A Robust Restricted Boltzmann Machine for Binary Image Denoising

Rafael Pires†, Alexandre L. M. Levada
Gustavo B. Souza
Department of Computing
Federal University of São Carlos
São Carlos - SP, Brazil
{rafapires,gustavo.botelho}@gmail.com
alexandre@dc.ufscar.br

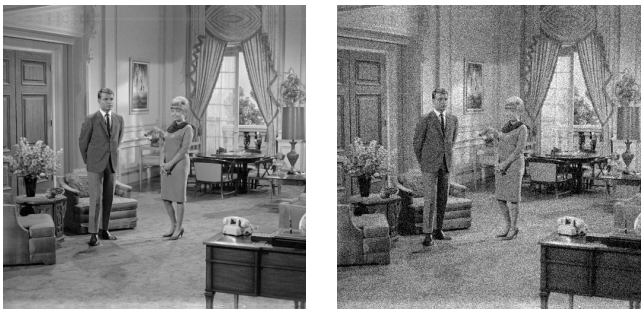
Luís A. M. Pereira†
Institute of Computing
University of Campinas
Campinas - SP, Brazil
luis.pereira@ic.unicamp.br

Daniel F. S. Santos†, João P. Papa
Department of Computing
São Paulo State University
Bauru - SP, Brazil
danielfssantos@gmail.com
papa@fc.unesp.br

***Abstract*—During the image acquisition process, some level of noise is usually added to the real data mainly due to physical limitations of the acquisition sensor, and also regarding imprecisions during the data transmission and manipulation. Therefore, the resultant image needs to be processed in order to attenuate its noise without loosing details. Machine learning approaches have been successfully used for image denoising. Among such approaches, Restricted Boltzmann Machine (RBM) is one of the most used technique for this purpose. Here, we propose to enhance the RBM performance on image denoising by adding a posterior supervision before its final denoising step. To this purpose, we propose a simple but effective approach that performs a fine-tuning in the RBM model. Experiments on public datasets corrupted by different levels of Gaussian noise support the effectiveness of the proposed approach with respect to some state-of-the-art image denoising approaches.**

I. Introduction

Noise is an undesirable artifact in images, being often caused by physical limitations of the image acquisition sensor or by unsuitable environmental conditions. These issues, however, are often unavoidable in practical situations, which turns the noise in images a prevalent problem. While qualitatively noise confers bad aspect to images (Figure 1), quantitatively it can impose difficulties to computational tasks such as edge detection, segmentation, and image classification. Hence, denoising is an important field in digital image processing.



(a) (b)

Fig. 1. An image can be visually ruined when contaminated by noise: (a) original noiseless image, and (b) its noise-contaminated version.

†These authors contributed equally to this paper.

Denoising an image is challenging because the noise is related to the high-frequency content of the image, that is, the details [1]. The goal, therefore, is to find a compromise between suppressing noise as much as possible and not losing too much image details. The most common approaches for image denoising are the filter-based ones such as the Median, Kuan, Wiener and BM3D filter [2]. They are simple and efficient; however, their effectiveness is highly dependent on the prior knowledge about the type of noise (e.g., Gaussian, salt-and-pepper, speckle) and its statistical properties (e.g., mean and variance) [1].

Machine learning approaches have been successfully used for image denoising (e.g., [4]–[6]). They are a natural option for the filter-based approaches because they are less affected by the non-specification of noise generative mechanism. The goal is to learn the noise characteristics without any prior knowledge related to its type and statistical properties. Among such approaches, Neural Networks have been one of the most explored techniques to model the image denoising problem [7]–[9]. The recent development in deep neural network architectures (e.g., deep Convolution Neural Networks) has also moved more eyes from the machine learning and computer vision communities to the image denoising task (e.g., [10]). Autoencoders are another prominent members of neural networks used for denoising [11]–[13], highlighting Restricted Boltzmann Machines (RBMs) [14]–[16].

