

## QUALITY ASSESSMENT OF MODELS AND DEEP LEARNING METHODS FOR SUPER-RESOLUTION IMAGE FORMATION

N. NEDASHKOVSKAYA, A. LANKO

**Abstract.** This article examines evaluation metrics for the results of super-resolution image generation in solving the SISR task. The study comprises two experiments: the implementation of custom network architectures for SRGAN, VDSR, and SRCNN, and fine-tuning of pre-trained SRGAN, VDSR, and SRCNN models. An algorithm for assessing the quality of models and deep learning methods for generating super-resolution images is suggested. The VDSR model performed best in terms of pixel, structural, and perceptual metrics, as well as training time and visual confirmation by a human, highlighting that residual learning is more effective than recursive learning under the conditions of the two conducted experiments. Threshold values for practically acceptable and high-quality results were determined through visual analysis of many generated images and their corresponding quality metrics, including those reported by other researchers.

**Keywords:** single image super-resolution, quality assessment, generative models, deep learning methods, convolutional neural network, residual learning, recursive learning, fine-tuning of pre-trained models, perceptual metric, LPIPS, multicriteria decision analysis, DIV2K dataset, thresholds for practically acceptable and high-quality generated images.

### INTRODUCTION

The task of Single Image Super-Resolution (SISR) involves the formation of highly detailed versions of low-resolution images [1]. Despite significant progress in modern imaging technologies, this task remains relevant due to such factors as image quality deterioration after transmission through communication channels and hardware failures, image compression for compact storage on data carriers, and the inability to use professional equipment in certain natural conditions.

The goal of SISR methods is to create high-quality images by restoring or adding details missing in the original low-resolution images. To achieve this, generative models and deep learning methods are used [2].

Generative models form new parts by simulating the data distribution in the training selection [2]. Among them, the most common for SISR are modifications of generative adversarial networks (GAN); diffusion models are more complex and efficient, the use of streaming models and autoencoders is also known [3].

Deep learning methods analyze important features of training images to reconstruct image details [2]. These include convolutional neural networks (CNN), recurrent neural networks (RNN), and residual neural networks (ResNet) [3]. It is important to note that they are often part of architecture of generative models that implement a particular learning principle. For example, the generator and discriminator in a GAN are deep neural networks.

SISR models are trained by learning pairs of low- and high-resolution images from the training selection. The effectiveness of super-resolution image gen-

eration is assessed based on a set of indicators, which must include both quantitative and perceptual metrics. An important step in evaluating the results of SISR is the visual analysis of the generated images by a human.

It should be noted that SISR algorithms are complex and time-consuming, so they require powerful computing resources, and model optimization is still the main focus of researchers' work on this topic. That is why, when choosing the optimal model, technical indicators are added to the evaluation criteria, including time of training, training cost, and the availability of a hardware accelerator in the form of a graphics processing unit (GPU) [4].

## PROBLEM STATEMENT

Let us introduce the notation  $H$  for height,  $W$  for width, and  $C$  for the number of image channels (e.g. RGB). Let  $I_{LR} \in R^{H \times W \times C}$  be a low-resolution image, and  $I_{HR} \in R^{H \times W \times C}$  be its corresponding high-resolution image. The goal of the SISR problem is to find the following mapping

$$f: I_{LR} \rightarrow I_{HR}, \quad (1)$$

that will ensure the most accurate recovery of the details of the  $I_{HR}$  image based on the information from the  $I_{LR}$ .

Mapping (1) is a formalization tool, as it can describe different processes depending on the resolution enhancement method. That is why we will further consider the implementation of (1), the model  $f_{\theta} \in F$ , where  $\theta$  are the model parameters,  $F$  is the set of all SISR models. The target super-resolution image is the output of  $f_{\theta}$  and the result of solving the problem:

$$I_{SR} = f_{\theta}(I_{LR}).$$

An important step in the process of training models from  $F$  is to solve the optimization problem

$$\min_{\theta} L(I_{HR}, I_{SR}),$$

where  $L(I_{HR}, I_{SR})$  is the model loss function. The objective is to find such model parameters  $\theta$  that the value of the loss function  $L$  is minimal.

In this paper, the task of multicriteria quality assessments of images generated (formed) by different models and deep learning methods is set. Let  $A = \{a_i \mid i = 1, 2, \dots, n\}$  be a set of super-resolution images  $I_{SR}$ , generated by different deep learning models based on a single low-resolution image  $I_{LR}$ ;  $C = \{c_j \mid j = 1, 2, \dots, m\}$  be a set of quality criteria for the generated images and technical characteristics of model training. In the following,  $a_i$  will be considered as alternatives, and  $c_j$  as decision criteria.

The task is to find the aggregated or global weights

$$W^{aggr} = \{w_i^{aggr} \mid i = 1, 2, \dots, n\} \quad (2)$$

of alternative generated (formed) images according to a set of criteria from  $C$  and selection of the best generated image.

The quality criteria for the generated images are:

- traditional quantitative metrics PSNR [5], SSIM [6], MSSIM [6] (1<sup>st</sup> group of criteria);
- perceptual indicators BRISQUE [7], NIQE [8], PIQUE [9], LPIPS [10] and their modifications (e.g., LR-PSNR) (2<sup>nd</sup> group).

The decision criteria also include technical characteristics (3<sup>rd</sup> group):

- training time and cost;
- availability of a hardware accelerator in the form of a graphics processing unit (GPU).

The purpose of the studied generative models and deep learning methods is to increase the resolution of images, scale them by 4, 8, or more times, and generate realistic and beautiful images based on a given low-resolution image for further display of the generated images on large screens and human perception. Therefore, another group of criteria (4<sup>th</sup> group) ensures that the generated image is evaluated directly by a human: effects of smoothing, blurring, edge lightening, and photorealism of the image.

