALS AND METHODS

The architecture of the LaMa-Fourier network is shown in Fig. 1 [4]. The network is inputted an image with missing areas and a mask with pixels need to be inpainted. Next, the image is reduced by a factor of 3 and passes through nine residual blocks (Fig. 1, a). After that, the image is enlarged to its original size and outputted [4].

In the residual block, the FFC is applied twice to the image and the result is added to the original image. FFC decomposes the image into local and global textures, which are further processed by convolution layers (Fig. 1, b). The global texture additionally passes through the spectral transform block. The outputs of the convolution layers are summed “cross over cross”. Then BN and the ReLU activation function are applied to them. The results of local and global texture processing are concatenated (Fig. 1, b) [4].

In the spectral transform block the image is Fourier transformed into the frequency domain, the real and imaginary parts are concatenated. Then the convolutional layer, BN and the ReLU activation function are applied sequentially (Fig. 1, c). The obtained result is splitted on the real and imaginary parts. Finally, the iFFT is applied, the result of which is the output of the block [4].

The LaMa-Fourier network is able to represent the general structure of images. But difficulties arise with the inpainting of fine high-frequency details and with the generation of the image edges. There may also be problems when reproducing complex textures, such as small leaves, thin fabric fibers, or detailed patterns (Fig. 1, d–f). Difficulties in ensuring similarity between the inpainted region and the existing texture may be related to the fact that the Fourier transform traditionally works better with low and medium frequencies than with very high ones [15].

As an alternative to the Fourier transform, to overcome the mentioned shortcomings, it is appropriate to use the wavelet transform [17, 18]. Then, it is necessary to define the block of the network to which changes will be made. Since the FFC decomposes the image into global and local texture, it was decided to improve only the network block, processing the global context of the image, namely, the spectral transform block. Applying the wavelet transform to the local context can increase the noise level in the image.

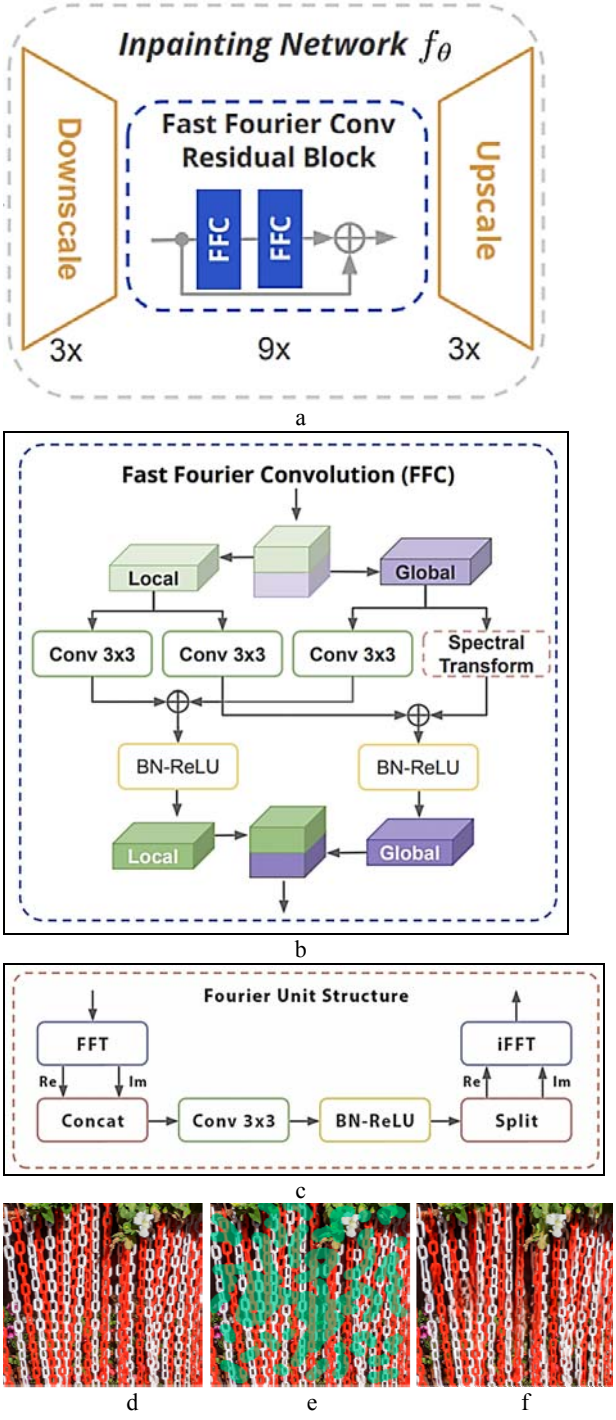


Figure 1 – LaMa-Fourier network architecture: a – Residual Block; b – FFC; c – Fourier Unit Structure [4]. LaMa-Fourier inpainting example: d – original image; e – original image and mask; f – inpainted image

To improve the block of spectral transform the Simple Wavelet Convolution Block elaborated by the authors is used instead of Fourier Unit Structure (Fig. 2). In this block, 3D wavelet transform of the image on two levels using the Daubechies wavelet db4 (Fig. 3, a) was initially performed.

DOI 10.15588/1607-3274-2024-1-19

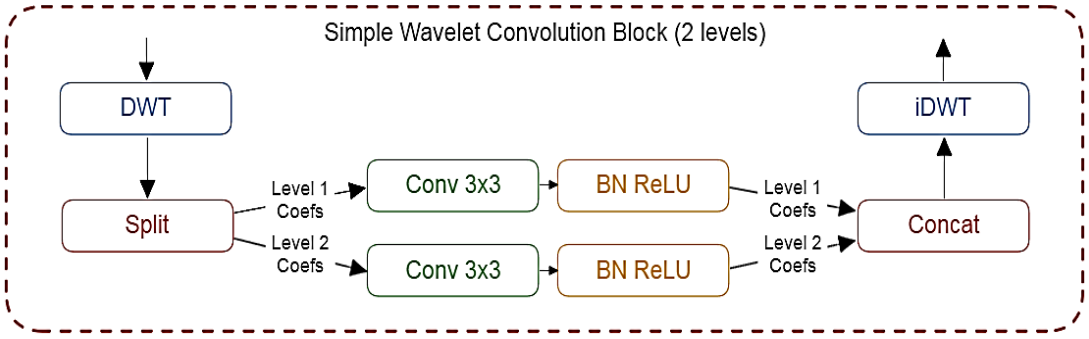


Figure 2 – Simple Wavelet Convolution Block

As a result of this transform, the eight subbands of coefficients at each level were obtained [19]. These are LLL, LLH, LHL, HLL, LHH, HLH, HHL, HHH subbands (Fig. 3, b). The main advantage of using a 3D wavelet transform, as opposed to applying a 2D transform separately for each image channel, is its ability to analyze inter-channel image correlation. This allowing to take into account color details on an inpainted image.

