



A new weighted Caputo fractional-order total variation for Cauchy noise removal with Bayesian optimization

Hamdi Braiek^{1,2}

Received: 16 November 2025 / Accepted: 3 May 2026

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2026

Abstract

In this paper, we propose a novel variational framework for image denoising corrupted by impulsive heavy-tailed Cauchy noise. The proposed model integrates a weighted fractional-order total variation (FOTV) regularization with an adaptive spatial weighting strategy and a nonconvex Cauchy data-fidelity term. The fractional regularizer, formulated in the Caputo sense, introduces a tunable smoothness control through the fractional order. It enables a flexible balance between local edge preservation and global texture smoothing, while the adaptive weight modulates the regularization strength according to local image structures. To efficiently minimize the resulting nonconvex and nonsmooth energy, we derive a split Bregman iterative algorithm coupled with a half-quadratic reformulation of the Cauchy fidelity, leading to tractable weighted least-squares subproblems solved via the conjugate gradient method. Furthermore, a Bayesian optimization framework is incorporated to automatically tune the key model parameters based on the PSNR metric. This data-driven calibration improves convergence robustness and ensures optimal trade-offs between noise suppression and detail preservation without manual intervention. We also establish the existence of minimizers of the model in the weighted fractional bounded-variation space. Numerical experiments on synthetic and benchmark images show that the proposed model achieves superior robustness to outliers and better preservation of fine structures compared to classical TV, FOTV, and nonlocal variational models. Quantitative results in terms of PSNR, SSIM, SNR, and relative error confirm the effectiveness, adaptivity, and stability of the proposed approach.

Keywords Caputo fractional derivative · Weighted fractional-order TV · Split Bregman · Bayesian optimization · Cauchy noise · Image denoising

1 Introduction

Image denoising remains a core problem in imaging science. It aims to recover a clean image from an observed, corrupted one. Classical methods assume additive Gaussian noise. Total variation (TV) regularization is a standard tool for that case. TV preserves edges while removing small-scale noise. However, TV often yields piecewise-constant solutions and can remove fine texture [1–3]. Classical variational approaches

to image denoising have long been framed within a Bayesian maximum a posteriori (MAP) perspective, whereby a noise model and a prior on the clean image lead to an energy minimization formulation. Early work in this direction, such as the variational MAP estimation framework of Hamza and Krim [4], established a principled connection between probabilistic image priors and variational energy minimization for denoising. In this framework, a Markov random field prior is combined with an appropriate noise likelihood to derive an energy functional whose minimizer corresponds to the MAP estimate of the clean image. Subsequent developments include formulations that unify probabilistic and variational estimators [5].

In many practical imaging systems, the noise deviates significantly from the Gaussian assumption. Heavy-tailed distributions such as the Cauchy noise better describe impulsive or outlier-contaminated data. Such noise occurs in sonar, radar, ultrasound, and other medical imaging modalities [6, 7]. Since Cauchy noise has undefined variance and long

The author has legally changed their name from Hamdi Houichet to Hamdi Braiek.

✉ Hamdi Braiek
hamdi.houichet@gmail.com

¹ ESPRIT School of Engineering, 18 rue de l'usine, 2035 Ariana, Tunisia

² Laboratory for Mathematical and Numerical Modeling in Engineering Science, National Engineering School of Tunis, B.P. 37, Belvédère 1002, Tunis, Tunisia

tails, Gaussian-based fidelity terms fail to yield reliable reconstructions. Robust data-fidelity terms derived from the Cauchy or other nonconvex likelihoods are therefore essential for effective denoising in such settings.

Fractional-order total variation (FOTV) extends the classical TV by replacing integer-order derivatives with fractional ones [8, 9]. The fractional order introduces a tunable degree of smoothness control, offering a flexible balance between edge preservation and texture smoothing. This helps model medium-range dependencies and reduces the staircasing artifacts typical of first-order TV.

