Image Denoising Using Variational Convolutional Neural Network

Radin Hamidi Rad, Kandasamy Illanko
{radin, killanko}@ryerson.ca
Department of Electrical and Computer Engineering
Ryerson University, Toronto, Canada.

***Abstract*—We propose a novel architecture for image reconstruction by utilizing convolutional neural networks (CNN). The main idea is to use such networks to learn statistical distribution of pixels and extract concepts in form of feature vectors. Most of the image noises are come from additive Gaussian Noise (AWGN). By changing ordinary CNN architecture and also using generative models these additive noise can be detected and deducted from the disrupted images. Using variational neurons as the fundamental elements of the proposed architecture will let us to denoise both low-level and high-level distorted images. Contrary to the existing discriminative denoisers, our proposed model will need significantly less amount of training data to achieve acceptable performance and also it needs less time for calculation. We ran various set of experiments on proposed and competitive models. This conducted evaluations show that our model is efficient and making it practical denoising applications.**

***Index Terms*—Denoising, Image Classification, Convolutional Neural Networks, Artificial Intelligence**

I. Introduction

Images are inevitably mixed with noise during different stages. From taking the photo in real world, during transmission and compression and storage phases. The real-world noise that mixes with the image in photo shooting process [1] is signal dependent and origins from five main reasons: photon shot, fixed pattern, dark current, readout and quantization [2]–[4]. Some if these noises are part of the system nature. For example, quantization error in inevitable and hard to avoid. These noises will affect further processes on the images or videos. Therefore a denoising stage is needed With the presence of noise. Nowadays, this stage plays a very important role in digital media industry. Almost all the image processing systems have a module for image denoising before their image processing tasks such as video processing and image editing. Image denoising system witnessed lots of improvements in the both efficiency and computation speed in past two decades. Before applying neural network based solutions for this topic classic intelligence system approaches are more common to used to solve the problem. Features including non-spatial and feature similarities were mostly used to analyze nature of an image to determine if it has noise or not [5]. Furthermore, matrix decomposition technique is introduced to the industry. Considering an image as a matrix, we can decompose the matrix in order to determine the singular values. Then we can smooth and deduct from unwanted noise and reconstruct the image. Technically by decomposing the image we can easily from high frequency noises, regarding to the fact that we have access to singular values. The same techniques used to be used for other tasks, including background removal [6], [7], anomaly detection [8], etc. By introducing convolutional neural networks, a new era of image analysis has began. Convolutional neural network tries to capture the repetitive patterns in images [9]. By using convolutional cells we decompose pixels inside the sliding window. This will help us to capture the pattern hidden inside the spectrum of pixel values in the window. Furthermore, by passing these extracted values, we will be able to learn a rich representation of images [10]. This representation that is called feature vector can be used as the input to deep neural network in order to learn different tasks, including image classification, object detection and segmentation, and etc.

In our work we tried to take advantage of variational neurons in order to capture the probability distribution in pixels of an image. The objective of our work in this paper is to present an approach that would satisfy the main requirement of an effective image denoising system: maximal reconstruction ability. Current solutions suffer from slow learning rate, need of huge train data and non-scalability. To address these limitations, we present an approach that learns feature representations over a set of images using a variational autoencoder architecture. Instead of learning representations over the regular convolutional neural network, we consider each of images as instance to learn variational neurons over the Gaussian distribution.

Our contributions can be enumerated as follows:

- We formally define problem of image denoising and proposed a novel approach based on convolutional neural network. This neural network will use variational autoencoders setup to capture Gaussian distributions over the pixel values.
- We will test our model with wide range of noise levels to study robustness of model against different type of noises.
- We evaluate our proposed model over the well-known image dataset to compare the results. We will show that proposed model will outperform current methods that are based on other approaches

In the next section, we provide a literature review about most important paper works in this domain. The problem statement and our proposed models will be discussed in section III. In section IV we will demonstrate the experiments performed and will analyse the evaluation results.

II. Related Work

There have been several attempts to solve denoising problem using machine learning. In this section, we briefly review and discuss the major researches conducted that are relevant to this work.

Jain et al. in [11], Jain and Seung, tried to use convolutional neural networks (CNNs) for image denoising. They have mentioned that theses type of neural networks have similar or even better representation power than the Markov Random Field models. Markov Random Field (MRF) models are based on Markov probablisitc rule. They try to guess and remove noises based on calculation of probablities based on Markov chain.