RBMs are two-layers bidirectional neural networks that can be seen as a probabilistic graph model [3]. The most common used version of RBMs, known as Bernoulli-Bernoulli Restricted Boltzmann Machine (BB-RBM), has binary-valued units (or neurons) in its layers [14]. In the context of image denoising, the BB-RBM goal is to build a probabilistic model on a set of noise-contaminated images by using a single layer. The posterior probability of each noise-contaminated image is estimated. Finally, a noiseless image is then reconstructed by sampling from a conditional probability over the posterior probability [16]. BB-RBM performs well on low level amount of noise; however, as the noise increases, its performance may degrade. We hypothesize that this shortcoming may be related to the way BB-RBM model is learned, that is, in a fully

unsupervised fashion.

Here, we propose to enhance the BB-RBM performance on image denoising by adding a posterior supervision before its final denoising phase (i.e., the decoding phase). Indeed, we demonstrate that this process enhances the BB-RBM capacity of learning the noise model. To this purpose, we propose a simple but effective approach that performs a fine-tuning in the BB-RBM parameters concerning the decoding phase in order to reduce the noise levels. Experiments on public datasets corrupted by different levels of Gaussian noise support the effectiveness of the proposed approach with respect to some state-of-the-art image denoising methods.

II. Description of the Proposed Approach

In this section, some fundamental concepts about BB-RBM are first presented in Section II-A, for the further discussion of the proposed approach in Section II-B.

A. Bernoulli-Bernoulli Restricted Boltzmann Machine

RBMs are energy-based stochastic neural networks composed of two layers of units, a visible and a hidden, whose the learning phase is performed in an unsupervised fashion [14], [17]. They were originated based on the classical Boltzmann Machine (BM) [18] with the restriction that no connections among neurons of the same layer are allowed. This restriction allows training RBM in a significantly lower complex way in comparison to BM without high losses of performance. Roughly speaking, RBMs can be seen as a bipartite graph (Figure 2).

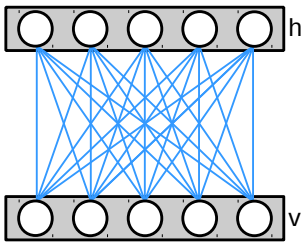


Fig. 2. The RBM architecture as a bipartite graph. Each neuron in the visible layer (**v**) is connected to all the neurons in the hidden layer (**h**). Note that there is no connection among units of the same layer.

In this paper, the proposed approach models the image denoising problem using the well-known Bernoulli-Bernoulli Restricted Boltzmann Machine whose both layers of units follow the Bernoulli distribution. The goal, therefore, is to model the image distribution content avoiding the noise distribution, which follows a Gaussian distribution in this work.

BB-RBM has a visible $\mathbf{v} = \{v_i\}_{i=1}^{m}$ and a hidden layer $\mathbf{h} = \{h_i\}_{i=1}^{n}$, $\forall v_i, h_i \in \{0, 1\}$ (i.e., binary-valued units). A matrix $\mathbf{W} \in \mathbb{R}^{m \times n}$ models the weights among the units in **v** and those in **h**; thus, an entry w_{ij} is the weight between the visible unit v_i and the hidden unit h_j. Furthermore, there are visible biases $\mathbf{b_v} = \{b_{v_i}\}_{i=1}^{m}$ associated to each visible unit, and hidden biases $\mathbf{b_h} = \{b_{h_i}\}_{i=1}^{n}$ associated to each hidden unit. The training goal, therefore, is to find the set of parameters $\eta = \{\mathbf{W}, \mathbf{b_v}, \mathbf{b_h}\}$ that maximizes the likelihood of a set of observations $\mathcal{V} = \{\mathbf{v}_i\}_{i=1}^{N}$ (e.g., N images) as follows:

$$\hat{\eta} = \arg\max_{\eta} \prod_{i}^{N} P(\mathbf{v}_i; \eta). \tag{1}$$

The marginal probability of an observation **v** is given by:

$$P(\mathbf{v}) = \sum_{\mathbf{h}} P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}, \tag{2}$$

in which

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j - \sum_{i=1}^{m} \sum_{j=1}^{n} v_i h_j w_{ij} \tag{3}$$

is the energy function, and Z stands for the well-known partition function, which is computed as follows:

$$Z = \sum_{\mathbf{v}} \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}, \tag{4}$$

Roughly speaking, Z is a normalization factor. Computing Z, however, is an intractable task because it is needed to compute all the possible configurations involving the visible and hidden units. Consequently, computing the joint probability $P(\mathbf{v}, \mathbf{h})$ is also intractable.