The nonlocal behavior induced by fractional-order derivatives is closely related to that of explicit nonlocal variational models, such as nonlocal total variation (NLTV), which exploit patch-based or graph-based similarities to preserve repetitive textures and long-range dependencies [10, 11]. While NLTV-type approaches have demonstrated strong performance in texture preservation, their effectiveness strongly depends on the robustness of similarity measures, which can be severely degraded in the presence of heavy-tailed impulsive noise such as Cauchy noise. In addition, the construction and manipulation of nonlocal operators entail high computational and memory costs [12, 13]. In contrast, fractional-order regularization provides an implicit and computationally efficient way to capture nonlocal interactions without relying on explicit patch similarity computations. Recent studies demonstrate that fractional-order models enhance both visual quality and quantitative performance in image restoration tasks [14–16]. FOTV has also been successfully applied in super-resolution reconstruction [16], image registration [17], and texture or hyperspectral image classification, where its nonlocal nature improves the discriminative representation of spatial structures.

Spatially adaptive weighting further improves the flexibility and performance of fractional-order models. Weighted FOTV allows the regularization strength to vary across the image, enabling strong smoothing in homogeneous regions while preserving fine structures and sharp transitions in textured or high-contrast areas. The weights may be constructed using local gradients, structure tensors, or related features, leading to enhanced edge-aware regularization. Recent studies on weighted or hybrid fractional-order TV models report superior visual fidelity and quantitative performance compared to their unweighted counterparts [18, 19]. Building upon these advances, the weighted Caputo fractional-order total variation model proposed in this work remains local in formulation while implicitly capturing nonlocal interactions through fractional derivatives, providing a computationally efficient framework for robust Cauchy noise removal.

Solving variational models that combine nonconvex data-fidelity terms with fractional or nonlocal regularizers is computationally demanding. The split Bregman (SB) framework [2, 20, 21] offers an efficient solution strategy by

decoupling nonsmooth and nonlocal terms into simpler subproblems, each admitting closed-form or fast solvers. It has proven highly effective for L1-type and fractional-order regularizations.

Despite these advances, model performance still depends heavily on the selection of hyperparameters, such as the regularization weight, fractional order, and penalty terms in the SB framework. Manual tuning is time-consuming and suboptimal. To address this, we incorporate Bayesian Optimization (BO) [22, 23] for automated hyperparameter selection. BO provides a principled probabilistic framework to optimize the parameters of complex, nonconvex objectives with limited evaluations. This strategy ensures a consistent balance between fidelity and regularization terms across different images and noise levels.

In this paper, we propose a novel weighted FOTV model with a Cauchy-based robust data-fidelity term, optimized through a Bayesian optimization-driven parameter tuning scheme. The proposed method is implemented using a SB solver combined with a half-quadratic reformulation of the Cauchy fidelity, resulting in a tractable sequence of weighted least-squares problems solved efficiently via the conjugate gradient method.

The main contributions of this research work are the following

- (i) We propose a new variational formulation that integrates a robust Cauchy data-fidelity term with a spatially adaptive weighted FOTV regularizer formulated in the Caputo sense.
- (ii) We establish the existence of minimizers for model in the weighted fractional bounded-variation space.
- (iii) We derive an efficient SB algorithm combined with a half-quadratic reformulation to handle the nonconvex and nonsmooth nature of the model.
- (iv) We introduce a Bayesian optimization framework (BO) for automatic tuning of hyperparameters.
- (v) We conduct comprehensive experiments on synthetic, natural, and medical images, including grayscale and color cases.

In particular, the proposed method combines a spatially adaptive weighting strategy with a BO-based parameter tuning scheme. While the weighting mechanism adapts the data-fidelity term locally according to image structures in order to preserve edges and fine details while suppressing outliers, BO globally optimizes the model parameters to best exploit this adaptivity. This joint design significantly enhances denoising performance and robustness.

The remainder of the paper is organized as follows. Section 2 introduces Cauchy noise statistics, fractional derivatives, and formulates the proposed weighted FOTV-Cauchy model. Section 3 presents the SB solver and half-quadratic

reformulation. Section 4 describes the discretization strategy and the Bayesian optimization framework for hyperparameter tuning. Section 5 reports numerical experiments and comparisons for grayscale and color images. Finally, Sect. 6 concludes the paper and outlines possible extensions.