In [12], the popular multi-layer perceptron (MLP) neural network model was proposed as a potential solution for image denoising. In this architecture, images will be read as a raw matrix. There is no convolutional nor spatial analysis for pattern recognition. Therefore, performance of the model is limited to model learning power. The drawback of using this model is that trained model can denoise for a certain noise level. It was the first attempt to solve the denoising problem using discriminative denoising method.

Xie et al. in [13], stacked sparse autoencoder are used to handle denoising problem Stacked sparse autoencoders (SSAE) are able to handle Gaussian noises, however, they are hard to train due to the fact that they use different cost function that drops learning speed dramatically. Kullback-leibler divergence is used to calculate the error on each epoch. This cost function calculates the skew in divergence of input and ideal data and tries to fix the divergence by back propagating errors accordingly.

In [14], authors used a trainable nonlinear reaction diffusion (TNRD) architecture. The proposed model is a feed-forward deep neural network that is unfolded by a fixed number of gradient descent inference steps. This deep neural network (DNN) architecture can remove the noise from images with certain range of noise levels. However it does not have accuracy of [12].

Authors of [15] used a deconvolutional neural networks in a recursively branched paradigm. The proposed model, Recursively Branch Neural Network (RBDN) tries to perform regression image-to-image based in order to remove the noise from the pixels. Therefore, pixel values will be result of regression calculated based on similar images on neighbor branches.

Tai et al. [16] tried to handle the denoising image problem by proposing a very deep persistent memory network (MemNet). In this model, a memory block tries to mine persistent memory though an adaptive learning process. In other words, memory blocks will learn and save relations during learning process by iterating through the samples. This model lacks generalization and testing domain is limited to the training dataset, although it shows an acceptable performance for the wide range of noise levels.

Fine tuning the convolutional neural network is one of other attempts to improve convolutional neural networks. Zhang et al. [17] tries to use detailed convolutional layers in deep neural network to improve the feature extraction phase. After building feature vector, extracted features will be passed to a seven layer architecture for denoising purposes.

Also, Zhang er al. in [25], introduced the FFDNet, a fast and flexible solution for image denoising, based on convolutional neural networks. The proposed solution will work on down-sampled sub-images instead of the whole image. This trick will let them achieving a good trade-off between inference speed and denoising performance. Also, due to sub-sampling technique, they are able to remove noises with various spatial properties. The drawback of this model is the lack of ability in recognising noises that uniformly affect the whole image in the same way. Since they are use smaller scope to analyze image, they may miss the holistic inference.

In [18], Mao et al. trained an extremely deep neural network containing fully connected convolutional layers in order to capture more generalized concepts from images. Generally, by adding to the layers of a deep neural network, model would be able to extract features from concepts that are extracted in previous layer. This method let the model to perform well even on non-trained datasets, thanks to its very deep convolutional architecture.

Lefkimmiatis et al. [19] and Qiao et al. [20] tried to solve denoising problem using a proximal gradient-based approach. The proposed model uses non-local similarity of images to denoise an image. By using proximal gradient technique, input image will incorporate with predicted similar images. After that, model uses candidates as an inference to remove the noises from target image.

Authors in [21], proposed a fully trainable patch-based neural network for image denoising. This model is based on Gaussian Conditional Random Field inference (GCRF). Te inference will help the model to predict noise distribution over the target image, however, system suffers low performance when it comes to non-Gaussian noises.

III. System Model and the Problem Statement

Given a set of noise-free images, we employ a variational Bayesian neural network to form the model. We introduce uncertainty on the pixel distribution parameters (weights) and subsequently in the final predictions using variational Bayesian neural architecture, which not only yields a better regularisation via probabilistic weights, but also leads to a richer representation and more accurate predictions [23]. The mentioned setup become noticeable when we witness a remarkable sparsity in the input data. Despite variety of usages in other computer science fields, such as language modeling [24], this architecture to best of our knowledge, is the most effective approach used for image denoising. We show that variational neural networks are well-suited and a closer proxy for removing noises from images.