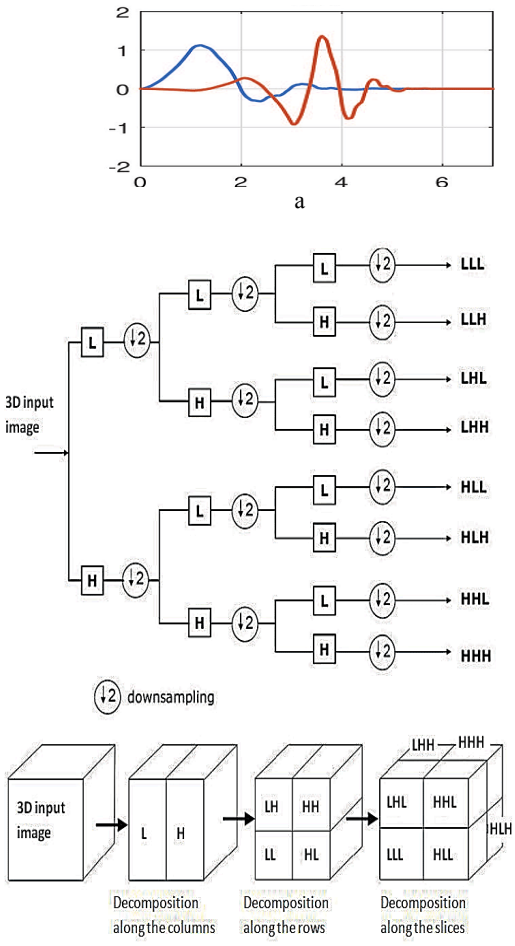
The obtained coefficients of 3D wavelet transform are splitted so that each subband represents a separate feature of the image. Convolutional layer, BN, and ReLU activation function are sequentially applied to the results of splitting of coefficients on each level of wavelet transform (Fig. 2). The number of convolutions in the convolutional layer was equal to the number of subbands at the corresponding level of the wavelet transform. After applying the ReLU activation function, the obtained subbands of wavelet coefficients are concatenated and the iDWT is applied to them, the result of which is the output of the block. Note that the wavelet coefficients at different levels were processed separately. This made it possible to reduce the computational complexity of calculating the network outputs while preserving the impact of each level to image inpainting.

The LaMa network uses the loss function, which is specially designed to solve the problem of filling large missing regions. This loss function L_{final} combines L_2, L_P and L_D to ensure the realism, semantic integrity and structural continuity of the inpainted regions, which corresponds to the human perception of image [4]:

$$L_{final} = kL_2 + aL_P + bL_D.$$

The gradient penalty [4] is not used to reduce amount of computation. The MSE between the original and restored images was used to estimate L_2 pixel loss [20]. For perceptual loss of L_P, LPIPS is used, which evaluates the perceptual similarity between the inpainted and original images using a pre-trained neural network [21].

The discriminator is used to estimate competition loss L_D. This additional CNN is trained in parallel with the basic network to distinguish between real and generated images. Based on this evaluation, the discriminator tunes the basic network coefficients to improve the realism of the generated images. Then, the L_D are estimation of the error in the global and local textures computed from the discriminator output [22].



b

Figure 3 – Elements of LaMa-Wavelet network architecture: a – Daubechies db4 scaling function (blue) and wavelet (red) [16]; b – subbands of 3D wavelet transform coefficients [19]

4 EXPERIMENTS

At the first stage of the experiment, the LaMa-Wavelet network was trained to inpaint test database images. The Google Colab environment with a pre-configured NVIDIA Tesla T4 GPU, which has 16 GB of GDDR6 memory and 2,560 CUDA cores, was used. The T4 is optimized for machine learning computing and supports

DOI 10.15588/1607-3274-2024-1-19

NVIDIA Turing Tensor Cores to accelerate tensor operations. The Google Colab environment uses processors to process data and to interact with the GPU. The Google Colab environment provides about 60 GB of RAM. But no more than 4 GB was used for training, because the system is more demanding of video memory and most of the RAM is used only for mask generation. Google Drive data storage was used to store the training and testing datasets to easy access to them.

The LaMa-Wavelet was trained using the Adam optimizer with the following parameters of the L_{final} loss function: k=10, a=100, b=30.

While network training, an initial image and a mask are selected. A mask is a binary image on which black pixels correspond to pixels of the missing region of the initial image, and white pixels correspond to known pixels. Pixels, which will be inpainted later, were removed from the input image according to the mask. Next, the obtained image and the mask were fed to the input of the trained network. The image with filled missing regions was outputted by the network. The original image was superimposed on this generated image according to the mask. Thus, the result was an image in which the known pixels were copied from the original image, and the pixels of the missing regions were generated by the network.

The 16,000 images from databases [23, 24] were used to train the LaMa-Wavelet network. These images were scaled to a size of 256x256 pixels and randomly splitted into training and testing sets in the ratio of 95% to 5%. For each image, with a probability of 0.5, either a mask of 1–4 rectangles with sides of 30–150 pixels, or a mask of 1–5 straight lines 10–200 pixels long, 1–100 pixels wide and with a slope from 0 to 2π was generated. The sizes of the masks were variable, from narrow (10% of the image pixels) to large (80% of the image pixels). This ensured that the network was trained at different levels of inpainting complexity. Masks were generated with a random uniform distribution over the entire area to ensure uniform coverage of different image areas.

Evaluation of the results of the first stage of the experiment was performed by the FID score [25]. It was calculated using the Inception v3 network for two sets of images, specifically, real image set and set of images with generated regions. When this network obtained the features from the real and generated image sets, the FID is calculated as

$$FID=\| m_r - m_g\|^2 + \mathrm{tr}(R_r + R_g - 2(R_r R_g)^{1/2}).$$

FID measures the distance between the feature distributions of real images and images inpainted by the network. Lower FID value means that the feature distributions are closer, indicating more similarity between the generated and real images. This metric takes into account both the variability and the quality of the generated images [25].

The original LaMa-Fourier network is balanced in terms of image inpainting quality and processing time. Training of this model was completed according to a standard protocol, providing a reliable baseline for comparison [26]. After 128 epochs the training of the LaMa-Fourier network was completed. But the dependence of loss function from epoch for the LaMa-Wavelet still showed a downward trend. This indicated the possibility of further improvement of the loss provided the training is continued.

Continued training of the LaMa-Wavelet to 212 epochs was intended to approach or even exceed the FID value of the trained LaMa-Fourier network. Training throughout 212 epochs reduced the FID of the LaMa-Wavelet to about 8 on the training set and to 24 on the validation set. This equalizes it with the FID of the LaMa-Fourier network. The similarity of the FID for the LaMa-Fourier and LaMa-Wavelet networks indicates that the improvement in image inpainting by the LaMa-Wavelet can only be achieved through long training.

At the second stage of the experiment, the images of testing set are inpainted using the LaMa-Fourier and LaMa-Wavelet networks. Then the inpainted images were compared with the original images. For this, three separate categories of masks were formed, namely, narrow, medium, and large, covering 15%, 40%, 65% of the image area, respectively. Each of these mask categories represented a different level of image inpainting complexity. The test set included 2000 images from the Places2 dataset [23]. One mask from each category was generated for each image. After element-by-element multiplication of images on masks, the inpainting was performed using LaMa-Fourier and LaMa-Wavelet networks.

To compare the inpainted images with the original images, the FID was first calculated. It has been observed that the FID evaluates the overall similarity of the generated and original images, but does not focus on the recovering of edges or details. To solve this problem, two additional indexes, PSNR and SSIM, are applied [20].

