

# Evolutionary Neural Architecture Search for Image Restoration

Gerard Jacques van Wyk  
*Department of Computer Science*  
*University of Pretoria*  
 Pretoria, South Africa  
 gerard.wyk@gmail.com

Anna Sergeevna Bosman  
*Department of Computer Science*  
*University of Pretoria*  
 Pretoria, South Africa  
 annar@cs.up.ac.za

**Abstract**—Convolutional neural network (CNN) architectures have traditionally been explored by human experts in a manual search process that is time-consuming and ineffectively explores the massive space of potential solutions. Neural architecture search (NAS) methods automatically search the space of neural network hyperparameters in order to find optimal task-specific architectures. NAS methods have discovered CNN architectures that achieve state-of-the-art performance in image classification among other tasks, however the application of NAS to image-to-image regression problems such as image restoration is sparse. This paper proposes a NAS method that performs computationally efficient evolutionary search of a minimally constrained network architecture search space. The performance of architectures discovered by the proposed method is evaluated on a variety of image restoration tasks applied to the ImageNet64x64 dataset, and compared with human-engineered CNN architectures. The best neural architectures discovered using only 2 GPU-hours of evolutionary search exhibit comparable performance to the human-engineered baseline architecture.

**Index Terms**—artificial neural networks, neural architecture search, convolutional neural networks, genetic algorithms, image restoration

## I. INTRODUCTION

It would be hugely inefficient to set many thousands of weight parameters in a modern neural network (NN) by hand, yet many hyperparameters of NN architectures are currently hand-crafted by human experts. Early convolutional neural networks (CNNs) [1] contained few hyperparameters, and a small variety of primitive building blocks connected together in simple topologies. In contrast, modern CNN architectures are embedded within an ever-growing search space of possible designs, as researchers continuously invent novel gradient-based optimisation algorithms [2], [3], differentiable weight-containing layers [4], activation functions [5]–[7], normalisation methods [8], network topology schemes [9], and many other forms of algorithmic and architectural improvements.

As the search space of potential neural architectures grows, it becomes less likely that any given pre-existing network architecture is still the best solution to the problem it was originally designed for. Neural architecture search (NAS) methods attempt to automate the process of finding optimal neural architectures for any given task. NAS methods generally achieve this by treating neural architecture design as an optimisation problem, where the objective is to discover architectures with

minimal validation loss for a given task. NAS methods have been demonstrated to be able to discover neural architectures that yield performance comparable or superior to human-engineered neural architectures in the domain of image classification [10]–[14] and image restoration [15]. Most existing NAS methods achieve success by severely limiting the search space of possible solutions.

Image restoration problems are a broad class of image-to-image regression problems where the objective is to reconstruct an original image from a corrupted input. Image restoration problems are of significant scientific and commercial interest. An example scientific application of image restoration is reversal of optical distortions present in optical imaging instruments such as microscopes and telescopes.

This study aims to describe and evaluate a novel NAS method to automate the process of finding optimal CNN architectures for arbitrary image restoration problems. Novel contributions of this work are summarised as follows:

- A neural architecture search space is proposed that is both highly expressive and highly searchable.
- Adaptive average pooling is effectively employed to eliminate topological constraints.
- The feasibility of performing NAS for image-to-image architectures under significant memory and computational time constraints is demonstrated.

The rest of the paper is structured as follows: Section II discusses the background and related work. Section III describes the proposed search space. Section IV describes the proposed evolutionary algorithm. Section V discusses the experimental setup. Section VI presents the empirical results. Section VII summarises the findings of this study, and suggests topics for future research.

## II. BACKGROUND AND RELATED WORK

Image restoration is a subset of image-to-image regression problems where the objective is to accurately reconstruct an original image, given a corrupted input. Image restoration problems that have been investigated in the context of deep learning include: image denoising [16], image deblurring [17], single-image superresolution [18], and compressive sensing [19], among others. Given a dataset of corrupted and original images, a model can be trained to accept corrupted

images as input, and produce restored images as output. The original images are used as target outputs in the process of supervised learning.