We assume the generative part of the model similar to the deep version of latent Gaussian models that has layers of latent variables [24]. Let $\mathcal{P} = \{\mathcal{P}_i\}$ and $\mathcal{N} = \{\mathcal{N}_j\}$ be the sets of pixels and noises, respectively, $(\mathbf{p}, \mathbf{n})$ is a set of noisy pixels, i.e. corrupted image. $\mathbf{n} \subseteq \mathcal{N}; \mathbf{n} \neq \emptyset$, which has been formed

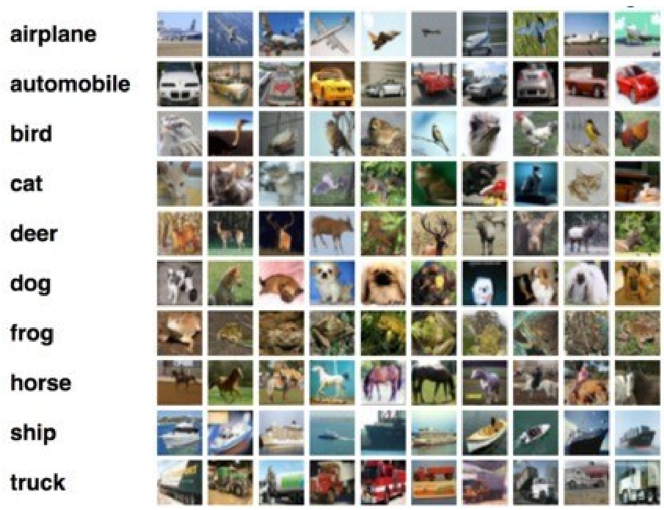

Fig. 1. The CIFAR-10 dataset snapshot.

with respect to pixels $\mathbf{p} \subseteq \mathcal{P}; \mathbf{p} \neq \emptyset$, and $\mathcal{I} = \{(\mathbf{p}, \mathbf{n})\}$ indexes an image. Our task is to learn $f : \mathcal{P}(\mathcal{P}) \rightarrow \mathcal{P}(\mathcal{N})$, the function that maps parameters $\boldsymbol{\theta}$ from pixels set to noisy pixels set, such that $f(\mathbf{p}; \boldsymbol{\theta}) = \mathbf{n}$.

A. Variational Inference

We aim at optimizing the maximum a posteriori probability of $\boldsymbol{\theta}$ in $f(\cdot, \boldsymbol{\theta})$, i.e., $p(\boldsymbol{\theta}|\mathcal{I})$ where f is a multi-layer variational neural network, $\mathcal{I}$ is the target image whose pixels $(\mathbf{p}, \mathbf{n})$ consist of an input pixel subset $\mathbf{p}$ and a target expert subset $\mathbf{n}$, which are assumed to be drawn independently from a joint distribution $p(\mathbf{p}, \mathbf{n})$, and $\boldsymbol{\theta}$ are real-valued parameters, or weights. By Bayes theorem,

$$p(\boldsymbol{\theta}|\mathcal{I}) \propto p(\mathcal{I}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad \text{where} \quad P(\mathcal{I}|\boldsymbol{\theta}) = \prod_{(\mathbf{p},\mathbf{n})\in\mathcal{I}} p(\mathbf{n}|\mathbf{p}, \boldsymbol{\theta}) \tag{1}$$

and $p(\boldsymbol{\theta})$ is the prior probability of weights. It can be inferred that maximizing $p(\mathcal{T}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ will maximize the the posteriori estimate of $\boldsymbol{\theta}$. The true prior probability of weights $p(\boldsymbol{\theta})$, although, it cannot be calculated using analytical methods or sampled, therefore, we approximate the equation by a more feasible distribution $q(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma})$ with multivariate diagonal Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$. The elements of $\boldsymbol{\sigma}$ will be diagonal co-variance matrix that in our problem means that neural network weights $\boldsymbol{\theta}$ are assumed to be completely uncorrelated. Contrary to non-variational models, where weights are parameterized, herein each hidden layer weight matrix$\theta_i{\in}\boldsymbol{\theta}$ is derived from a Gaussian distribution that has an individual mean μ_i and variance σ_i^2. This will increase the number of parameters to learn. They are doubled when training our Bayesian neural network via variational inference.

