

UDC 004.8:004.93'1(045)
DOI:10.18372/1990-5548.89.21472

¹Olena Chumachenko,
²Anatoliy Kot

STRUCTURAL-PARAMETRIC SYNTHESIS OF NEURAL NETWORKS FOR IMAGE RESTORATION UNDER SYNTHETIC OBJECT-TARGETED DAMAGE

Educational and Research Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" Kyiv, Ukraine

E-mails: ¹eliranvik@gmail.com ORCID: 0000-0003-3006-7460

²anatoly.kot@gmail.com, ORCID 0000-0002-7490-8834

Abstract—The article is devoted to the structural-parametric synthesis of neural networks for guided recovery on pairs "damaged - original". The genetic algorithm is used as an optimization algorithm. The training data is synthesized using object-oriented random erasure. The data are built with object-targeted Random Erasing. The train / val / test splits are disjoint: 5362/670/670. On the holdout ($n = 670$; valid letterbox region, without padding) Wider U-Net yields 25.33 ± 4.50 dB (1.88M; 116.6 GMACs), and Warm Start genetic algorithm yields 25.23 ± 4.39 dB (1.15M; 97.3 GMACs). By the paired Wilcoxon test, the gap behind Wider U-Net is 0.10 dB, and the advantage over Baseline genetic algorithm is 0.25 dB in this run. Both synthesized networks are substantially better than Telea and identity. The level width in the synthesis space is limited to 128 filters, therefore Wider U-Net (256) lies outside it. The result is read as a competitive trade-off, without a claim of absolute PSNR leadership and without generalizing warm start beyond this experiment.

Keywords—Structural-parametric synthesis, image restoration, inpainting, neural networks, genetic algorithm, random erasing, peak signal-to-noise ratio, structural similarity index, encoder-decoder, holdout.

I. INTRODUCTION

Image quality affects subsequent scene analysis. The practical motivation is partial damage or occlusion of objects in the imagery. Resources are often limited. Onboard UAV scenarios are only an example of such a setting.

If an informative region is damaged, the analysis becomes harder. Therefore image restoration (inpainting) is reasonable. It is the reconstruction of lost or corrupted pixels closer to the original. Supervised restoration requires an image-to-image network. It has to fill the damaged region from context. The choice of architecture and training settings affects both quality and computational cost.

Different approaches exist. Classical inpainting methods fill the hole by local smoothness or by patches. They do not learn from a large paired sample. Learning-based methods build the damaged-to-original mapping as a convolutional network. Most often this is an encoder-decoder with skip connections. Such architectures are usually designed by hand. Methods of automatic architecture search develop separately. They are mostly oriented towards classification of clean images, not towards restoration under object-targeted damage. In practice a contradiction arises. A wider network often gives higher quality. At the same time it is heavier in the number of parameters and operations (GMACs).

This is critical under resource constraints.

There is also the line of structural-parametric synthesis of neural networks. It allows the model structure and the training hyperparameters to be formed jointly in a single space of candidates. The structure includes depth, level width and the presence of skips. This is an alternative to separate manual tuning. In this way a restorer can be obtained automatically that is close in quality to a strong hand-designed U-Net and more compact in params / GMACs.

Given this, a study of structural-parametric synthesis for image restoration under object-targeted damage is expedient. The subject is the search for a trade-off between quality, model size and computational cost, not the maximization of PSNR without regard to resources.

II. RELATED WORK

Convolutional neural networks have shown high efficiency in computer vision tasks and related applications. By the type of architectural modifications they are conventionally divided into seven classes: networks with an emphasis on spatial exploitation; networks with an emphasis on depth; multi-path networks; networks with many width-wise connections; feature-map based networks; channel-boosted networks; attention-based networks. The encoder-decoder restorer for image restoration

belongs to this family of convolutional architectures.

Structural-parametric synthesis of neural networks means the joint formation of the network topology and of its training parameters within a single search process. This differs from the scheme “first set the architecture by hand, then tune the hyperparameters separately”. Structure and parameters can be coupled degrees of freedom of one design task. The theoretical framework of such an approach for hybrid artificial neural networks was developed in the works of Zgurovsky, Sineglazov and Chumachenko. The problem statement of this paper relies on it.

In the formation of topologies of hybrid artificial neural networks, structural synthesis can be presented as a separate scientific and applied task rather than as purely heuristic layer selection [1]. Ways of building an architecture from heterogeneous blocks and connections can be systematized. The quality of the system can then be determined already at the stage of topology choice. For image restoration it follows that an encoder-decoder restorer can also be treated as an object of structural synthesis. Depth, level width and skip connections are the controlled variables. This is not merely a fixed template with subsequent weight training.

Multicriteria optimization methods in the context of hybrid neural networks can be classified and analysed with respect to reconciling several conflicting quality indicators [2]. Scalar and vector statements, trade-offs between model accuracy and complexity, and mechanisms for selecting non-equivalent variants provide the apparatus for tasks in which a single criterion is not enough. For a restorer under limited resources this can correspond to the statement “restoration quality – model size – inference cost”. Synthesis is then oriented not to the absolute PSNR maximum but to an acceptable point on the trade-off curve.

An intelligent system for determining the degree of damage of objects in images can be built on the basis of generative adversarial networks – in particular, for assessing the degree of bark beetle damage to trees [3]. Deep models can be effective in tasks related to damage of visual content. The practical statement naturally relies on “damaged image – reference information” pairs. At the same time such a system can be aimed at assessing the degree of damage rather than at synthesizing a pixel restoration network. The gap between damage analysis and structural-parametric synthesis of a restorer outlines the subject area of the proposed study.

