RWRM: RESIDUAL WASSERSTEIN REGULARIZATION MODEL FOR IMAGE RESTORATION

RUIQIANG HE\textsuperscript{a,b}, XIANGCHU FENG\textsuperscript{a,*}, XIAOLONG ZHU\textsuperscript{a}, HUA HUANG\textsuperscript{a} AND BINGZHE WEI\textsuperscript{a}

\textsuperscript{a}School of Mathematics and Statistics, Xidian University, Xi’an 710126, China
\textsuperscript{b}Department of Mathematics, Xinzhou Teachers University, Xinzhou 034000, China

ABSTRACT. Existing image restoration methods mostly make full use of various image prior information. However, they rarely exploit the potential of residual histograms, especially their role as ensemble regularization constraint. In this paper, we propose a residual Wasserstein regularization model (RWRM), in which a residual histogram constraint is subtly embedded into a type of variational minimization problems. Specifically, utilizing the Wasserstein distance from the optimal transport theory, this scheme is achieved by enforcing the observed image residual histogram as close as possible to the reference residual histogram. Furthermore, the RWRM unifies the residual Wasserstein regularization and image prior regularization to improve image restoration performance. The robustness of parameter selection in the RWRM makes the proposed algorithms easier to implement. Finally, extensive experiments have confirmed that our RWRM applied to Gaussian denoising and non-blind deconvolution is effective.

1. Introduction. Image restoration, which aims to estimate the original clean image $x$ from the observed image $y$, is a classical but still active research topic in low level vision [22] [31]. A widely used data model [6] [49] [33] can be formulated as $y = Kx + n$, where $n$ is additive white Gaussian noise (AWGN), and $K$ is a degenerate operator, typically an identity operator in image denoising, a blurring operator in image deblurring, a down-sampling operator in image super-resolution, a random projection operator in compressed sensing [7] [1]. These inverse problems are of great concern because of their indispensable steps in many practical applications.

For several decades, new approaches in image restoration have been emerging. The statistical modeling based on natural image priors plays an important role in these image restoration tasks. Models built by mining image priors mainly include nonlocal self-similarity (NSS) models [33] [5] [30] [18] [32], sparse representation models [22] [28] [8], gradient domain models [25] [44], and Markov random field (MRF) models [19] [20]. In the NSS models, there are some state-of-the-art representative methods such as BM3D [5], NCSR [6].

In recent years, another type of method based on image prior modeling is to utilize histogram approximation. It is well known that a histogram is a discrete approximation form of the probability density distribution of a continuous random variable. For an image, its histogram is defined as the statistical relationship between the image gray levels and the corresponding pixel numbers [10]. In many
fields of image processing, the histograms of pixel or gradient domains can provide effective information about image statistics. For instance, a texture enhanced image denoising algorithm GHP [49] is presented by enforcing the denoised image gradient histogram to be close to the original image reference gradient histogram. GHP [49] achieves good image restoration effect at the expense of complex reference gradient histogram estimation.

Similarly, in UniHIST [24], the Wasserstein distance is used to minimize the difference between the histogram of the restored image and the reference histogram of the original image in the pixel and gradient fields. For UniHIST denoising [24], the images involved are pattern images, not natural images, and the reference histogram is assumed to be known. In [39], a variational model with the image histogram prior is proposed, and two equivalent convex relaxations of the given problem are discussed. In addition, in [29] [40] [9], various variants of image histogram prior are employed for image color transfer, image synthesis and texture reconstruction, respectively. It can be found that the image prior information is crucial in these applications.

Different from the above image restoration models that only utilize image prior regularizations, in this paper a residual Wasserstein regularization model (RWRM) is designed to regularize the residual image \( n = y - Kx \), i.e., the difference between the degraded observation \( y \) and image formation model \( Kx \). It turns out that our residual Wasserstein and image prior integration is effective in boosting image restoration performance.

In our RWRM, the benefits of using the residual Wasserstein regularization are two-fold. First, without the complicated reference histogram estimation like GHP [49], the reference Gaussian residual histogram is readily available and is stable. Second, complex manual parameter adjustments in many non-convex optimization problems are not required. This is because the parameter selection is robust, which facilitates the experiment implementation.

In this work, we employ the Wasserstein distance to force the observed image residual histogram close to the reference Gaussian residual histogram. Further, the residual regularization and the image prior regularization are coupled to restore the image. In terms of restoration performance, our RWRM is beyond the popular BM3D [5], NCSR [6] and GHP [49]. We do not claim that our RWRM can achieve the best restoration results. However, we want to illustrate from the residual estimation perspective that the proposed RWRM can be combined with other variational methods to improve the image restoration quality.

The main contributions of this paper can be summarized as follows:

(1) We propose a residual Wasserstein regularization model (RWRM) in which the residual histogram constraint is embedded within a class of variational problems. Many existing image priors can be incorporated into the model to enhance the output image quality.

(2) The histogram matching by Wasserstein distance effectively minimize the discrepancy between the estimated residual histogram and the reference residual histogram. Further, the RWRM is applied to Gaussian denoising and non-blind deconvolution.

(3) Histogram matching and Chambolle duality projection algorithm are alternated to solve the RWRM. The parameter selection robustness makes our optimization algorithms easy to implement.
The rest of this paper is organized as follows. Section 2 provides a survey of the related work. Section 3 proposes the residual Wasserstein regularization model (RWRM) for image restoration. Section 4 presents the application of the RWRM in Gaussian denoising. The RWRM applied to non-blind deconvolution is solved in Section 5. Finally, Section 6 concludes the work.

2. Related work.

2.1. Literature analysis. Image restoration methods can be divided into two categories: model-based approaches and learning-based methods. Model-based approaches make full use of image and noise prior models to reconstruct the clean image, while learning-based methods manage to learn a mapping function from the degraded image to the clean image [38] [17] [14]. Here we briefly review some approaches that impose the Wasserstein constraint in image processing.

In the context of matching histogram, Woodford et al. [45] utilized a MAP estimation framework, namely marginal probability field (MPF) model to match a histogram of low-level image features such as gradients and texels for a range of image applications, including image segmentation, texture denoising and texture synthesis. Heeger and Bergen [16] implemented texture synthesis of an image that matches the texture histogram of a given image. Schmidt et al. [35] reconstructed images by matching the gradient distribution with an efficient Gibbs sampler, and achieved an excellent denoising effect. It can be found that these methods require multiple samples and complex calculations to restore a clean image.

In [3], Cho et al. developed iterative distribution reweighting (IDR) that enforces a global constraint on gradients in the process of image restoration. The approach minimizes the KL divergence between the restored image gradient distribution and the estimated reference distribution. IDR enhances the visual details of the reconstructed images with a high precision parameter density fitting.

Swoboda et al. in [39] presented a histogram prior based variational model. They mainly discussed two convex relaxations of the given original problem, and gave an equivalent proof of these two relaxations. By minimizing the Wasserstein distance between the discrete distributions of some features, a variational approach for image synthesis and image restoration [40] was presented. In [9], a high-resolution patch of the image texture was assumed to be known. Gheche et al. employed the Wasserstein distance to compare the histograms of the intensity values, the gradient and the Laplacian, so that the synthesized image has similar detail information as the high-resolution patch. In [29], the sliced Wasserstein distance between two d-dimensional N-point clouds was utilized as a statistical fidelity term for color transfer application.

