Multi-Level Generative Chaotic Recurrent Network for Image Inpainting

Cong Chen
Virginia Tech

Amos Abbott
Virginia Tech

Daniel Stilwell
Virginia Tech

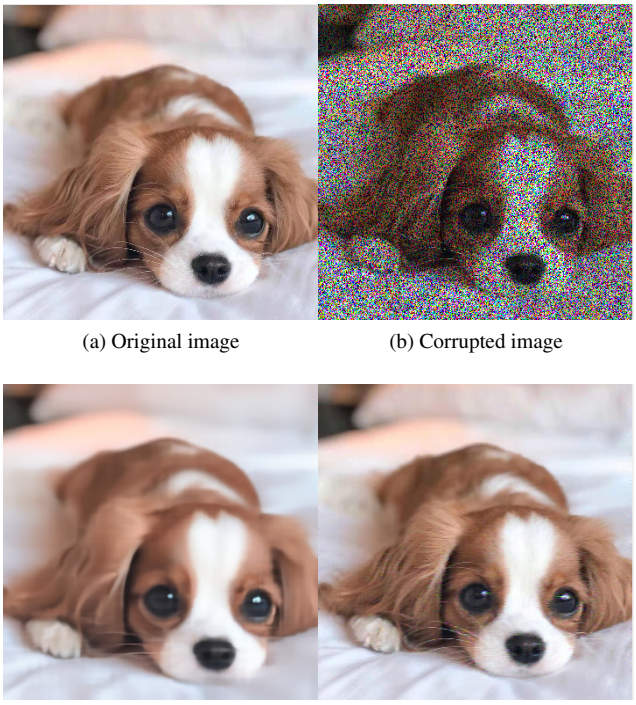
Abstract

This paper presents a novel multi-level generative chaotic Recurrent Neural Network (RNN) for image inpainting. This technique utilizes a general framework with multiple chaotic RNN that makes learning the image prior from a single corrupted image more robust and efficient. The proposed network utilizes a randomly-initialized process for parameterization, along with a unique quad-directional encoder structure, chaotic state transition, and adaptive importance for multi-level RNN updating. The efficacy of the approach has been validated through multiple experiments. In spite of a much lower computational load, quantitative comparisons reveal that the proposed approach exceeds the performance of several image-restoration benchmarks.

1. Introduction

A growing number of applications require the acquisition of high-quality images under noisy conditions or in situations where portions of an image are missing. For such applications, image restoration techniques such as inpainting are essential.

Historically, the image restoration problem has been addressed by modeling the natural image, and the image degradation process has been modeled by edge statistics [8, 31] and noise mask construction [1, 42]. Recent image restoration methods have utilized deep learning methods. Utilizing tools such as convolutional neural networks (CNN) [20, 22, 36] and generative adversarial networks (GAN) [10, 21], restoration methods have been developed that yield better results than previous hand-crafted methods (e.g., [41, 44, 43]). However, problems with deep learning methods include the need for significant tuning of hyper-parameters, often with requirements for large training sets. Furthermore, some image detail information may be lost with some CNN-based methods because convolutional layers consider local image regions only, failing to learn contextual dependencies, so that the image becomes blurry during the restoration process [35, 37].



(a) Original image (b) Corrupted image

(c) Deep prior [36] (d) Proposed method

Figure 1: Sample results of the proposed technique. The original image is shown in (a), and a corrupted version is shown in (b). A baseline result from [36] is shown in (c), and some blurring is evident. The result in (d) is from the novel generative chaotic RNN procedure that is presented in this paper. For this case, inpainting is performed using a generative recurrent image prior.

In contrast to state-of-the-art methods that utilize learning to construct an image prior, this paper presents an untrained multi-level chaotic generative Recurrent Neural Network (RNN) for image restoration that can capture significant image statistics independent of learning. An example of our approach is shown in Fig. 1. The memory mechanism of RNN is capable of contextual correlation analysis, and here we utilized a Chaotic Recurrent Network (CRN) [2, 27]. The CRN concept is well known in dynamic

system modeling. The input layer and transition matrix are randomly and sparsely initialized using normal distributions, and then fixed. During learning, only the output layer is updated. The chaotic state transition structure in a CRN enables learning of context while resisting different types of noise. We have found that the approach yields good performance for the image inpainting task when data is missing randomly. For data encoding, the proposed technique first utilizes a max-pooling layer on the image prior to the RNN. Then the encoder scans the down-sampled image from four directions as four global image region sequences. After this iterative process, a fully connected layer is utilized as the decoder to produce the restored image from the four generative RNN outputs.

To reinforce the robust performance of the proposed chaotic RNN, selection of the optimal state transition matrix for the corrupted image is essential. However, a single transition matrix may have limited inference capability due to its fixed sparsity and chaos level in the latent space. To overcome this problem, we developed a multi-level structure with multiple chaotic state transition matrices for each direction. These transition matrices are combined in an intermediate layer. The use of multiple transition matrices allows recovery of missing points for an image, using combined knowledge that is obtained from limited training. Since training is applied only to the output layer, the computational burden remains light.