Beyond this theoretical line, methods of automatic architecture search are developing

actively. They are close in idea to structural-parametric synthesis but are mostly oriented towards classification. The search for transferable architectures for scalable image recognition can be implemented as computationally expensive reinforcement-learning NAS [4]. The selection criterion there is accuracy on clean data. Therefore the results do not transfer to inpainting directly. Regularized evolution can also be aimed at classifiers at a high search cost [5]. Differentiable NAS can accelerate the structure search along the gradient [6]. The candidate space remains closer to a sequential classifier than to an encoder-decoder with full joint optimization of hyperparameters. Weight sharing can reduce the cost of evaluating architectures but does not account for the specifics of occlusions [7]. Genetic formation of a convolutional neural network (CNN) allows the network structure to be evolved [8]. However, each new candidate is usually trained from scratch, without knowledge transfer between generations. Multi-objective NAS can reconcile model accuracy and complexity without posing the task of restoration under object occlusion [9]. The combination of evolutionary search with parameter reuse in a one-shot scheme comes close to structural-parametric synthesis, but again within classification [10]. The NSGA-II algorithm can provide the classical apparatus of elitist multi-objective selection and is consistent with the analysis of multicriteria methods [2]; by itself, however, it does not define an encoder-decoder space for inpainting [11].

A separate line is parametric synthesis of networks for image restoration. The architecture is mostly fixed by hand. “Synthesis” is then reduced to training the weights and, sometimes, tuning hyperparameters on paired data - without a joint topology search in the sense of [1]. Classical exemplar-based inpainting can fill the hole with patches without deep learning [12]. A residual CNN for denoising allows the parameters to be trained for a predefined structure [13]. U-Net with skip connections can be regarded as the de facto standard for image-to-image settings [14]. The topology is set manually, after which parametric training is performed. Context Encoder extends inpainting with an adversarial component while keeping the manual design of the encoder-decoder [15]. Partial Convolutions can improve handling of a known mask through mask-aware convolutions, but they also do not perform a structural search [16]. Diffusion models can provide high perceptual quality at the price of substantial computation and again rely on a fixed architecture [17]. Restoration quality is conventionally assessed by PSNR and

SSIM as full-reference metrics [18]. It is precisely these that can enter a multicriteria statement in the spirit of [2]. For modelling damage, Cutout is used as a regularizer [19], as is Random Erasing with erasure of a random region [20]. In their basic form the region is often not tied to a target object. Damage analysis tasks close to [3] motivate an object-oriented statement.

Thus, the supporting apparatus of structural-parametric synthesis and multicriteria selection is provided by [1], [2]. An example of deep damage analysis in images is [3]. Methods of automatic architecture search and modern restorers with a fixed topology broaden the review. They, however, do not solve the task of joint synthesis of the structure and the hyperparameters of an encoder-decoder for object-targeted damage under the quality / params / GMACs criteria with a final check on an independent holdout.

III. PROBLEM STATEMENT

Image restoration (inpainting) is the reconstruction of lost or corrupted pixels from the preserved context. Let $x \in [0,1]^{(H \times W \times 3)}$ be an undamaged RGB image, $\tilde{x}$ its corrupted version, and $m \in [0,1]^{H \times W}$ the binary mask of the damaged region ($m_{i,j} = 1$ inside the hole). In the supervised setting the restorer f_θ is trained on triples $(x, \tilde{x}, m)$ to recover the original. With a known mask, the natural input is the concatenation of the damaged frame and the mask:

$$\hat{x} = f_\theta([\tilde{x}, m]), \quad \hat{x} \in [0,1]^{(H \times W \times 3)}. \quad (1)$$

The goal is to bring $\hat{x}$ close to x over the whole frame and especially inside. The restorer performs no object classification.

The sources of damage vary: noise, blur, erasure, occlusion. For a controlled study of partial object occlusion, a model based on Random Erasing [20] is convenient. Inside the region of interest (for example, the union of bounding boxes B) a rectangle R is chosen and its pixels are replaced by independent noise $u_{i,j} \sim U[0,1]^3$:

$$\tilde{x}_{i,j} = \begin{cases} u_{i,j}, & (i,j) \in R, \\ x_{i,j}, & (i,j) \notin R, \end{cases} \quad m_{i,j} = \begin{cases} 1, & (i,j) \in R, \\ 0, & (i,j) \notin R. \end{cases} \quad (2)$$

If $R \subset B$, the damage is object-targeted. The hole lies on the target object rather than on arbitrary background.

Classical inpainting methods (diffusion, exemplar-based [12], Telea and others) fill the hole by local smoothness or by patches without learning

from a large paired sample. Learning-based approaches build the mapping (1) as a CNN. For image-to-image tasks the typical topology is an encoder-decoder with skip connections (U-Net [14]). The encoder compresses the spatial scale and accumulates context. The decoder restores the resolution. Skip connections carry fine details from the corresponding encoder levels. Depth, level width (the number of filters) and the presence of skips are structural degrees of freedom. Learning rate, optimizer, dropout and so on are parametric ones. Together they determine both the restoration quality and the resources (the number of weights, the number of operations).

Training a restorer is reduced to minimizing a loss between $\hat{x}$ and x . A pixel-wise error is often combined with structural similarity. One common form is a weighted sum of L1 and the SSIM complement:

$$L = \alpha \|\hat{x} - x\|_1 + (1 - \alpha)(1 - \text{SSIM}(\hat{x}, x)), \quad (3)$$

where $\alpha \in (0,1)$ balances intensity accuracy against structure preservation. If $\mathcal{L}$ is computed over the whole frame rather than over the mask only, the loss is not mask-aware. The priority of the damaged region is not set explicitly.