PSNR compares the original and reconstructed image in terms of differences between them at the pixel level. In the context of image inpainting, PSNR indicates how well edges and fine details are reconstructed. It is estimated as the ratio between the maximum possible power of an original image and the MSE between original and inpainted images if minimum intensity level supposes to be 0 [20]:

$$PSNR = 10\log_{10}((L-1)^2/MSE).$$

SSIM provides a perceptually relevant estimation of image quality considering the differences in image structure, texture, and contrast. This is critical for preserving the natural appearance of image edges and textures. The SSIM is calculated based on the luminance term, contrast term and structural or correlation term as [20]:

DOI 10.15588/1607-3274-2024-1-19

$$SSIM(v, w)=l\,(v, w)^{\alpha}\ c\,(v, w)^{\beta}\ s\,(v, w)^{\gamma},$$
$$l\,(v, w) = (2m_v m_w + C_1)/(m_v^2 + m_w^2 + C_1),$$
$$c\,(v, w) = (2\sigma_v \sigma_w + C_2)/(\sigma_v^2 + \sigma_w^2 + C_2),$$
$$s\,(x, y)=(\sigma_{vw} + C_3)/(\sigma_v \sigma_w + C_3),$$

If α=β=γ=1, then the index is in normalized scale with values between 0 to 1.

The PSNR and SSIM allowed a more detailed estimation of the inpainting performance, especially in terms of edge quality and structural integrity.

At the third stage of the experiment, the results of the inpainting of specific images by LaMa-Fourier and LaMa-Wavelet networks were compared. 3–5 images were selected with different complexity of background, namely, uniform background, background with structural texture (with repeating patterns) [26], complex background with repeating objects. To demonstrate how well the networks performed in texture, color, and edge recovery, each image was processed using the original LaMa-Fourier and LaMa-Wavelet network trained on 212 epochs. Inpainted images were evaluated visually and using PSNR and SSIM.

5 RESULTS

At the first stage of the experiment, the results of training of LaMa-Fourier and LaMa-Wavelet networks were evaluated using the FID score on training and validation sets. In addition, the training epoch time and image inpainting time were estimated. Image inpainting time is averaged for a set of 25 images of size 1024x1024 pixels (Table 1).

Table 1 – The LaMa-Fourier and LaMa-Wavelet training results

CNN	FID on training set	FID on validation set	Epoch time, minutes	Image inpainting time, seconds
LaMa-Fourier	8.2	25	40	2.2
LaMa-Wavelet	9.2	32	150	6.6

At the second stage of the experiment, the dependencies of the FID, PSNR, and SSIM from the size of the missing areas were researched (Table 2).

The PSNR of the CelebA-HQ [27] and Plases2 [23] datasets images inpainted by the methods known from the literature are given in Table 3 [28]. Note, however, that the results of Table 3 were obtained under significantly different experimental conditions and are used as collating data.

Table 2 – The LaMa-Fourier and LaMa-Wavelet testing results

CNN	Epochs	FID	PSNR	SSIM
Narrow masks				
LaMa-Fourier	128	21.8	25.68	0.7811
LaMa-Wavelet	128	24.7	26.58	0.8066
LaMa-Wavelet	212	21.3	26.82	0.8088
Medium masks				
LaMa-Fourier	128	24.3	25.04	0.8232
LaMa-Wavelet	128	31.1	25.88	0.8367
LaMa-Wavelet	212	24.8	26.19	0.8394
Large masks				
LaMa-Fourier	128	32.7	22.16	0.7857
LaMa-Wavelet	128	39.7	22.96	0.7973
LaMa-Wavelet	212	32.1	23.48	0.7999

Table 3 – The PSNR of images from the CelebA-HQ [27] and Plases2 [23] datasets inpainted by the methods known from the literature [28]

CNN, reference, publication year	CelebA-HQ images 256x256 pixels		
	Narrow masks	Medium masks	Large masks
LaMa-Fourier [4], 2021	22.7	34.1	27.8
CoModGAN [14], 2021	35.9	48.4	64.4
DeepFill v2 [13], 2019	37.0	45.3	43.0
EdgeConnect [12], 2019	29.2	40.5	34.7
	Places images 512x512 pixels		
LaMa-Fourier [4], 2021	12.7	11.7	12.0
CoModGAN [14], 2021	16.3	12.4	10.4
DeepFill v2 [13], 2019	17.9	18.3	22.1
EdgeConnect [12], 2019	18.9	21.9	30.5

At the third stage of the experiment, the results of the inpainting of specific images by the LaMa-Fourier and LaMa-Wavelet networks were obtained (Fig. 4–9). The corresponding values of the SSIM and PSNR are shown in Table 4.

Table 4 – The PSNR and SSIM of specific images inpainted by LaMa-Fourier and LaMa-Wavelet

Image	LaMa-Fourier		LaMa-Wavelet	
	PSNR	SSIM	PSNR	SSIM
Fig. 4, a	35.04	0.90	35.30	0.90
Fig. 4, d	34.58	0.87	35.82	0.89
Fig. 4, g	24.54	0.90	35.10	0.90
Fig. 4, j	38.20	0.98	40.01	0.99
Fig. 4, m	22.91	0.92	22.52	0.91
Fig. 4, p	32.88	0.97	33.19	0.96
Fig. 5, a	28.84	0.96	28.96	0.95
Fig. 5, d	24.89	0.94	24.78	0.93
Fig. 6, a	23.67	0.88	22.96	0.88
Fig. 6, d	19.43	0.78	22.07	0.72
Fig. 7, a	20.39	0.88	22.39	0.87
Fig. 7, d	19.59	0.88	19.47	0.87
Fig. 8	13.61	0.84	15.03	0.85

6 DISCUSSIONS

Analysing Table 1 it should be noted the follow. The LaMa-Wavelet network requires more time for training and image inpainting after training. Namely, the LaMa-Wavelet network requires 3.75 times more time per training epoch as compared with LaMa-Fourier. After training it takes 3 times longer to inpaint one image with LaMa-Wavelet than with LaMa-Fourier network. It is noticed that a significant part of the training time is spent on calculating the wavelet transform.