To our knowledge, this paper is the first to investigate a multi-level chaotic RNN for the image inpainting problem from a single noisy image. A strength of the approach is that the core formulation and training process are straightforward. In spite of the relative simplicity, its performance in the extraction and restoration of hidden image information is comparable to deep neural network image inpainting methods.

The main contributions of this paper are as follows:

- A robust and efficient generative chaotic RNN with a quad-directional encoder has been introduced for image restoration from a single corrupted image without any pre-training or modeling of the image.
- Compared to existing deep CNN image restoration methods, the structure of the proposed network architecture is compact, which leads to a significant reduction in computational demands without sacrificing restoration quality.
- The proposed network has a fast convergence rate because only the output layer needs to be trained. The computed loss is low from the beginning of training.
- The chaotic recurrent state transition framework, with multi-level adaptive state importance levels, has resulted in good generalization performance for various noise types.

Experimental results have demonstrated that the proposed method produces outcomes that are comparable to learning-based methods.

2. Related Work

Deep CNNs are powerful models for image restoration problems with state-of-the-art performance. These learning-based methods have shown success in the inverse image reconstruction problem, and the model is parameterized well by sufficient pre-training from the data set. Initially, CNN-based image completion techniques were applied only to compensate for a small, thin noise mask [19, 32, 39]. Later, these techniques were applied to other image restoration problems such as single image super-resolution [7, 23, 15, 49, 50, 51], image inpainting with small missing regions [17, 26, 30, 43], and more recently image inpainting with large regions of missing data [46, 33, 11, 13, 40]. Yang et al. also proposed a technique that combines the CNNs with certain optimization techniques in image restoration [41]. However, these methods also invariably need to be trained from a large image data set to estimate the image prior. Furthermore, it is unclear whether learning-based network parameterization increases the quality of image prior learning [36, 48]. There are some RNN-based methods that currently cannot perform image inpainting and require pre-training [14, 25], motivating our future work.

Another type of technique to tackle the image inpainting problem is generative adversarial networks (GAN) [10], which are an extension of CNN techniques with two networks trained simultaneously. A generator performs image restoration, and the discriminator tries to distinguish the original images from recovered images. Many researchers have applied GANs to the image restoration problem, including compression artifact removal, and JPEG-related loss during image transmission [12, 9]. Shaham et al. used a single image for pyramid GAN training for image synthesis [34]. Iizuka et al. [17] introduced GANs in image inpainting with good results. However, the training process for GAN is unstable, and a large training set is always required by GAN.

There is also a trend of techniques that are designed to restore images without pre-training using large image datasets. Some methods utilize dictionary-based fitting for corrupted image patches [28, 47], and some use convolutional model fitting such as sparse coding [47, 45]. Also, some restoration techniques have utilized statistical models that are similar to the shallow layer CNN [29, 5] for inference of missing image parts. All these methods are more efficient in parameterization, but the results are not as good. The method that is proposed in this paper displays results that are comparable with trained neural network methods but without the disadvantage of requiring pre-