Zuo et al. [49] showed a texture-enhanced image denoising approach, which adds a gradient penalty term to a non-local patch-based image model. This method improves the image texture structure while removing noise by enforcing the gradient histogram of the latent denoised image to be as close as possible to the reference gradient histogram of the original clear image. Recently, Mei et al. [24] presented a unified method to enforce histogram constraints in pattern image denoising and non-blind deconvolution. The critical factor in their work is to minimize the difference between the histograms of the restored image and the reference histograms of the original image in pixel and gradient domains by employing the Wasserstein distance.

In W-LDMM [15], a Wasserstein driven low-dimensional manifold model was proposed to implement image denoising and image inpainting. W-LDMM makes
full use of the complementarity of Wasserstein distance noise estimation and low-dimensional manifold. In [11], convergence in the Wasserstein metric with applications to image restoration was discussed. In the paper [47], a low-dose CT image denoising method was introduced utilizing Wasserstein distance by generative adversarial networks (GANs). In [46], a single-image super-resolution method combining Wasserstein generative adversarial networks and deep residual network was proposed, which aims to generate realistic images with finer texture details. Similarly, in [41], Wasserstein GANs were also used for CT image reconstruction from a sparse Angle.

It can be found that the existing image restoration models almost all use various image prior information, but rarely dig out the residual information. As discussed in Section 1, considering the simplicity and validity of the Gaussian residual histogram, we propose the residual Wasserstein regularization model (RWRM) to restore images. We all know that Gaussian residual histogram is readily available, and it is stable under certain condition. By means of Wasserstein distance histogram matching, the estimated residual histogram can be close to the reference residual histogram. Moreover, we subtly impose residual histogram constraint on image restoration models for higher quality image restorations including Gaussian denoising, non-blind deconvolution. In the following we will focus on the theory of Wasserstein distance as the dissimilarity measure between two histograms.

2.2. **Wasserstein distance.** It is well known that Wasserstein distance [26] [29] [27] originated from the theory of optimal transport [43] is defined as the minimal cost which must be paid to convert one histogram into the other. Wasserstein distance is a natural metric between two histograms. Intuitively, if we regard two histograms as two piles of sand on the ground, each grain of sand is an observation sample. To quantify the difference between these two distributions, we need to measure the minimum amount of sand that has to be moved so that these two distributions are exactly the same. Therefore, Wasserstein distance is the minimal total travelled ground distance that is weighted by the amount of sand moved.

For the two probability distributions $p$ and $q$ defined on the real number field $R$, the Wasserstein distance of $p$ and $q$ can be expressed as the solution to the following Monge problem [43]:

$$W(p, q) = \inf_{\phi} \int_{-\infty}^{\infty} (x - \phi(x))^2 p(x)dx$$

where the random variable $x$ obeys the distribution $p$, while the variable $\phi(x)$ follows another distribution $q$. The infimum in Eq. (1) is for all the determined functions $\phi : R \rightarrow R$ that map any random variable $x$ into another variable $\phi(x)$. One can see that the Wasserstein distance is a true statistical metric between two given probability distributions, and will be zero if and only if $p$ and $q$ are the same distribution.

Given a probability distribution $p$ on $R$, we can define its cumulative distribution function (cdf) by the formula $F_p(x) = \int_{-\infty}^{x} p(\tau)d\tau$, where $F$ is called a histogram equalization transform [13]. In addition, the percentile function of the probability distribution $p$ (the pseudo-inverse of $F_p$) is defined as follows $F^{-1}_p(t) = \inf\{x \in R : F_p(x) > t\}$. In the comprehensive introduction of optimal transport theory [43], a fundamental result is that the optimal $\phi$ in Eq. (1) has the following closed-form solution:
\[ \phi_{p \to q}(x) = F_q^{-1}(F_p(x)). \]

Noting that the Wasserstein distance used to measure the difference between the two distributions is expressed as an expected value, we can introduce another form of the Wasserstein distance

\[ W(p, q) = \inf_{\phi} E(x - \phi(x))^2 \]

By representing \( x = (x_1, x_2, \cdots, x_n)^T \) as \( n \) independent samples derived from the distribution \( p \), and treating histogram \( h_q \) as the discrete approximation form of another distribution \( q \), we can introduce the equivalent discrete definition of the Wasserstein distance as follows:

\[ \hat{W}(p, q) = \hat{W}(h_x, h_q) = \min_{\hat{\phi}} \frac{1}{n} \sum_{i=1}^{n} (x_i - \xi_i)^2 \]

where the function \( \hat{\phi} \) maps \( x_i \) to \( \xi_i = \hat{\phi}(x_i) \), such that the transformed samples \( \xi = (\xi_1, \xi_2, \cdots, \xi_n)^T \) meet the histogram \( h_q \). Similar to the continuous situation just described, the optimal \( \hat{\phi} \) for Eq. (4) has also a closed-form solution [43]:

\[ \xi_i = \hat{\phi}_{h_x \to h_q}(x_i) = F_q^{-1}(F_{h_q}(x_i)) \]

where \( F_{h_x} \) and \( F_{h_q}^{-1} \) are the cumulative distribution function and the percentile function that are from the histograms \( h_x \) and \( h_q \) respectively. For an input image \( x = (x_1, x_2, \cdots, x_n)^T \), the histogram matching operation \( \hat{\phi} \) can ensure that the output sample \( \xi = (\xi_1, \xi_2, \cdots, \xi_n)^T \) better matches the given histogram \( h_q \). Eqs. (4) and (5) constitute the theoretical basis of the coming image restoration.

3. The proposed residual Wasserstein regularization model for image restoration. In this section, we present the proposed residual Wasserstein regularization model (RWRM) and do some necessary analysis. Then, the basic approach of solving the RWRM is given.

3.1. The residual Wasserstein regularization model (RWRM). The observed version \( y \) of a underlying clean image \( x \) is usually defined by the model:

\[ y = Kx + n \]

where \( n \) is the AWGN with zero mean and variance \( \sigma^2 \). \( K \) is a linear operator which accounts for some damage to the image. Based on the maximum-a-posteriori (MAP) estimation, the desired image \( x \) can be obtained by solving the optimization problem

\[ \arg\min_x \frac{1}{2\sigma^2} \|y - Kx\|_2^2 + \lambda R(x) \]

where the specific form of the regularization term \( R(x) \) depends on the utilized image priors, and \( \lambda \) is a scale parameter.

Different from most image restoration models which only contain image prior regularization, we deal with image restoration by regularizing the residual image \( n = y - Kx \), i.e., the difference between the degraded observation \( y \) and image formation model \( Kx \). Fully utilizing residual regularization strategy, we propose
the following framework of residual Wasserstein regularization model (RWRM) for image restoration:

\[
\begin{align*}
\text{arg min}_{n,x} & \quad \frac{1}{2\sigma^2} \|n\|_2^2 + \lambda R(x) + \frac{\beta}{2} \hat{W}(h_n, h_g) \\
\text{s.t.} & \quad y = Kx + n
\end{align*}
\]

where \( h_n \) is the estimated residual histogram, and \( h_g \) is the reference Gaussian residual histogram (ground truth). \( \hat{W}(h_n, h_g) \) denotes the Wasserstein distance between \( h_n \) and \( h_g \). \( \beta \geq 0 \) is the Wasserstein regularization parameter.

It is not difficult to find that the framework (8) actually embeds the residual Wasserstein regularization into the exiting image restoration models. It combines residual estimate and image prior regularization organically. For various image priors \( R(x) \), for instance, total variation (TV) [33], wavelet sparsity [23], nonlocal total variation (NLTV) [12] and sparse-representation modeling [22] [8], we can exploit different image restoration models. It can be seen that the proposed RWRM framework (8) has extensive restoration capabilities.