Restoration quality is assessed by full-reference metrics. For the range $[0,1]$

$$\text{PSNR}(\hat{x}, x) = 10 \log_{10} \left(\frac{1}{\text{MSE}(\hat{x}, x)} \right), \quad (4)$$

where MSE is the mean squared error over the channels and pixels of the evaluated region. Structural similarity index (SSIM) [18] compares local means, variances and covariance. It reflects structural similarity better than plain mean squared error (MSE). Separately, mask-PSNR is introduced – peak signal-to-noise ratio (PSNR) computed on $\{m = 1\}$ only. Under limited computational resources, quality is compared against model cost: the number of parameters and the estimated number of MACs / GMACs per frame of fixed size. This is a proxy of inference cost, not a measured onboard latency.

For image-to-image restoration the encoder and the decoder are parts of a convolutional neural network. The encoder takes the input $[\tilde{x}, m]$ and successively reduces the spatial scale through convolutions and pooling. Context is thus accumulated in the form of latent features. The decoder restores the spatial size through upsampling and convolutions. The output $\hat{x}$ is formed at the end. The general scheme is shown in Fig. 1.

Without additional connections, details from the early encoder levels can be lost after compression. Therefore skip connections between the corresponding encoder and decoder levels are often added. They pass the feature maps directly. Such a topology corresponds to U-Net. The scheme with skips is shown in Fig. 2.

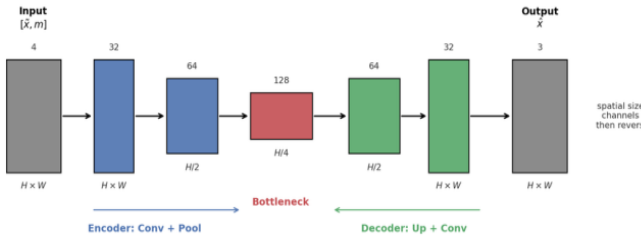


Fig. 1. Scheme of an encoder-decoder convolutional network for image restoration

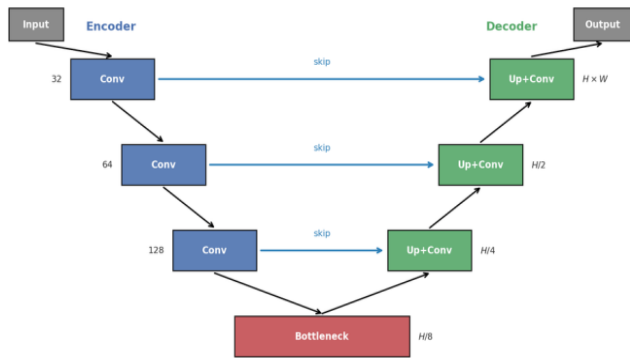


Fig. 2. Scheme of a U-Net-like network with skip connections between encoder and decoder

In such a network the structural degrees of freedom are depth, level width (the number of filters) and the presence of skips. The parametric ones are learning rate, optimizer, dropout and so on. It is exactly these that are chosen jointly in structural-parametric synthesis.

Hence, the formal problem of this study is structural-parametric synthesis. From paired data with object-targeted damage (2) an encoder-decoder restorer (1) trained by (3) has to be built. The structure and the hyperparameters are chosen jointly. On an independent sample a competitive trade-off between PSNR/SSIM (4) and resources (params, GMACs) relative to strong hand-designed U-Net baselines has to be obtained. The implementation of the synthesis in a discrete candidate space is described in Section IV.

The object of the study is structural-parametric synthesis of neural networks for image restoration under partial object-targeted damage. The experiments were carried out on an open vehicle dataset (see the caveat about motivation in Section I).

The aim of the work is to check whether such synthesis of an encoder-decoder restorer on “damaged – original” pairs yields models close in

quality to a strong hand-designed Wider U-Net on an independent holdout while being more compact in the number of parameters and in estimated GMACs. The thin U-Net serves as a lighter but weaker reference. A separate onboard latency benchmark was not performed. Computational cost is compared through GMACs per 512×512 frame.

Hypothesis: under a protocol with disjoint train / val / test and metrics computed on the valid letterbox region, structural-parametric synthesis selects a competitive point on the quality / size / GMACs curve without a claim of absolute PSNR leadership.

Objectives of the study: (1) to substantiate an approach to structural-parametric synthesis of an encoder-decoder restorer CNN for supervised inpainting under object-targeted damage; (2) to ensure the correctness of the comparison through a protocol with disjoint splits, evaluation on the valid region (without padding) and per-image statistics on the holdout; (3) to find out whether the synthesized networks form a competitive quality / size / GMACs point relative to the thin and the Wider U-Net and to trivial baselines (identity, Telea); (4) to assess whether different synthesis variants (in particular warm start) change the selected topology and the restoration quality on the holdout.

The following assumptions are adopted. The setting is supervised image-to-image restoration. A noise rectangle inside a bounding box models partial occlusion. RGB and a binary mask are sufficient input features. Identical genetic algorithm (GA) parameters ensure a correct comparison of the synthesis variants. The final conclusions rest on the holdout only.

Simplifications of the setting: no IR/TV/multispectral data; the proxy synthesis stage at 64×64 ; a single rectangular occlusion; full-res retrain and holdout at letterbox 512; a single dataset; the synthesis space limits the level width to the set $\{16, 32, 64, 128\}$.

IV. APPROACH

Structural-parametric synthesis is implemented by a genetic algorithm. The chromosome encodes the encoder-decoder structure and the training hyperparameters at the same time.

The paired sample is built as follows. The original images are the target, the damaged ones are the input (model (1) – (2)). The dataset is Military and Civilian Vehicles Classification (Kaggle) with XML Pascal VOC annotations. A pair shares the same file name. Pixels are normalized from $[0, 255]$ to $[0, 1]$. Proxy synthesis stage: 64×64 . Full-res retrain / holdout: letterbox 512 (long side ≤ 512 , padding to a square; size divisible by 8). The disjoint splits are given in Table I (splits/, seed = 13,

$\approx 80/10/10$). The test split enters neither the fitness (6) nor early stopping.