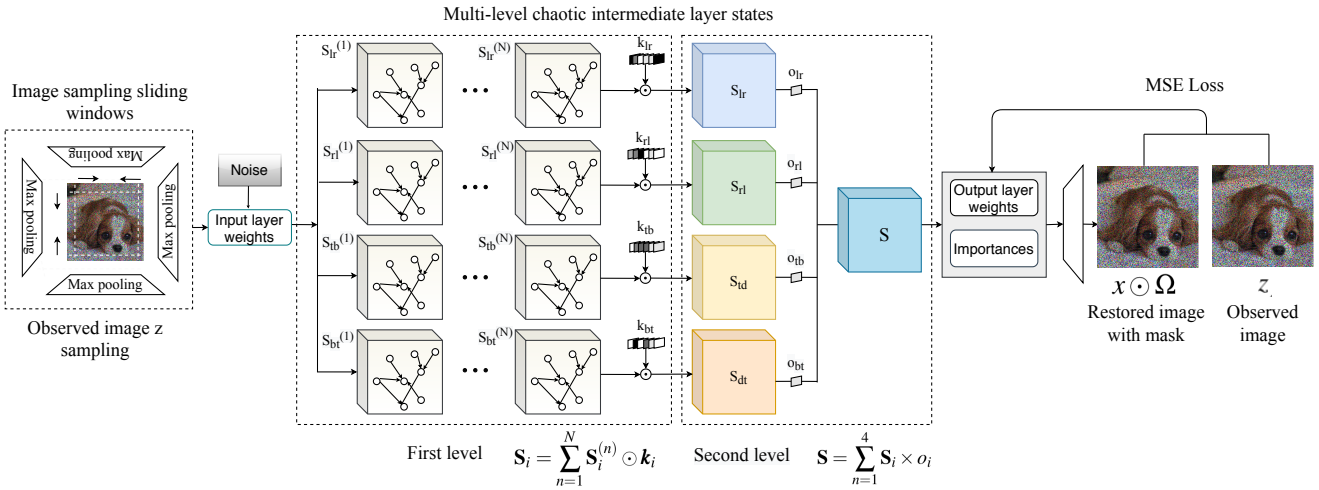


Figure 2: The overall multi-level generative quad-directional chaotic recurrent network. The input image is scanned over four directions prior to being fed into four directional RNNs. Max-pooling layers are added for subsampling the input image. Four rows in the chaotic intermediate layer are corresponded to four directional RNNs. In each row, multiple chaotic state transition matrices are composed of the following directional state transition matrix, weighted by adaptive importance $\boldsymbol{k}$. In the second level, four directional state transition matrices are composed together of $\boldsymbol{S}$, based on their corresponding weights $\boldsymbol{o}$. The output layer weights and all importances are jointly optimized by the loss between output image and visible ground truth image.

training from a large data set.

3. Method

Given an observed image z, the task of image restoration can be generally formulated as

$$x^* = \arg\min_x (E(x;z) + p(x)), \tag{1}$$

where x represents a possible restored image, and E is the fitness between x and the noisy observed image z. The image generator can be selected using a neural network, denoted as $x = g_\theta(z;w)$, where θ is the set of neural network parameters and w is a random code noisy vector. Here p is the image prior, which is used as a regularizer to help the restored image become more realistic. We choose $p(x)$ as a implicit prior θ^* that is learned from a neural network.

$$\theta^* = \arg\min_\theta E(g_\theta(z;w);z) \tag{2}$$

Both θ and w are randomly initialized. Therefore, the selection of the generative neural network $g_\theta(z;w)$ and the optimal θ are the key for successful image restoration.

3.1. Inpainting

Now we consider the objective of inpainting an image. We define

$$E(x;z) = \|x \odot \Omega - z\|_2 \tag{3}$$

$$= \|g_\theta(z;w) \odot \Omega - z\|_2, \tag{4}$$

where Ω represents the binary mask with 1 for the observed pixels, and 0 otherwise. The symbol $\odot$ is the Hadamard product. The image's missing parts due to the mask will be inferred by the proposed network after training.

3.2. Quad-directional Generative Chaotic RNN.

In this paper, we consider using the chaotic RNN [2, 27] as the image generative network $g_\theta(z;w)$. Our proposed RNN-based generative neural network is plotted in Fig. 2. It is composed of four decoupled RNNs which respectively scan the image in four directions: left to right, right to left, top to bottom and bottom to top as Fig. 3. The scanned images are fed into the RNNs to generate intermediate states along with the random noise input according to

$$\boldsymbol{s}_n(p) = \sigma(\boldsymbol{W}_{in}\boldsymbol{n}(p) + \boldsymbol{W}_{feed}\boldsymbol{z}(p) + \boldsymbol{W}_{rec}\boldsymbol{s}_n(p-1)) \tag{5}$$

where $\boldsymbol{s}_n(p)$ represents the RNN state belonging to the nth RNN at the pth row/column of the image. The term $\boldsymbol{z}(p)$ refers to max-pooling results, operating on the sliding window which samples the image centered at the pth column/row of the corrupted image. The length of the window is set uniformly and denoted as w. Random noise input of the RNN is indicated by $\boldsymbol{n}(p)$. The inference is based on teacher forcing [38], which is common for training RNN: ground truth samples are fed back into the network as a condition to keep the prediction close to the ground truth sequence. Here we consider the corrupted image as the ground truth in inference. The input layer is random noise, $\boldsymbol{W}_{in}$; the recurrent layer $\boldsymbol{W}_{rec}$ contains the RNN state from