CNNs have a strong prior for the structure of images [20], and have been demonstrated to be highly adept at a wide variety of image processing tasks, including various image restoration tasks [16]–[19]. Whilst CNNs eliminate the need to design problem-specific image restoration algorithms, they introduce the problem of finding optimal hyperparameter values and designing an appropriate network topology.

To the best of author’s knowledge, the only research to date on the topic of NAS methods for image restoration is by Suganuma et al. [15], where a neural architecture search space is defined that incorporates a symmetrical convolutional autoencoder architecture constraint in order to reduce the search space complexity. Size mismatches are avoided by allowing skip connections only between layers of the same size. While this search space is highly searchable, it can only represent a very limited section of the true underlying architectural search space. Notable restrictions of their approach include only having ReLU activation functions, only being able to express convolutional autoencoder architectures with skip connections, lack of normalisation layers, and no searchable convolutional layer hyperparameters.

In general, the existing NAS approaches [10]–[14] have two major limitations:

- 1) A restrictive set of hyperparameter values available to the search algorithm.
- 2) A lack of an efficient approach to deal with size and dimensionality mismatches between successive convolutional blocks.

This study proposes an expressive search space by providing an extensive list of hyperparameters and modules available to the evolutionary search algorithm. Additionally, dimensionality constraints are alleviated by applying adaptive average pooling between mismatching modules to enforce valid architectures.

### III. NEURAL ARCHITECTURE SEARCH SPACE

The neural architecture search space is arguably the core of any NAS method, as it defines the set of all architectures that could possibly be discovered by the search process. Designing a good architecture search space presents a difficult trade-off, where adding more degrees of freedom increases the number of (potentially superior) unique architectures that can be represented, simultaneously increasing the difficulty of the optimisation problem. If the architecture search space is overly constrained, then high performance architectures can not exist within the search space. This section describes the hyperparameters and search space constraints of the proposed neural architecture representation.

#### A. Network topology representation

To represent the connected topology of NN architectures, an acyclic directed graph with a single input node and a single output node is used in this study. An adjacency matrix is used

to represent the graph. Enforcing a single input and a single output structure is justified in the context of image restoration tasks, as the final model is expected to receive a single image as input and produce a single image as output.

Each node in the graph constitutes a NN primitive. A primitive is defined as any optionally parameterised differentiable function that receives a tensor input and produces a tensor output. In order to define the network topology as a graph, connective primitives are required that can take multiple previous nodes as input. Previous NAS research generally made use of layer concatenation or elementwise addition. For the proposed architecture representation, the following connective primitives are used: depthwise concatenation, elementwise addition, and elementwise multiplication. Thus, each primitive has one or two predecessor nodes in the adjacency matrix, depending whether it is a single-input primitive or a connective primitive.

The connective operations listed above all share an important constraint: they can only be applied to two inputs of matching dimensionality. This can become a hindrance to an evolutionary algorithm, as crossover and random mutations may introduce dimensionality mismatches, thus generating invalid architectures. This study proposes a novel method to alleviate this constraint: If the shapes of the two input nodes are incompatible, then one of the inputs is resized using adaptive average pooling to ensure compatibility. Adaptive average pooling is simply an average pooling operation that, given an input and output dimensionality, calculates the correct kernel size necessary to produce an output of the given dimensionality from the given input. This simple operation is expected to make the search space significantly more searchable and expressive, since potentially invalid architectures, instead of being discarded, will be converted to valid architectures.

The output of the last executed primitive is taken as the output of the network. This is usually the last primitive in the graph, unless the generated graph exceeded the memory limit, in which case an earlier primitive’s output may be used.

#### B. Neural network primitives

A shortcoming of the existing NAS methods is the use of a small variety of unique NN primitives [10]–[15]. To counter this deficiency, the proposed architecture search space made use of a diverse set of activation functions, normalisation layers, and convolutional layer hyperparameters sourced from recent advancements in neural architecture design. Expanding the set of primitives increases the descriptive power of the network architecture representation at the cost of searchability.

The following activation functions were available to the proposed NAS algorithm: ReLU [21], PReLU [7], ELU [6], SELU [5], hyperbolic tangent (tanh), sigmoid, and softmax.

The following normalisation layers were used: batch normalisation [8], instance normalisation [22], and local response normalisation [1].