TABLE I. DATA AND DISJOINT SPLIT SCHEME (PROTOCOL GROUPED_DISJOINT_V1_SALVAGE)

Stage	Count	Comment
Original files	6772	all JPEG/PNG
Paired original $\cap$ damaged on disk	6772	including 70 fallback
Names in the splits	6702	without fallback; experiment pool
Train / Val / Test	5362 / 670 / 670	training / fitness / holdout only
Intersection train $\cap$ val $\cap$ test	0	leakage-free

The synthetic damage is built as an object-targeted variant of Random Erasing [20]. In its classical form Random Erasing is an augmentation for classification. A rectangle is chosen at random in the frame and filled with a constant or with noise. This reduces the dependence of the model on local patterns. In this work the same idea is applied differently. It is not a classifier regularizer but a controlled model of partial object occlusion for supervised inpainting under equations (1), (2). The “original – damaged” pair is formed once at the data preparation stage and does not change afterwards during training.

The algorithm for processing one frame is the following. All object bndbox entries are read from the XML Pascal VOC annotation and their union bounding box $B = (x_1, y_1, x_2, y_2)$ is built, with width $b_w = (x_2 - x_1)$ and height $b_h = (y_2 - y_1)$. If $b_w < 4$ or $b_h < 4$, or the XML is missing, a fallback scheme is used (below). Otherwise, inside B a rectangle R of area $s \cdot |B|$ is sought, where the area fraction is $s \sim U[0.20; 0.40]$, with aspect ratio $r = h / w \sim U[0.3; 3.3]$. The height and width of the damaged region are set as

$$h = \text{round}(\sqrt{s|B|r}), \quad w = \text{round}(\sqrt{s|B|/r}). \quad (5)$$

An attempt is rejected if $w \geq b_w$ or $h \geq b_h$. An admissible top-left corner (x, y) is drawn uniformly so that $R \subset B$ (at most 100 attempts). The pixels inside R are replaced by independent RGB noise $u \sim U\{0, \dots, 255\}^3$. Outside R the frame stays the original. In parallel a binary mask m is formed: $m = 1$ on R and $m = 0$ otherwise. The damage is thus localized on the object (or on the union of objects) and not on arbitrary background. This is exactly what object-targeted Random Erasing means.

The fallback scheme (global Random Erasing) uses the same s and r but relative to the area of the whole frame $H \times W$. The rectangle is placed at an arbitrary location of the image. There are 70 such frames on disk. They are not included in splits/. For the 6702 names in train/val/test the damage is object-targeted. Synthesis, full retraining and the holdout are performed on these 6702 names only. The mask is fed as the fourth input channel of the restorer.

A Su-57 frame (1d98a84b05bcc76dc3b12b7bc183d31e.jpg) is used as an illustration. The original is shown in Fig. 3.



Fig. 3. Original Su-57 image

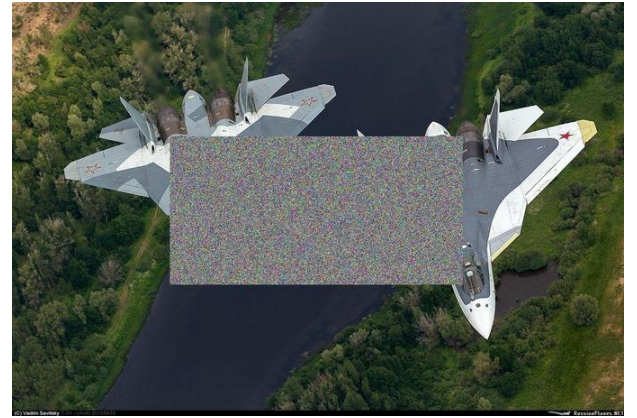


Fig. 4. The same frame after object-targeted Random Erasing

In Figure 4 the noise lies on the object region rather than on the background. Restoration is performed by (1). The structure and the hyperparameters of the restorer are encoded by a chromosome c in the discrete synthesis space C (Table II). The structural genes are depth, filters, skip_connections. The parametric ones are learning_rate, batch_size, optimizer, lr_scheduler, dropout. The network f_c is a U-Net-like encoder-decoder. An encoder level contains two Conv2D (3×3 , ReLU), an optional Dropout and MaxPool (except on the last level). A decoder level contains UpSampling2D, an optional skip Concatenate and two Conv2D. The output is Conv2D(3, 1×1 ,

sigmoid). The topology key is $d\{depth\}_f\{...\}_s\{01...\}$. The level width is limited to $\{16, 32, 64, 128\}$, therefore a level with 256 filters (Wider U-Net) lies outside C .

TABLE II. CHROMOSOME GENES OF THE STRUCTURAL-PARAMETRIC SYNTHESIS OF THE RESTORER CNN

Gene	Range / values	Role
depth	2 ... 4	structure: encoder/decoder levels
filters[i]	{16, 32, 64, 128}	structure: level width
skip_connections[i]	{T, F}	structure: concat encoder→decoder
learning_rate	$[10^{-4}, 10^{-3}]$	parameters: training step
batch_size	{16, 32, 64}*	parameters; *full-res: 2
optimizer	adam, sgd, rmsprop	parameters: optimizer
lr_scheduler	none, cosine	parameters: cosine annealing
dropout	[0, 0.3]	parameters: after conv blocks

An example of the encoding is shown in Fig. 5. The structural genes set the topology, the parametric ones set the training regime. Together they define a single candidate whose quality is evaluated by (6).