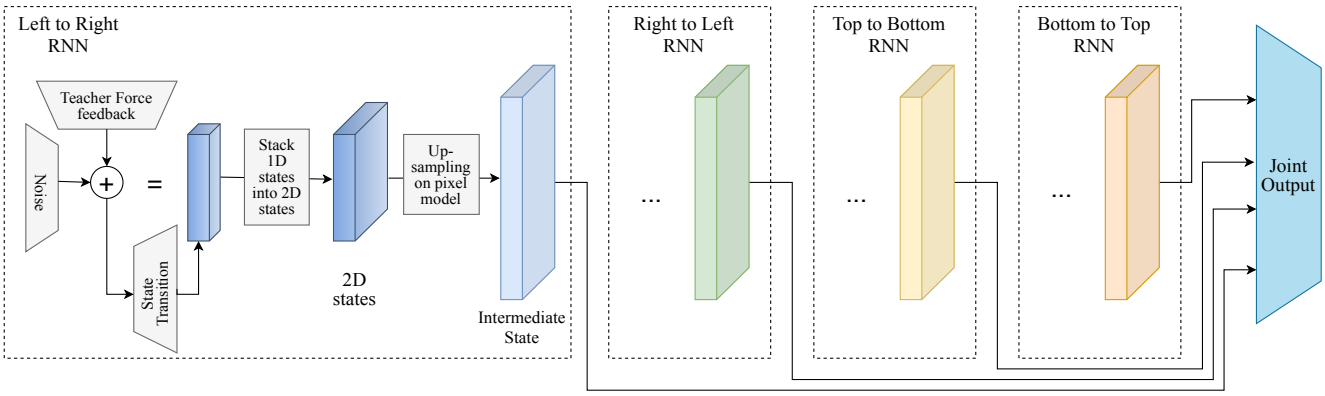


Figure 3: The quad-directional generative RNN framework. The proposed network is composed of four independent generative RNNs. The solid lines represent the forward path of the proposed network and the dashed lines separate the four independent generative RNNs Inside each RNN, the sequential feature is generated by the noise input along with the teacher-forced feedback. Sequential features are stacked and up-sampled to represent intermediate state at pixel level. The joint output layer is a fully connected layer that takes the concatenation of the outputs from four generative RNNs to be the restored image.

the previous step; and $\boldsymbol{W}_{feed}$ is the feedback layer from the sequence data in the corrupted image as condition for prediction. The activation function is σ. All the weights of these layers are randomly initialized, and then fixed. Only the output layer is updated during the learning process.

Through the random sparsity of the chaotic RNN, the image is encoded as a high-dimensional sequence (Fig. 3). The output layer weights can be learned by fitting the output to the observation. After the output weights have been learned, the missing part of this image is reproduced by the underlying neural network. Here, we only adjust the output layers of the neural network according to the gradients.

The intermediate states from the quad-directional RNNs are concatenated as follows,

$$\boldsymbol{S} = [\boldsymbol{S}_{lr}; \boldsymbol{S}_{rl}; \boldsymbol{S}_{tb}; \boldsymbol{S}_{bt}] \tag{6}$$

where $\boldsymbol{S}_{lr} = [\boldsymbol{s}_1(1), \boldsymbol{s}_1(2), \cdot, \boldsymbol{s}_1(N)]$, and $\boldsymbol{s}_1$ is from the definition in (5) with N representing the number of columns of the image. The intermediate states $\boldsymbol{S}_{rl}$, $\boldsymbol{S}_{tb}$ and $\boldsymbol{S}_{bt}$ are defined similarly. The generated output is constructed using output layer weights $\boldsymbol{W}_{out}$:

$$\boldsymbol{X} = \boldsymbol{S}^T \boldsymbol{W}_{out}. \tag{7}$$

Therefore, the output weights of the RNN are learned by solving

$$\min_{\boldsymbol{W}_{out}} \quad \|\boldsymbol{Z} - \Omega \odot (\boldsymbol{S}^T \boldsymbol{W}_{out})\|_2. \tag{8}$$

here $\boldsymbol{Z}$ is the corrupted observed image. The above problem can be solved by a standard gradient descent as follows:

$$\boldsymbol{W}^{(k)} = \boldsymbol{W}^{(k-1)} + \mu(k) \nabla_{\boldsymbol{W}^{(k)}} E, \tag{9}$$

where E is the objective in (3). The gradient of $\boldsymbol{W}_{out}$ is calculated by

$$\nabla_{\boldsymbol{W}_{out}} E = -2\boldsymbol{S}(\boldsymbol{Z} - \Omega \odot (\boldsymbol{S}^T \boldsymbol{W}_{out})). \tag{10}$$

Moreover, to enhance the convergence, we chose the step-size update rule from [3], which is given by

$$\mu(k) = \frac{Tr(|(\boldsymbol{W}^{(k)} - \boldsymbol{W}^{(k-1)})^T (\nabla_{\boldsymbol{W}^{(k)}} E - \nabla_{\boldsymbol{W}^{(k-1)}} E)|)}{\|\nabla_{\boldsymbol{W}^{(k)}} E - \nabla_{\boldsymbol{W}^{(k-1)}}\|_2^2}. \tag{11}$$

3.3. Multi-level Generative RNN with Adaptive State Importance.

Recall that we intend to use chaotic RNN for image generation. Ulyanov et al. [36] observed that different CNN networks in the generative framework are optimal for one type of image and noise type. Based on this observation, and the fact that a chaotic network can provide various transition states with different sparsity and chaos levels, we designed an unified framework with multiple transition states $\mathbf{S}$ inside the intermediate layer. The index for each direction is i. Inside each direction, we introduced adaptive importance vector $\boldsymbol{k}_i$ for $\mathbf{S}_i$, importance $\boldsymbol{k}_i$ is a n-dimensional vector for n number of $\mathbf{S}_i^{(n)}$. The entries of the importance vector represent the degree of importance of the corresponding $\mathbf{S}_i^{(n)}$. With that composition, each directional state transition $\mathbf{S}_i$ can be represented as:

$$\mathbf{S}_i = \sum_{n=1}^{N} \mathbf{S}_i^{(n)} \odot \boldsymbol{k}_i \tag{12}$$

The second level importance $\boldsymbol{o}_i$ for four directional state transition $\mathbf{S}_i$ combine into a single state transition. The

(a) Original image (b) Corrupted image (c) [36] (d) [6] (e) Proposed method

(f) Original image (g) Corrupted image (h) [36] (i) [6] (j) Proposed method

Figure 4: Image inpainting comparison with different approaches.