Without loss of generality, in our work we choose the classic total variation [33] as the regularization term to illustrate the effectiveness of the proposed RWRM framework. In other words, in Eq. (8), \( R(x) = \|\nabla x\|_1 \). Therefore, we mainly analyze the following residual Wasserstein regularization model (RWRM) for image restoration:

\[
\begin{align*}
\text{arg min}_{n,x} & \quad \frac{1}{2\sigma^2} \|n\|_2^2 + \lambda \|\nabla x\|_1 + \frac{\beta}{2} \hat{W}(h_n, h_g) \\
\text{s.t.} & \quad y = Kx + n,
\end{align*}
\]

where \( \nabla x \) denotes the discrete gradient of the true image \( x \), and \( \|\nabla x\|_1 \) represents the sum of the length of gradient vector at each point.

The reasons of using the Wasserstein distance for residual regularization are two-fold. On the one hand, the reference Gaussian residual histogram is readily available and is stable under certain condition. On the other hand, the parameter selection in the RWRM is robust, making specific experiments easy to implement (see Fig. 11).

3.2. The basic approach of solving the RWRM. In order to solve the proposed RWRM (9), we first introduce an auxiliary variable \( \xi \), and the Eq. (9) is split into:

\[
\begin{align*}
\text{arg min}_{n,x,\xi} & \quad \frac{1}{2\sigma^2} \|n\|_2^2 + \lambda \|\nabla x\|_1 + \frac{\beta}{2} \|\xi - n\|_2^2 \\
\text{s.t.} & \quad y = Kx + n, h\xi = h_g
\end{align*}
\]

This problem can be solved by alternately minimizing Eq. (10) with fixed \( n, x \) (\( n, x \) are simultaneous) and \( \xi \) respectively.

Specifically, fixing \( n, x \), we get the following optimization problem:

\[
\begin{align*}
\text{arg min}_{\xi} & \quad \frac{\beta}{2} \|\xi - n\|_2^2 \text{s.t.} h\xi = h_g
\end{align*}
\]

Using the solution of the Wasserstein distance, the optimal \( \xi \) is given by the histogram matching operation as follows:

\[
\xi_i = \phi_{h_n \rightarrow h_g}(n_i)
\]

where \( i \) denotes the \( i-th \) element.
Fixing $\xi$, the $n, x$ problem in Eq. (10) reduces to

$$\arg \min_{n, x} \frac{1}{2\sigma^2} \| n \|^2_2 + \lambda \| \nabla x \|_1 + \frac{\beta}{2} \| \xi - n \|^2_2$$

subject to $y = Kx + n$.

The above problem can be written as an unconstrained optimization problem:

$$\arg \min_x \frac{1}{2\sigma^2} \| y - Kx \|^2_2 + \lambda \| \nabla x \|_1 + \frac{\beta}{2} \| y - Kx - \xi \|^2_2$$

Furthermore, Eq. (14) can be converted into

$$\arg \min_x \frac{1 + \beta \sigma^2}{2\lambda \sigma^2} \| \frac{y + \beta \sigma^2 (y - \xi)}{1 + \beta \sigma^2} - Kx \|^2_2 + \| \nabla x \|_1$$

For simplicity of notations, we set $\gamma = \frac{1 + \beta \sigma^2}{2\lambda \sigma^2}$, $z = \frac{y + \beta \sigma^2 (y - \xi)}{1 + \beta \sigma^2}$. Then Eq. (15) is rewritten as

$$\hat{x} = \arg \min_x \frac{\gamma}{2} \| z - Kx \|^2_2 + \| \nabla x \|_1$$

For different image restoration tasks, different approaches can be utilized to solve Eq. (16). For example, for Gaussian denoising, Chambolle’s duality projection algorithm [2] can be employed to solve this problem (16). For non-blind deconvolution, the optimization problem can be solved by the proximal forward-backward splitting (PFBS) method [4].

Once $\hat{x}$ in Eq. (16) is obtained, we can estimate the residual $n$ by

$$n = y - K\hat{x}.$$
TV-L2 model. Therefore, we present an approach to integrate residual Wasserstein regularization with TV-L2 model for Gaussian denoising.

4.2. The solution of RWRM for Gaussian denoising. According to the analysis in Section 3.3, when solving the RWRM applied to Gaussian denoising, we also need to specifically solve the following optimization problem:

\[
\hat{x} = \arg \min_x \gamma \frac{1}{2} \| z - x \|^2_2 + \| \nabla x \|_1
\]

where \( \gamma = \frac{1+\beta \sigma^2}{\lambda^2 \sigma^2} \), \( z = \frac{y + \beta \sigma^2 (y - \xi)}{1+\beta \sigma^2} \).

The problem (19) can be efficiently solved by the Chambolle’s duality projection algorithm [2]. Hence, the solution of (19) is obtained by

\[
\hat{x} = z - \frac{\text{div} p}{\gamma},
\]

where the vector \( p \) can be solved by the fixed-point method below: initializing \( p = 0 \) and iterating

\[
p = p + \frac{\tau (\text{div} p - \gamma z)}{1 + \frac{\tau | \text{div} p - \gamma z |}}
\]

with \( \tau \leq \frac{1}{8} \) to ensure convergence, and please see [2] for more details.

To sum up, in solving the proposed model Eq. (18) for Gaussian denoising, we need to alternate the two main steps: histogram matching operation Eq. (12) and Chambolle’s duality projection Eq. (20). The solution process of RWRM applied to Gaussian denoising can be summarized as Algorithm 1.

4.3. Experimental results for Gaussian denoising. We do not claim that our RWRM has the best restoration performance. We only want to show that in the RWRM the residual Wasserstein regularization can be combined with the TV constraint to improve the output image quality. In this section, we compare RWRM with TV-L2 and some popular methods such as BM3D [5], NCSR [6], GHP [49] and learning-based method LDCNN [48], and finally illustrate the robustness of parameter selection.

![Figure 1](image_url)
Algorithm 1 Gaussian Denoising by RWRM

1. Input: Degraded image $y$, $n=\text{eye(size}(y))$, $k = 1$, $\lambda$, $\beta$, $\tau$.
2. while $k \leq J$ do
3. Update $\xi$ by the histogram matching operation:
   $$\xi_i = \hat{\phi}_{h_n \rightarrow h_g}(n_i)$$
4. Update $\gamma$:
   $$\gamma = \frac{1 + \beta\sigma^2}{\lambda\sigma^2}$$
5. Update $z$:
   $$z = \frac{y + \beta\sigma^2(y - \xi)}{1 + \beta\sigma^2}$$
6. Fixed-point iteration: initializing $p = 0$ and iterating
   $$p = \frac{p + \tau\nabla(\text{div} p - \gamma z)}{1 + \tau|\nabla(\text{div} p - \gamma z)|}$$
7. Implement Chambolle’s duality projection:
   $$\hat{x} = z - \frac{\text{div} p}{\gamma}$$
8. Estimate residual:
   $$n = y - \hat{x}$$
9. $k \leftarrow k + 1$
10. end while
11. Output: Restored image $\hat{x}$.

4.3.1. Comparison between the proposed RWRM and the TV-L2. To verify the performance of the RWRM for Gaussian denoising, we apply it to ten natural images, whose scenes are shown in Fig. 1. All the experimental images are $512 \times 512$ grayscale images whose gray level is from 0 to 255. The noisy images are generated by adding AWGN with $\sigma = 25$ to the clean images. We first discuss the parameter settings in our algorithm, and then evaluate the performance with PSNR values.