For spatial resolution altering primitives,  $2 \times 2$  max pooling [16] and nearest neighbour upscaling were employed.

*Convolutional primitive:* For the convolutional primitive, another divergence was made from previous NAS research.

Rather than have several convolutional primitive block types each with pre-set parameters, such as  $3 \times 3$  depthwise convolution block or a  $3 \times 3$  spatial convolution block, this work proposes a single convolutional block type that takes several constrained parameters. The number of input channels is determined by the number of channels in the predecessor node, determined topologically. The number of output channels is constrained to 7 options: same as input channels,  $2 \times$  input channels, input channels/2,  $4 \times$  input channels, input channels/4, 3, and 32. The kernel size of a convolutional layer can be  $1 \times 1$ ,  $3 \times 3$ , or  $5 \times 5$ . The stride can be 1 or 2. A convolutional layer can be transposed or regular. A convolutional layer can be a separable depthwise convolution [23]. Weight normalisation can be set to True or False. The number of parameters of the topological representation is reduced by grouping convolutional blocks together with an activation function and a normalisation layer, since such combination is typically present in human-engineered architectures.

### C. Gradient optimisation hyperparameters

Existing NAS methods [10]–[15] generally do not include gradient optimisation hyperparameters in their NN architecture representations. The proposed approach included a parameter to specify the optimiser type between a choice of Adam [2], RMSprop [3], and stochastic gradient descent (SGD) with momentum. A parameter was also included to specify the initial learning rate, constrained to the continuous range of  $[1 \times 10^{-1}, 1 \times 10^{-5}]$ , and a learning rate decay parameter, constrained to the continuous range of  $[0, 1]$ . Optimising the training algorithm for a specific task is important to the success of any NN architecture, thus this minor increase in search space complexity is considered worthwhile.

### D. Network constraints

A number of constraints were imposed on the evolved network representation to limit the search space, and to ensure execution within the predefined computational budget.

*Memory usage:* During a NN’s execution, layers were computed and added to a stack of intermediate tensors that were available as potential inputs to all successive layers as determined by the architecture topology. The execution ended either when the graph was completed, or when memory usage exceeded the set memory limit. In either case, the last tensor in the stack was resized to the target output shape. The memory usage of a NN was estimated as approximately equal to the sum of elements across all the layers at the moment of execution.

*Time to execute:* If a NN took more than 50 seconds to execute the first 1000 iterations of gradient descent, then gradient optimisation was halted for that individual.

## IV. EVOLUTIONARY ALGORITHM FOR NEURAL ARCHITECTURE SEARCH

A simple evolutionary algorithm approach was taken. A population of size  $N$  was randomly initialised, and the number of allowed gradient iterations was set to  $i$ . For each generation,

all individuals were trained for  $i$  iterations on the training set, then the fitness of each individual was evaluated on 1000 minibatches from the validation set. Then, the population was sorted by fitness, and the worst half of the population was killed. If the size of the resulting population was below the minimum population size  $n$ , the entire population was cloned. Then,  $E$  number of elites, or best individuals, were copied directly to the next generation. Crossover operation with probability  $p_c = 0.5$  was applied to the rest of the individuals (second parent was randomly chosen from the population), and resulting offspring were used to replace the parents. Uniform crossover was used for the gradient optimisation hyperparameters, and random single-point crossover was used for the primitives. Finally, all individuals were mutated. This process repeated until the predefined computational budget expired.

To promote convergence, an initially large population of  $N$  individuals was gradually reduced to the minimal size  $n$ . Thus, exploration was emphasised at the beginning of the search, with a strong shift towards exploitation at the end of the search.

A high mutation rate of 50% mutation probability was required to aggressively search the architecture space in the few generations ( $< 20$ ) of evolutionary search within the target 2 hour time limit. The following mutation rules were used:

*Graph:* The adjacency matrix that represents the NN topology was mutated by randomly flipping a bit below the diagonal of the matrix. This mutation randomly connects or disconnects primitives, thus it was possible to generate primitives with no input connections. In this case, the said primitive was connected to the nearest preceding primitive. After the mutations took place, graph pruning was performed to remove primitives with no causal connection to the final output.