Fortunately, this issue can be overcome by using Gibbs sampling since the conditional probabilities in terms of **h** and **v** can be easily computed as follows:

$$P(v_i = 1 | \mathbf{h}) = \phi \left(b_{v_i} + \sum_{j=1}^{n} w_{ij} h_j \right) \tag{5}$$

and

$$P(h_j = 1 | \mathbf{v}) = \phi \left(b_{h_j} + \sum_{i=1}^{m} w_{ij} v_i \right), \tag{6}$$

in which $\phi(\cdot)$ is the sigmoid function.

In practice, the BB-RBM model is trained by sampling from $P(h_j = 1|\mathbf{v})$ and $P(v_i = 1|\mathbf{h})$ alternated. An approximation to the gradient descent, called the Contrastive Divergence (CD), is often used to this end [14].

B. Proposed approach

BB-RBM learns its hyper-parameters in an unsupervised fashion. Here, we describe the proposed approach that adds a posterior supervision to the BB-RBM model, and how it can be used to suppress noise in images.

Let $\mathcal{V} = \{\mathbf{v}_i\}_{i=1}^{N}$ be a set of noisy images, and $\mathcal{V}' = \{\mathbf{v}'_i\}_{i=1}^{N}$ be the set of their respectively noiseless version. Furthermore, let $\mathbf{W}$ and $\mathbf{b_v}$ be the weight matrix and the visible bias learned by training the BB-RBM on the set of noisy images $\mathcal{V}$, respectively. The proposed approach attempts at fine tuning $\mathbf{W}$ and $\mathbf{b_v}$ such that the difference between $\mathcal{V}$ and $\mathcal{V}'$ is then reduced. To this end, the proposed approach models the problem of hyper-parameter fine-tuning in terms of the cross entropy, as follows:

$$J(\theta) = -\frac{1}{N}\sum_{i=1}^{N} \mathbf{v}'_i \log(z(\mathbf{v}_i;\theta)) + (1-\mathbf{v}'_i)\log(1-z(\mathbf{v}_i;\theta)), \tag{7}$$

where

$$z(\mathbf{v}_i;\theta) = \phi\left(\hat{\mathbf{b}}_{\mathbf{v}} + \widehat{\mathbf{W}}^{\top}\mathbf{v}_i\right), \tag{8}$$

in which $\theta = \{\widehat{\mathbf{W}}, \hat{\mathbf{b}}_{\mathbf{v}}\}$ is the set of parameters of interest, that is, $\widehat{\mathbf{W}}$ and $\hat{\mathbf{b}}_{\mathbf{v}}$ stand for the weight matrix and the visible bias after the fine-tuning process, respectively. In order to minimize Equation 7, we use the gradient descent algorithm as implemented in Algorithm 1.

Algorithm 1: Proposed algorithm.

Input : $\mathcal{V}'$, $\mathcal{V}$, $\mathbf{W}$, $\mathbf{b}_{\mathbf{v}}$, α (learning rate), and T (number of iterations)

Output: $\widehat{\mathbf{W}}$ and $\hat{\mathbf{b}}_{\mathbf{v}}$

for $t \in \{1, 2, \ldots, T\}$ **do**

  $\theta^{t+1} \leftarrow \theta^t + \alpha\frac{\partial J(\theta^t)}{\partial \theta^t}$

After finding $\widehat{\mathbf{W}}$ and $\hat{\mathbf{b}}_{\mathbf{v}}$, we can use them to suppress the noise in a test set (i.e., noisy images) $\mathcal{V}_{\text{test}}$ as follows:

$$\widehat{\mathcal{V}}_{test} = \mathcal{V}_{\text{test}}\widehat{\mathbf{W}}^{\top} + \hat{\mathbf{b}}_{\mathbf{v}}, \tag{9}$$

in which $\widehat{\mathcal{V}}_{test}$ is the filtered (reconstructed) version of $\mathcal{V}_{\text{test}}$.