DOI 10.15588/1607-3274-2024-1-19

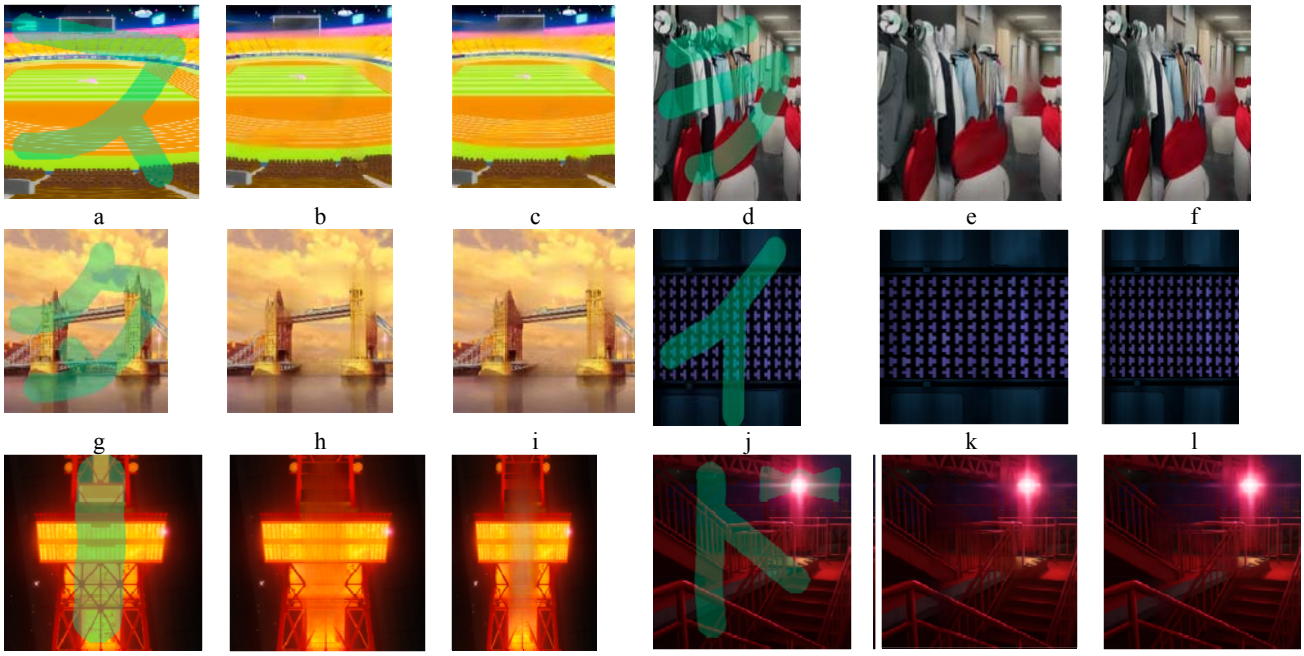


Figure 4 – Images with random masks: a, d, g, j, m, p – original image and mask; b, e, h, k, n, q – image inpainted with LaMa-Fourier; c, f, i, l, o, r – image inpainted with LaMa-Wavelet

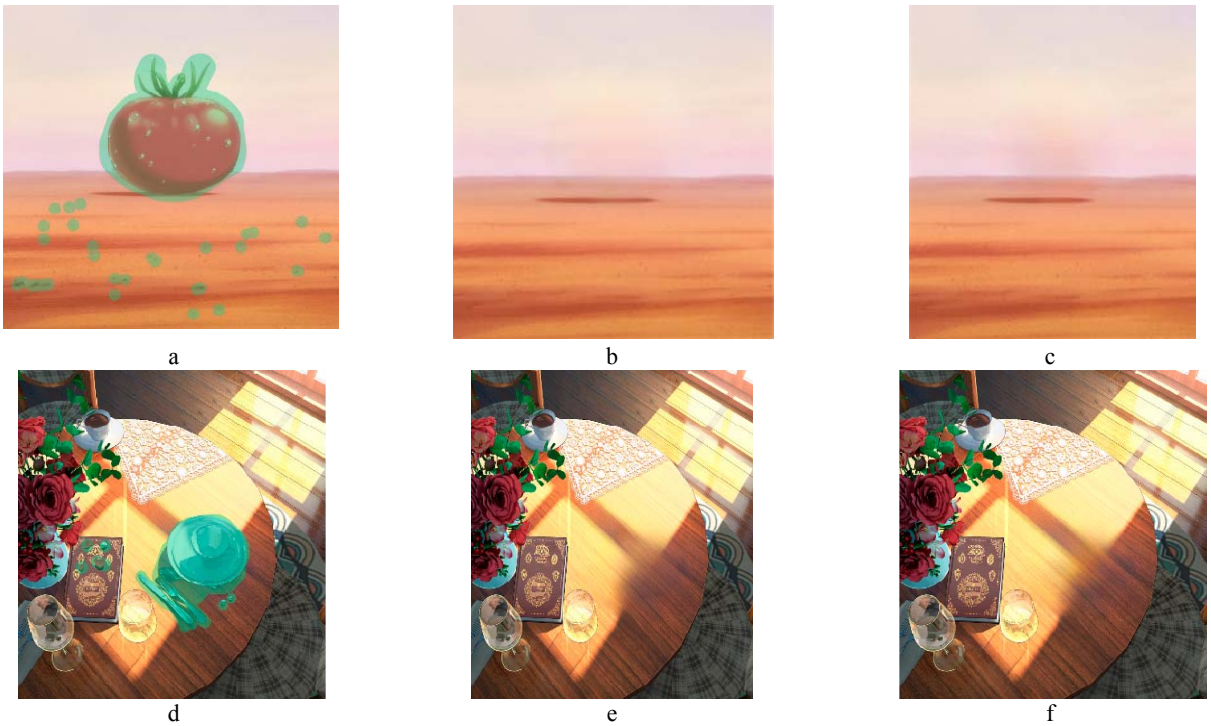


Figure 5 – Images with homogeneous background: a, d – original image and mask; b, e – image inpainted with LaMa-Fourier; c, f – image inpainted with LaMa-Wavelet

DOI 10.15588/1607-3274-2024-1-19

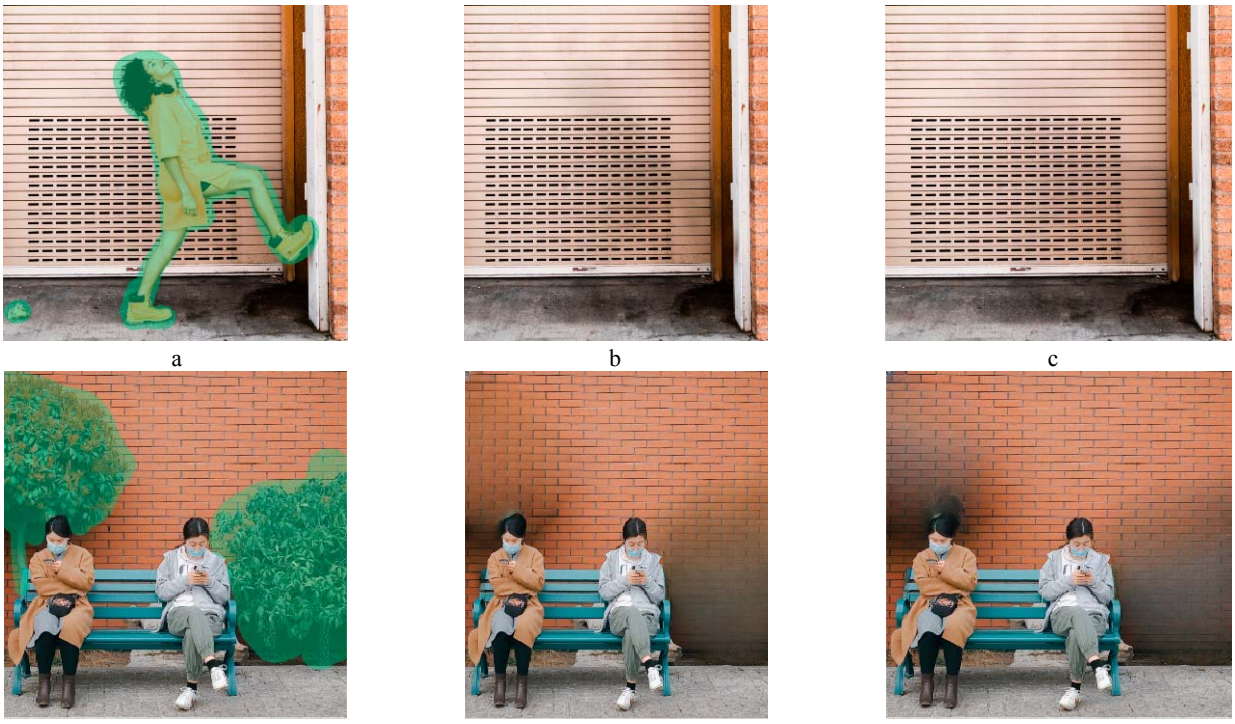


Figure 6 – Images with periodical background: a, d – original image and mask; b, e – image inpainted with LaMa-Fourier; c, f – image inpainted with LaMa-Wavelet