The coefficients of relative importance of decision criteria are determined by decision support methods [11–13] using expert pairwise comparison judgements depending on the application. The interdependence between individual decision criteria and the need to take into account fuzzy judgements provided by an expert require the use of hybrid methods [14; 15].

## MATERIALS AND METHODS

### Deep learning models for generating super-resolution images

The following models were used in the study, representing generative and deep learning methods.

1. **SRGAN (Super-Resolution Generative Adversarial Network)** is a generative adversarial network for increasing the resolution, where the generator creates super-resolution images, and the discriminator is trained to recognize real and generated images. The generator is optimized using a combination of loss functions: adversarial loss for plausibility and content loss for pixel accuracy. Full implementations also use a perceptual loss function to improve textures [16].

2. **VDSR (Very Deep Super Resolution)** is a very deep convolutional neural network for resolution enhancement tasks [17]. Its main advantage is usage of residual connections, which allow the model to learn from the difference between the input low-resolution image and the corresponding super-resolution image. This reduces the risk of gradient vanishing during training, accelerates convergence and increases training stability. Due to a large number of convolutional layers, VDSR effectively captures both fine textures and complex structures of objects in the image, which ensures high-quality results.

3. **DRCN (Deeply-Recursive Convolutional Network)** uses the concept of recursive blocks, where the same set of parameters is applied repeatedly. This allows for significant depth without increasing the number of model parameters, which reduces its computational complexity and memory requirements. As a result, DRCN effectively recovers the details of a high-resolution image while maintaining resource efficiency. The network also uses methods of averaging the

output results, supervised skip connections, which increase the stability and accuracy of recovery of details [18].

4. **SRCNN (Super-Resolution Convolutional Neural Network)** is a convolutional neural network for resolution enhancement that performs the following three sequential operations: interpolation of the input image to high resolution, feature extraction using convolutional layers, and reconstruction of the super-resolution image [19]. The model is simple and efficient, but limited in depth and ability to reconstruct complex textures. In this study, it is used as a discriminator in our implementation of SRGAN, as well as a separate pre-trained model in the framework of retraining experiments.

Two types of blocks were also used in the networks:

- 1) a **residual block** to maintain the stability of the gradients;
- 2) a **recursive block** that repeats convolutional layers with the same weights multiple times to enhance the selected features and create a more complex architecture.

The architecture of the implemented models [20] is shown in Table 1, and the architecture of their component blocks is further explained in Table 2.

**Table 1.** Architecture of the implemented models in-house

Model		Architecture
SRGAN	Generator SRResNet	Consists of an initial 9×9 convolutional layer, 5 residual blocks (ResidualBlock), an intermediate 3×3 convolutional block, a resolution upscaling block (2 3×3 convolutional layers with PixelShuffle), and a final 9×9 convolutional layer
	Discriminator SRCNN	Consists of 8 3×3 convolutional layers with increasing number of channels with normalization (BatchNorm2d) and LeakyReLU activation, 1 adaptive averaging layer and 2 final fully connected layers. The filter size for all convolutional layers is 3×3
VDSR		Consists of an initial convolutional layer, 18 convolutional layers with ReLU activation, and an output layer that adds the residual to the input image. The filter size for all convolutional layers is 3×3
DRCN		Consists of an input convolutional layer, a recursive block (RecursiveBlock) that is repeated a specified number of times (16), and an output convolutional layer. The filter size for all convolutional layers is 3×3

**Table 2.** Architecture of the model components

Model	Architecture
ResidualBlock	Contains 2 3×3 convolutional layers, a normalization layer (BatchNorm2d) after each convolutional layer, and a PReLU activation function after the 1st layer
RecursiveBlock	Contains 1 3×3 convolutional layer with ReLU activation

**Algorithm for training and evaluation models from scratch**

The following algorithm for training SRGAN, VDSR, and DRCN models for generating super-resolution images and evaluation of these models in terms of quantitative and perceptual indicators is suggested:

1. Splitting the set into training and validation samples. In the case of using the DIV2K set [1], this stage is skipped, since the images are already distributed in the set.

2. Initialization of model weights using the methods of Kaiming He [21] or Xavier Glaurot [22], depending on the characteristics of the model to be trained.
3. Training on a given number of epochs (200 for the generating model with a batch size of 16; and 100 epochs for deep learning methods with a batch size of 32) on the training set with tracking the values of the loss function (adversarial loss (MSE+BCE) for the generating model, MSE for deep learning methods).
4. Saving model weights in case of training interruption or early stopping.
5. Calculating the training time of models.
6. Evaluation of the results on the test sample: calculation of the quantitative indicators PSNR, SSIM, MSSIM and the perceptual indicator LPIPS of the generated images. The pre-trained VGG network19 is used to calculate the LPIPS metric. The average value of the indicators for each model is presented for 10 random images.

### Algorithm for training models using pre-training technology

An algorithm for training of pre-trained models for the formation of super-resolution images is suggested, which consists of the following steps:

1. Careful selection of a pre-trained model, which must be aimed at the same task and preferably trained on a large universal data set.
2. Loading the weights for the selected model, with the values of which training will continue.
3. Determine the number of epochs for which the model should be retrained.
4. Fine-tuning the model: freezing layers (usually the initial ones) and adding new ones which extract high-level features (residual blocks, convolutions with small kernels, normalization layers, Upsampling or PixelShuffle), using a low learning rate to ensure its stability, combining the main loss with the perceptual loss to focus on the visual quality of the generated images.
5. Applying early stopping in case of signs of model overfitting according to metrics PSNR, SSIM, MSSIM and a perceptual metric LPIPS.

The experiment on retraining of pre-trained models was conducted on 20 epochs. The purposes of the experiment are: to improve the result of image generation, as well as to check whether it is possible to obtain a result better than that of other researchers [23], and whether overfitting is occur.

### Quantitative and perceptual metrics and indicators

The quality of SISR models is traditionally evaluated based on metrics and indicators that compare the SR image generated by the model with the original HR image from a labeled test image set [24].