B. Objective Function

To estimate the true posterior $p(\boldsymbol{\theta})$ by $q(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma})$, we tried to use the Kullback-Leibler divergence that is oftenly used for sparse data. It is calculated between q and p with regard to the Gaussian mean and variance vectors as suggested by Graves [26]:

$$\mathrm{KL}(q(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma})||p(\boldsymbol{\theta}|\mathcal{T})) = \int q(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma}) \ \log[\frac{q(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma})}{p(\boldsymbol{\theta}|\mathcal{T})}]\mathrm{d}\boldsymbol{\theta} \tag{2}$$

$$= \mathbb{E}_{q(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\sigma})} \ \log[\frac{q(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma})}{p(\mathcal{T}|\boldsymbol{\theta})p(\boldsymbol{\theta})}p(\mathcal{T})] \tag{3}$$

In order to minimize $\mathrm{KL}(q(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma})||p(\boldsymbol{\theta}|\mathcal{T}))$, we need to minimize the first two terms in Eq. 3, This equation is known as "variational free energy", regarding that the log marginal likelihood log $p(\mathcal{T})$ does not depend on $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.

Since the Kullback-Leibler divergence is always non-negative, L is a lower bound on log $p(\mathcal{T})$, i.e., $\mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \mathcal{T}) \leq$ log $p(\mathcal{T})$ and is also known as evidence lower bound (elbo).

$$\mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \mathcal{I}) = \log\, p(\mathcal{I}) - \mathrm{KL}(q(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma})||p(\boldsymbol{\theta}|\mathcal{I})) \tag{4}$$

C. Model Architecture

We now explain our proposed variational convolutional neural network. We will predict a denoised image $\mathbf{i} \subseteq \mathcal{I}$ for a given image subset $\mathbf{p} \subseteq \mathcal{P}$ by a mapping function $f(\mathbf{p}; \boldsymbol{\theta})$ using a our architecture. We will use hidden layer $\mathbf{h}$ of size d, without dropout and loss of generality to multiple hidden layers. The input layer will be: $v_{\mathcal{N}}(\mathbf{N})$ and output layer will be: $v_{\mathcal{S}}(\mathbf{s})$:

$$\mathbf{h} = \pi_1(\boldsymbol{\theta}_1 v_{\mathcal{N}}(\mathbf{n}) + \mathbf{b}_1) \tag{5}$$

$$v_{\mathcal{S}}(\mathbf{s}) = \pi_2(\boldsymbol{\theta}_2 \mathbf{h} + \mathbf{b}_2) \tag{6}$$

$$\boldsymbol{\theta} = \boldsymbol{\theta}_1 \cup \boldsymbol{\theta}_2 \cup \mathbf{b}_1 \cup \mathbf{b}_2 \tag{7}$$

Network of dense variational hidden layers $\mathbf{h}_i$ of size d_i where $i \in \{1, 2, ..., \mathrm{I}\}$, with input layer $v_{\mathcal{N}}(\mathbf{n})$ and output layer $v_{\mathcal{S}}(\mathbf{s})$.

In Eq. 7, $\pi_.$ is a nonlinear activation function, $\boldsymbol{\theta}{\sim}\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ that Gaussian distribution properties are estimated by minimizing previously introduced "variational free energy" Eq. 3, $v_{\mathcal{N}}(\mathbf{n})$ will be representation of input noisy image and that is subset of an Image $\mathbf{I}$, and $v_{\mathcal{P}}(\mathbf{p})$ will be representation of denoised image that is subset $\mathbf{P}$ given an image $(\mathbf{n}, \mathbf{p}){\in}\mathcal{I}$.

IV. Experiments

A. Dataset

1) Dataset: We choose CIFAR-10 [22] 1, as the base dataset for model development and intuition. Since most of the baselines methods are used a common set of images for illustration, we also used those images for comparison purposes. This image set includes, Lena, Boat, Starfish, Parrot, and House [17].

B. Setup

To be filled.

C. Results and Discussion

To be filled.

V. Conclusion

To be filled.

References

[1] G. Healey and R. Kondepudy, “Radiometric ccd camera calibration and noise estimation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 3, pp. 267–276, 1994

[2] S. Nam, Y. Hwang, Y. Matsushita and S. J. Kim, ”A Holistic Approach to Cross-Channel Image Noise Modeling and Its Application to Image Denoising,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 1683-1691, doi: 10.1109/CVPR.2016.186.

[3] Y. Tsin, V. Ramesh, and T. Kanade, “Statistical calibration of ccd imaging process,” in Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE InternationalConference on, 2001, pp. 480–487 vol.1.

[4] S. J. Kim, H. T. Lin, Z. Lu, S. Ssstrunk, S. Lin, and M. S. Brown, “A new in-camera imaging model for color computer vision and its application.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 12, pp. 2289–302, 2012.