1) Parameter Settings

In the following we will discuss the parameter settings in two cases: $\beta = 0$ and $\beta \neq 0$.

When $\beta = 0$, the proposed model is the TV-L2 minimization problem. Here there is only one model parameter $\lambda > 0$ which is introduced to control the trade-off between the regularity of the desired image $x$ and the fidelity of the noisy observation $y$. In the TV-L2 model, the adjustment of the regularization parameter $\lambda$ is challenging. This is because it is difficult to find a standard formula for estimating $\lambda$. The general method is to adjust the parameter $\lambda$ to ensure that the solution satisfies the discrepancy principle. Specifically, let the variance of the output residual $y - \hat{x}$ as much as possible equal to the noise variance $\sigma^2$. The main problem here is that $\lambda$ is the weight of the image regularization, and the image regularization metrics vary from image to image.
In this work, we present a method for selecting the regularization parameter $\lambda$ in the TV-L2 model. As one can see from the Fig. 2, when $\lambda$ is small, the TV-L2 removes the noise but blurs the image edges. When $\lambda$ is large, the TV-L2 retains the edges but introduces noise. Faced with this dilemma, the TV-L2 model with a moderate regularization parameter $\lambda$ can be implemented to remove noise while retaining details such as image edges. Specifically, for different noisy images, we experimentally choose different parameters $\lambda$, which are the integer multiples of 100 ranging from 1500 to 1800. In Fig. 3 (a), taking the image “Boat” as an example, the relationship graph between PSNR and the parameter $\lambda$ is presented. How to choose the optimal $\lambda$ is important, and the readers interested in the topic can also refer to the literatures [37] [34] for more details.

![Image](image1.png)

**Figure 2.** TV-L2 denoising results on noisy image ($\sigma = 25$) with different regularization parameters $\lambda$. One can find that the TV-L2 model with a moderate $\lambda \in [1500, 1800]$ can optimize the denoising effect.

When $\beta \neq 0$, the RWRM model contains two parameters $\lambda$ and $\beta$. For a fair comparison, according to the above method, we choose the regularization parameter $\lambda$, which can achieve the best denoising effect when $\beta = 0$. In the RWRM, the Wasserstein regularization parameters $\beta$ is easy to adjust. We believe that $\beta$ should be proportional to the parameter $\lambda$, and inversely proportional to the noise variance $\sigma$. Based on specific experimental experience, we employ $\beta = \frac{\lambda}{0.8\sigma}$ to estimate the parameter $\beta$. The graph of PSNR affected by the parameters $\lambda$ and $\beta$ is presented in Fig. 3 (b). One can see that the optimal PSNR is almost the same under different parameter settings. In addition, we take the identity matrix as the initialization residual $n$ in **Algorithm 1**.

In Gaussian denoising, histogram matching and Chambolle dual projection are alternately performed 5 times, so the iteration number is set to $J = 5$. We perform parameter estimation according to the formula $\beta = \frac{\lambda}{0.8\sigma}$. We set $\lambda = 1800$. For example, for $\sigma = 60$, then $\beta = 37.5$. In the experiment, we take $\tau = 1/8$ to ensure the convergence of fixed point iteration in Chambolle dual projection algorithm.

2) Analysis of experimental data

In Gaussian denoising, we make use of the reference Gaussian residual histogram $h_g$. For all the experimental images, we employ a 256-bin histogram $h_g$. The peak signal to noise ratio (PSNR) and structural similarity index measurement (SSIM) are evaluation metrics for our image restoration. The larger the PSNR and SSIM, the better the restoration effect is. Firstly, the definition of the mean square error (MSE) is given as follows,
FIGURE 3. Relationships between PSNR and model parameters taking the image “Boat”. (a) In the TV-L2 model \( \beta = 0 \), the effect of the parameter \( \lambda \) on PSNR. (b) In our RWRM, the effect of Wasserstein regularization parameter \( \beta \) on PSNR in the cases of three different \( \lambda \).

\[
MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} [x(i,j) - x_0(i,j)]^2,
\]

where \( M \times N \) is the size of the image, \( x_0 \) and \( x \) represent the original and denoised images, respectively. On this basis, the definition of PSNR can be expressed as

\[
PSNR = 10\log_{10} \frac{255^2}{MSE}.
\]

Secondly, SSIM is mainly used to measure the degree of similarity between two images. The higher the value of SSIM, the closer the content of the two images is. SSIM is defined as following

\[
SSIM = \frac{(2\mu_A\mu_B + C_1)(2\sigma_{AB} + C_2)}{\mu_A^2 + \mu_B^2 + C_1 + \sigma_A^2 + \sigma_B^2 + C_2},
\]

where \( \mu_A \) and \( \mu_B \) are the means of two different images \( A \) and \( B \), respectively. \( \sigma_A^2 \) and \( \sigma_B^2 \) are the variances of \( A \) and \( B \). \( \sigma_{AB} \) is the covariance of \( A \) and \( B \), \( C_1 = (0.01 \times 255)^2 \) and \( C_1 = (0.03 \times 255)^2 \) are constants used to maintain stability.

In Table 1, we show the denoised PSNR and SSIM of the RWRM and the TV-L2. It is gratifying that the average PSNR and SSIM values of RWRM are 1.13 and 0.0743 higher than TV-L2, respectively. In addition, the PSNR comparison curves of the two methods on the ten images are shown in Fig. 4. It can be found that the PSNR of the denoised image obtained by the RWRM is significantly better than that of the TV-L2 model. These indicate that the introduction of residual Wasserstein regularization based on the TV-L2 model can significantly improve the denoising performance. This is also the reason why the PSNRs of RWRM and TV-L2 in Fig. 4 are positively linearly correlated.

In Fig. 5, we show the distribution curve of the estimated residual using TV-L2, RWRM. They are also compared with real Gaussian distribution curve (ground truth). It can be found that RWRM has obvious advantages in residual distribution estimation. This is due to the fact that Wasserstein distance distribution estimation works.

3) Subjective evaluation
Table 1. Denoising PSNR comparison on the ten test images in Fig. 1: the TV-L2 model and our RWRM.

| Images  | Airfield | Lena  | Peppers | Plane | Couple | Girl | Lake | Boat | Lolo | Martha | Avg. |
|---------|----------|-------|---------|-------|--------|------|------|------|------|--------|------|
| TV-L2   | 25.55    | 29.23 | 28.70   | 28.02 | 31.49  | 29.91| 27.19| 27.92| 25.52| 29.44  | 28.30|
|         | 0.6689   | 0.7410| 0.7233  | 0.7198| 0.8300 | 0.7502| 0.7129| 0.7136| 0.6212| 0.7458 | 0.7227|
| RWRM    | 26.55    | 30.16 | 30.47   | 29.60 | 32.28  | 31.02| 27.90| 28.71| 26.90| 30.74  | 29.43|
|         | 0.7008   | 0.8183| 0.8070  | 0.8361| 0.8515 | 0.8244| 0.7635| 0.7990| 0.7500| 0.8195 | 0.7970|

Figure 4. PSNR comparison curve on the ten images about the proposed RWRM ($\beta \neq 0$) and TV-L2 ($\beta = 0$) in Gaussian denoising.