*Primitives:* Network primitives were mutated by adding a primitive, deleting a primitive, or mutating an existing primitive’s hyperparameters. Primitive hyperparameters were mutated by being reinitialised on a per-parameter basis according to the mutation probability.

## V. EXPERIMENTAL SETUP

This section describes the experimental setup of the study. Section V-A describes the human-engineered CNN architecture used as the baseline. Section V-B describes the dataset and the image restoration tasks used in the experiments. Section V-C discusses the fitness function used to evaluate the individuals. Section V-D lists the hyperparameter values used for the evolutionary search. Section V-E lists the hardware and software used to conduct the experiments.

### A. Human-engineered baseline network

In order to evaluate the performance of the architectures produced by the proposed NAS method, a human-engineered architecture is required as a performance baseline. This baseline NN should ideally be the state-of-the-art architecture for the set of image restoration problems being investigated. For this purpose, a modified version of the U-Net [9] architecture was used with PReLU activations, batch normalization, the

Adam optimiser, an initial learning rate of 0.01 halved every 2000 iterations, and a squeeze-and-excitation module [4] inserted after batch normalisation in each convolutional block.

For each problem, the baseline CNN was trained for 20,000 iterations using minibatch training with a batch size of 8.

### B. Dataset and image restoration tasks

Due to the nature of image restoration problems, any image dataset can be converted to an image restoration dataset by generating input-output pairs required for learning. Corrupted input images are produced by applying a task-specific image degradation function to the original images, and the original images are used as target output. In order to evaluate the performance of the proposed NAS method and the human-engineered architecture, the following image restoration tasks were used as benchmarks: single image superresolution, uniform random noise image denoising, Gaussian random noise image denoising, image deblurring, compressive sensing, and checkerboard rendering reconstruction.

*Single image superresolution:* Due to the baseline architecture expecting inputs and outputs to be of the same size, the low-scale input images were resized with nearest neighbour upscaling before being given to the network as input.

*Compressive sensing:* Given a random 25% of image pixels, the network was supposed to reconstruct the missing values.

*Checkerboard rendering reconstruction:* Checkerboard rendering reconstruction is a technique used to optimise real-time graphics upscaling in computer graphics [24]. To the best of author’s knowledge, this is the first time deep NNs have been used to learn a checkerboard rendering reconstruction filter.

To evaluate the performance of various neural architectures for the above image restoration tasks, an image dataset is required. In choosing a dataset, the following attributes were considered as desirable: a large sample count, a large diversity of natural images, colour images, and a resolution that is large enough to be representative of real world data, while being small enough to quickly train and evaluate a large number of candidate architectures. We eliminated the original version of ImageNet [25] for having a resolution that is too large for the given time and memory constraints. Taking all the desirable attributes into account, the ImageNet64x64 dataset [26] was chosen. The Imagenet64x64 training set consists of 1,281,167 training images divided into 1000 classes. As the name suggests, it is the ImageNet dataset downsampled to a resolution of  $64 \times 64$  pixels.

### C. Evaluating the performance of the proposed architectures

In order to evaluate a network architecture on a given problem, the dataset was split into training, validation, and test sets. The training set was used for gradient-based optimisation of each individual NN, the validation set was used to estimate a NN’s performance on unseen images, and a separate test set acted as a measure of performance of a given NN on unseen data. The test set was not observed until the final experiments, or used in any kind of gradient-based or evolutionary optimisation.

For the training/validation/test set split, the ImageNet64x64 dataset was sequentially split into training, validation, and test sets using the ratios of 0.6/0.2/0.2. The same splits were retained across all experiments.

The chosen loss function across all experiments was the mean squared error (MSE), as it is commonly used for image restoration. In the experimental results, the MSE loss is expressed in the form of peak signal to noise ratio (PSNR). PSNR is a reparameterisation of MSE calculated as:

$$\text{PSNR} = 10 \log_{10}(1/\text{MSE})$$

A higher PSNR indicates higher quality of image restoration.

### D. Evolutionary architecture search

For each image restoration task, the evolutionary architecture search method trained a population of 32 individuals for 20,000 iterations of gradient-based optimisation per individual on the training set, and evaluated the performance of each individual using 1000 minibatches of size 8 from the validation set. After 2 hours, architecture search was halted, and the performance of the best discovered architecture was evaluated.