[5] A. Buades, “A review of image denoising algorithms, with a new one,” Siam Journal on Multiscale Modeling and Simulation, vol. 4, no. 2, pp. 490–530, 2005.

[6] M. Garcia-Fernandez et al., ”SVD-Based clutter removal technique for gpr,” 2017 IEEE International Symposium on Antennas and Propagation USNC/URSI National Radio Science Meeting, San Diego, CA, 2017, pp. 2369-2370, doi: 10.1109/APUSNCURSINRSM.2017.8073227.

[7] A. Rajwade, A. Rangarajan, and A. Banerjee, “Image denoising using the higher order singular value decomposition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 4, pp. 849–862, 2013.

[8] X. Zang, G. Li, Z. Li, N. Li and W. Wang, ”An Object-Aware Anomaly Detection and Localization in Surveillance Videos,” 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), Taipei, 2016, pp. 113-116, doi: 10.1109/BigMM.2016.33.

[9] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. ”Imagenet classification with deep convolutional neural networks.” In Advances in neural information processing systems, pp. 1097-1105, 2012.

[10] Sainath, Tara N., Abdel-rahman Mohamed, Brian Kingsbury, and Bhuvana Ramabhadran. ”Deep convolutional neural networks for LVCSR.” In 2013 IEEE international conference on acoustics, speech and signal processing, pp. 8614-8618, 2013.

[11] V. Jain and H. S. Seung, “Natural image denoising with convolutional networks,” Adv. Neural Inf. Process. Syst. 21 - Proc. 2008 Conf., pp. 769–776, 2009.

[12] H. C. Burger, C. J. Schuler, and S. Harmeling, “Image denoising: Can plain neural networks compete with BM3D?,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 2392–2399, 2012.

[13] J. Xie, L. Xu, and E. Chen, “Image denoising and inpainting with deep neural networks,” Adv. Neural Inf. Process. Syst., vol. 1, pp. 341–349, 2012.

[14] Y. Chen and T. Pock, ”Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1256-1272, 1 June 2017, doi: 10.1109/TPAMI.2016.2596743.

[15] V. Santhanam, V. I. Morariu and L. S. Davis, ”Generalized Deep Image to Image Regression,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 5395-5405, doi: 10.1109/CVPR.2017.573.

[16] Y. Tai, J. Yang, X. Liu and C. Xu, ”MemNet: A Persistent Memory Network for Image Restoration,” 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 4549-4557, doi: 10.1109/ICCV.2017.486.

[17] K. Zhang, W. Zuo, S. Gu and L. Zhang, ”Learning Deep CNN Denoiser Prior for Image Restoration,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 2808-2817, doi: 10.1109/CVPR.2017.300.

[18] X. J. Mao, C. Shen, and Y. Bin Yang, “Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections,” Adv. Neural Inf. Process. Syst., no. Nips, pp. 2810–2818, 2016.

[19] S. Lefkimmiatis, ”Non-local Color Image Denoising with Convolutional Neural Networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 5882-5891, doi: 10.1109/CVPR.2017.623.

[20] P. Qiao, Y. Dou, W. Feng, R. Li, and Y. Chen, “Learning non-local image diffusion for image denoising,” MM 2017 - Proc. 2017 ACM Multimed. Conf., pp. 1847–1855, 2017.

[21] R. Vemulapalli, O. Tuzel and M. Liu, ”Deep Gaussian Conditional Random Field Network: A Model-Based Deep Network for Discriminative Denoising,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 4801-4809, doi: 10.1109/CVPR.2016.519.

[22] A. Krizhevsky, “Learning Multiple Layers of Features from Tiny Images,” 2009.

[23] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” 2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc., no. Ml, pp. 1–14, 2014.

[24] J. Su, S. Wu, D. Xiong, Y. Lu, X. Han, and B. Zhang, “Variational recurrent neural machine translation,” 32nd AAAI Conf. Artif. Intell. AAAI 2018, pp. 5488–5495, 2018.

[25] K. Zhang, W. Zuo and L. Zhang, ”FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising,” in IEEE Transactions on Image Processing, vol. 27, no. 9, pp. 4608-4622, Sept. 2018, doi: 10.1109/TIP.2018.2839891.

[26] A. Graves, “Practical variational inference for neural networks,” in Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2011, pp. 2348–2356.