The classical **PSNR** (Peak Signal-to-Noise Ratio) metric has limitations for evaluating structured data such as images, as it assumes pixel independence. PSNR measures the difference between pixels of a pair of images as a ratio between the maximum possible signal strength and noise. For example, blurring an image can cause a large perceptual change and at the same time a small change in the  $L_2$  measure. SSIM [6] index assesses structural similarity of two images.

The perceptual distance estimates the similarity of high-level features of two images similar to human visual perception. Perceptual indicators such as BRISQUE [7], NIQE [8], PIQUE [9], LPIPS [10], and others have been suggested. Let us describe some of them in more detail.

**SSIM** (Structural Similarity Index Measure) evaluates the similarity of two images  $x$  and  $y$  based on three image components: brightness, contrast, and structure [6]:

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma,$$

where  $\alpha, \beta, \gamma > 0$  are the coefficients of relative importance of the three components, are the parameters.

The SSIM satisfies the symmetry properties  $SSIM(x, y) = SSIM(y, x)$ ; boundedness  $SSIM(x, y) \leq 1$ ; and unique maximum:  $SSIM(x, y) = 1$  if and only if  $x = y$ .

Later, the authors of [6] move on to a following simplified expression:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (3)$$

where  $\mu_x$  is the average image intensity value  $x$ ;  $\sigma_x$  is the standard deviation for image  $x$ , which serves as an unbiased estimate of its contrast;  $\sigma_{xy}$  is the covariance between two images  $x$  and  $y$ , which is the basis for comparing image structures after subtracting brightness and normalizing variance, and also use the following modified estimates of local statistics  $\mu_x$ ,  $\sigma_x$  та  $\sigma_{xy}$ :

$$\mu_x = \sum_{i=1}^N v_i x_i, \quad \sigma_x = \left( \sum_{i=1}^N v_i (x_i - \mu_x)^2 \right)^{1/2};$$

$$\sigma_{xy} = \sum_{i=1}^N v_i (x_i - \mu_x)(y_i - \mu_y)$$

with a circularly symmetric normalized Gaussian weight function  $v = \{v_i | i = 1, 2, \dots, N\}$  with a standard deviation of 1.5 samples,  $\sum_{i=1}^N v_i = 1$ , and a sliding window approach that ensures the property of local isotropy of the quality maps.

The constants  $C_1$  і  $C_2$  are included in (3) to avoid instability when the expressions  $\mu_x^2 + \mu_y^2$  і  $\sigma_x^2 + \sigma_y^2$  are practically zero.  $C_1 = (K_1 L)^2$  and  $C_2 = (K_2 L)^2$  are defined, where  $L$  is the dynamic range of pixel values, e.g.,  $L = 255$  for 8-bit grayscale images, and  $K_1 \ll 1$  and  $K_2 \ll 1$  are small constants, for example,  $K_1 = 0.01$ ,  $K_2 = 0.03$  [6].

In practice, in cases where a single overall measure of quality of the entire image is required, the average value of SSIM indices (3) over a set of image pixels called MSSIM is suggested, which aggregates the structural similarity between the reference and distorted images. MSSIM is calculated as the arithmetic mean of  $SSIM(x_j, y_j)$  over the image content in the  $j$ -th local window [6].

In this paper, a weighted average of different samples in the SSIM index map is proposed:

$$WM\_SSIM(X, Y) = \sum_{j=1}^M w_j SSIM(x_j, y_j),$$

where  $M$  is the number of local windows in the image,  $x_j$  and  $y_j$  are the content of the reference  $X$  and distorted  $Y$  images at the  $j$ -th local window, and  $w_j$  are weighting coefficients for different samples (e.g. different image textures attract a person’s attention with varying degrees). Weights  $w_j$  are calculated depending on the practical problem by analyzing decision hierarchies or networks with the consideration of human assessments [11; 12; 14].

**LPIPS** (Learned Perceptual Image Patch Similarity) is a perceptual metric that aimed at evaluating the visual perception of an image by a person at the level of details and uses deep neural networks to assess the visual similarity of a pair of features based on extracted features [10]:

$$LPIPS(I_{SR}, I_{HR}) = \sum_l w_l \cdot \frac{\|\phi_l(I_{SR}) - \phi_l(I_{HR})\|_2^2}{H_l \cdot W_l \cdot C_l},$$

where  $\phi_l(I_{SR})$  is an activation of VGG or another deep network on the  $l$ -th layer for the image  $I_{SR}$ ;  $H_l, W_l, C_l$  are the height, width and number of channels of the  $l$ -th feature map;  $w_l$  is a weighting factor that adjusts the contribution of different layers.

An explanation of the values for each indicator is provided in Table 3. Through visual analysis of a large number of generated images and the corresponding values of quality indicators, thresholds for practically acceptable and high-quality results were obtained, which are given in the last two columns of Table 3.

**Table 3.** Indicator analysis criteria for the SISR task [20]

Indicator	Value range	Practically acceptable result	High-quality result
PSNR↑	[0; 1]	>20	>30
MSSIM↑	[0; 1]	>0.7	>0.9
LPIPS↓	[0; 1]	<0.3	<0.1

For an objective evaluation of the models, it is necessary to add the training time of the models to the indicator analysis. Attention should also be paid to the fact that the indicator values are not worse than the bicubic increase (scaling LR to HR), as this will indicate extremely poor quality of the models even if practically acceptable values are obtained.