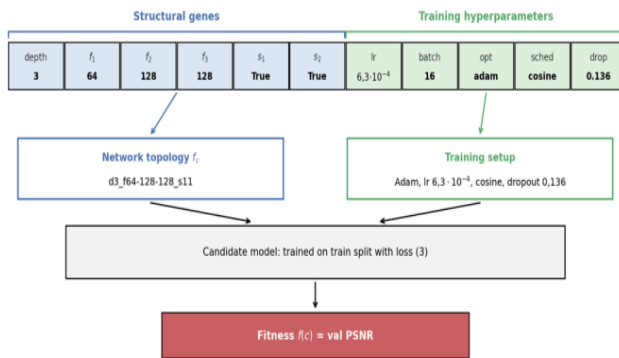


Fig. 5. Encoding of the restorer structure and hyperparameters by a single chromosome

The candidate selection criterion during synthesis is val PSNR after training on train with the loss (3) at $\alpha = 0.84$. The final conclusions come from the holdout (test) only:

$$f(c) = \text{PSNR}(f_c(\tilde{X}_{\text{val}}), X_{\text{val}}), c^* = \arg \max_{c \in C} f(c). \quad (6)$$

Structural similarity index, mask-PSNR, params and time are recorded additionally. The search in C is performed by a genetic algorithm (Table III). The GA was chosen because of the discreteness of the structural genes and the need to optimize them jointly with the parametric ones. Run parameters of the 64×64 proxy synthesis: pop=12, gen=15, max

epochs=50, seed=13 (about 180 chromosome evaluations). The exact total synthesis cost in GPU hours was not recorded by a separate measurement (see the limitations in Section V). As a reference point, the full-res retrain of one model takes ≈ 1.1 –3.1 h on an NVIDIA L4. Baseline GA is synthesis with candidates trained from random weights. Warm Start GA uses the same operators, with the offspring weights taken from the parent by a prefix match of the encoder levels.

TABLE III. GENETIC OPERATORS OF THE SYNTHESIS

Operator	Parameter	Description
Selection	$k = 2$	tournament: max fitness out of two
Crossover	$P_c = 0.7$	depth / filters / skips – from one parent; lr, dropout – average
Mutation	$P_m = 0.4$	depth ± 1 ; filter→random; toggle skip; lr $\times(0.5 - 2)$
Elitism	1	the best individual passes unchanged

The hand-designed baselines are trained on the same pairs and the same loss (3). Thin U-Net [14]: depth=3, filters=[32, 64, 128], skip=[T,T], Adam, lr=0.0003, cosine, dropout=0.1 (471, 907 params). Wider U-Net: [64, 128, 256] with the same hyperparameters (1, 883, 843 params; the 256 level is outside the synthesis space). Trivial baselines on the same holdout: identity (output = damaged input) and the classical Telea inpainting (OpenCV).

The holdout metrics are per-image PSNR/SSIM by (4) on the valid letterbox region (without padding), plus mask-PSNR. Mean±std and 95% CI are reported. Methods are compared by the paired Wilcoxon test on per-image PSNR/SSIM. The estimated computational cost is reported as GMACs per 512×512 frame. LPIPS was not computed. The protocol has two stages: (i) 64×64 proxy synthesis driven by the fitness (6); (ii) full-res retrain from scratch of the best synthesized chromosomes and of both U-Nets on train=5362, with early stopping on val=670 and the final evaluation on test=670 only (Table IV). Implementation: TensorFlow 2.15, GPU NVIDIA L4 (24 GB). The headline numbers are those of the holdout (Section V, Tables VIII and IX). The val values during retrain (Section V, Tables VI and VII) serve as training control only.

TABLE IV. PARAMETERS OF THE FULL-RES RETRAIN AND OF THE HOLDOUT (LETTERBOX 512)

Parameter	Value
max_side	512 px (letterbox)
batch (train / val)	2 / 2
epochs	30, early stopping patience=5
train stream	tf.data, prefetch(2)
train / val / test	5362 / 670 / 670
final metric	holdout PSNR/SSIM on the valid test region

V. RESULTS

The results are presented below following the objectives of Section III: the course of the structural-parametric synthesis, the correctness of the protocol, the quality / size / GMACs trade-off and the effect of the synthesis variants.

At the proxy stage (64×64) the GA selects an encoder-decoder by val PSNR. The dynamics of the best PSNR over generations are given in Table V and in Fig. 6. These numbers describe the course of the synthesis. They do not replace the holdout.

TABLE V. SYNTHESIS CONVERGENCE: BEST VAL PSNR OVER GENERATIONS (PROXY 64×64)

Method	Gen 0	Gen 7	Gen 14
Baseline GA	24.26 dB	25.72 dB	27.45 dB
Warm Start GA	23.55 dB	29.60 dB	31.49 dB

Figure 6 shows the dynamics of the best val PSNR over GA generations for Baseline GA and Warm Start GA.

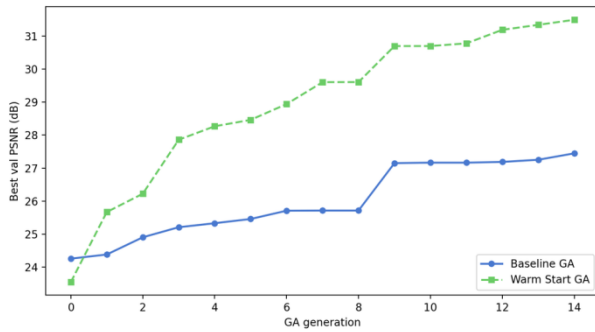


Fig. 6. Dynamics of the best val PSNR over GA generations (64×64 proxy synthesis)

The synthesis converges. Both variants raise the best val PSNR from gen 0 to gen 14. At the proxy stage Warm Start looks faster (gen 7 – 29.60 dB against 25.72 dB). Part of the advantage is explained by the inherited weights under the same epoch budget per evaluation. After an identical full-res retrain, the holdout comparison is given below (Tables VIII and IX). The proxy PSNR gap does not transfer one to one to full resolution.