Figure 5. The distribution comparison for the final residuals of TV-L2 (green), RWRM (red), and the ground truth (blue). One can see that RWRM residual estimate is more accurate than TV-L2.

The experimental output images of the RWRM and the TV-L2 are presented in Fig. 6 and Fig. 7. One can see that the RWRM improves the visual quality in the following aspects: firstly, we acquire the clearer restored results than the TV-L2; secondly, in some cases, the TV-L2 might blur image structures, while our method can recover better detail features such as image edges. In a word, we can draw a conclusion that in the proposed RWRM the residual Wasserstein constraint is effective in Gaussian denoising.

4.3.2. Comparison with some popular methods. Here, we compare the proposed RWRM with some popular methods, including BM3D [5], NCSR [6], GHP [49]
Figure 6. Denoising results on the three images Peppers, Plane and Martha. Here (c) and (d) are the result images generated by the TV-L2 and our RWRM respectively. One can see that our approach clearly yields better visual results.

Figure 7. Local magnification comparison of denoised images. One can see that the RWRM can recover better image edge details.
operating environment. Our denoising method takes 16s to process a 256 × 256 image, which is much faster than NCSR (321s) and GHP (324s). BM3D processes a 256 × 256 image with 2s, but it should be noted that BM3D is mainly implemented by the C language. It should be mentioned that the start-of-the-art method UniHIST [24] only deals with pattern image denoising, not natural images. Therefore, there is no comparison with UniHIST.

In order to show the quantitative experimental results, we use the eight images shown in Fig. 8 as test images. We set five cases with noise standard deviation $\sigma = 60, 70, 80, 90$ and 100 respectively. The PSNR and SSIM results of the four denoising methods are presented in Table 2. One can see that the RWRM has significantly better denoising performance than NCSR [6] and GHP [49]. The RWRM overall performance can be comparable to BM3D [5]. However, we should note that BM3D is a block-matching denoising algorithm in which the computation amount is very large. As the noise level increases, the denoising performance of the RWRM is getting better than BM3D.

On the other hand, Fig. 9 shows the output images of the four denoising methods acting on test image 8 when the noise standard deviations are 60, 80, and 100, respectively. We can find that the proposed RWRM can get better image recovery. This is due to the fact that the RWRM makes the estimated residual histogram better approximate the true Gaussian residual histogram. The residual estimate is accurate, which helps to restore the image.

![Figure 8. Eight test images. From left to right, they are labeled as 1 to 8 respectively.](image)

Finally, we compare the image denoising effect between the proposed RWRM and the learning-based method LDCNN [48]. The test images used in the comparison experiment are 8 images as shown in Fig. 8. Table 3 shows the PSNR and SSIM results of the proposed RWRM and learning-based method LDCNN [48] in image denoising. Fig. 10 shows the visual effects of both methods on the test image 5 with $\sigma = 40$. It can be found that, although the numerical index of RWRM is not as good as LDCNN, the visual effects of the two methods have their own advantages and disadvantages, implying that the proposed method is effective.

4.3.3. Robustness of parameter selection. A great advantage of our RWRM is the robustness of parameter selection. More specifically, when $\beta = 0$ in the proposed model, the PSNR of the denoised image varies greatly with change of parameter $\lambda$. However, when $\beta \neq 0$ ($\beta$ is invariable), the PSNR of the denoising image varies very little with the change of $\lambda$.

The five subgraphs shown in Fig. 11 are parameter robustness curves, which are respectively obtained by denoising the five different test images in Fig. 8.
Table 2. PSNR (dB) and SSIM values of BM3D [5], NCSR [6], GHP [49] and our RWRM in Gaussian denoising.

| σ   | Methods   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | Avg.      |
|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----------|
| 60  | BM3D      | 24.77| 23.33| 29.95| 23.36| 27.43| 24.80| 23.27| 27.81| 25.59     |
|     | NCSR      | 24.62| 23.25| 29.27| 23.23| 27.04| 24.74| 23.32| 27.59| 25.38     |
|     | GHP       | 24.74| 23.32| 29.40| 23.32| 27.18| 24.71| 23.32| 27.88| 25.44     |
|     | RWRM      | 24.70| 23.25| 29.68| 23.32| 27.48| 24.75| 23.30| 27.88| 25.50     |
|     | PSNR      | 0.490| 0.481| 0.793| 0.463| 0.602| 0.535| 0.505| 0.691| 0.570     |

| σ   | Methods   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | Avg.      |
|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----------|
| 70  | BM3D      | 24.37| 22.93| 29.26| 23.03| 26.97| 24.43| 22.84| 27.34| 25.15     |
|     | NCSR      | 24.21| 22.84| 28.99| 22.86| 26.60| 21.26| 22.56| 27.13| 24.93     |
|     | GHP       | 24.35| 22.89| 28.93| 22.93| 26.75| 21.31| 22.65| 27.12| 24.99     |
|     | RWRM      | 24.36| 22.90| 29.07| 22.94| 27.10| 24.42| 22.55| 27.51| 25.11     |
|     | PSNR      | 0.464| 0.451| 0.788| 0.437| 0.588| 0.518| 0.481| 0.681| 0.551     |

| σ   | Methods   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | Avg.      |
|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----------|
| 90  | BM3D      | 23.75| 22.32| 28.12| 22.40| 26.30| 21.11| 22.47| 26.92| 24.77     |
|     | NCSR      | 23.65| 22.16| 28.10| 22.29| 25.88| 23.62| 21.82| 26.30| 24.23     |
|     | GHP       | 23.68| 22.15| 27.99| 22.29| 25.97| 23.65| 21.72| 26.21| 24.21     |
|     | RWRM      | 23.64| 22.21| 28.27| 22.40| 26.33| 21.84| 22.17| 26.87| 24.49     |
|     | PSNR      | 0.445| 0.428| 0.782| 0.416| 0.576| 0.504| 0.458| 0.672| 0.535     |

| σ   | Methods   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | Avg.      |
|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----------|
| 100 | BM3D      | 23.48| 22.07| 27.69| 22.27| 26.00| 21.55| 21.85| 26.16| 21.12     |
|     | NCSR      | 23.43| 21.89| 27.71| 22.05| 25.58| 23.35| 21.50| 25.92| 23.94     |
|     | GHP       | 23.35| 21.71| 27.47| 21.93| 25.56| 23.26| 21.22| 25.69| 23.77     |
|     | RWRM      | 23.64| 22.00| 28.00| 22.20| 26.32| 23.67| 21.60| 26.61| 24.26     |
|     | PSNR      | 0.421| 0.392| 0.766| 0.385| 0.560| 0.484| 0.419| 0.660| 0.511     |

Table 3. Denoising PSNR and SSIM results of RWRM and learning-based method LDCNN [48]

| σ   | Methods   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | Avg.      |
|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----------|
| 40  | RWRM      | 25.76| 24.38| 30.74| 24.30| 28.50| 25.71| 24.28| 28.88     |
|     | LDCNN     | 26.14| 24.83| 31.58| 24.81| 28.66| 26.05| 24.92| 29.24     |

| σ   | Methods   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | Avg.      |
|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----------|
| 50  | RWRM      | 25.06| 23.66| 30.11| 23.64| 27.92| 25.15| 24.51| 28.31     |
|     | LDCNN     | 25.47| 24.09| 30.90| 24.12| 28.07| 25.45| 24.17| 28.56     |
|     | PSNR      | 0.5522| 0.5520| 0.8178| 0.5415| 0.6330| 0.5994| 0.5894| 0.7132      |
Figures 9 and 10. The denoising output results (PSNR and SSIM) of BM3D [5], NCSR [6], GHP [49] and our RWRM on the test image 8. From the first to the third row, the Gaussian noise standard deviations of the corresponding noisy images are 60, 80, and 100, respectively.