**Algorithm for assessing the quality of models and deep learning methods in terms of multiple quantitative and qualitative criteria**

Generative models and deep learning methods, which are studied, are aimed at increasing the resolution of images, scale them by 4 or more times, and as a result generate realistic and beautiful images for further human perception. Therefore, it is necessary to add another group of qualitative decision criteria, including effects of smoothing, blurring, edge lightening, and photorealism of the image. In terms of these criteria, we evaluate the set of images (decision alternatives) generated by different generative models and deep learning methods. Evaluation is made directly by a human using one of the pairwise comparison methods [11–15]. The decision support (DS) problem of multiple criteria evaluation of decision alterna-

tives can be solved using a systematic approach and methodology based on hierarchical and network models [25]. On their basis, an algorithm to solve the problem is suggested, which has the following five stages:

1. Determine interdependencies among decision criteria and decision alternatives. A hierarchy or DS network is formed, which includes the overall goal — selection of the best generated image, qualitative decision criteria: effects of smoothing, blurring, edge lightening, and photorealism of the image, and decision alternatives: image\_SRGAN, image\_VDSR and image\_DRCN (Fig. 1).

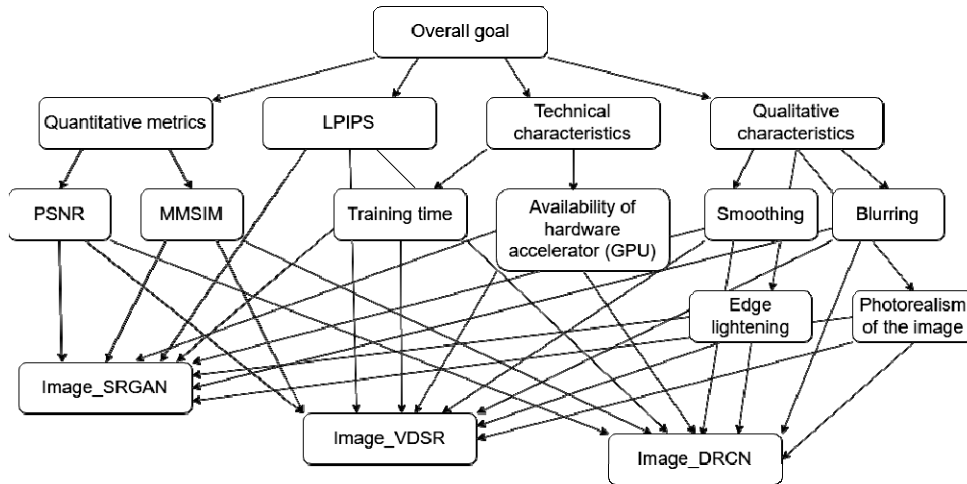


Fig. 1. An hierarchy for assessing the quality of images generated by different models

2. The importance of the decision criteria in relation to the main goal is assessed by experts using the pairwise comparison method on a special scale. Based on the results of the assessment, pairwise comparison matrices (PCMs) are constructed, and the quality of expert opinions is analyzed and, if necessary, improved using the method of evaluation and consistency improvement. The most inconsistent expert opinion is founded. As a result, for all elements of the hierarchy or the DS network, we obtain a set of PCMs of acceptable quality.

3. The coefficients of relative importance (local weights) of the elements of the hierarchy or the DS network are calculated based on the PCMs.

4. The local weights are aggregated using different methods depending on whether the decision criteria are independent (hierarchy case), interdependent (hierarchy case with a loop at the criterion level), or whether there are feedbacks from alternatives to decision criteria (DS network case).

5. The sensitivity analysis of aggregated results (2) is performed.

The purposes of the algorithm are: to calculate local weights for decision alternatives (image\_SRGAN, image\_VDSR, and image\_DRCN) in terms of each decision criteria, as well as to calculate aggregated weights and perform their sensitivity analysis.

## RESULTS OF THE EXPERIMENTS

### Dataset

The DIV2K dataset [1] was introduced as part of the NTIRE 2017 Challenge on Single Image Super-Resolution, held during the CVPR Workshops 2017 conference. It was created to enhance the effectiveness of solving the SISR problem by

addressing the limitations of existing datasets, namely insufficient scene diversity and the limited number of images.

DIV2K consists of a labeled set of 1000 pairs of low-resolution (LR) and high-resolution (HR) color images. The dataset is divided into three subsets: 800 samples for training, 100 samples for testing, and 100 samples for validation. Historically, the test set was designed for contestants to evaluate their models after training, while the validation set was reserved for organizers to determine the winners. The validation set initially included only LR images, and participants were required to generate their super-resolution (SR) counterparts. Once the HR versions of the validation set were made publicly available, both the test and validation sets could be utilized to assess model performance (Fig. 2).

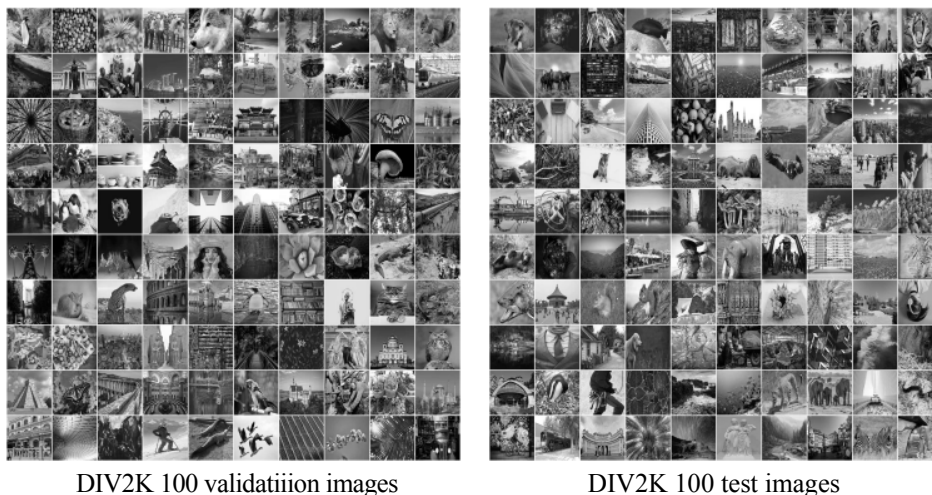


Fig. 2. Example of images for model evaluation from the DIV2K set [1]

The low-resolution (LR) images in the DIV2K dataset are derived from the original high-resolution (HR) images using either bicubic downscaling or more advanced methods that simulate real-world degradations. These methods include modeling blurring caused by motion, introducing fractional noise, and applying distortions due to uneven pixel mapping, among others.