2 Weighted fractional-order total variation model

2.1 Cauchy fidelity term

Let $\Omega \subset \mathbb{R}^2$ be the image domain. Let $f : \Omega \rightarrow \mathbb{R}$ be the observed noisy image. We assume an additive Cauchy noise model:

$$f(x) = u(x) + \eta(x), \text{ with } \eta(x) \sim \text{Cauchy}(0, \gamma),$$

where $\gamma > 0$ is the scale parameter. The objective is to find a restored image $u : \Omega \rightarrow \mathbb{R}$ from f . The Cauchy distribution has heavy tails. From the Cauchy log-likelihood the data-fidelity term reads [6, 7, 24]:

$$\mathcal{D}(u; f) = \int_{\Omega} \log \left(1 + \frac{(f(x) - u(x))^2}{\gamma^2} \right) dx.$$

This loss is nonquadratic and robust to large outliers. It is widely used for Cauchy noise restoration.

Proposition 1 *Let $\gamma > 0$ and define the Cauchy log-loss $\phi_{\gamma}(r) = \log(1 + \frac{r^2}{\gamma^2})$. Then, the following statements hold:*

- (i) $\phi_{\gamma} \in C^2(\mathbb{R})$ with $\phi'_{\gamma}(r) = \frac{2r}{\gamma^2 + r^2}$, and $\phi''_{\gamma}(r) = \frac{2(\gamma^2 - r^2)}{(\gamma^2 + r^2)^2}$.
- (ii) ϕ_{γ} is convex for $|r| \leq \gamma$ and concave for $|r| > \gamma$.
- (iii) As $|r| \rightarrow \infty$, $\phi_{\gamma}(r) \sim 2 \log |r| - 2 \log \gamma$.
- (iv) $|\phi''_{\gamma}(r)| \leq 2/\gamma^2$.

All statements follow from direct differentiation and standard calculus inequalities. We omit the proof.

2.2 Caputo fractional derivative

We work with fractional order $\alpha \in (1, 2)$. We first give the definition of the directional Caputo derivative and Caputo gradient in 2D.

Definition 1 [25] Let $\Omega = (a_x, b_x) \times (a_y, b_y)$. For $u : \Omega \rightarrow \mathbb{R}$ and $\alpha \in (1, 2)$, define directional Caputo derivatives

$${}^C D_x^{\alpha} u(x, y) := \frac{1}{\Gamma(2-\alpha)} \int_{a_x}^x (x-t)^{1-\alpha} \partial_t^2 u(t, y) dt,$$

$${}^C D_y^{\alpha} u(x, y) := \frac{1}{\Gamma(2-\alpha)} \int_{a_y}^y (y-s)^{1-\alpha} \partial_s^2 u(x, s) ds.$$

The Caputo fractional gradient is the vector

$$\nabla^{\alpha, C} u := ({}^C D_x^{\alpha} u, {}^C D_y^{\alpha} u)^T.$$

2.3 Weighted fractional total variation regularization

In this subsection, we introduce the definition and main properties of the weighted fractional-order total variation (FOTV_w^{α}), which generalizes both the classical TV [3] and the unweighted fractional total variation (FOTV^{α}) [8]. This operator combines spatial adaptivity through a weight function and nonlocal behavior through the fractional-order differentiation.

Definition 2 Let $w : \Omega \rightarrow \mathbb{R}_+$ be a bounded positive weight and $0 < l = n - 1 < \alpha < n$. The weighted fractional total variation of $u \in L^1(\Omega)$ is defined as

$$\text{FOTV}_w^{\alpha}(u) = \sup_{v \in \mathcal{V}_w^{\alpha}} \int_{\Omega} (-u \operatorname{div}^{\alpha} v(x)) dx, \quad (1)$$

where for all $x \in \Omega$

$$\mathcal{V}_w^{\alpha} = \left\{ \mathbf{v} = (v_1, v_2) \in C_0^l(\Omega, \mathbb{R}^2) \mid \|\mathbf{v}(x)\|_{L^{\infty}(\Omega, \mathbb{R}^2)} \leq w(x) \right\}.$$

and $\operatorname{div}^{\alpha} \mathbf{v} = \sum_{i=1}^2 \frac{\partial^{\alpha} v_i}{\partial x_i^{\alpha}}$ denotes the fractional divergence operator of order α . Here, $\frac{\partial^{\alpha} v_i}{\partial x_i^{\alpha}}$ represents a fractional derivative along the x_i direction in the Caputo sense.