### E. Hardware and software used

All experiments were performed on a system with an i7-8700k, a single GTX 1080 ti GPU with 11GB of VRAM, and 16GB of system memory. The dataset was stored on a SSD. The project was implemented in the PyTorch deep learning framework [27].

## VI. RESULTS

This section presents the results of the experiments conducted. The mean PSNR values for all experiments are summarised in Table I.

It is evident from the PSNR values that the human-engineered architecture performed better than the best evolved architecture on most image restoration problems, with the exception of the single image superresolution, where both

TABLE I: Mean PSNR results for image restoration tasks

Image Restoration Task	Data Subset	Baseline PSNR	Evolved PSNR
Single Image Superresolution ( $2 \times 2$ upscaling)	Training	23.7278	23.7280
	Validation	23.7251	23.7336
	Test	23.7315	23.7321
Denoising (Uniform noise $\sim [-0.5, 0.5]$ )	Training	23.5569	21.3105
	Validation	23.5585	21.3039
	Test	23.5605	21.3073
Denoising (Gaussian noise, $\sigma = 0.2$ )	Training	23.6653	21.7955
	Validation	23.6611	21.7894
	Test	23.6656	21.7929
Deblurring (Gaussian kernel, $\sigma = 2, r = 3$ )	Training	25.7073	23.8966
	Validation	25.7085	23.8922
	Test	25.7142	23.8916
Compressive Sensing	Training	27.0249	21.7724
	Validation	27.0210	21.7687
	Test	27.0217	21.7712
Checkerboard Rendering Reconstruction	Training	26.9629	25.6542
	Validation	26.9642	25.6549
	Test	26.9637	25.6539

architectures performed on par. However, the evolved architectures performed on a comparable level, which is impressive given the time and complexity constraints imposed on the evolutionary process. While the human-engineered architecture was heavily overparameterised, the evolved architectures were forced to learn to perform the same task with a significantly smaller number of total parameters. Fig. 1 and 2 illustrate the performance of the human-engineered and the evolved architectures on the six image restoration tasks applied to a set of 8 images selected from the publicly available Kodak image suite [28]. It is evident from the visual inspection that the evolved architectures performed adequately. In multiple cases, the difference in performance is barely noticeable.

Another interesting result is the tight coupling between the training, validation, and test set accuracies, i.e. very little to no overfitting. The minimal variance between the dataset splits is likely due to the large sample count of the dataset, and the good generalisation ability of the network architectures investigated.

#### A. Properties of evolved architectures

Example evolved architecture parameters are presented in Tables II, III, IV, and V. PyTorch syntax is used to describe the primitives.

Table II shows one of the larger evolved architectures, with a total of 14 nodes. Multiple convolutional blocks of varying dimensionality were evolved, and combined using adaptive average pooling. This architecture performed as well as the human engineered architecture, but was smaller in size.

Table III shows that for the denoising task, a very compact CNN architecture emerged. Same applies to the checkerboard reconstruction, shown in Table V. Both have employed concatenation of the earlier layer signals with the later layer signals, similar to the U-Net architecture.

Table IV shows an example of a simple feed-forward architecture evolved for the compressive sensing task. In this case, no concatenations took place, and sigmoidal functions were used in a number of layers. Compressive sensing task provided only 25% of valid inputs, which may have caused concatenation of the input signals to the hidden layer signals to be ineffective.

It is hard to draw definite conclusions from the small sample size of high performance architectures discovered by the proposed NAS method, but certain tendencies can be observed nevertheless. The high frequency of Adam and RMSprop as opposed to SGD training, the high frequency of transposed convolutional layers, and the high frequency of rectifier-based activation functions are in line with the existing best practices among CNN practitioners. Perhaps the most interesting property of the evolved architectures is the sheer diversity of network configurations that were produced. This serves as evidence for the expressivity of the proposed neural architecture search space.