The dataset includes images reduced by scaling factors of 2 ( $\times 2$ ), 3 ( $\times 3$ ), and 4 ( $\times 4$ ). Greater downscaling significantly diminishes image quality (Fig. 3) while also reducing the time required for model training. The classical approach to Single Image Super-Resolution (SISR) typically employs LR images generated through a 4-fold reduction of the original HR images using bicubic interpolation.



Fig. 3. Demonstration of image quality deterioration with a 2 and 4 times reduction in resolution

After its introduction in 2017, the DIV2K dataset has been extensively used to evaluate various super-resolution (SR) models, including in studies conducted in 2019 [26], 2020 [23], and 2023 [27].

### Training process and results

In the first experiment (Section 3.2), we trained our own implementations of the SRGAN, VDSR, and DRCN models from scratch using the DIV2K dataset. The optimization processes of their respective loss functions during training are illustrated in Figs. 4 and 5, while the metric values obtained are presented in Table 4.

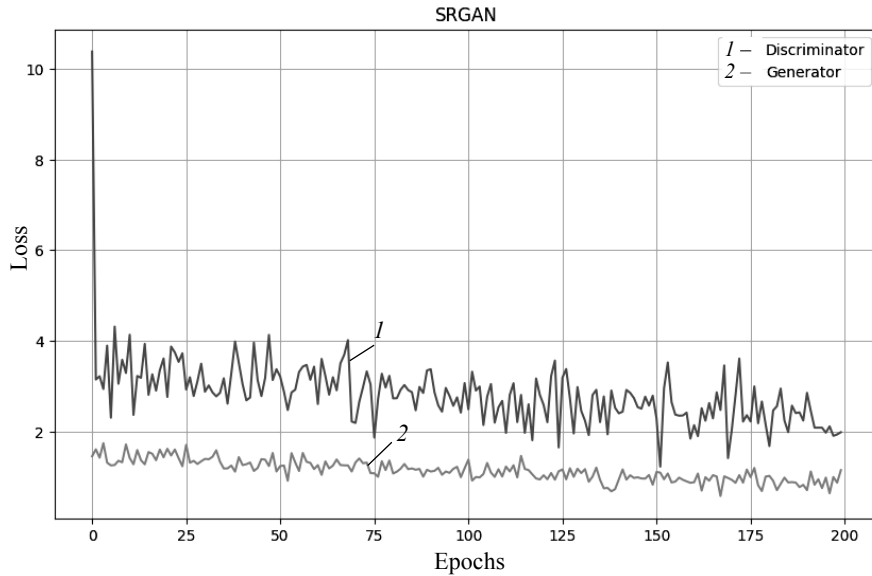


Fig. 4. The process of optimising the loss functions of the generator and discriminator of the SRGAN model [20]

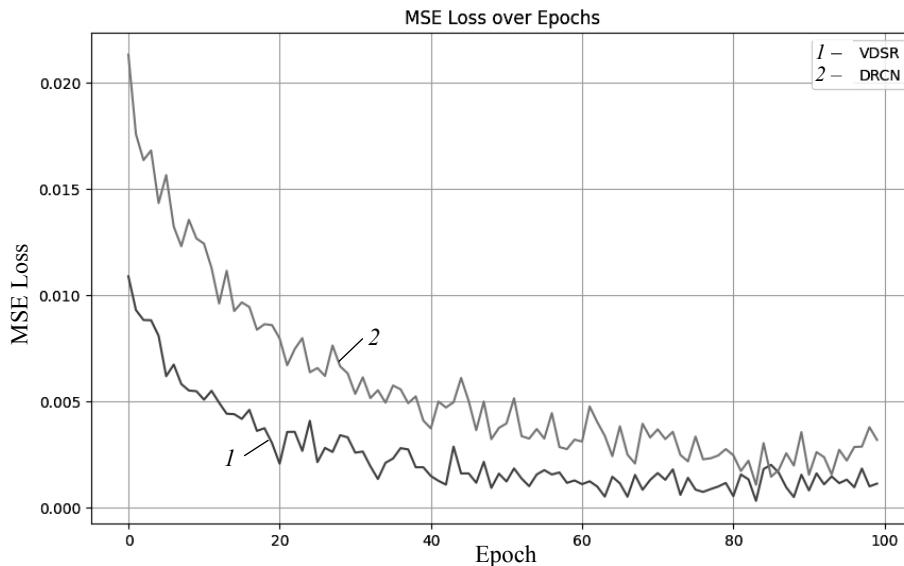


Fig. 5. The process of optimising the loss functions of VDSR and DCRN networks

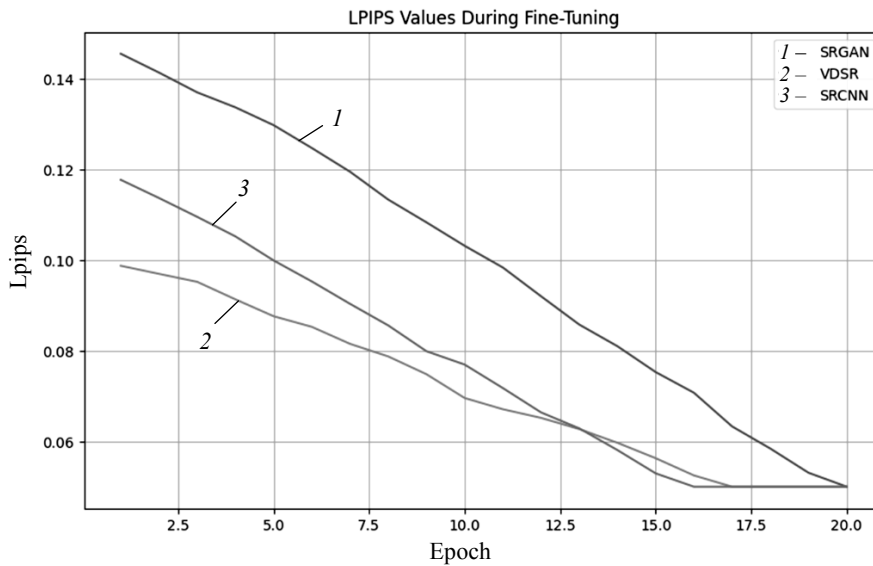
The second experiment (Section 3.3) involved retraining the previously trained SRGAN, VDSR, and DRCN models. The results of this retraining process are provided in Table 5, and the evolution of perceptual quality, as measured by the LPIPS metric, is shown in Fig. 6. For the pre-trained models, we used implementations of SRGAN [16; 28], VDSR [17; 29], and SRCNN [19; 30].