Moreover, the images are respectively added to AWGN with $\sigma = 60, 70, 80, 90, 100$ during denoising.
From Fig. 11, we can find the fact that if the original parameter $\lambda$ (when $\beta = 0$) is inaccurate, the result may be much worse. However, adding the Wasserstein histogram constraint, the result is not sensitive to the change of parameter $\lambda$. Therefore, the proposed method is robust to the parameter choices, which makes our RWRM easier to be implemented based on existing image restoration methods. This also shows that in RWRM the residual Wasserstein regularization term is more important.

5. Application of RWRM in non-blind deconvolution. In this section, we first present the RWRM applied to non-blind deconvolution, then provide the solution process of the model, and finally give the experimental evaluation.

5.1. The proposed RWRM for non-blind deconvolution: Non-blind deconvolution tries to restore a latent clear image $x$ from its blurry and noisy observation $y = Kx + n$, where $K$ is a known blurring kernel, and $n$ is AWGN with variance $\sigma^2$. Unlike most image restoration models which only employ image priors, we combine residual regularization and image prior to present the following residual Wasserstein regularization model (RWRM) for non-blind deconvolution:

\[
\arg\min_{n,x} \frac{1}{2\sigma^2} \|n\|_2^2 + \lambda \|\nabla x\|_1 + \frac{\beta}{2} \hat{W}(h_n, h_g)
\]

subject to $y = Kx + n$,

where $h_n$ is the estimated residual histogram in the observed image, $h_g$ is the real Gaussian histogram (ground truth), and $\hat{W}(h_n, h_g)$ denotes the Wasserstein distance between $h_n$ and $h_g$. 
5.2. The solution to the RWRM for non-blind deconvolution. According to the analysis in Section 3.2, when solving the RWRM applied to non-blind deconvolution, we need to further explicitly solve the following optimization problems:

\[ \arg \min_x \frac{1}{2} \| Kx - z \|_2^2 + \frac{1}{\gamma} \| \nabla x \|_1 \]

where \( \gamma = \frac{1 + \beta \sigma^2}{\lambda \sigma^2} \), \( z = y + \beta \sigma^2 (y - \xi) \).

It is well known that solving the problem (23) is sometimes difficult, since it involves inverting the operator \( K \). By combining the principles of splitting and decomposition, we employ a proximal forward-backward splitting (PFBS) strategy [4] to deal with this problem. The PFBS algorithm [4] applied to the problem (23) can be described as

\[ x^{k+1} = \text{Prox}_{\delta J}(x^k - \delta K^T (Kx^k - z)) \]

where the positive number \( 0 < \delta < \frac{2}{\| K^T K \|} \), and \( J(x) = \frac{1}{\gamma} \| \nabla x \|_1 \). Here the proximal operator is defined as follows

\[ \text{Prox}_{\delta J}(v) = \arg \min_u (\delta J(u) + \frac{1}{2} \| u - v \|_2^2) \]

In the other words, by introducing an auxiliary variable \( v^k \), the minimization problem (23) can be solved by alternately iterating the following two-step proximal forward-backward splitting (PFBS) [4]:

\[ v^{k+1} = x^k - \delta K^T (Kx^k - z) \]

\[ x^{k+1} = \arg \min_x \frac{1}{2} \| x - v^{k+1} \|_2^2 + \frac{\delta}{\gamma} \| \nabla x \|_1 \]

We know that Eq. (27) can be also rewritten as follows:

\[ x^{k+1} = \arg \min_x \theta \frac{1}{2} \| x - v^{k+1} \|_2^2 + \| \nabla x \|_1 \]

where \( \theta = \frac{2}{\delta} \).

The problem (28) can be efficiently solved by Chambolle’s duality projection algorithm [2]. So the solution of (28) is obtained by

\[ x^{k+1} = v^{k+1} - \frac{\text{div} p}{\theta}, \]

where the vector \( p \) can be solved by fixed-point method: initializing \( p = 0 \) and iterating

\[ p = \frac{p + \tau \nabla (\text{div} p - \theta v^{k+1})}{1 + \tau | \nabla (\text{div} p - \theta v^{k+1}) |} \]

with \( \tau \leq 1/8 \) to ensure convergence.

Based on the analysis in Section 3.3 and above, the solution process of the proposed RWRM for non-blind deconvolution is summarized in Algorithm 2.

5.3. The experimental results of non-blind deconvolution. For non-blind deconvolution, we first compare the proposed RWRM with TV-L2, and then compare it with some representative methods.
Algorithm 2 Non-Blind Deconvolution by RWRM.

1. Input: Degraded image $y$, $n = \text{eye}(\text{size}(y))$, $j = 1$, $k = 1$, $x^1$, $\lambda$, $\beta$, $\delta$, $\tau$.
2. while $j \leq J$ do
3. Update $\xi$ by the histogram matching operation:
   $$\xi_i = \hat{\phi}_{n_i \rightarrow n}(n_i)$$
4. while $k \leq M$ do
5. Update $z$:
   $$z = \frac{y + \beta \sigma^2 (y - \xi)}{1 + \beta \sigma^2}$$
6. Update $v$ by the gradient descent:
   $$v^{k+1} = x^k - \delta K^T (K x^k - z)$$
7. Update $\theta$:
   $$\gamma = \frac{1 + \beta \sigma^2}{\lambda \sigma^2}, \theta = \frac{\gamma}{\delta}$$
8. Fixed-point iteration: initializing $p = 0$ and iterating
   $$p = \frac{p + \tau \nabla (\text{div} p - \theta z)}{1 + \tau |\nabla (\text{div} p - \theta z)|}$$
9. Implement Chambolle’s duality projection algorithm:
   $$x^{k+1} = v^{k+1} - \frac{\text{div} p}{\theta}$$
10. $k \leftarrow k + 1$
11. end while
12. Estimate residual:
   $$n = y - K x^k$$
13. $j \leftarrow j + 1$
14. end while
15. Output: Restored image $x^k$.

5.3.1. Comparison between the proposed RWRM and TV-L2. For the RWRM applied to non-blind deconvolution, we first use the ten images shown in Fig. 12 as test images to show the experimental results. Here, we mainly illustrate the role of the residual Wasserstein regularization by comparing the RWRM ($\beta \neq 0$) with the TV-L2 ($\beta = 0$). Likewise, we first discuss the experimental parameter settings, then analyze the experimental data, and finally give the subjective evaluation.

1) Parameter settings: In our non-blind deconvolution, the experimental blurry images are synthesized by first utilizing a commonly used $5 \times 5$ Gaussian blur kernel with a standard deviation 1.5 and then adding AWGN with $\sigma = 25$.

When $\beta = 0$, there is only one model parameter $\lambda$. Through some experiments, we find that in general the larger the value of $\lambda$, the higher the PSNR of the restored images will be. However, the value of $\lambda$ that is too large will result in a tendency
Figure 12. Ten test images for non-blind deconvolution. From left to right, they are Plane, Goldhill, Couple, Peppers, Martha, Boat, Girl, Lena, Bacteria and Brain respectively.

to decrease the PSNR. Here we apply a strategy similar to the one in Gaussian denoising to select $\lambda$. For different images, we choose the optimal parameters $\lambda$, which are the integer multiples of 100 ranging from 1700 to 2000. In Fig. 13, taking test image “Bacteria” as an example, we have shown the curve of the PSNR with respect to the parameter $\lambda$.