## VII. CONCLUSIONS AND FUTURE WORK

This paper proposed a novel NAS method, comprised of an expressive yet compact search space, and a simple, rapidly

convergent evolutionary algorithm. Adaptive average pooling was employed to alleviate topological constraints caused by mismatching dimensions of successive convolutional blocks. The performance of the proposed method was evaluated on a variety of difficult image restoration tasks, applied to the ImageNet64x64 dataset. The performance of the discovered NN architectures was compared with a high-performance human-engineered CNN architecture. The NN architectures discovered by the proposed NAS method using only 2 GPU-hours yielded lesser, but comparable performance to the human-engineered baseline architecture. The discovered architectures were significantly smaller in size than the baseline architecture, yet yielded adequate performance, and demonstrated a diversity of topological structures. Thus, the proposed NAS method was capable of discovering compact, usable solutions under significant memory and computational time constraints.

The obvious first step in future work would be to drastically increase the computational budget. If 2 hours on a single consumer GPU can reliably yield usable results, then a large GPU cluster with a larger time budget should be able to significantly surpass the results presented in this paper. The ability of the proposed NAS method to find compact architectures makes it a good option for researchers and practitioners with limited computing resources.

There is a vast amount of potential in applying the proposed method to other image restoration problems, such as aperture synthesis in radio astronomy, as the current solutions are severely outdated hand-designed deconvolution algorithms. Improved image restoration techniques in this domain would unlock useful scientific data.

It would also be interesting to perform an in-depth study of the best architectures evolved, as the analysis of the evolved architectures may yield significant insights about NN architectures and search spaces at large.

## REFERENCES

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [2] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations*, May 2015.
- [3] G. Hinton, N. Srivastava, and K. Swersky, "RMSProp: Divide the gradient by a running average of its recent magnitude," *Neural networks for machine learning, Coursera lecture 6e*, 2012.
- [4] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2018.
- [5] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-Normalizing neural networks," in *Advances in Neural Information Processing Systems*, Dec. 2017.
- [6] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," in *Proceedings of the International Conference on Learning Representations*, May 2016.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.
- [8] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the International Conference on Machine Learning*, 2015.





Fig. 1: Visual comparison of the baseline and evolved NN architecture performance. Top rows: ground truth, middle rows: corrupted input, bottom rows: neural network output.

- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.
- [10] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le, "Regularized evolution for image classifier architecture search," *ArXiv e-prints*, Feb. 2018.
- [11] H. Pham, M. Y. Guan, B. Zoph, Q. V. Le, and J. Dean, "Efficient neural architecture search via parameter sharing," *ArXiv e-prints*, Feb. 2018.
- [12] E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. Le, and A. Kurakin, "Large-Scale evolution of image classifiers," *ArXiv e-prints*, Mar. 2017.
- [13] H. Liu, K. Simonyan, O. Vinyals, C. Fernando, and K. Kavukcuoglu, "Hierarchical representations for efficient architecture search," in *Proceedings of the International Conference on Learning Representations*, Nov. 2018.
- [14] L. Xie and A. Yuille, "Genetic CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, Oct. 2017.
- [15] M. Suganuma, M. Ozay, and T. Okatani, "Exploiting the potential of standard convolutional autoencoders for image restoration by evolutionary search," *ArXiv e-prints*, Mar. 2018.
- [16] J. Nagi, F. Ducatelle, G. A. D. Caro, D. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *Proceedings of the IEEE International Conference on Signal and Image Processing Applications*, Nov. 2011, pp. 342–347.
- [17] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, no. 2, 2017, p. 3.
- [18] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proceedings of the European Conference on Computer Vision*. Springer, 2014, pp. 184–199.
- [19] A. Mousavi, A. B. Patel, and R. G. Baraniuk, "A deep learning approach to structured signal recovery," in *Proceedings of the Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2015, pp. 1336–1343.
- [20] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," *ArXiv e-prints*, Nov. 2017.
- [21] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the International Conference on Machine Learning*, 2010, pp. 807–814.
- [22] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *ArXiv e-prints*, Jul. 2016.
- [23] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings on the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 1800–1807.
- [24] T. McFerron and A. Lake, "White paper: Checkerboard rendering for real-time upscaling on intel integrated graphics," Intel Corporation, Tech. Rep., August 2018.
- [25] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
- [26] P. Chrabaszcz, I. Loshchilov, and F. Hutter, "A downsampled variant of ImageNet as an alternative to the CIFAR datasets," *ArXiv e-prints*, Jul. 2017.
- [27] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *NIPS Workshop*, 2017.
- [28] R. Franzen, "Kodak lossless true color image suite," 1999, accessed 30-March-2019. [Online]. Available: <http://r0k.us/graphics/kodak/>