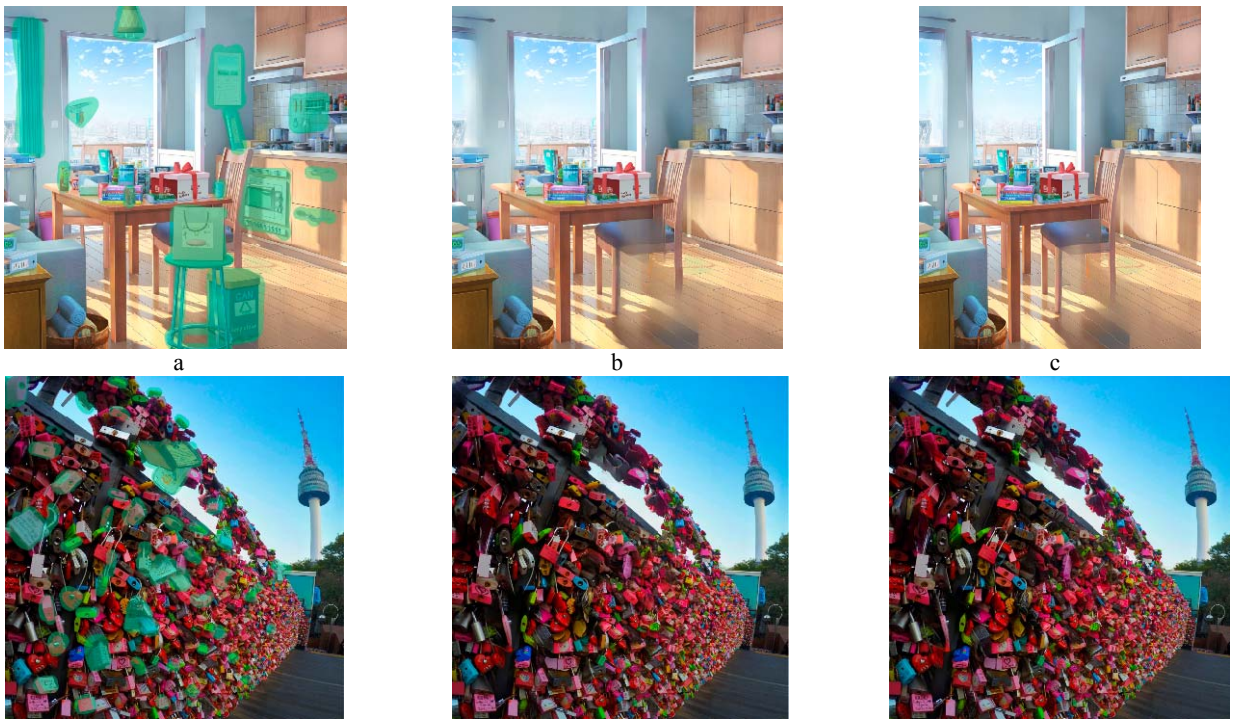
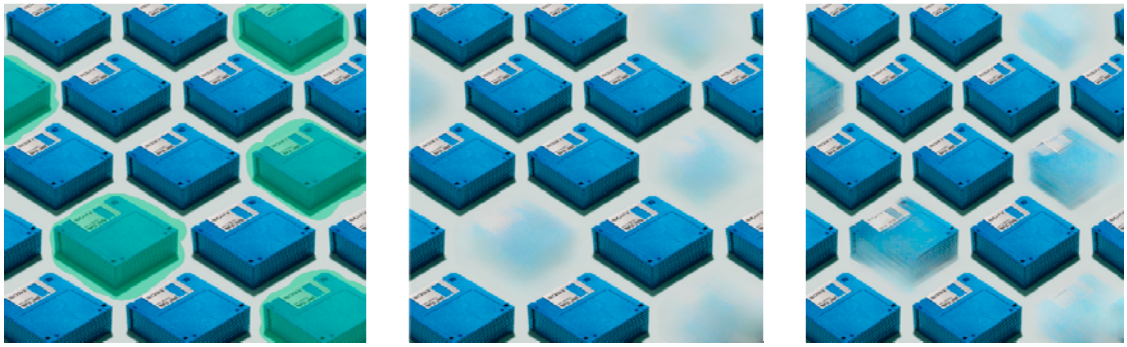


Figure 7 – Images with complex background: a, d – original image and mask; b, e – image inpainted with LaMa-Fourier; c, f – image inpainted with LaMa-Wavelet

DOI 10.15588/1607-3274-2024-1-19



a b c

Figure 8 – Image with generation artifacts: a – original image and mask; b – image inpainted with LaMa-Fourier; c – image inpainted with LaMa-Wavelet

The LaMa-Wavelet network is worse in terms of the training quality than LaMa-Fourier. Specifically, the FID on training and validation sets of images has increased by 12% and 28% respectively.

The analysis of the Table 2 showed that FID values obtained by the LaMa-Fourier and LaMa-Wavelet networks on test set of images are similar for narrow, medium, and large masks. Thus LaMa-Wavelet shows higher generalization ability than LaMa-Fourier. The PSNR of images inpainted using the LaMa-Wavelet exceeds the results obtained using the LaMa-Fourier network for narrow and medium masks in average by 1.15 dB (4.5%), for large masks in average by 1.3 dB (6%). The LaMa-Wavelet can enhance SSIM in average by 2–4% depending on a mask size.

In addition, it was noted that the quality improvement of inpainted images is largely determined by their content and properties. Thus, for structural textures [26] or objects against the background of such textures, the PSNR of images inpainted using the LaMa-Wavelet exceeds the results obtained using the LaMa-Fourier by 1.2–2.7 dB (3.5–14%). For objects on uniform background, the PSNR of images reconstructed with the LaMa-Wavelet is improved by 1.8–9.6 dB (9.8–41%) compared to the results obtained using the LaMa-Fourier. The improvement range of SSIM is less. This may mean that the LaMa-Wavelet is more capable of restoring fine details and edges of objects than inpainted the image structure.

Analysing of the LaMa-Wavelet network performance for different mask sizes it was noticed that there is a tendency for higher quality of inpainting of images with a low number of details and straight lines, especially with narrow masks. The inpainting of complex textures such as grass, leaves, branches or large numbers of people is difficult for both LaMa-Fourier and LaMa-Wavelet networks. The deformation of the image color in the large filled regions appears more frequently if the LaMa-Wavelet network has been used.

In the case of narrow masks, both networks show similar quality. However the LaMa-Wavelet shows a significant improvement in the inpainting of large missing areas relative to the LaMa-Fourier.

Let compares the results of the inpainting of specific images by the LaMa-Fourier and LaMa-Wavelet networks. At first several images with randomly selected narrow and medium masks were visually analyzed (Fig. 4). It was once again confirmed that, in general, LaMa-Wavelet shows better texture inpainting than LaMa-Fourier, but when the size of the masks increases, color deformation begins to appear. Also, the LaMa-Wavelet combines parallel lines less if they are close, and better preserves the bends of curved lines.

For images with a uniform background (Fig. 5) the numeric estimates are similar. There is a more pronounced background color defect at the place of the tomato and cup if the LaMa-Wavelet is applied.

The repeating background in the images (Fig. 6) is well reproduced by both networks. But when using larger masks, you can see, both visually and numerically, that LaMa-Wavelet generates better edges of texture elements, but loses in the generation of image color.