Figure 13. The graph of the PSNR with $\lambda$ in the RWRM ($\beta = 0$) applied to non-blind deconvolution.

When $\beta \neq 0$, the proposed model has two parameters $\lambda$ and $\beta$. For the scale parameter $\lambda$, we still choose the optimal value when $\beta = 0$. In our approach, we utilize Wasserstein regularization parameter $\beta = \frac{\lambda}{0.45\sigma}$, where $\sigma$ is noise standard deviation. We still employ the image “Bacteria” to illustrate the choice of parameter $\beta$. As shown in Fig. 14, with the increase of the parameter $\beta$ ($\lambda = 1800$), the PSNR of the restored image increases rapidly, and then decreases slowly. It can be found from Fig. 14 that the optimal parameter $\beta = 160$.

In solving the RWRM, the step size $\delta$ that is too large or too small in the gradient descent will increase the iteration number. The step size in the vicinity of 1 is appropriate. Therefore, we set the step size $\delta$ to 1.1. Initial input image $x^1$ is obtained by Tikhonov regularization [42] on degraded image $y$. In addition, for the initialization noise in Algorithm 2, we select the identity matrix as large as the observation image.

In the implementation of the Algorithm 2, the number of outer cycles $J = 5$, and the number of inner cycles $M = 90$. In the experiment, we set $\lambda = 1800$. According to the formula $\beta = \frac{\lambda}{0.45\sigma}$, we adaptively select the parameter $\beta$. In fixed point iteration, $\tau = 1/8$ is chosen to ensure its convergence.

2) Analysis of experimental data
RWRM: Residual Wasserstein regularization model for image restoration

For the TV-L2 ($\beta = 0$) and our RWRM ($\beta \neq 0$), the quantitative experimental results are shown in Table 4. One can see that the RWRM has a higher PSNR on each of the experimental images. It is gratifying that the average PSNR of the ten images in RWRM is 0.65 dB higher than the TV-L2. In addition, in Fig. 15, compared with the TV-L2, the PSNR increase graph of the RWRM on ten images is plotted. It can be concluded that it is effective to introduce the residual Wasserstein regularization on the basis of the TV-L2.

3) Subjective evaluation

In order to reflect the deblurring visual effects, in Fig. 16 we show the result images of the TV-L2 and the RWRM. It can be found that compared with the TV-L2 effect, our RWRM can produce more visually appealing restoration results. And furthermore, we take the local details and zoom in to show in Fig. 17. One can see that the RWRM can well enhance the image detail features such as edges, which are often over-smoothed by other deconvolution methods.
Figure 16. The results of non-blind deconvolution on the images Couple, Martha and Bacteria. Here (a) is the blurred and noisy images, (b) and (c) are the result images of the TV-L2 and the RWRM, respectively. One can see that the RWRM has obvious advantages in terms of both visual and numerical aspects.

Figure 17. Local enlarged images on non-blind deconvolution. One can see that the RWRM can recover better edge information.
5.3.2. **Comparison with some representative methods.** We compare RWRM with other two representative methods NL-H\(^1\) [21] and NL-TV [21] on another eight test images shown in Fig. 18. In the experiments, the blurry images are synthesized by first utilizing a commonly used 3 × 3 Gaussian blur kernel with a standard deviation 2 and then adding AWGN with \(\sigma = 15\). The experimental results are shown in Table 5. It can be found that from the perspective of PSNR and SSIM, our RWRM has advantages over competing methods. Fig. 19 shows a visual comparison of the three approaches. One can see that RWRM can retain better details such as edges, which implies the effectiveness of the proposed method.

![Figure 18. Another eight test images for the non-blind deconvolution. From left to right, they are labeled as 1 to 8 respectively.](image)

| Images | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | Avg. |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|------|
| NL-H\(^1\) | 25.64 | 26.17 | 26.48 | 27.11 | 26.03 | 26.36 | 26.47 | 25.30 | 26.70 |
|        | 0.7800 | 0.8218 | 0.7982 | 0.7690 | 0.8481 | 0.8044 | **0.8785** | 0.7999 | 0.8114 |
| NL-TV  | 25.83 | 29.91 | 26.42 | 26.96 | 26.74 | 26.44 | 27.40 | 25.50 | 26.90 |
|        | 0.7752 | 0.8209 | 0.7944 | 0.7649 | 0.8504 | 0.8048 | 0.8631 | 0.7972 | 0.8089 |
| RWRM   | **26.24** | **30.95** | **26.83** | **27.88** | **27.98** | **26.90** | **28.24** | **26.32** | **27.67** |
|        | 0.7993 | 0.8358 | 0.8096 | 0.7800 | 0.8707 | 0.8148 | 0.8599 | **0.8217** | **0.8240** |

![Figure 19. Visual effect comparison of non-blind deconvolution by three methods NL-H\(^1\), NL-TV and the proposed RWRM.](image)

(a) Blurred (b) NL-H\(^1\)(30.17/0.8218) (c) NL-TV(29.91/0.8209) (d) RWRM(30.95/0.8358)
6. Conclusion and future work. In this paper, we have introduced a residual Wasserstein regularization model (RWRM) which can integrate the residual histogram constraint with some image priors for image restoration. Concretely, with the Wasserstein distance, we minimize the discrepancy between the estimated residual histogram and the reference residual histogram. The witty cooperation of the residual Wasserstein regularization and total variation improves the output image quality in Gaussian denoising and non-blind deconvolution.

Our RWRM is different from other histogram-constrained image restoration methods in the following two aspects. Firstly, one of the advantages of RWRM is its simplicity, without complex reference histogram estimation in pixel or gradient domain. Secondly, the robustness of model parameter selection and the fast implementation of the corresponding algorithms make the RWRM easier to perform image restoration.

Based on the effectiveness of the RWRM, the basic idea of our future work is to combine the residual Wasserstein regularization with other image priors for image restoration. One strategy is to utilize residual Wasserstein regularization and low dimensional manifold [36] for other image restoration applications, such as blind inpainting and super-resolution.

Acknowledgments. The authors would like to thank the anonymous reviewers for their valuable comments. This paper is supported by the National Natural Science Foundation of China (Grant Nos. 61772389, 61941111), Scientific Research Project of Xinzhou Normal University 2019KY07, and the Key Construction Disciplines Project of Xinzhou Normal University XK201501.