III. Experimental design

A. Datasets

Four public datasets were use to evaluate the robustness of the proposed approach:

MNIST: contains $60,000$ training images and $10,000$ test images of handwritten digits [19]. A subset of $10,000$ images was sampled from the training set; hereinafter this subset will be called *noiseless training set*. Three copies of the noiseless training set, referred to as *pre-training noisy sets*, were created and they were respectively contaminated with Gaussian noise with variance (σ) of 0.1, 0.2, and 0.3. For testing purposes, a subset of $1,000$ images was sampled from the test set and three copies of this subset, referred to as *noisy test sets*, were created and contaminated with the same noise level used for the pre-training noisy sets.

USPS: contains $11,000$ images of handwritten digits [20]. The dataset was split into a noiseless training set containing $10,000$ images and a test set containing $1,000$ images. Three pre-training noisy sets were created from the noiseless training in the same fashion performed for MNIST. Additionally, three noisy test sets were created from the test set also in the same fashion performed for MNIST.

Semeion: contains $1,593$ images of handwritten digits [21]. The dataset was split into a training set containing $1,000$ images and a test set containing 593 images. A noiseless training set was created from the training set by increasing the number of images from $1,000$ to $10,000$. For this end, we created 10 copies of the training set. One copy was kept without any noise, whereas in the remaining nine copies we applied the Gaussian noise with a small variance of $0.0001, 0.0002, \ldots, 0.0009$, respectively. Furthermore, three pre-training noisy sets were created from the noiseless training in the same fashion conducted for MNIST and USPS datasets. Finally, three noisy test sets were created from the test set also in the same fashion performed for MNIST and USPS. Note that MNIST contains 28×28-sized images, wheres Semeion and USPS contain 16×16-sized images. Hence, we centered the images from Semeion and USPS datasets into a 28×28-black-squared window to obtain all the digits with same size.

Caltech: contains $4,100$ training images and $2,307$ test images of object's silhouette [22]. A noiseless training set was created from the training set by increasing the number of images from $4,100$ to $8,200$. For this purpose, we created two copies of the training set. One copy was kept without any noise, whereas in the remaining one we applied a Gaussian noise with a small variance of 0.0001. Besides, three pre-training noisy sets were created from the noiseless training set in the same fashion conducted for the previous dataset. Finally, three noisy test sets were created from the test set also in the same fashion conducted for the previous dataset.

B. Experimental settings

The proposed approach was compared against four techniques:

- **BM3D**: denoising method based on an enhanced sparse representation [2];
- **ND** (No-denoising): simple comparison between a noise-contaminated image and its noiseless version;
- **BB-RBM**: standard Bernoulli-Bernoulli Restricted Boltzmann Machine; and
- **Wiener Filter:** filter with a 3×3-window-size [1].

We evaluated the proposed approach in three different experimental settings:

- **Standard**: the training (including the pre-training with BB-RBM) and the test phases were performed on the *same* dataset. BB-RBM was pre-trained on each one of the pre-training sets and the proposed approach was trained on the noiseless training set (recall the training of the proposed approach consists in a pre-training and a training phases). The denoising was performed on the noisy test sets. Here, the proposed approach was compared against all baselines. Techniques ND, BM3D, and the Wiener filter were performed only on the noisy test sets. BB-RBM was trained on the pre-training noisy sets and tested on the noisy test sets.
- **Cross-dataset**: the training and the test phases were performed on *different* datasets, and only the digit datasets (MNIST, USPS and Semeion) were considered in this setting. Both BB-RBM and the proposed approach were pre-trained on each one of the pre-training sets, and the proposed approach was trained on the noiseless training set. The denoising was performed on the noisy test

set of a different dataset. Here, the proposed approach was compared against BM3D and BB-RBM. BM3D was applied only on the noisy test sets, and BB-RBM was evaluated on the noisy test set of another dataset not used for its training step.

- **Transfer learning**: the training and the test phases were performed on the *same* dataset, and only the digit datasets (MNIST, USPS and Semeion) were considered in this setting. Both BB-RBM and the proposed approach were pre-trained on each one of the pre-training sets of a given dataset, and only the proposed approach was trained (fine-tuned) on the noiseless training set of a *different* dataset. The test (denoising) was performed on the noisy test sets belonging to the same dataset in which the approaches were pre-trained on. We compared the performance of the proposed approach in the transfer learning setting against the performances achieved by BB-RBM and the proposed approach in the standard setting.