**Table 4.** Values of quality indicators of the generated super-resolution images for our own model implementations at 4-fold image magnification [20]

Model	Indicator			Training time (h)
	PSNR↑	MSSIM↑	LPIPS↓	
Bicubic	25.80	0.74	0.46	–
SRGAN	24.50	0.71	0.33	32
VDSR	26.73	0.77	0.31	16
DRCN	26.41	0.76	0.37	25

**Table 5.** Values of quality indicators of images enlarged by 4 times as a result of retraining of pre-trained models

Model	Indicator			Training time (min)
	PSNR↑	MSSIM↑	LPIPS↓	
Bicubic	25.80	0.74	0.46	–
EDSR [31]	28.98	0.83	0.270	–
RRDB [32]	29.44	0.84	0.253	–
ESRGAN [32]	26.22	0.75	0.124	–
pre-trained SRGAN	26.9	0.79	0.16	27
pre-trained VDSR	28.9	0.84	0.1	11
pre-trained SRCNN	27.5	0.81	0.12	2



*Fig. 6.* Change in the perceptual quality of LPIPS images enlarged by a factor of 4 when retraining pre-trained SRGAN, VDSR and SRCNN models

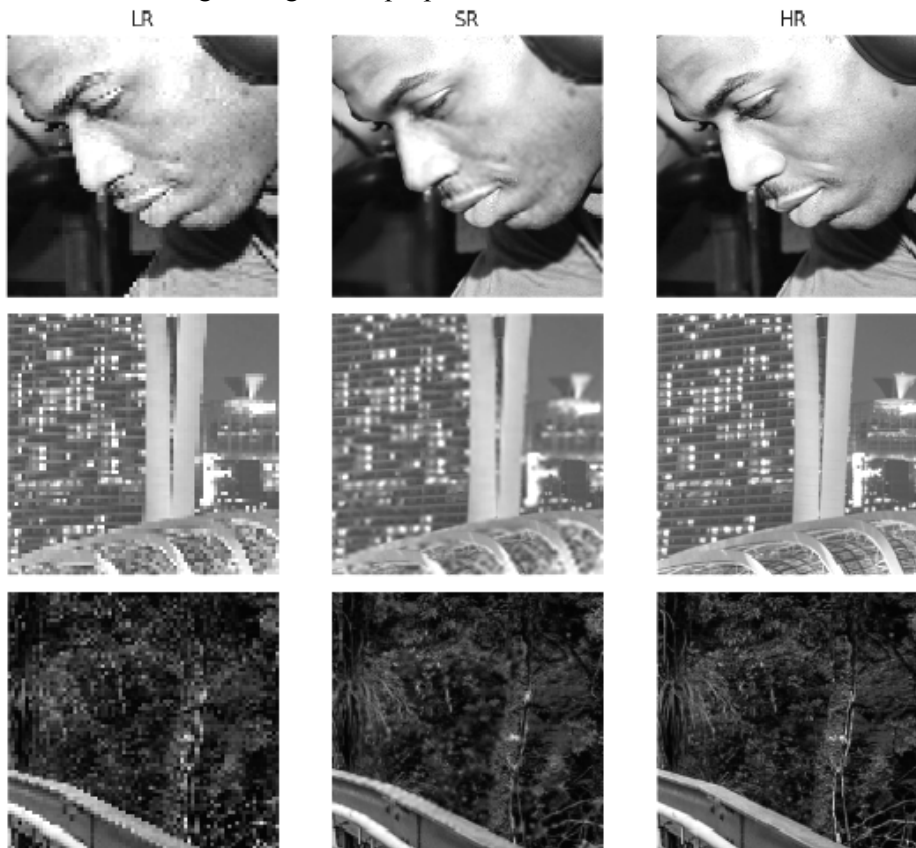
The software solutions for these experiments were developed in the Jupyter Notebook environment using Python, along with the PyTorch library for model development and the matplotlib library for visualization. The models were trained on a PC equipped with an Nvidia GeForce RTX 4060 GPU accelerator.

### ANALYSIS OF THE RESULTS AND DISCUSSION

The results of the first experiment (Section 3.2, Figs. 4, 5, Table 4) demonstrate practically acceptable outcomes for all considered models, with VDSR perform-

ing the best. This highlights, in particular, that residual learning proved to be more effective than recursive learning. The SRGAN architecture, in this experiment, was too simplistic for the given task, as generating new details often outperforms feature refinement.

A comparison of the results in Table 4 with those obtained by other researchers [23] indicates that the metrics in Table 4 are worse than those reported for other SISR models [23]. However, the visual comparison of the generated super-resolution (SR) images with their low-resolution (LR) and high-resolution (HR) counterparts (Fig. 7) shows satisfactory results, provided that the models were trained using the algorithm proposed in Section 3.2.