The splits are 5362 / 670 / 670 with no overlap of names (see Table I). Test entered neither the synthesis fitness nor early stopping. All final rankings come from the holdout on the valid region only (Tables VIII and IX). During the full-res retrain, val PSNR/SSIM were recorded as training control (Tables VI and VII), not as a basis for the conclusions of the paper.

TABLE VI. FULL-RES RETRAIN: INTERNAL VAL QUALITY (N = 670; NOT A HOLDOUT)

Method	Val PSNR ↑ (dB)	Val SSIM ↑
U-Net (thin)	23.52	0.915
Wider U-Net	23.88	0.922
Baseline GA	23.69	0.917
Warm Start GA	23.85	0.922

TABLE VII. FULL-RES RETRAIN: PARAMS AND TRAINING Time

Method	Params	Train time (h)
U-Net (thin)	0.47M	1.1
Wider U-Net	1.88M	2.8
Baseline GA	1.15M	1.9
Warm Start GA	1.15M	3.1

Val was used for early stopping. Test was evaluated separately (Table VIII). The retrain val metrics do not necessarily follow the valid-region protocol of the holdout. The test values are the ones taken for publication. Time is wall-time on an NVIDIA L4.

The comparison on the independent test set (valid letterbox region, without padding; mean padding share $\approx 32\%$) is given in Tables VIII and IX. Mean $\pm$ std of the per-image metrics is reported ($n = 670$).

All the neural networks were trained from scratch on train=5362. Test was not used for selection. LPIPS was not computed. Wider U-Net lies outside the synthesis space (filters up to 128). The comparison with U-Net is not equal-budget with respect to the synthesis cost. GMACs are given per 512×512 frame.

TABLE VIII. HOLDOUT: RESTORATION QUALITY (N = 670, VALID REGION, MEAN $\pm$ STD)

Method	PSNR ↑ (dB)	SSIM ↑	PSNR ↑ (mask)
Identity	19.24 $\pm$ 3.15	0.849 $\pm$ 0.068	10.18 $\pm$ 1.74
Telea	22.97 $\pm$ 3.54	0.901 $\pm$ 0.046	13.96 $\pm$ 2.35
U-Net (thin)	24.73 $\pm$ 4.09	0.918 $\pm$ 0.041	16.11 $\pm$ 3.61
Baseline GA	24.98 $\pm$ 4.18	0.919 $\pm$ 0.041	16.15 $\pm$ 3.46
Warm Start GA	25.23 $\pm$ 4.39	0.924 $\pm$ 0.039	16.35 $\pm$ 3.70
Wider U-Net	25.33 $\pm$ 4.50	0.924 $\pm$ 0.040	16.40 $\pm$ 3.74

The highest mean PSNR on this holdout belongs to the trained Wider U-Net (25.33 dB). For these models Warm Start GA falls behind by 0.10 dB (paired Wilcoxon on per-image PSNR, $p \approx 3 \times 10^{-6}$). At the same time it has about 39% fewer parameters and about 17% fewer GMACs (Table IX: 97.3 against 116.6). Both synthesized models are above the thin U-Net. Warm Start GA is above Baseline GA by 0.25 dB ($p \approx 2 \times 10^{-24}$) in this experiment. This does not generalize warm start as a method beyond this run. All the neural networks are substantially better than identity (about +6 dB) and Telea (about +2 dB). The models perform useful restoration rather than merely copying the undamaged pixels.

TABLE IX. HOLDOUT: MODEL SIZE AND COMPUTATIONAL COST

Method	Params	GMACs
Identity	—	—
Telea	—	—
U-Net (thin) [14]	0.47M	29.3
Baseline GA	1.15M	97.3
Warm Start GA	1.15M	97.3
Wider U-Net (outside the synthesis space)	1.88M	116.6

Qualitative examples for the same input (Fig. IV) are given in Figs 7–10. The quantitative conclusions are taken from Tables VIII and IX and from the Wilcoxon test.

Figure 7 shows the restoration by the thin U-Net.

Figure 8 shows the restoration by the Baseline GA model.



Fig. 7. Su-57 restoration by the thin U-Net (input – Fig. 4)



Fig. 8. Su-57 restoration by the Baseline GA model (input – Fig. 4)



Fig. 9. Su-57 restoration by the Warm Start GA model (input – Fig. 4)



Fig. 10. Su-57 restoration by the Wider U-Net model (input – Fig. 4)

Figure 9 shows the restoration by the Warm Start GA model.

Figure 10 shows the restoration by the Wider U-Net model.

Thin U-Net: depth=3, filters=[32, 64, 128], skips enabled. Wider U-Net: [64,128,256] (outside the synthesis space). Baseline GA and Warm Start GA converged to the same topology [64, 128, 128]. It is wider than the thin U-Net and narrower than the Wider U-Net at the last level. They differed in lr and dropout (6.3×10^{-4} / 0.136 against 4.4×10^{-4} / 0.060). That is, with a shared structure the parametric genes differed. Retrain: batch=2, epochs=30 (Table IV). Params and GMACs are given in Table IX.

On this holdout the difference of 0.25 dB between these trained models under a shared topology points to the effect of the selected hyperparameters and of the warm start trajectory in this run, not to a different architecture. The compactness relative to Wider U-Net is partly built in by the limitation of the synthesis space (max 128 filters).