When $\alpha = 1$, FOTV_w^{α} reduces to the classical weighted total variation. When $w(x) \equiv 1$, the definition coincides with the FOTV^{α} [8, 26]. We define the space of functions of weighted α -bounded variation as

$$BV_w^{\alpha}(\Omega) = \left\{ u \in L^1(\Omega) : \text{FOTV}_w^{\alpha}(u) < \infty \right\},$$

equipped with the norm

$$\|u\|_{BV_w^{\alpha}(\Omega)} = \|u\|_{L^1(\Omega)} + \text{FOTV}_w^{\alpha}(u).$$

Lemma 1 *The space $BV_w^\alpha(\Omega)$, endowed with the norm $\|\cdot\|_{BV_w^\alpha(\Omega)}$, is a Banach space.*

Proof The proof follows directly from standard arguments of completeness in BV^α spaces [8] combined with the boundedness of the weight $w(x)$, $0 < w_0 \leq w(x) \leq w_1 < \infty$. $\square$

We can also define the weighted fractional Banach-type space [8, 26]

$$W_{1,w}^\alpha(\Omega) = \left\{ u \in L^1(\Omega) \mid \|u\|_{W_{1,w}^\alpha(\Omega)} < \infty \right\},$$

where the weighted fractional norm is

$$\|u\|_{W_{1,w}^\alpha(\Omega)} = \int_{\Omega} |u(x)| dx + \int_{\Omega} w(x) |\nabla^{\alpha,C} u(x)| dx,$$

Proposition 2 *Assume $u \in W_{1,w}^\alpha(\Omega)$ and that the directional Caputo derivatives ${}^C D_x^\alpha u$ and ${}^C D_y^\alpha u$ exist in $L^1(\Omega)$. Then,*

$$\text{FOTV}_w^\alpha(u) = \int_{\Omega} w(x) |\nabla^{\alpha,C} u(x)| dx.$$

Similarly as [8], we can show:

Proposition 3 *The functional $\text{FOTV}_w^\alpha : L^1(\Omega) \rightarrow [0, +\infty)$ satisfies:*

1. FOTV_w^α is convex and positively homogeneous;
2. FOTV_w^α is lower semi-continuous with respect to $L^1(\Omega)$ convergence;
3. If w is bounded below by a positive constant, FOTV_w^α is coercive in $BV^\alpha(\Omega)$.

The weighted fractional total variation inherits the edge-preserving property of classical TV and reduces the staircasing effect due to the fractional-order operator. The weight function $w(x)$ provides local adaptivity: Regions with small w values (typically edges) experience less smoothing, while homogeneous areas (large w) are more regularized.

2.4 Proposed model and its well-posedness

In this subsection, we introduce an adaptive variational model for image restoration under Cauchy noise. The model combines a robust nonconvex fidelity term derived from the Cauchy log-likelihood with a weighted FOTV regularizer that captures both nonlocal smoothness and local adaptivity. We then consider the minimization problem

$$\min_{u \in BV_w^\alpha(\Omega)} \left\{ \mathcal{J}(u) = \lambda \mathcal{D}(u; f) + \text{FOTV}_w^\alpha(u) \right\}, \quad (2)$$

where $1 \leq \alpha \leq 2$ denotes the fractional order, $w : \Omega \rightarrow \mathbb{R}_+$ is a bounded positive weight, and $\lambda > 0$ is a regularization parameter.