Table 1: PSNR of image inpainting by different methods.

	Barbara	Boat	House	Lena	Peppers	C-man	Couple	Finger	Hill	Man	Montage
Deep prior	23.70	27.49	27.18	23.92	32.24	27.01	**29.07**	28.43	28.87	24.59	34.54
Bayesian	23.26	26.91	30.47	28.75	28.27	26.17	26.70	25.06	27.20	27.18	34.65
RNN	**28.30**	**28.30**	**31.70**	**32.30**	**38.52**	**35.58**	28.41	**28.65**	**29.57**	**30.24**	**36.13**

dimensionality of $\boldsymbol{o}$ is a 4-component vector. The resulting transition state matrix is

$$\mathbf{S} = \sum_{n=1}^{4} \mathbf{S}_i \times o_i. \tag{13}$$

Only the output layer and importance are updated, based on maximizing the likelihood of inference output and given a single corrupted image.

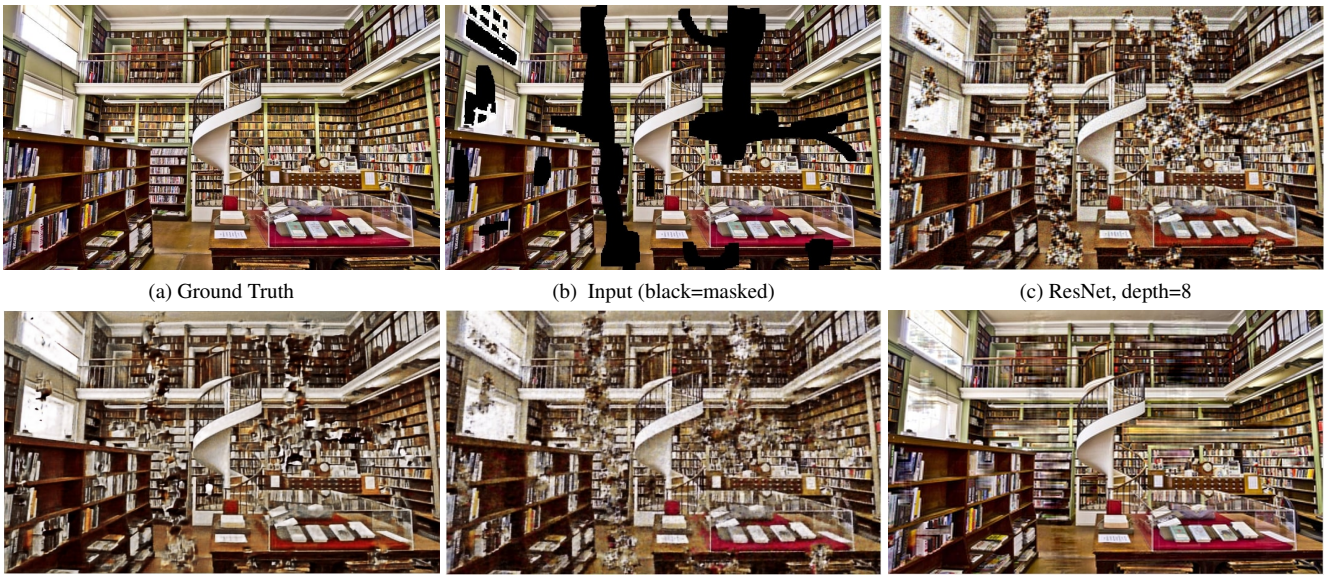
4. Experimental Results

4.1. Image Restoration Comparison with State-of-the-art Methods

Fig. 4 illustrates the pictures with a random noise mask according to the binary Bernoulli distribution. The task of image restoration is to recover the underlying image according to the intrinsic features of the image object. In our qualitative experiments, we use the standard image inpainting dataset [16] consisting of eleven grayscale images with 40% of the pixels randomly dropped.The proposed method's outcome is compared with the results from [36], which are based on deep image priors, and from [6], which improved [36] by using stochastic gradient Langevin.

We evaluated the performance of different methods in both effectiveness and efficiency. For a fair comparison, all of the experiments were conducted with 3000 iterations for the same data set, as was performed in [16]. As shown in Table 1, our method has outperformed these benchmark methods [6, 36], without any pre-training. In addition, we present two restored results for qualitative comparison in Fig. 4. These results show that, although the proposed method only learned from a single corrupted image without any prior, it can recover many details even with a severe rate of data removal. Contrary to the current GAN methods which are unstable and hard to train in many situations, the performance of proposed generative method is very robust and its loss is low even from the beginning of training. Though the proposed generative RNN has only a shallow structure, it has outperformed deep CNN-based approaches in both efficiency and effectiveness.