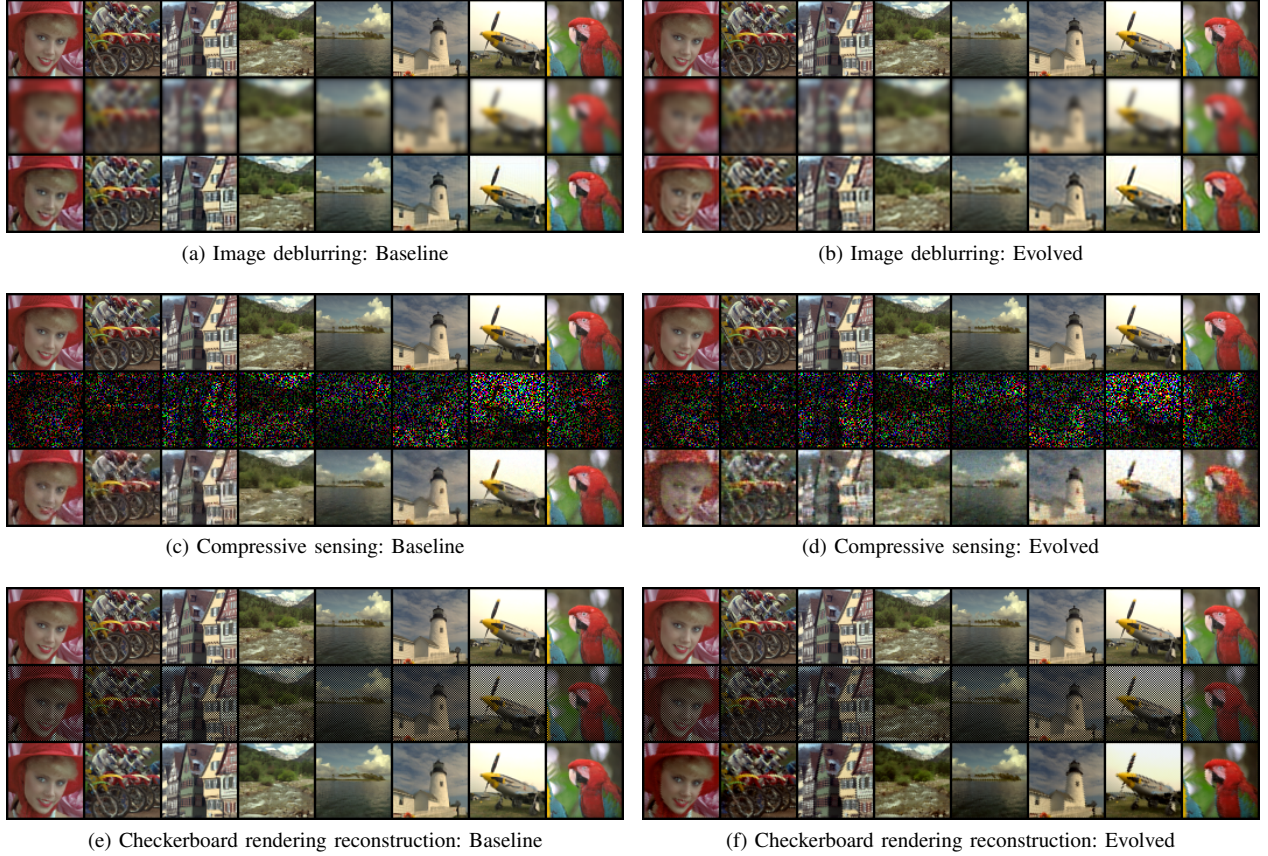


Fig. 2: Visual comparison of the baseline and evolved NN architecture performance. Top rows: ground truth, middle rows: corrupted input, bottom rows: neural network output.

TABLE II: Parameters of the best evolved architecture for the single image superresolution task.