The spatial weight $w(x)$ is designed to adapt the regularization strength to local image structures. It is defined as

$$w(x) = \exp(-|\nabla(G_\sigma * u)(x)|), \quad (3)$$

where G_σ denotes a Gaussian smoothing kernel with standard deviation $\sigma > 0$. This choice assigns smaller weights near edges and larger weights in homogeneous regions, thereby preserving sharp transitions while enforcing stronger smoothing in flat areas. The Gaussian smoothing improves robustness to noise and avoids spurious oscillations in the weight map.

Let the domain Ω be discretized on a uniform grid of $N \times M$ pixels with spacing h . Denote $\mathbf{U}_{i,j}$ and $\mathbf{F}_{i,j}$ the discrete image values and the discrete weight matrix $\mathbf{W} = (\mathbf{W}_{i,j})$ with $\mathbf{W}_{i,j} > 0$ the weights. In vectorized form, we stack the image column-wise into $\mathbf{u} \in \mathbb{R}^{NM}$. We consider a discrete approximation of the Caputo derivative for $1 < \alpha < 2$ by discretizing the integral in the Caputo definition and approximating u'' by second-order finite differences [27]. The resulting one-dimensional backward approximation in the x -direction and y -direction reads

$$({}^{D_x^{\alpha,C}} \mathbf{U})_{i,j} = \frac{h^{-\alpha}}{\Gamma(2-\alpha)} \sum_{m=0}^{i-1} a_m^{(\alpha)} \Delta_x^2 \mathbf{U}_{i-m,j}, \quad (4)$$

$$({}^{D_y^{\alpha,C}} \mathbf{U})_{i,j} = \frac{h^{-\alpha}}{\Gamma(2-\alpha)} \sum_{n=0}^{j-1} a_n^{(\alpha)} \Delta_y^2 \mathbf{U}_{i,j-n}, \quad (5)$$

where the centered second finite difference along the x - and y -axis is, respectively, given by

$$\begin{aligned} \Delta_x^2 \mathbf{U}_{k,j} &= \mathbf{U}_{k+1,j} - 2\mathbf{U}_{k,j} + \mathbf{U}_{k-1,j} \\ \Delta_y^2 \mathbf{U}_{i,k} &= \mathbf{U}_{i,k+1} - 2\mathbf{U}_{i,k} + \mathbf{U}_{i,k-1}, \end{aligned}$$

and the corresponding quadrature weights in the x - and y -directions are

$$a_\ell^{(\alpha)} = (\ell + 1)^{1-\alpha} - \ell^{1-\alpha}, \quad \text{with } \ell = m, n.$$

We then define the discrete Caputo fractional gradient magnitude at pixel (i, j) by

$$G_{i,j}(\mathbf{U}) = \sqrt{({}^{D_x^{\alpha,C}} \mathbf{U}_{i,j})^2 + ({}^{D_y^{\alpha,C}} \mathbf{U}_{i,j})^2}.$$

Therefore, using the discrete Caputo operators above the discrete energy reads

$$\mathcal{J}(\mathbf{U}) = \sum_{i=1}^N \sum_{j=1}^M \log\left(1 + \frac{(\mathbf{F}_{i,j} - \mathbf{U}_{i,j})^2}{\gamma^2}\right) + \lambda \sum_{i=1}^N \sum_{j=1}^M \mathbf{w}_{i,j} G_{i,j}(\mathbf{U}). \quad (6)$$

Theorem 1 *Let the discrete weight map satisfy $0 < w_{\min} \leq \mathbf{W}_{i,j} \leq w_{\max} < \infty$ for all i, j , and fix $\lambda > 0$, $\gamma > 0$. For a given observed image $\mathbf{F} \in \mathbb{R}^{N \times M}$, the discrete energy (6) $\mathcal{J} : \mathbb{R}^{N \times M} \rightarrow \mathbb{R}$ attains at least one global minimizer $\mathbf{U}^* \in \mathbb{R}^{N \times M}$.*

Proof It is clear that $\mathcal{J}(\mathbf{U})$ is continuous on $\mathbb{R}^{N \times M}$. Now, we show that $\mathcal{J}(\mathbf{U})$ is coercive. Assume, for contradiction, that $\mathcal{J}$ is not coercive. Then, there exists a sequence $\{\mathbf{U}^k\} \subset \mathbb{R}^{N \times M}$ such that $\|\mathbf{U}^k\|_2 \rightarrow \infty$ but $\mathcal{J}(\mathbf{U}^k) \leq C$ for some fixed $C > 0$ and all k . Write each $\mathbf{U}^k$ as $\mathbf{U}^k = s^k \mathbf{V}^k$, with $s^k := \|\mathbf{U}^k\|_2 \geq 0$, and $\|\mathbf{V}^k\|_2 = 1$. Then, we have $s_k \rightarrow \infty$.