The aim of the experiment is not to maximize PSNR at any cost but to check whether structural-parametric synthesis can produce a restorer for the quality / size / GMACs trade-off. On this holdout (Tables VIII and IX) the trained Wider U-Net has the highest mean PSNR (25.33 dB). Warm Start GA falls behind statistically significantly on per-image PSNR, but only by 0.10 dB, with 1.15M parameters against 1.88M and 97.3 GMACs against 116.6. The neural networks clearly outperform identity and Telea. In this experiment Warm Start GA is better than Baseline GA by 0.25 dB under the same topology. This does not prove the advantage of warm start in general without a multi-seed rerun. The practical conclusion is that for these models the synthesis yields a competitive point on the quality / size / GMACs curve. The compactness relative to Wider U-Net has to be read together with the fact that a level of 256 filters was outside the synthesis space.

The limitations of the study are as follows. The damage is synthetic object-targeted Random

Erasing, not real occlusion. The comparison with the hand-designed baselines is not equal-budget with respect to the synthesis cost (the proxy stage takes ≈ 180 chromosome evaluations; the total GPU hours were not measured by a separate log aggregate; the wall-time of the full-res retrain is given in Table VII). Therefore the claim “synthesis is cheaper than manual tuning” is not asserted quantitatively. The loss (3) is not mask-aware. The metrics are PSNR/SSIM without LPIPS [21]. GMACs are an analytical estimate, not an onboard latency. The 64×64 proxy ranks the full-res models only partially. There is no separate hand-designed baseline d3_f64-128-128_s11 with the U-Net hyperparameters. The paired Wilcoxon test concerns per-image differences between the particular trained models on this holdout, not a multi-seed advantage of the method in general. Future work: equal-budget tuning of the hand-designed baselines (including d3_f64-128-128), extension of the synthesis space to 256 filters, a mask-aware loss, multi-objective NSGA-II [11] up to the holdout, LPIPS [21], real irregular occlusions, a measured onboard latency, and publication of the GPU hours of the proxy synthesis from the logs.

VI. CONCLUSIONS

An approach to structural-parametric synthesis of neural networks (an encoder-decoder restorer) for supervised inpainting under object-targeted damage has been formulated and verified. It is a way to search for a quality / size / GMACs trade-off relative to a strong hand-designed Wider U-Net. The synthesis is implemented by a genetic algorithm with a shared chromosome of structure and hyperparameters.

A disjoint train/val/test protocol (5362 / 670 / 670) has been built, with the final evaluation on the holdout over the valid letterbox region (without padding) and per-image statistics (mean $\pm$ std, paired Wilcoxon).

On this holdout (Tables VIII and IX) Warm Start GA is close to Wider U-Net in PSNR at smaller params/GMACs and is substantially better than Telea / identity. In this experiment it is also above Baseline GA under a shared topology (without generalizing warm start beyond this run). This confirms the expediency of structural-parametric synthesis as an alternative to manual restorer tuning for the quality / size / GMACs trade-off, provided the boundaries of the synthesis space are acknowledged honestly.

REFERENCES

- [1] M. Z. Zgurovsky, V. M. Sineglazov, and O. I. Chumachenko, “Formation of Hybrid Artificial Neural Networks Topologies. Artificial Intelligence Systems Based on Hybrid Neural Networks: Theory and Applications,” *Studies in Computational Intelligence*, vol. 904, pp. 175–232, 2021. Cham: Springer, https://doi.org/10.1007/978-3-030-48453-8_3.
- [2] M. Z. Zgurovsky, V. M. Sineglazov, and O. I. Chumachenko, “Classification and Analysis of Multicriteria Optimization Methods. Artificial Intelligence Systems Based on Hybrid Neural Networks: Theory and Applications,” *Studies in Computational Intelligence*, vol. 904, pp. 59–174, 2021. Cham: Springer. https://doi.org/10.1007/978-3-030-48453-8_2.
- [3] V. Sineglazov, M. Hordun, and S. Junttila, “An intelligent system for determining the degree of tree bark beetle damage based on the use of generative-adversarial neural networks,” *Plant-Environment Interactions*, vol. 5, no. 6, 2024, Article e70015. <https://doi.org/10.1002/pei3.70015>.
- [4] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning Transferable Architectures for Scalable Image Recognition,” *CVPR*, pp. 8697–8710, 2018. <https://doi.org/10.1109/CVPR.2018.00907>.
- [5] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le, “Regularized Evolution for Image Classifier Architecture Search,” *AAAI*, vol. 33, no. 1, pp. 4780–4789, 2019. <https://doi.org/10.1609/aaai.v33i01.33014780>.
- [6] H. Liu, K. Simonyan, and Y. Yang, “DARTS: Differentiable Architecture Search,” *ICLR*, 2019. arXiv:1806.09055. <https://doi.org/10.48550/arXiv.1806.09055>.
- [7] H. Pham, M. Guan, B. Zoph, Q. Le, and J. Dean, “Efficient Neural Architecture Search via Parameter Sharing,” *ICML, PMLR* 80, pp. 4095–4104, 2018. arXiv:1802.03268. <https://doi.org/10.48550/arXiv.1802.03268>.
- [8] L. Xie and A. Yuille, “Genetic CNN,” *ICCV*, pp. 1388–1397, 2017. <https://doi.org/10.1109/ICCV.2017.154>.
- [9] Z. Lu, I. Whalen, V. Boddeti, Y. Dhebar, K. Deb, E. Goodman, and W. Banzhaf, “NSGA-Net: Neural Architecture Search using Multi-Objective Genetic Algorithm,” *GECCO*, pp. 419–427, 2019. <https://doi.org/10.1145/3321707.3321729>.
- [10] N. Sinha and K.-W. Chen, “Evolving Neural Architecture Using One Shot Model,” *GECCO*, pp. 910–918, 2021. <https://doi.org/10.1145/3449639.3459275>.
- [11] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, 2002, pp. 182–197. <https://doi.org/10.1109/4235.996017>.
- [12] A. Criminisi, P. Perez, and K. Toyama, “Region Filling and Object Removal by Exemplar-Based