As the complexity of the image background increases, it can be seen that the LaMa-Wavelet network begins to lack context for the correct estimation of the expected generation (Fig. 7). In this regard, the quality of the edges of the inpainted regions reduces. However, under conditions of high complexity of the texture, the possibility of visual assessment is lost, since objects become, in principle, difficult to distinguish even for a person. Also, in such a scenario, the ability of the LaMa-Wavelet network to continue repeating textures becomes more of a problem than an advantage. This network starts mixing different textures in an attempt to continue them. This is seen in the example in Figure 8, where large duplicate objects need to be removed from an image. Instead of removing objects, the LaMa-Wavelet network generates an average between the background and the original texture in an attempt to restore the texture. It can also be seen that the LaMa-Wavelet network is more seek to forced inpainting of textures than the LaMa-Fourier.

Thus, analysis of specific image inpainting confirmed the practical effectiveness of LaMa-Wavelet network as compared with LaMa-Fourier. In particular, when removing objects on a homogeneous background, both networks show similar results. However, on complex backgrounds with repeating elements, the LaMa-Wavelet

DOI 10.15588/1607-3274-2024-1-19

is often more effective in restoring textures, despite some cases of texture mixing.

CONCLUSIONS

The actual scientific and applied problem of an inpainting of the fine details and object edges has been solved when missing regions of images are filled by CNN.

The scientific novelty is the proposed method of natural image inpainting with LaMa-Wavelet network. Due to applying wavelet transform, the image inpainting with large masks based on the LaMa network is improved. Specifically, the quality of reconstruction of image edges and fine details is increased.

The practical significance of obtained results is that the software realizing the proposed LaMa-Wavelet network is developed, as well as experiments to research its image inpainting performance are conducted. The experimental results allow to recommend the proposed LaMa-Wavelet for use in practice, as well as to determine effective conditions for the application of this network.

Prospects for further research is reducing the computing time by using fast transforms. It is also nessesary to identify classes of images for the inpainting of which it is advisable to use LaMa-Wavelet.

ACKNOWLEDGEMENTS

The authors express their deep gratitude to V. N. Krylov, Doctor of Technical Sciences, Professor of the Department of Applied Mathematics and Information Technologies, National University “Odessa Polytechnic” for valuable and constructive advice and comments while working on this paper.

REFERENCES

1. Ma Y., Liu X., Bai S. et al. Region-wise generative adversarial image inpainting for large missing areas, *IEEE Transactions on Cybernetics*, 2023, Vol. 53, № 8, pp. 5226–5239. DOI: 10.1109/TCYB.2022.3194149.
2. Petrov K. E., Kyrychenko V. V. Removal of rain components from single images using a recurrent neural network, *Radio Electronics, Computer Science, Control*, 2023, № 2, 91–102. DOI: 10.15588/1607-3274-2023-2-10
3. Kolodochka D. O., Polyakova M. V. Comparative analysis of convolutional neural networks for filling missing image regions, *Science and education: problems, prospects and innovations: 7th International Scientific and Practical Conference, Kyoto, Japan, 1–3 April, 2021: proceedings.* CPN Publishing Group, 2021, pp. 562–570.
4. Suvorov R., Logacheva E., Mashikhin A. et al. Resolution-robust large mask inpainting with Fourier convolutions, *Applications of Computer Vision: IEEE Workshop/Winter Conference, WACV, Waikoloa, Hawaii, 4–8 January, 2022 : proceedings*. IEEE, 2022, pp. 2149–2159. DOI: 10.1109/WACV51458.2022.00323
5. Leoshchenko S. D., Oliynyk A. O., Subbotin S. O., Hoffman E. O., Kornienko O. V. Method of structural adjustment of neural network models to ensure interpretability, *Radio electronics, computer science, management*, 2021, № 3, pp. 86–96. DOI: 10.15588/1607-3274-2021-3-8
6. Polyakova M. V. RCF-ST: Richer Convolutional Features network with structural tuning for the edge detection on natural images, *Radio electronics, computer science, management*, 2023, № 4, pp. 122–134. DOI: 10.15588/1607-3274-2023-4-12
7. Pathak D., Krahenbuhl P., Donahue J., Darrell T., Efros A. A. Context encoders: feature learning by inpainting, *Computer Vision and Pattern Recognition: IEEE Conference, CVPR, Las Vegas, NV, USA, 27–30 June, 2016 : proceedings.* IEEE, 2016, pp. 2536–2544. DOI: 10.1109/CVRP.2016.278
8. Liu G., Reda F. A., Shih K. J. et al. Image inpainting for irregular holes using partial convolutions, *Computer Vision: European Conference, ECCV, Munich, Germany, 8–14 September, 2018 : proceedings.* IEEE: 2018, pp. 85–100. DOI: 10.1007/978-3-030-01252-6_6
9. Yang C., Lu X., Lin Z. et al. High-resolution image inpainting using multi-scale neural patch synthesis, *Computer Vision and Pattern Recognition: IEEE Conference, CVPR, Honolulu, HI, USA, 21–26 July 2017 : proceedings.* IEEE, 2017, pp. 6721–6729. DOI: 10.1109/CVPR.2017.434
10. Iizuka S., Simo-Serra E., Ishikawa H. Globally and locally consistent image completion, *ACM Transactions on Graphics*, 2017, Vol. 36, № 4, pp. 107:1. DOI: 10.1145/3072959.3073659
11. Yu J., Yang J., Shen X., Lu X., Huang T. S. Generative image inpainting with contextual attention. *Computer Vision and Pattern Recognition Workshops: IEEE/CVF Conference, CVPRW, Salt Lake City, UT, USA, 18–22 June, 2018 : proceedings.* IEEE, 2018, pp. 5505–5514. DOI: 10.1109/CVPRW.2018.00577
12. Nazeri K., Ng E., Joseph T., Qureshi F., Ebrahimi M. EdgeConnect: structure guided image inpainting using edge prediction, *Computer Vision Workshop: IEEE/CVF International Conference, ICCVW, Seoul, Korea (South), 27–28 October, 2019 : proceedings.* IEEE, 2019, pp. 2462–2468. DOI: 10.1109/ICCVW.2019.00408
13. Yu J., Lin Z., Yang J. et al. Free-form image inpainting with gated convolution, *Computer Vision: IEEE/CVF International Conference, ICCV, Seoul, Korea (South), 27 October – 2 November, 2019 : proceedings.* IEEE, 2019, pp. 4471–4480. DOI: 10.1109/ICCV.2019.00457
14. Zhao S., Cui J., Sheng Y. et al. Large scale image completion via co-modulated generative adversarial networks, *Learning Representations: International Conference, ICLR, Vienna, Austria, 4 May 2021: processings* [Electronic resource]. Access mode: https://arxiv.org/abs/2103.10428. DOI: 10.48550/arXiv.2103.10428
15. Gonzalez R. C., Woods R. E. Digital Image Processing. NY, Pearson, 4th Edition, 2017, 1192 p.
16. Daubechies I. Ten Lectures on Wavelets, Philadelphia, SIAM Press, 1992, 352 p.
17. Li Q., Shen L., Guo S., Lai Z. WaveCNet: wavelet integrated CNNs to suppress aliasing effect for noise-robust image classification, *IEEE Transactions on Image Processing*, 2021, Vol. 30, pp. 7074–7089. DOI:10.1109/TIP.2021.3101395
18. Liu P., Zhang H., Zhang K., Lin L., Zuo W. Multi-level Wavelet-CNN for image restoration, *Computer Vision and Pattern Recognition Workshops: IEEE/CVF Conference, CVPRW, Salt Lake City, UT, USA, 18 – 22 June, 2018 : proceedings*. IEEE, 2018, pp. 2149–2159. DOI: 10.1109/CVPRW.2018.00121