We have also investigated the typical inpainting problem in which large blocks of data are missing, as seen in Fig. 5. Our method performs surprisingly well and arguably surpasses all of the other methods except for [36]. In other examples, Fig. 6, the images are corrupted by center crop, and by severe random crop with 80 percent missing data. The restoration result of the proposed method is more realistic than the deep neural network methods [36] and other GAN-based methods, as it provided more semantically plausible while other methods have noticeable incon-



(a) Ground Truth (b) Input (black=masked) (c) ResNet, depth=8

(d) U-Net, depth=5 (e) Deep image prior (f) Proposed method

Figure 5: Inpainting comparison between the proposed method with other deep learning methods with different depths and architectures. The figure shows that the proposed method has better inpainting results.

(a) Ground truth (b) Corrupted image (c) Single-code GAN (d) Optimal feature map (e) Deep image prior (f) Ours

Figure 6: Image region inpainting using different network and settings. (c) GAN inversion through single latent code [24], (d) GAN inversion through optimizing feature maps [4], (e) DIP [36], (f) ours.

sistencies. Table 2 shows quantitative results for these inpainting cases. It indicates good performance of the proposed technique in image context correlation extraction, and unknown area inferencing, with results comparable to the tranditional self-similarity based methods.

4.2. Ablation Study

An experiment was conducted to investigate the selection of convergence rate and number of neurons. We can evaluate the neuron-number effect qualitatively. In Fig. 7, we see that from left to right, as the neuron number increases, the resulting image contains more details as well as fewer artifacts. As shown in Fig. 8, the quad-directional encoder can provide much more detail and higher PSNR

Table 2: PSNR quantitative comparison of different image inpainting methods in cases with large missing regions. Central cropping removed 64 x 64 box and ramdom cropping removed 80 percent pixels.

Method	Center Crop	Random Crop
(a) Single-code GAN [24]	10.37	12.79
(b) Optimized feature map[4]	14.75	18.72
(c) Deep image prior[36]	17.92	18.02
(d) Ours	**20.91**	**21.52**

(a) Number of neurons: 100 (b) Number of neurons: 500 (c) Number of neurons: 1000 (d) Number of neurons: 4000 (e) Ground truth

Figure 7: Image inpainting with different neuron counts.

(a) One-RNN: 25.4 (b) Bi-RNN (horizontal): 28.9 (c) Bi-RNN (vertical): 29.8 (d) Quad-RNN: 32.0 (e) Ground truth

Figure 8: Inpainting with different directional RNNs and corresponding PSNR. (a) Single-directional RNN, scanning from left to right. (b) Bi-directional RNNs with horizontal scans. (c) Bi-directional RNNs with vertical scans. (d) Quad-directional RNNs with horizontal and vertical scans.

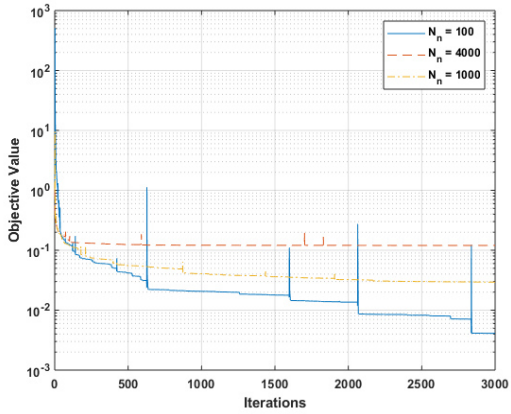


Figure 9: Convergence of the proposed algorithm with different numbers of neurons.

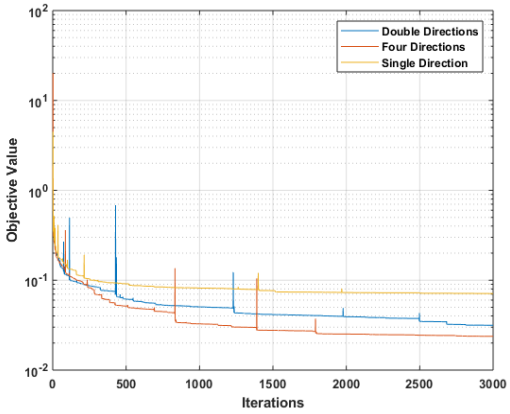
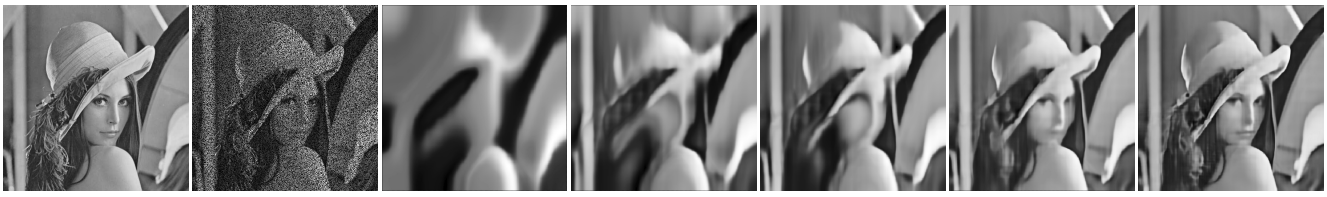


Figure 10: Convergence of the proposed algorithm with different configurations of directional RNNs.

than other encoding configurations. A typical learning curve of our proposed RNN-based completion is given in Fig. 9. The vertical axis of the plot indicates reconstruction error. Meanwhile, the step size of the gradient descent is initialized as a fixed value, 10^{-4}. The curve shows a quick descent at the beginning, followed by a slow convergence at the end. The results plotted in Fig. 9 suggest that the number of neurons can determine the lower bound that the objective can achieve.