Because $\mathbf{V}^k \in \mathbb{R}^{N \times M}$ has unit ℓ^2 norm, at least one component of $\mathbf{V}^k$ has magnitude at least $1/\sqrt{NM}$. Concretely, there exists an index $(\bar{i}, \bar{j})$ with $|\mathbf{V}_{\bar{i},\bar{j}}^k| \geq \frac{1}{\sqrt{NM}}$. For that pixel, we estimate the fidelity contribution of $\mathbf{U}^k$ to get

$$\begin{aligned} \phi_{\bar{i},\bar{j}}(\mathbf{U}_{\bar{i},\bar{j}}^k) &= \log\left(1 + \frac{(\mathbf{F}_{\bar{i},\bar{j}} - s^k \mathbf{V}_{\bar{i},\bar{j}}^k)^2}{\gamma^2}\right) \\ &\geq \log\left(1 + \frac{(|s^k| |\mathbf{V}_{\bar{i},\bar{j}}^k| - |\mathbf{F}_{\bar{i},\bar{j}}|)^2}{\gamma^2}\right) \\ &\geq \log\left(1 + \frac{(s^k/\sqrt{NM} - |\mathbf{F}_{\bar{i},\bar{j}}|)^2}{\gamma^2}\right). \end{aligned}$$

As $s^k \rightarrow \infty$, the lower bound above behaves like $2 \log s^k + O(1)$. Hence, $\phi_{\bar{i},\bar{j}}(\mathbf{U}_{\bar{i},\bar{j}}^k) \rightarrow +\infty$ as $k \rightarrow \infty$. Because all terms in $\mathcal{J}$ are nonnegative, the divergence of a single pixel fidelity term forces $\mathcal{J}(\mathbf{U}^k) \rightarrow +\infty$, contradicting the uniform bound hypothesis of $\mathcal{J}$. Therefore, $\mathcal{J}$ is coercive. Since $\mathcal{J}$ is continuous and coercive on the finite-dimensional space $\mathbb{R}^{N \times M}$, any sublevel set $\{\mathbf{U} : \mathcal{J}(\mathbf{U}) \leq C\}$ is closed and bounded, hence compact. By the Weierstrass extreme value theorem, $\mathcal{J}$ attains its minimum on such a compact sublevel set. Thus, there exists at least one global minimizer $\mathbf{U}^* \in \mathbb{R}^{N \times M}$. $\square$

3 SB solver for the Caputo weighted FOTV

We solve the discrete minimization problem (6) using the SB algorithm. Introduce an auxiliary vector field $\mathbf{d} = (d^{(x)}, d^{(y)})$ to decouple the nonlocal fractional gradient from the non-differentiable ℓ_1 term:

$$\mathbf{d}_{i,j} \approx \nabla^{\alpha,C} \mathbf{U}_{i,j} = (D_x^{\alpha,C} \mathbf{U}_{i,j}, D_y^{\alpha,C} \mathbf{U}_{i,j})^T.$$