REFERENCES

[1] E. J. Candes and T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Transactions on Information Theory, 52 (2006), 5406–5425.
[2] A. Chambolle, An algorithm for total variation minimization and applications, Journal of Mathematical Imaging and Vision, 20 (2004), 89–97.
[3] T. S. Cho, C. L. Zitnick, N. Joshi, S. B. Kang, R. Szeliski and W. T. Freeman, Image restoration by matching gradient distributions, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (2012), 683–694.
[4] P. L. Combettes and V. Wajs, Signal recovery by proximal forward-backward splitting, Multiscale Modeling and Simulation, 4 (2005), 1168–1200.
[5] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, IEEE Transactions on Image Processing, 16 (2007), 2080–2095.
[6] W. Dong, L. Zhang, G. Shi and X. Li, Nonlocally centralized sparse representation for image restoration, IEEE Transactions on Image Processing, 22 (2013), 1620–1630.
[7] D. L. Donoho, Compressed sensing, IEEE Transactions on Information Theory, 52 (2006), 1289–1306.
[8] M. Elad and M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, IEEE Transactions on Image Processing, 15 (2006), 3736–3745.
[9] M. El Gheche, J.-F. Aujol, Y. Berthoumieu and C.-A. Deledalle, Texture reconstruction guided by the histogram of a high-resolution patch, IEEE Trans. Image Process, 26 (2017), 549–560.
[10] W. Feller, An Introduction to Probability Theory and Its Applications II, John Wiley & Sons, 1968.
[11] A. L. Gibbs, Convergence in the wasserstein metric for markov chain monte carlo algorithms with applications to image restoration, Stochastic Models, 20 (2004), 473–492.
[12] G. Gilboa and S. Osher, Nonlocal operators with applications to image processing, Multiscale Modeling and Simulation, 7 (2008), 1005–1028.
[13] R. C. Gonzalez, R. E Woods and S. L Eddins, Digital Image Processing Using MATLAB, Prentice Hall Press, 2007.
[14] S. Harmeling, C. J. Schuler and H. C. Burger, Image denoising: Can plain neural networks compete with bm3d? In IEEE Conference on Computer Vision and Pattern Recognition, (2012), 2392–2399.

[15] R. He, X. Feng, W. Wang, X. Zhu and Chunyu Yang, W-lldmm: A wasserstein driven low-dimensional manifold model for noisy image restoration, Neurocomputing, 371 (2020), 108–123.

[16] D. J. Heeger and J. R. Bergen, Pyramid-based texture analysis/synthesis, In International Conference on Image Processing, 1995. Proceedings, (1995), 229–238.

[17] V. Jain and H. Sebastian Seung, Natural image denoising with convolutional networks, In International Conference on Neural Information Processing Systems, (2008), 769–776.

[18] D. Krishnan and R. Fergus, Fast image deconvolution using hyper-laplacian priors, In International Conference on Neural Information Processing Systems, (2009), 1033–1041.

[19] X. Lan, S. Roth, D. Huttenlocher and M. J Black, Efficient belief propagation with learned higher-order markov random fields, In European Conference on Computer Vision, pages 269–282. Springer, 2006.

[20] S. Z. Li, Markov Random Field Modeling in Image Analysis, Springer-Verlag London, Ltd., London, 2009.

[21] Y. Lou, X. Zhang, S. Osher and A. Bertozzi, Image recovery via nonlocal operators, Journal of Scientific Computing, 42 (2010), 185–197.

[22] J. Mairal, M. Elad and G. Sapiro, Sparse representation for color image restoration, IEEE Transactions on Image Processing, 17 (2008), 53–69.

[23] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, Inc., San Diego, CA, 1998.

[24] X. Mei, W. Dong, B. G. Hu and S. Lyu, Unihist: A unified framework for image restoration with marginal histogram constraints, In Computer Vision and Pattern Recognition, pages 3753–3761, 2015.

[25] S. Osher, M. Burger, D. Goldfarb, J. Xu and W. Yin, An iterative regularization method for total variation-based image restoration, Multiscale Modeling and Simulation, 4 (2005), 460–489.

[26] O. Pele and M. Werman, Fast and robust earth mover’s distances, In IEEE International Conference on Computer Vision, (2010), 460–467.

[27] G. Peyré, J. Fadili and J. Rabin, Wasserstein active contours, In IEEE International Conference on Image Processing, (2013), 2541–2544.

[28] J. Portilla, V. Strela, M. J. Wainwright and E. P. Simoncelli, Image denoising using scale mixtures of gaussians in the wavelet domain, IEEE Transactions on Image Processing, 12 (2003), 1338–1351.

[29] J. Rabin and G. Peyré, Wasserstein regularization of imaging problem, In IEEE International Conference on Image Processing, (2011), 1541–1544.

[30] A. Rajwade, A. Rangarajan and A. Banerjee, Image denoising using the higher order singular value decomposition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35 (2012), 849–862.

[31] W. H. Richardson, Bayesian-based iterative method of image restoration, Journal of the Optical Society of America, 62 (1972), 55–59.

[32] Y. Romano, M. Protter and M. Elad, Single image interpolation via adaptive nonlocal sparsity-based modeling, IEEE Transactions on Image Processing, 23 (2014), 3085–3098.

[33] L. I. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, In Eleventh International Conference of the Center for Nonlinear Studies on Experimental Mathematics: Computational Issues in Nonlinear Science. Phys. D, 60 (1992), 259–268.

[34] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier and F. Lenzen, Variational Methods in Imaging, Springer, 2009.

[35] U. Schmidt, Q. Gao and S. Roth, A generative perspective on mrfs in low-level vision, In Computer Vision and Pattern Recognition, 2010, pages 1751–1758.

[36] O. Stanley, Z. Shi and W. Zhu, Low dimensional manifold model for image processing, SIAM Journal on Imaging Sciences, 10 (2017), 1669–1690.

[37] D. Strong and T. Chan, Edge-preserving and scale-dependent properties of total variation regularization, Inverse Problems, 19 (2003), 165–187.

[38] K. Suzuki, I. Horiba and N. Sugie, Efficient approximation of neural filters for removing quantum noise from images, IEEE Transactions on Signal Processing, 50 (2002), 1787–1799.

[39] P. Swoboda and C. Schnorr, Convex variational image restoration with histogram priors, SIAM Journal on Imaging Sciences, 6 (2013), 1719–1735.
[40] G. Tartavel, G. Peyré and Y. Gousseau, Wasserstein loss for image synthesis and restoration, *SIAM Journal on Imaging Sciences*, 9 (2016), 1726–1755.

[41] F. Thaler, K. Hammmernik, C. Payer, M. Urschler and D. Stern, Sparse-view ct reconstruction using wasserstein gans, 2018, pages 75–82.

[42] M. Vauhkonen, D. Vadasz, P. A. Karjalainen, E. Somersalo and J. P. Kaipio, Tikhonov regularization and prior information in electrical impedance tomography, *IEEE Transactions on Medical Imaging*, 17 (1998), 285–93.

[43] C. Villani, *Optimal Transport: Old and New*, volume 338, Springer-Verlag, Berlin, 2009

[44] Y. Weiss and W. T. Freeman, What makes a good model of natural images? In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, (2007) pages 1–8.

[45] O. J. Woodford, C. Rother and V. Kolmogorov, A global perspective on map inference for low-level vision, In *IEEE International Conference on Computer Vision*, 2009, pages 2319–2326.

[46] F. Wu, B. Wang, D. Cui and L. Li, Single image super-resolution based on wasserstein gans, *Chinese Control Conference (CCC)*, 2018.

[47] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun and G. Wang, Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss, *IEEE Transactions on Medical Imaging*, 37 (2018), 1348–1357.

[48] K. Zhang, W. Zuo, S. Gu and L. Zhang, Learning deep cnn denoiser prior for image restoration, 2017, pages 2808–2817.

[49] W. Zuo, L. Zhang, C. Song, D. Zhang and H. Gao, Gradient histogram estimation and preservation for texture enhanced image denoising, *IEEE Transactions on Image Processing*, 23 (2014), 2459–2472.

Received October 2019; revised August 2020.

E-mail address: ruiqianghe@sina.com (R. He)
E-mail address: xcfeng@mail.xidian.edu.cn (X. Feng)
E-mail address: zxl001@aliyun.com (X. Zhu)
E-mail address: hhuang@cqwu.edu.cn (H. Huang)
E-mail address: bingzhewei@126.com (B. Wei)