The hyper-parameters used to train the BB-RBM and the proposed approached are summarized in Table I. These values were empirically chosen.

TABLE I
HYPER-PARAMETERS USED IN THREE EXPERIMENTAL SETTINGS.

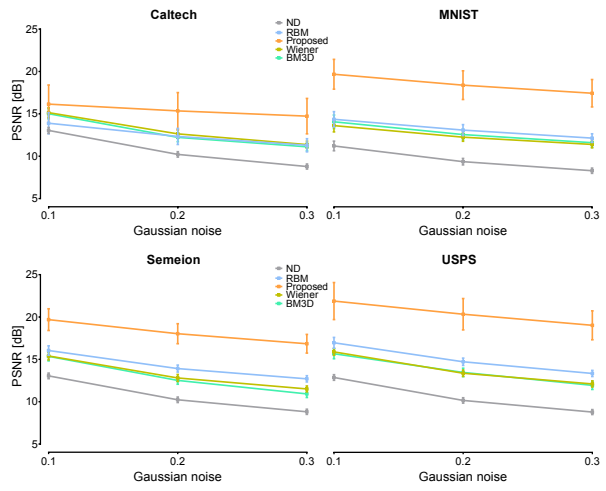
	Parameters	Standard	Cross-dataset	Transfer learning
Pre-training	Learning for weights	0.005	0.005	0.005
	Learning for visible units	0.005	0.005	0.005
	Learning for hidden units	0.005	0.005	0.005
	Momentum (min)	0.5	0.5	0.5
	Momentum (max)	0.9	0.9	0.9
	Weight decay	0.0002	0.0002	0.0002
	Epochs	50	50	50
	Batch size	100	100	300
Training	Learning rate	0.1	0.1	0.001
	Batch size	100	300	300
	Number interations	5	1	1

Figure 3 displays the proposed pipeline for image denoising. As aforementioned, the approach adopted in this work is divided in two steps: pre-training and training. While the former is in charge of an unsupervised learning process, the latter one makes use of clean images to fine-tune the weights learned in the previous step.

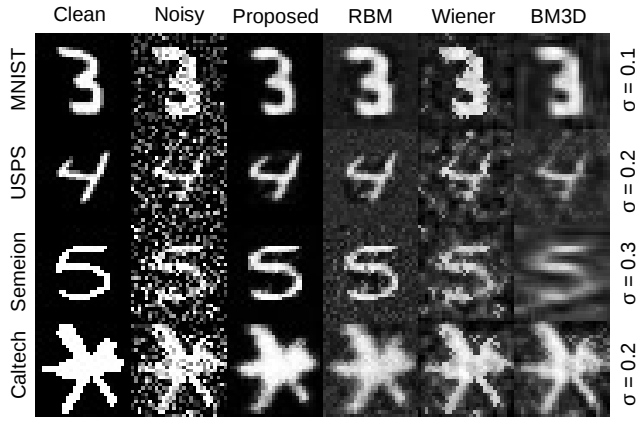
IV. Experimental Section

At first, we evaluated the proposed approach using the standard setting, as described in the previous section. Figure 4a displays the Peak signal-to-noise Ratio (PSNR) of the compared techniques over different levels of Gaussian noise for the datasets considered in this work. On training and testing in the same dataset, the proposed approach overcame all baselines used in the experiments. Notice the proposed approach has been less affected to increasing noise levels when compared to the other methods. These quantitative results impacted positively in the visual quality of the images, as one can observe in Figure 4b.

Figure 5a depicts the cross-dataset experiment, in which the proposed approach performed better once again. In most



(a)



(b)

Fig. 4. Standard setting: (a) quantitative results over Caltech, MNIST, USPS, and Semeion datasets; and (b) qualitative results over MNIST, USPS, and Semeion datasets. Notice that σ denotes the variance of the Gaussian noise.

cases, the proposed approach overcame BM3D, mainly at higher noise levels. It is important to highlight that BM3D is considered one the best denoising techniques to date. With respect to BB-RBM, the proposed approach has been slightly better in reducing the noise regarding the quantitative and qualitative results, as displayed in Figure 5b.