In addition, we investigate the effectiveness of the different configurations of the directional data encoder for RNNs quantitatively. In Fig. 10, we see that the fully quad-directional encoder has the fastest convergence rate compared to single or bi-directional RNNs. The result corroborates our intuition that the quad-directional RNN can capture spatial correlations within the image. We chose 2x2 max-pooling for feature downsampling, and obtained improved performance by 10% compared to using the original image, while a bigger stride lost more details. After using a multi-level structure with adaptive importance, as



(a) Ground truth (b) Corrupted image (c) DIP [36]: 100 (d) DIP [36]: 300 (e) DIP [36]: 500 (f) DIP [36]: 1000 (g) DIP [36]: 1500

(h) Ground truth (i) Corrupted image (j) Ours: 100 (k) Ours: 300 (l) Ours: 500 (m) Ours: 1000 (n) Ours: 1500

Figure 11: Image inpainting with different iteration amounts. The first row contains results from deep image prior [36]: the PSNR value after 1500 iterations is 26.63. The second row is the results from our method: the PSNR value reached 30.66 after 1000 iterations, and 31.01 after 1500 iterations.

shown in the first level of Fig. 2, the inpainting performance of our method improved by 5%.

4.3. Model Efficiency

As shown in Table 3, the proposed framework has a more compact network structure regarding to number of coefficients. Our run time including training for each picture is less than a minute, whereas CNN methods such as [36] require more than 14 minutes on the same GPU to reach 32 PSNR (10 times faster). A more detailed comparison with [36] is provided in Table 4 and Fig. 11. Because most GAN methods require pretraining, that comparison is not discussed here.

Table 3: Model size of different methods.

Network	ResNet	U-Net	DIP	Ours
Model size	0.15M	1.6M	3.0M	0.38M

Table 4: Number of iterations required to reach particular PSNR values in image inpainting.

PSNR	28.5	30	30.5	31	32
DIP [36]	8900	8950	9000	9050	9150
Ours	**400**	**500**	**1000**	**1500**	**2000**

4.4. Discussion

A successful image inpainting algorithm should handle random noise in the data with good stability. To accommodate this need, the proposed network exhibits chaotic RNN state activity prior to training. The chaotic activity of the network enables a strong generalization performance for various types of noise and image content, and it handles the chaotic attributes of the data. The network maintains a low loss from the beginning. Because training is performed for the output layer only [18], the result is computational efficiency and smaller network size compared to CNN methods.

The essential element for the training procedure is the feedback loop that transmits the output back to the network. This loop enables the network's learning capability without a large training set. The procedure has demonstrated rapid weight modification at the beginning of the training process. Overall, our network training procedure has exhibited more stability than backpropagation training of similar-sized convolutional networks.

5. Conclusion

This paper has presented a quad-directional multi-level generative chaotic RNN-based image inpainting method that produces experimental results that are competitive with recent deep methods. The proposed network can effectively learn the image prior by first utilizing a quad-directional encoder to generate mid-level features of the image as four global image region sequences. A multi-level generative chaotic RNN for each direction can then make a robust inference regardless of the random noise. The proposed method does not require any pre-training or modeling of image degradation processes. In particular, the proposed method produces results when inpainting large parts of an image that are comparable with deep methods. Restored images generated by the proposed technique contain proper levels of detail, whereas recent benchmark methods tend to introduce blur into the image. The proposed network structure is computationally efficient and does not require large training sets, allowing the proposed RNN structure to be implemented in hardware-limited systems, such as portable devices.

References

[1] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.

[2] Kazuyuki Aihara, T Takabe, and Masashi Toyoda. Chaotic neural networks. *Physics letters A*, 144(6-7):333–340, 1990.

[3] Jonathan Barzilai and Jonathan M Borwein. Two-point step size gradient methods. *IMA journal of numerical analysis*, 8(1):141–148, 1988.

[4] David Bau, Hendrik Strobelt, William Peebles, Bolei Zhou, Jun-Yan Zhu, Antonio Torralba, et al. Semantic photo manipulation with a generative image prior. *arXiv preprint arXiv:2005.07727*, 2020.

[5] Cong Chen, Abel Gawel, Stephen Krauss, Yuliang Zou, A Lynn Abbott, and Daniel J Stilwell. Robust unsupervised cleaning of underwater bathymetric point cloud data.