*Fig. 7. Visual comparison of the generated SR images with the high-resolution (HR) original and low-resolution (LR) input image for the proprietary implementation of the VDSR model*

The results of the second experiment (Section 3.3, Table 5), which employed pre-training techniques, are comparable to those achieved by other researchers [23]. Specifically, the VDSR model, implemented and fine-tuned using the algorithm proposed in this study, achieved an MSSIM value of 0.84, which is on par with the RRDB model [32] and surpasses the MSSIM values of other models developed and fine-tuned in this study: SRGAN (MSSIM = 0.79), SRCNN (MSSIM = 0.81), as well as EDSR [31] and ESRGAN [32].

In terms of the perceptual quality metric LPIPS, the VDSR model trained with the proposed algorithm outperformed other SRGAN and SRCNN models implemented in this study, as well as the EDSR [31], RRDB [32], and ESRGAN [32] models.

The second experiment (Section 3.3) revealed no signs of overfitting, and the generated SR images demonstrated high quality compared to the input LR-HR pairs (Fig. 8). The VDSR model consistently produced the best visual results, underscoring the advantage of feature enhancement when addressing SISR tasks for highly detailed data and complex real-world scenes.



*Fig. 8.* Visual comparison of the generated SR images with the high-resolution (HR) original and low-resolution (LR) input image for the VDSR model trained with the suggested algorithm

## CONCLUSIONS

This study presents an algorithm for the comprehensive evaluation of image super-resolution results based on quantitative metrics, perceptual indicators, technical characteristics, and aspects of human image perception. Threshold criteria for practically acceptable and high-quality results were determined through visual analysis of many generated images and their corresponding quality metrics, including those obtained by other researchers.

The VDSR model was identified as the optimal one (among those considered) in terms of pixel, structural, and perceptual metrics, as well as training time. The absence of overfitting and the quality of super-resolution images generated by VDSR were visually confirmed on selected test set samples depicting various

shapes, textures, and color combinations. Overall, deep learning methods demonstrated superiority over generative models in the conducted experiments based on the results of the comprehensive evaluation.

## REFERENCES

1. E. Agustsson, R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017*. doi: <https://doi.org/10.1109/cvprw.2017.150>
2. Z. Wang, J. Chen, S.C.H. Hoi, "Deep Learning for Image Super-resolution: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3365–3387, 2020. doi: <https://doi.org/10.1109/tpami.2020.2982166>
3. R. Timofte et al., "NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017*. doi: <https://doi.org/10.1109/cvprw.2017.149>
4. T. Ausare, "Ultimate Guide to Selecting a GPU for Deep Learning. Latest AI, ML & GPU Updates," *NeevCloud*. Available: <https://blog.neevcloud.com/ultimate-guide-to-selecting-a-gpu-for-deep-learning>
5. F.A. Fardo, V.H. Conforto, F.C. de Oliveira, P.S. Rodrigues, *A Formal Evaluation of PSNR as Quality Measurement Parameter for Image Segmentation Algorithms*. 2016. doi: <https://doi.org/10.48550/arXiv.1605.07116>
6. Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, Eero P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, issue 4, pp. 600–612, 2004. doi: <https://doi.org/10.1109/TIP.2003.819861>
7. A. Mittal, A. Moorthy, A. Bovik, "Referenceless image spatial quality evaluation engine," in *45th Asilomar Conference on Signals, Systems and Computers*, vol. 38, pp. 53–54, 2011. doi: <https://doi.org/10.1109/ACSSC.2011.6190099>
8. A. Mittal, R. Soundararajan, A.C. Bovik, "Making a "completely blind" image quality analyser," *IEEE Signal Process. Lett.*, vol. 20, issue 3, pp. 209–212, 2013. doi: <https://doi.org/10.1109/LSP.2012.2227726>
9. N. Venkatanath, D. Praneeth, Bh. Maruthi Chandrasekhar, S.S. Channappayya, S.S. Medasani, "Blind image quality evaluation using perception based features," *2015 Twenty First National Conference on Communications (NCC), Mumbai, India, 2015*, pp. 1–6. doi: <https://doi.org/10.1109/NCC.2015.7084843>
10. R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018*, pp. 586–595. doi: <https://doi.org/10.1109/CVPR.2018.00068>
11. N.I. Nedashkovskaya, "Method for weights calculation based on interval multiplicative pairwise comparison matrix in decision-making models," *Radio Electronics, Computer Science, Control*, no. 3, pp. 155–167, 2022. doi: <https://doi.org/10.15588/1607-3274-2022-3-15>
12. N.I. Nedashkovskaya, "Estimation of the accuracy of methods for calculating interval weight vectors based on interval multiplicative preference relations," *IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC), 2022*. doi: <https://doi.org/10.1109/SAIC57818.2022.9922977>
13. N.I. Nedashkovskaya, "Method for Evaluation of the Uncertainty of the Paired Comparisons Expert Judgements when Calculating the Decision Alternatives Weights," *Journal of Automation and Information Sciences*, vol. 47, issue 10, pp. 69–82, 2015. doi: <https://doi.org/10.1615/JAutomatInfScien.v47.i10.70>