Finally, the proposed approach was evaluated in the transfer learning setting. We observed the proposed approach with transfer learning performed worse than its version with clean data; however, it performed better than BB-RBM in all situation, mainly at higher noise levels, as one can observe in Figures 6a and Figures 6b.

V. Discussion and Conclusions

In this work, we demonstrated that a posterior supervision of the BB-RBM decoding phase enhances its capacity of image denoising. To this end, we proposed a technique that performs a fine-tuning in the BB-RBM hyper-parameters concerning the decoding phase. The reasoning behind this idea is to reduce the

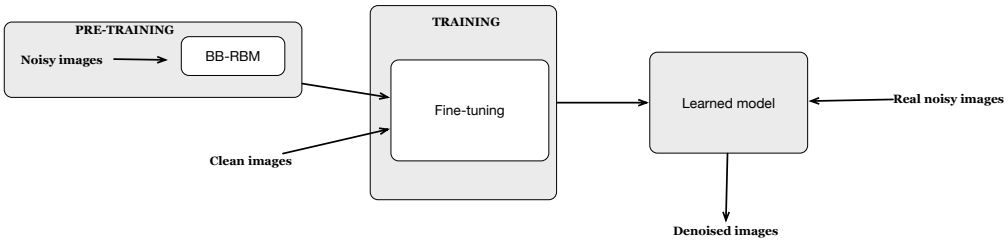


Fig. 3. Proposed pipeline for image denoising based on Restricted Boltzmann Machines.

error between the visible conditional probabilities and a given clean-image dataset. After applying the proposed approach, we can use the new set of parameters as an image denoising model to reduce the noise not suppressed by the BB-RBM, thereby enhancing the visual quality of a noise-contaminated image not seen in the learning phase. Experiments on public datasets corrupted by different levels of Gaussian noise support the effectiveness of the proposed approach in comparison to state-of-the-art image denoising methods.

Due to its natural ability to reconstruct images, RBMs have been studied in the context of image denoising (e.g., [12]). However, they may fail to accomplish such a task if the level of noise is considerably high. Consequently, its performance degrades to results similar to the ones achieved by approaches that do not need any kind of learning phase, such as filter-based methods (Figures 4a and b). This drawback may be related to its learning process, which is fully unsupervised. However, the proposed approach is able to fine-tune the BB-RBM parameters with the aid of clean image data, thus enhancing significantly the visual image quality (Figure 4b).

Furthermore, the proposed approach enhances the quality of images even more than BB-RBM in the cross-dataset setting, that is, training in one dataset and testing in another one. The results suggest the proposed approach is less sensitive to be applied to another dataset (Figure 5a and b). We observe, however, that the performance of the proposed approach in the cross-dataset setting is lower than the one observed in the standard setting (Figure 5a). This is likely because the dataset may have a certain difference among their marginal probabilities [23].

In certain cases, however, we may not have enough clean images data to perform the fine-tuning as in the standard and cross-dataset settings. What we can do then is somehow share the knowledge of another dataset that contains clean images [11]. In our experiments, we evaluated a simple transfer learning strategy to exploit the knowledge of another related dataset. In this setting, the proposed approach enhanced the image quality even more than BB-RBM (Figure 6a and b). This result suggested the proposed approach can handle the absence of clean image data by simply transferring knowledge from another dataset. We observe, however, that the performance of this strategy was lower than the one using clean image data from the same dataset (Figure 6a and b). This was probably because we did not use any strategy to reduce statistical differences among the datasets before applying the fine-tuning.

In regard to future works, we intend to evaluate the proposed approach in the context of gray-level images, as well as to transfer knowledge from non-related datasets.

ACKNOWLEDGMENT

The authors are grateful to FAPESP grants #2014/12236-1, #2016/19403-6 and #2015/09169-3, Capes, and CNPq grant #306166/2014-3.

REFERENCES

[1] R. Gonzalez and R. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3d filtering," in *Electronic Imaging 2006*. International Society for Optics and Photonics, 2006, pp. 606 414–606 414.