DOI 10.15588/1607-3274-2024-1-19

19. Bobulski J. Multimodal face recognition method with two-dimensional hidden Markov model, *Bulletin of the Polish Academy of Sciences, Technical Sciences*, 2017, Vol. 65, № 1, pp. 121–128. DOI: 10.1515/bpasts-2017-0015
20. Sara U., Akter M., Uddin M. S. Image quality assessment through FSIM, SSIM, MSE and PSNR – a comparative study, *Journal of Computer and Communications*, 2019, Vol. 7, № 3, pp. 8–18. DOI: 10.4236/jcc.2019.73002
21. Johnson J., Alahi A., Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. *Computer Vision – ECCV 2016. Lecture Notes in Computer Science / Leibe, B., Matas, J., Sebe, N., Welling, M. (eds)*. Springer, Cham, 2016, Vol. 9906, pp. 694–711. DOI: 10.1007/978-3-319-46475-6_43
22. Wang T.-C., Liu M.-Y., Zhu J.-Y. et al. High-resolution image synthesis and semantic manipulation with conditional GANs, *Computer Vision and Pattern Recognition: IEEE Conference, CVPR, Salt Lake City, UT, USA, 18–23 June, 2018 : proceedings*. IEEE, 2018, pp. 8798–8807. DOI: 10.1109/CVPR.2018.00917
23. Places365 Scene Recognition Demo [Electronic resource]. Access mode: http://places2.csail.mit.edu/
24. Safebooru [Electronic resource]. Access mode: https://safebooru.org/index.php?page=post&s=list&tags=no_humans+landscape
25. Heusel M., Ramsauer H., Unterthiner T., Nessler B., Hochreiter S. GANs trained by a two time-scale update rule converge to a local nash equilibrium, *Neural Information Processing Systems: 31st Annual Conference, NIPS, Long Beach, California, USA, 4–9 December, 2017 : proceedings.* Neural Information Processing Systems Foundation, Inc., 2017, pp. 6629–6640. DOI: 10.18034/ajase.v8i1.9
26. Polyakova M. V., Krylov V. N., Ishchenko A. V. Elaboration of the transform with generalized comb scaling and wavelet functions for the image segmentation, *Eastern-European Journal of Enterprise Technologies*, 2014, Vol. 2, № 2 (71), pp. 33–37. DOI: 10.15587/1729-4061.2014.27791
27. CelebA-HQ [Electronic resource]. Access mode: https://paperswithcode.com/dataset/celeba-hq
28. Supplementary material [Electronic resource]. Access mode: https://bit.ly/3zhv2rD/lama_supmat_2021.pdf

Received 11.01.2024.
Accepted 28.02.2024.

УДК 004.93

LAMA-WAVELET: РЕКОНСТРУКЦІЯ ЗОБРАЖЕНЬ З ВИСОКОЮ ЯКІСТЮ ВІДНОВЛЕННЯ ДЕТАЛЕЙ І КРАЇВ ОБ'ЄКТІВ

Колодочка Д. О. – студент Інституту комп'ютерних систем Національного університету «Одеська політехніка», Одеса, Україна.

Полякова М. В. – д-р техн. наук, доцент, професор кафедри прикладної математики та інформаційних технологій Національного університету «Одеська політехніка», Одеса, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто проблему реконструкції зображень в системах комп'ютерної графіки та комп'ютерного зору. Предметом дослідження є згорткові нейронні мережі глибокого навчання для реконструкції зображень.

Мета роботи. Покращення якості реконструйованих зображень в системах комп'ютерного зору та комп'ютерної графіки шляхом застосування вейвлет-перетворення в архітектурі нейронної мережі LaMa-Fourier.

Метод. Базова мережа LaMa-Fourier окремо обробляє глобальний та локальний контекст зображення. Пропонується вдосконалити для цієї мережі блок обробки глобального контексту зображення, а саме блок спектрального перетворення. Для цього замість Fourier Unit Structure використовується розроблений авторами Simple Wavelet Convolution Block, у якому спочатку виконується тривимірне вейвлет-перетворення зображення на двох рівнях. Отримані коефіцієнти розбиваються так, що кожна субполоса представляє окрему ознаку зображення. Згортковий шар, пакетна нормалізація та функція активації ReLU послідовно застосовуються до субполос коефіцієнтів на кожному рівні вейвлет-перетворення. Отримані субполоси вейвлет-коефіцієнтів конкатенуються і до них застосовується зворотне вейвлет-перетворення, результат якого передається на вихід блоку. Окрема обробка вейвлет-коефіцієнтів на різних рівнях зменшує обчислювальну складність, зберігаючи при цьому вплив контексту кожного рівня на реконструкцію зображення. Отриману нейронну мережу названо LaMa-Wavelet. Показники FID, PSNR, SSIM та візуальний аналіз були використані для оцінки якості зображень, реконструйованих мережею LaMa-Wavelet.

Результати. Запропоновану мережу LaMa-Wavelet програмно реалізовано та досліджено для вирішення проблеми реконструкції зображень. PSNR зображень, відновлених за допомогою мережі LaMa-Wavelet, перевищує результати, отримані за допомогою мережі LaMa-Fourier для малих і середніх масок у середньому на 4,5%, для великих масок – у середньому на 6%. Застосування LaMa-Wavelet може збільшити SSIM на 2–4% залежно від розміру маски. Але реконструкція одного зображення за допомогою LaMa-Wavelet займає в 3 рази більше часу, ніж за допомогою мережі LaMa-Fourier. Аналіз конкретних зображень демонструє, що обидві мережі показують схожі результати реконструкції однорідного фону. На складних фонах із повторюваними елементами LaMa-Wavelet часто ефективніше відновлює текстури.

Висновки. Отримана мережа LaMa-Wavelet дозволяє покращити відновлення великих областей зображень за рахунок застосування вейвлет-перетворення в архітектурі мережі LaMa. А саме, підвищується якість реконструкції країв зображення та дрібних деталей.

КЛЮЧОВІ СЛОВА: реконструкція зображення, вейвлет-перетворення, мережа LaMa, вейвлет Добеші, початкова відстань Фреше, вейвлет-згортка.

ЛІТЕРАТУРА

1. Region-wise generative adversarial image inpainting for large missing areas / [Y. Ma, X. Liu, S. Bai et al.] // IEEE Transactions on Cybernetics. – 2023. – Vol. 53, № 8. – P. 5226–5239. DOI: 10.1109/TCYB.2022.3194149.
2. Петров К. Е. Видалення компонентів дощу з одиночних зображень з використанням рекурентної нейронної