14. N.D. Pankratova, N.I. Nedashkovskaya, "Hybrid Method of Multicriteria Evaluation of Decision Alternatives," *Cybernetics and Systems Analysis*, vol. 50, no. 5, pp. 701–711, 2014. doi: <https://doi.org/10.1007/s10559-014-9660-2>
15. N.I. Nedashkovskaya, "Investigation of methods for improving consistency of a pairwise comparison matrix," *Journal of the Operational Research Society*, vol. 69, no. 12, pp. 1947–1956, 2018. doi: <https://doi.org/10.1080/01605682.2017.1415640>
16. C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 21–26 July 2017*, pp. 105–114. doi: <https://doi.org/10.1109/cvpr.2017.19>
17. J. Kim, J.K. Lee, K.M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*, pp. 1646–1654. doi: <https://doi.org/10.1109/cvpr.2016.182>
18. J. Kim, J.K. Lee, K.M. Lee, "Deeply-Recursive Convolutional Network for Image Super-Resolution," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*, pp. 1637–1645, 2016. doi: <https://doi.org/10.1109/cvpr.2016.181>
19. C. Dong et al., "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016. doi: <https://doi.org/10.1109/tpami.2015.243928>
20. A.A. Lanko, N.I. Nedashkovskaya, "Generative models and methods of deep learning for the SISR problem," *System sciences and informatics: collection of reports of the 3rd All-Ukrainian scientific and practical conference "System sciences and informatics", November 25–29, 2024, Kyiv. K.: IASA KPI, 2024*, pp. 176–181. Available: [http://mmsa.kpi.ua/sites/default/files/systemni\\_nauky\\_ta\\_informatyka\\_2024.pdf](http://mmsa.kpi.ua/sites/default/files/systemni_nauky_ta_informatyka_2024.pdf)
21. K. He et al., "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015*, pp. 1026–1034. doi: <https://doi.org/10.1109/iccv.2015.123>
22. X. Glorot, Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS), Sardinia, Italy, 13–15 May 2010*, PMLR, vol. 9, pp. 249–256. Available: <http://proceedings.mlr.press/v9/glorot10a.html>
23. A. Lugmayr et al., "SRFlow: Learning the Super-Resolution Space with Normalizing Flow," *Computer Vision – ECCV 2020, Cham, 2020*, pp. 715–732. doi: [https://doi.org/10.1007/978-3-030-58558-7\\_42](https://doi.org/10.1007/978-3-030-58558-7_42)
24. Q. Jiang et al., "Single Image Super-Resolution Quality Assessment: A Real-World Dataset, Subjective Studies, and an Objective Metric," *IEEE Transactions on Image Processing*, vol. 31, pp. 2279–2294, 2022. doi: <https://doi.org/10.1109/tip.2022.3154588>
25. N.I. Nedashkovskaya, "A system approach to decision support on basis of hierarchical and network models," *System Research and Information Technologies*, no. 1, pp. 7–18, 2018. doi: <https://doi.org/10.20535/srit.2308-8893.2018.1.01>
26. A. Ignatov et al., "PIRM challenge on perceptual image enhancement on smartphones: report," *Conference on Computer Vision (ECCV) Workshops, 2019*. doi: [https://doi.org/10.1007/978-3-030-11021-5\\_20](https://doi.org/10.1007/978-3-030-11021-5_20)
27. Dandan Gao, Dengwen Zhou, "A very lightweight and efficient image super-resolution network," *Expert Systems with Applications*, vol. 213, Part A, 1, March 2023, 118898. doi: <https://doi.org/10.1016/j.eswa.2022.118898>
28. "GitHub - tensorlayer/SRGAN: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *GitHub*. Available: <https://github.com/tensorlayer/SRGAN>
29. "GitHub - twtygqyy/pytorch-vdsr: VDSR (CVPR2016) pytorch implementation," *GitHub*. Available: <https://github.com/twtygqyy/pytorch-vdsr>.

30. “GitHub - Lornatang/SRCNN-PyTorch: Pytorch framework can easily implement srcnn algorithm with excellent performance,” *GitHub*. Available: <https://github.com/Lornatang/SRCNN-PyTorch>
31. B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, “Enhanced deep residual networks for single image super-resolution,” *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017*, pp. 1132–1140. doi: <https://doi.org/10.1109/CVPRW.2017.151>
32. X. Wang et al., “ESRGAN: Enhanced super-resolution generative adversarial networks,” *Computer Vision – ECCV 2018 Workshops: Munich, Germany, September 8-14, 2018, Proceedings, Part V*, pp. 63–79. doi: [https://doi.org/10.1007/978-3-030-11021-5\\_5](https://doi.org/10.1007/978-3-030-11021-5_5)

Received 27.12.2024

### INFORMATION ON THE ARTICLE

**Anna A. Lanko**, ORCID: 0009-0005-8370-5739, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e mail: [lanko.anna@iill.kpi.ua](mailto:lanko.anna@iill.kpi.ua)

**Nadezhda I. Nedashkovskaya**, ORCID: 0000-0002-8277-3095, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e mail: [nedashkovskaya.nadezhda@iill.kpi.ua](mailto:nedashkovskaya.nadezhda@iill.kpi.ua)

**ОЦІНЮВАННЯ ЯКОСТІ МОДЕЛЕЙ ТА МЕТОДІВ ГЛИБОКОГО НАВЧАННЯ ДЛЯ ФОРМУВАННЯ СУПЕРРОЗДІЛЬНИХ ЗОБРАЖЕНЬ / Н.І. Недашківська, А.А. Ланько**

**Анотація.** Розглянуто метрику оцінювання результатів генерації суперроздільних зображень під час розв’язання задачі SISR. Дослідження включає два експерименти: власну реалізацію мережевих архітектур для SRGAN, VDSR і SRCNN, і точне налаштування попередньо навчених моделей SRGAN, VDSR і SRCNN. Запропоновано алгоритм оцінювання якості моделей і методів глибокого навчання для генерації суперроздільних зображень. Модель VDSR продемонструвала найкращі результати з точки зору піксельного, структурних і перцептивних показників, а також часу навчання та візуального підтвердження якості згенерованого зображення людиною, підкреслюючи, що залишкове навчання є більш ефективним, ніж рекурсивне навчання за умов двох проведених експериментів. Порогові значення для прийнятних і високоякісних результатів визначено шляхом візуального аналізу багатьох згенерованих зображень і відповідних показників якості, включно з тими, про які повідомляли інші дослідники.

**Ключові слова:** задача SISR, оцінювання якості, генеративні моделі, методи глибокого навчання, згортова нейронна мережа, залишкове навчання, рекурсивне навчання, тонке налаштування попередньо навчених моделей, перцептивна метрика, LPIPS, багатокритеріальний аналіз розв’язань, набір даних DIV2K, порогові значення для прийнятних і високоякісних згенерованих зображень.