[3] A. Fischer and C. Igel, "Training restricted boltzmann machines: An introduction," *Pattern Recognition*, vol. 47, no. 1, pp. 25–39, 2014.

[4] J. Qiao and J. Liu, *A SVM-Based Blur Identification Algorithm for Image Restoration and Resolution Enhancement*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 28–35.

[5] D. Li, "Support vector regression based image denoising," *Image and Vision Computing*, vol. 27, no. 6, pp. 623–627, 2009.

[6] Y. Xia, C. Sun, and W. X. Zheng, "Discrete-time neural network for fast solving large linear l1 estimation problems and its application to image restoration," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 5, pp. 812–820, 2012.

[7] Y.-T. Zhou, R. Chellappa, A. Vaid, and B. K. Jenkins, "Image restoration using a neural network," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 7, pp. 1141–1151, 1988.

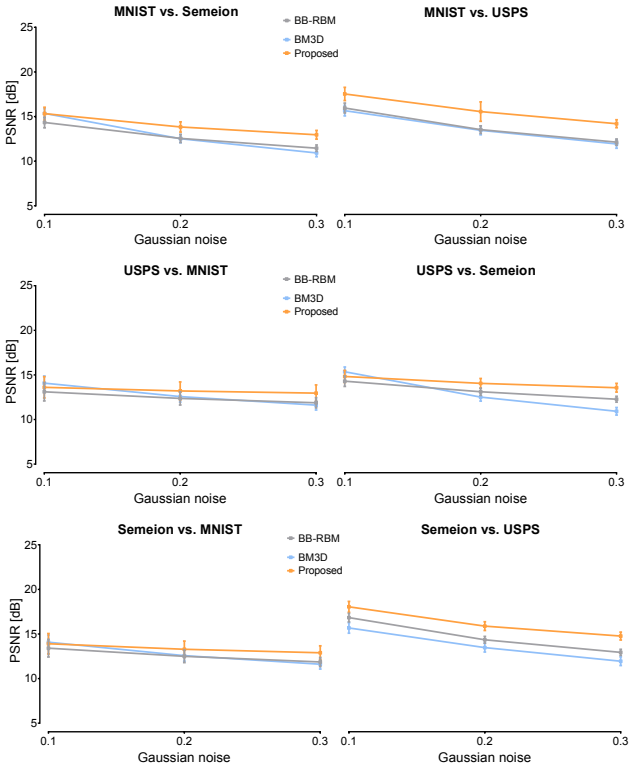
[8] J. Paik and A. Katsaggelos, "Image restoration using a modified hopfield network," *IEEE Transactions on Image Processing*, vol. Jan. Vol. 1, pp. 49–63, 1992.

[9] Y. Sun and S. Y. Yu, "A modified hopfield neural network used in bilevel image restoration and reconstruction," in *International Symposium on Information Theory Application*, vol. 3, 1992, pp. 1412–1414.

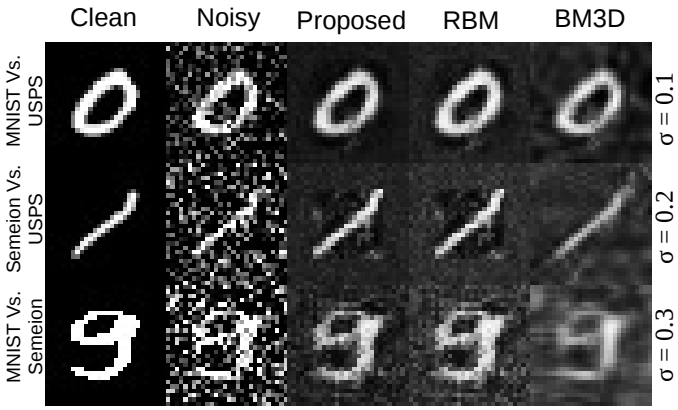
[10] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with bm3d?" in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2392–2399.

[11] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 341–349.

[12] Y. Tang, R. Salakhutdinov, and G. E. Hinton, "Robust boltzmann machines for recognition and denoising," in *IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '12. IEEE, 2012, pp. 2264–2271.