DOI 10.15588/1607-3274-2024-1-19

мережі / К. Е. Петров, В. В. Кириченко // Радіоелектроніка, інформатика, управління. – 2023. – № 2. – С. 91–102. DOI: 10.15588/1607-3274-2023-2-10
3. Колодочка Д. О. Порівняльний аналіз згорткових нейронних мереж для реконструкції відсутніх областей зображень / Д. О. Колодочка, М. В. Полякова // Science and education: problems, prospects and innovations: 7th International Scientific and Practical Conference, Kyoto, Japan, 1–3 April, 2021: proceedings. – CPN Publishing Group, 2021. – P. 562–570.
4. Resolution-robust large mask inpainting with Fourier convolutions / [R.Suvorov, E. Logacheva, A. Mashikhin et al.] // Applications of Computer Vision: IEEE Workshop/Winter Conference, WACV, Waikoloa, Hawaii, 4 – 8 January, 2022 : proceedings. – IEEE, 2022. – P. 2149–2159. DOI: 10.1109/WACV51458.2022.00323
5. Метод структурного доналаштування нейромережевих моделей для забезпечення інтерпретабельності / [С. Д. Леощенко, А. О. Олійник, С. О. Субботін та ін.] // Радіоелектроніка, інформатика, управління. – 2021. – № 3. – С. 86–96. DOI: 10.15588/1607-3274-2021-3-8
6. Polyakova M. V. RCF-ST: Richer Convolutional Features network with structural tuning for the edge detection on natural images / M. V. Polyakova // Radio electronics, computer science, management. – 2023. – № 4. – P. 122–134. DOI: 10.15588/1607-3274-2023-4-12
7. Context encoders: feature learning by inpainting / [D. Pathak, P. Krahenbuhl, J. Donahue et al.] // Computer Vision and Pattern Recognition: IEEE Conference, CVPR, Las Vegas, NV, USA, 27–30 June, 2016 : proceedings. – IEEE, 2016. – P. 2536–2544. DOI: 10.1109/CVRP.2016.278
8. Image inpainting for irregular holes using partial convolutions / [G. Liu, F. A. Reda, K. J. Shih et al.] // Computer Vision: European Conference, ECCV, Munich, Germany, 8 – 14 September, 2018 : proceedings. – IEEE: 2018. – P. 85–100. DOI: 10.1007/978-3-030-01252-6_6
9. High-resolution image inpainting using multi-scale neural patch synthesis / [C. Yang, X. Lu, Z. Lin et al.] // Computer Vision and Pattern Recognition: IEEE Conference, CVPR, Honolulu, HI, USA, 21–26 July 2017 : proceedings. – IEEE, 2017. – P. 6721–6729. DOI: 10.1109/CVPR.2017.434
10. Iizuka S. Globally and locally consistent image completion / S. Iizuka, E. Simo-Serra, H. Ishikawa // ACM Transactions on Graphics. – 2017. – Vol. 36, № 4. – P. 107:1. DOI: 10.1145/3072959.3073659
11. Generative image inpainting with contextual attention / [J. Yu, J. Yang, X. Shen et al.] // Computer Vision and Pattern Recognition Workshops: IEEE/CVF Conference, CVPRW, Salt Lake City, UT, USA, 18–22 June, 2018 : proceedings. – IEEE, 2018. – P. 5505–5514. DOI: 10.1109/CVPRW.2018.00577
12. Edge Connect: structure guided image inpainting using edge prediction / [K. Nazeri, E. Ng, T. Joseph et al.] // Computer Vision Workshop: IEEE/CVF International Conference, ICCVW, Seoul, Korea (South), 27–28 October, 2019 : proceedings. – IEEE, 2019. – P. 2462–2468. DOI: 10.1109/ICCVW.2019.00408
13. Free-form image inpainting with gated convolution / [J. Yu, Z. Lin, J. Yang et al.] // Computer Vision: IEEE/CVF International Conference, ICCV, Seoul, Korea (South), 27 October – 2 November, 2019 : proceedings. – IEEE, 2019. – P. 4471 – 4480. DOI: 10.1109/ICCV.2019.00457
14. Large scale image completion via co-modulated generative adversarial networks / [S. Zhao, J. Cui, Y. Sheng et al.] // Learning Representations: International Conference, ICLR, Vienna, Austria, 4 May 2021: processings [Electronic resource]. – Access mode: https://arxiv.org/abs/2103.10428. DOI: 10.48550/arXiv. 2103.10428
15. Gonzalez R. C. Digital Image Processing / R. C. Gonzalez, R. E. Woods. – NY : Pearson, 4th Edition, 2017. – 1192 p.
16. Daubechies I. Ten Lectures on Wavelets / I. Daubechies. – Philadelphia, SIAM Press, 1992. – 352 p.
17. WaveCNet: wavelet integrated CNNs to suppress aliasing effect for noise-robust image classification / [Q. Li, L. Shen, S. Guo, Z. Lai] // IEEE Transactions on Image Processing. – 2021. – Vol. 30. – P. 7074–7089. DOI:10.1109/TIP.2021.3101395
18. Multi-level Wavelet-CNN for image restoration / [P. Liu, H. Zhang, K. Zhang et al.] // Computer Vision and Pattern Recognition Workshops: IEEE/CVF Conference, CVPRW, Salt Lake City, UT, USA, 18 – 22 June, 2018 : proceedings. – IEEE, 2018. – P. 2149–2159. DOI: 10.1109/CVPRW.2018.00121
19. Bobulski J. Multimodal face recognition method with two-dimensional hidden Markov model / J. Bobulski // Bulletin of the Polish Academy of Sciences, Technical Sciences. – 2017. – Vol. 65, № 1. – P. 121–128. DOI: 10.1515/bpasts-2017-0015
20. Sara U. Image quality assessment through FSIM, SSIM, MSE and PSNR – a comparative study / U. Sara, M. Akter, M. S. Uddin // Journal of Computer and Communications. – 2019. – Vol. 7, № 3. – P. 8–18. DOI: 10.4236/jcc.2019.73002
21. Johnson J. Perceptual losses for real-time style transfer and super-resolution / J. Johnson, A. Alahi, L. Fei-Fei // Computer Vision – ECCV 2016. Lecture Notes in Computer Science / Leibe, B., Matas, J., Sebe, N., Welling, M. (eds). – Springer, Cham, 2016. – Vol. 9906. – P. 694–711. DOI: 10.1007/978-3-319-46475-6_43
22. High-resolution image synthesis and semantic manipulation with conditional GANs / [T.-C. Wang, M.-Y. Liu, J.-Y. Zhu et al.] // Computer Vision and Pattern Recognition: IEEE Conference, CVPR, Salt Lake City, UT, USA, 18–23 June, 2018 : proceedings. – IEEE, 2018. – P. 8798–8807. DOI: 10.1109/CVPR.2018.00917
23. Places365 Scene Recognition Demo [Electronic resource]. – Access mode: http://places2.csail.mit.edu/
24. Safebooru [Electronic resource]. – Access mode: https://safebooru.org/index.php?page=post&s=list&tags=no_humans+landscape
25. GANs trained by a two time-scale update rule converge to a local nash equilibrium / [M. Heusel, H. Ramsauer, T. Unterthiner et al.] // Neural Information Processing Systems: 31st Annual Conference, NIPS, Long Beach, California, USA, 4–9 December, 2017 : proceedings. – Neural Information Processing Systems Foundation, Inc., 2017. – P. 6629–6640. DOI: 10.18034/ajase.v8i1.9
26. Polyakova M. V. Elaboration of the transform with generalized comb scaling and wavelet functions for the image segmentation / M. V. Polyakova, V. N. Krylov, A. V. Ishchenko // Eastern-European Journal of Enterprise Technologies. – 2014. – Vol. 2, № 2 (71). – P. 33–37. DOI: 10.15587/1729-4061.2014.27791
27. CelebA-HQ [Electronic resource]. – Access mode: https://paperswithcode.com/dataset/celeba-hq
28. Supplementary material [Electronic resource]. – Access mode: https://bit.ly/3zhv2rD/lama_supmat_2021.pdf

DOI 10.15588/1607-3274-2024-1-19