Optimizer: Adam  
Initial learning rate: 0.041888  
Learning rate decay factor: 0.136235

Node index	Input Node(s)	Node type
0		Input node
1	0	(conv): ConvTranspose2d(3, 12, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), output_padding=(1, 1), groups=3, bias=False) (activ): PReLU(num_parameters=1) (norm): BatchNorm2d(12, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
2	1	upsample
3	0, 2	mul - resize to first
4	3	(conv): Conv2d(3, 32, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1)) (activ): SELU()
5	4	(conv): ConvTranspose2d(32, 16, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), output_padding=(1, 1)) (activ): SELU() (norm): LocalResponseNorm(16, alpha=0.0001, beta=0.75, k=1)
6	0, 5	add - resize to second
7	1	(conv): ConvTranspose2d(12, 48, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), output_padding=(1, 1), bias=False) (activ): ReLU() (norm): InstanceNorm2d(48, eps=1e-05, momentum=0.1, affine=False, track_running_stats=False)
8	7	upsample
9	7, 8	mul - resize to second
10	7, 9	add - resize to second
11	10	(conv): Conv2d(48, 3, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), bias=False) (activ): SELU() (norm): Softmax2d()
12	6, 11	mul - resize to first
13	12	(conv): ConvTranspose2d(16, 8, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), output_padding=(1, 1)) (activ): SELU() (norm): BatchNorm2d(8, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
14	13, 13	add - resize to first

TABLE III: Parameters of the best evolved architecture for the image denoising (Gaussian noise) task.

Optimizer: RMSprop  
Initial learning rate: 0.038556  
Learning rate decay factor: 0.133418

Node index	Input Node(s)	Node type
0		Input node
1	0	(conv): Conv2d(3, 12, kernel_size=(5, 5), stride=(2, 2), padding=(2, 2), groups=3, bias=False) (activ): PReLU(num_parameters=1)
2	1	(conv): ConvTranspose2d(12, 3, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), output_padding=(1, 1), bias=False) (activ): Tanh() (norm): LocalResponseNorm(3, alpha=0.0001, beta=0.75, k=1)
3	0	(conv): Conv2d(3, 1, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
4	1, 3	concat - resize to second
5	2, 3	concat - resize to first
6	4	(conv): ConvTranspose2d(13, 3, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), output_padding=(1, 1)) (activ): ELU(alpha=1.0) (norm): LocalResponseNorm(3, alpha=0.0001, beta=0.75, k=1)

TABLE IV: Parameters of the best evolved architecture for the compressive sensing task.

Optimizer: Adam  
Initial learning rate: 0.055330  
Learning rate decay factor: 0.250831

Node index	Input Node(s)	Node type
0		Input node
1	0	(conv): Conv2d(6, 32, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), bias=False) (activ): Sigmoid() (norm): InstanceNorm2d(32, eps=1e-05, momentum=0.1, affine=False, track_running_stats=False)
2	1	(conv): ConvTranspose2d(32, 16, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), output_padding=(1, 1)) (activ): Sigmoid() (norm): BatchNorm2d(16, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
3	2	(conv): ConvTranspose2d(16, 16, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), output_padding=(1, 1)) (norm): BatchNorm2d(16, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
4	3	(conv): Conv2d(16, 4, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1)) (activ): ReLU() (norm): BatchNorm2d(4, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
5	4	(conv): ConvTranspose2d(4, 3, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), output_padding=(1, 1), bias=False) (activ): Tanh() (norm): BatchNorm2d(3, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)

TABLE V: Parameters of the best evolved architecture for the checkerboard rendering reconstruction task.

Optimizer: RMSprop  
Initial learning rate: 0.077964  
Learning rate decay factor: 0.390523

Node index	Input Node(s)	Node type
0		Input node
1	0	(conv): ConvTranspose2d(3, 3, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), output_padding=(1, 1), groups=3) (activ): PReLU(num_parameters=1)
2	0, 1	add - resize to second input
3	0, 2	concat - resize to second input
4	1, 3	add - resize to second input
5	4	(conv): Conv2d(6, 3, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False) (activ): ELU(alpha=1.0) (norm): BatchNorm2d(3, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
6	5	(conv): Conv2d(3, 3, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), bias=False) (activ): PReLU(num_parameters=1) (norm): BatchNorm2d(3, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)