[6] Zezhou Cheng, Matheus Gadelha, Subhransu Maji, and Daniel Sheldon. A bayesian perspective on the deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5443–5451, 2019.

[7] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014.

[8] Raanan Fattal. Image upsampling via imposed edge statistics. *ACM transactions on graphics (TOG)*, 26(3):95, 2007.

[9] Leonardo Galteri, Lorenzo Seidenari, Marco Bertini, and Alberto Del Bimbo. Deep generative adversarial compression artifact removal. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4826–4835, 2017.

[10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[11] Artur Grigorev, Artem Sevastopolsky, Alexander Vakhitov, and Victor Lempitsky. Coordinate-based texture inpainting for pose-guided human image generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12135–12144, 2019.

[12] Jun Guo and Hongyang Chao. One-to-many network for visually pleasing compression artifacts reduction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3038–3047, 2017.

[13] Xiaoguang Han, Zhaoxuan Zhang, Dong Du, Mingdai Yang, Jingming Yu, Pan Pan, Xin Yang, Ligang Liu, Zixiang Xiong, and Shuguang Cui. Deep reinforcement learning of volume-guided progressive view inpainting for 3d point scene completion from a single depth image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 234–243, 2019.

[14] Kazi Nazmul Haque, Mohammad Abu Yousuf, and Rajib Rana. Image denoising and restoration with cnn-lstm encoder decoder with direct attention. *arXiv preprint arXiv:1801.05141*, 2018.

[15] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018.

[16] Felix Heide, Wolfgang Heidrich, and Gordon Wetzstein. Fast and flexible convolutional sparse coding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5135–5143, 2015.

[17] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):107, 2017.

[18] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *science*, 304(5667):78–80, 2004.

[19] Rolf Köhler, Christian Schuler, Bernhard Schölkopf, and Stefan Harmeling. Mask-specific inpainting with deep neural networks. In *German Conference on Pattern Recognition*, pages 523–534. Springer, 2014.

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[21] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8183–8192, 2018.

[22] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[23] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[24] Zachary C Lipton and Subarna Tripathi. Precise recovery of latent vectors from generative adversarial networks. *arXiv preprint arXiv:1702.04782*, 2017.

[25] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-local recurrent network for image restoration. In *Advances in Neural Information Processing Systems*, pages 1673–1682, 2018.

[26] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 85–100, 2018.

[27] Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

[28] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11(Jan):19–60, 2010.

[29] Vardan Papyan, Yaniv Romano, Jeremias Sulam, and Michael Elad. Convolutional dictionary learning via local processing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5296–5304, 2017.

[30] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.

[31] Daniele Perrone and Paolo Favaro. Total variation blind deconvolution: The devil is in the details. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2909–2916, 2014.

[32] Jimmy SJ Ren, Li Xu, Qiong Yan, and Wenxiu Sun. Shepard convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 901–909, 2015.

[33] Min-cheol Sagong, Yong-goo Shin, Seung-wook Kim, Seung Park, and Sung-jea Ko. Pepsi: Fast image inpainting with parallel decoding network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11360–11368, 2019.

[34] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. Singan: Learning a generative model from a single natural image. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4570–4580, 2019.

[35] Pavel Svoboda, Michal Hradis, David Barina, and Pavel Zemcik. Compression artifacts removal using convolutional neural networks. *arXiv preprint arXiv:1605.00366*, 2016.

[36] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[37] Zhangyang Wang, Ding Liu, Shiyu Chang, Qing Ling, Yingzhen Yang, and Thomas S Huang. D3: Deep dual-domain based fast restoration of jpeg-compressed images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2764–2772, 2016.

[38] Ronald J Williams and David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280, 1989.

[39] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In *Advances in neural information processing systems*, pages 341–349, 2012.

[40] Wei Xiong, Jiahui Yu, Zhe Lin, Jimei Yang, Xin Lu, Connelly Barnes, and Jiebo Luo. Foreground-aware image inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5840–5848, 2019.

[41] Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, and Hao Li. High-resolution image inpainting using multi-scale neural patch synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6721–6729, 2017.

[42] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010.

[43] Raymond A Yeh, Chen Chen, Teck Yian Lim, Alexander G Schwing, Mark Hasegawa-Johnson, and Minh N Do. Semantic image inpainting with deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5485–5493, 2017.

[44] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5505–5514, 2018.

[45] Matthew D Zeiler, Dilip Krishnan, Graham W Taylor, and Rob Fergus. Deconvolutional networks. In *2010 IEEE Computer Society Conference on computer vision and pattern recognition*, pages 2528–2535. IEEE, 2010.

[46] Yanhong Zeng, Jianlong Fu, Hongyang Chao, and Baining Guo. Learning pyramid-context encoder network for high-quality image inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1486–1494, 2019.

[47] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer, 2010.

[48] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530*, 2016.

[49] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3271, 2018.

[50] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301, 2018.

[51] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018.