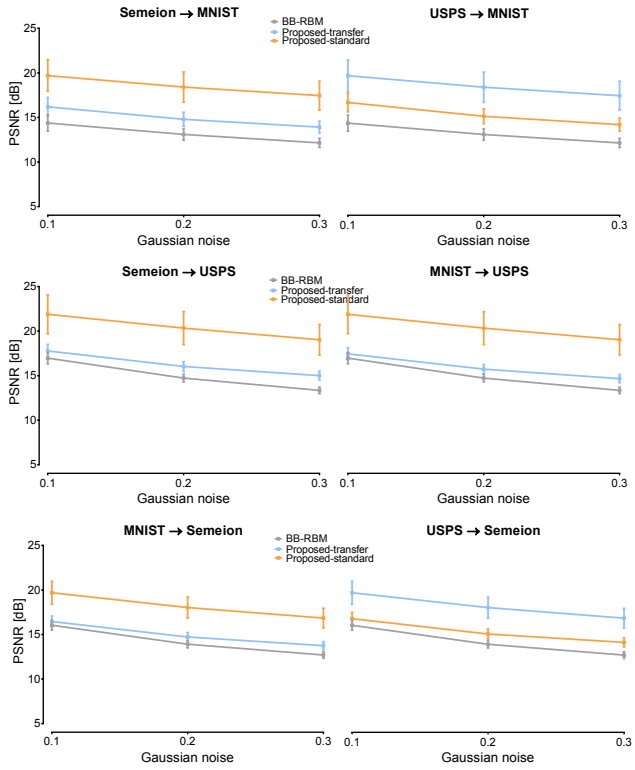


(a)

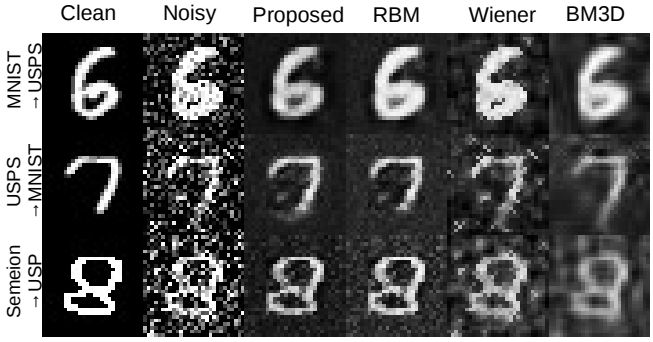


(b)

Fig. 5. Cross-dataset setting: (a) quantitative results over Caltech, MNIST, USPS, and Semeion datasets; and (b) qualitative results over MNIST, USPS, and Semeion datasets. Notice that σ denotes the variance of the Gaussian noise. The notation A vs. B (e.g., MNIST vs. USPS) means that the proposed approach and BB-RBM were trained on A and tested on B.



(a)



(b)

Fig. 6. Transfer learning setting: (a) quantitative results over Caltech, MNIST, USPS, and Semeion datasets; and (b) qualitative results over MNIST, USPS, and Semeion datasets. The notation A→B means the proposed proposed approach transferred the knowledge from A (noiseless images) and was tested on B (noise-contaminated images).

[13] R. Yan and L. Shao, "Image blur classification and parameter identification using two-stage deep belief networks," in *24th British Machine Vision Conference*, 2013, pp. 1–11.

[14] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[15] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[16] Y. Tang, R. Salakhutdinov, and G. Hinton, "Robust boltzmann machines for recognition and denoising," in *IEEE Conference on Computer Vision and Pattern Recognition, 2012, Providence, Rhode Island, USA*, 2012.

[17] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," DTIC Document, Tech. Rep., 1986.

[18] D. Waltz, *Connectionist models and their implications: readings from cognitive science*. Intellect Books, 1988.

[19] Y. LeCun, C. Cortes, and C. J. Burges, "The mnist database of handwritten digits," 1998.

[20] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2001.

[21] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[22] B. Marlin, K. Swersky, B. Chen, and N. Freitas, "Inductive principles for restricted boltzmann machine learning," in *International Conference*

on Artificial Intelligence and Statistics, 2010, pp. 509–516.
[23] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 1521–1528.