- Image Inpainting,” *IEEE Transactions on Image Processing*, vol. 13, no. 9, 2004, pp. 1200–1212. <https://doi.org/10.1109/TIP.2004.833105>.
- [13] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising,” *IEEE Transactions on Image Processing*, vol. 26, no. 7, 2017, pp. 3142–3155. <https://doi.org/10.1109/TIP.2017.2662206>.
- [14] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical DOI Image Segmentation,” *MICCAI*, pp. 234–241, 2015. https://doi.org/10.1007/978-3-319-24574-4_28.
- [15] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, “Context Encoders: Feature Learning by Inpainting,” *CVPR*, pp. 2536–2544, 2016. <https://doi.org/10.1109/CVPR.2016.278>.
- [16] G. Liu, F. A. Reda, J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, “Image Inpainting for Irregular Holes Using Partial Convolutions,” *ECCV*, pp. 89–105, 2018. https://doi.org/10.1007/978-3-030-01252-6_6.
- [17] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, “RePaint: Inpainting using Denoising Diffusion Probabilistic Models,” *CVPR*, pp. 11451–11461, 2022. <https://doi.org/10.1109/CVPR52688.2022.01117>.
- [18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, 2004, pp. 600–612. <https://doi.org/10.1109/TIP.2003.819861>.
- [19] T. DeVries and G. W. Taylor, “Improved Regularization of Convolutional Neural Networks with Cutout,” arXiv:1708.04552, 2017. <https://doi.org/10.48550/arXiv.1708.04552>.
- [20] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random Erasing Data Augmentation,” *AAAI*, vol. 34, no. 7, pp. 13001–13008, 2020. <https://doi.org/10.1609/aaai.v34i07.7000>.
- [21] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric,” *CVPR*, pp. 586–595, 2018. <https://doi.org/10.1109/CVPR.2018.00068>.

Received: July 19, 2026

Accepted: August 12, 2026

Published: September 18, 2026

Chumachenko Olena. ORCID 0000-0003-3006-7460. Doctor of Engineering Science. Professor.

Department of Artificial Intelligence, Educational and Research Institute for Applied System Analysis, National Technical University of Ukraine “Ihor Sikorsky Kyiv Polytechnic Institute,” Kyiv, Ukraine.

Education: Georgian Polytechnic Institute, Tbilisi, Georgia, (1980).

Research area: system analysis, artificial neural networks.

Publications: more than 80 papers.

E-mail: eliranvik@gmail.com Scopus: 55635565800

Kot Anatoly. ORCID 0000-0002-7490-8834. Senior Lecturer.

Artificial Intelligence Department, Educational and Research Institute for Applied System Analysis, National Technical University of Ukraine “Ihor Sikorsky Kyiv Polytechnic Institute,” Kyiv, Ukraine.

Education: National Technical University of Ukraine “Ihor Sikorsky Kyiv Polytechnic Institute”, (2025).

Research area: artificial Intelligence.

Publications: 13 publications.

E-mail: anatoly.kot@gmail.com Scopus: 57219503223

О. І. Чумаченко, А. Т. Кот. Структурно-параметричний синтез нейронних мереж для відновлення зображень у разі пошкодження синтетичними об'єктами

Статтю присвячено структурно-параметричному синтезу нейронних мереж для керованого відновлення на парах «пошкоджений – оригінал». В якості алгоритма оптимізації використано генетичний алгоритм. Навчальні дані синтезуються за допомогою об'єктно-орієнтованого випадкового стирання. Розподіл потяг / значення / тест не перетинається: 5362 / 670 / 670. На виході ($n = 670$; дійсна область letterbox, без доповнень) Wider U-Net дає $25,33 \pm 4,50$ дБ (1,88 млн; 116,6 GMAC), а Warm Start GA дає $25,23 \pm 4,39$ дБ (1,15 млн; 97,3 GMAC). За парним тестом Вілкоксона, розрив від Wider U-Net становить 0,10 дБ, а перевага над Baseline GA становить 0,25 дБ у цьому запуску. Обидві синтезовані мережі значно кращі за Telea та identique. Ширина рівнів у просторі синтезу обмежена 128 фільтрами, тому Wider U-Net (256) знаходиться поза ним. Результат інтерпретується як конкурентний компроміс, без претензії на абсолютне лідерство PSNR та без узагальнення теплого запуску за межі цього експерименту.

Ключові слова: структурно-параметричний синтез, відновлення зображень, інпайнінг, нейронні мережі, генетичний алгоритм, випадкове стирання, пікове співвідношення сигнал/шум, індекс структурної подібності, кодер-декодер, затримка.

Чумаченко Олена Іллівна. ORCID 0000-0003-3006-7460. Доктор технічних наук. Професор.

Кафедра штучного інтелекту, Навчально-науковий інститут прикладного системного аналізу, Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.

Освіта: Грузинський політехнічний інститут, Тбілісі, Грузія, (1980).

Напрямок наукової діяльності: системний аналіз, штучні нейронні мережі.

Кількість публікацій: більше 80 наукових робіт.

E-mail: eliranvik@gmail.com Scopus: 55635565800

Кот Анатолій Тарасович. ORCID 0000-0002-7490-8834. Старший викладач.

Кафедра штучного інтелекту, Навчально-науковий інститут прикладного системного аналізу, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Освіта: Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна, (2025).

Напрямок наукової діяльності: штучні нейронні мережі, програмування.

Кількість публікацій: 12 публікацій.

E-mail: anatoly.kot@gmail.com Scopus: 57219503223