The constrained problem becomes

$$\min_{\mathbf{U}, \mathbf{d}} \lambda \sum_{i,j} \log\left(1 + \frac{(\mathbf{F}_{i,j} - \mathbf{U}_{i,j})^2}{\gamma^2}\right) + \sum_{i,j} \mathbf{w}_{i,j} |\mathbf{d}_{i,j}| \quad (7)$$

$$\text{s.t. } \mathbf{d} - \nabla^{\alpha,C} \mathbf{U} = 0. \quad (8)$$

Then, we write the augmented Lagrangian version for the constrained problem (7)–(8) with penalty parameter $\beta > 0$:

$$\begin{aligned} \mathcal{L}(\mathbf{U}, \mathbf{d}; \mathbf{b}) &= \lambda \sum_{i,j} \log\left(1 + \frac{(\mathbf{F}_{i,j} - \mathbf{U}_{i,j})^2}{\gamma^2}\right) \\ &\quad + \sum_{i,j} \mathbf{w}_{i,j} |\mathbf{d}_{i,j}| \\ &\quad + \frac{\beta}{2} \sum_{i,j} \|\mathbf{d}_{i,j} - \nabla^{\alpha,C} \mathbf{U}_{i,j} + \mathbf{b}_{i,j}\|_2^2, \end{aligned} \quad (9)$$

where $\mathbf{b}$ is the scaled Bregman variable. The SB iteration alternates minimization in $\mathbf{U}$, in $\mathbf{d}$, and updates $\mathbf{b}$ [2]. It is difficult to minimize the energy (9) with respect to $\mathbf{U}$, $\mathbf{d}$, and $\mathbf{b}$ jointly. Thus, in each iteration, the original problem (9) can be split as follows

$$\mathbf{U}^{k+1} = \arg \min_{\mathbf{U}} \mathcal{L}(\mathbf{U}, \mathbf{d}^k; \mathbf{b}^k), \quad (10)$$

$$\mathbf{d}^{k+1} = \arg \min_{\mathbf{d}} \mathcal{L}(\mathbf{U}^{k+1}, \mathbf{d}, \mathbf{b}^k), \quad (11)$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k + \nabla^{\alpha,C} \mathbf{U}^{k+1} - \mathbf{d}^{k+1}. \quad (12)$$

3.1 U-subproblem

At each outer SB iteration k , the u -update (10) solves

$$\begin{aligned} \mathbf{U}^{k+1} &= \arg \min_{\mathbf{U}} \lambda \sum_{i,j} \phi_{\gamma}(\mathbf{F}_{i,j} - \mathbf{U}_{i,j}) \\ &\quad + \frac{\beta}{2} \sum_{i,j} \|\mathbf{d}_{i,j}^k - \nabla^{\alpha,C} \mathbf{U}_{i,j} + \mathbf{b}_{i,j}^k\|_2^2, \end{aligned} \quad (13)$$

Since ϕ_{γ} is globally nonconvex, the problem (13) is nonconvex and challenging to solve directly. To obtain a tractable surrogate, we employ a half-quadratic reformulation that introduces an auxiliary variable $\omega > 0$ and expresses the Cauchy potential as the minimum of a family of convex quadratic functions.

The half-quadratic theory, also known as variable-splitting or quadratic majorization, asserts that a wide class of differentiable penalties $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ admits the representation

$$\phi(r) = \min_{\omega > 0} \left\{ \frac{1}{2} \omega r^2 + \psi(\omega) \right\}, \quad (14)$$

for some convex function $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}$. Eliminating the auxiliary variable ω by inner minimization recovers the original ϕ . This decomposition is the theoretical basis of

the majorization–minimization, iteratively reweighted least squares, and half-quadratic frameworks [28–30].

Lemma 2 Let $\gamma > 0$. Then, ϕ_γ admits an exact half-quadratic representation (14), where the auxiliary potential $\psi : (0, \infty) \rightarrow \mathbb{R}$ can be chosen as

$$\psi(\omega) = -\ln \omega + \frac{\gamma^2}{2} \omega + \ln\left(\frac{2}{\gamma^2}\right) - 1. \quad (15)$$

Moreover, the minimizing weight is

$$\omega^*(r) = \frac{2}{\gamma^2 + r^2}, \quad r \neq 0, \quad (16)$$

which satisfies $0 < \omega^*(r) \leq 2/\gamma^2$.

Proof Define for fixed r the function of $\omega > 0$

$$\Phi_r(\omega) := \frac{1}{2} \omega r^2 + \psi(\omega).$$

We aim to choose ψ so that $\inf_{\omega > 0} \Phi_r(\omega) = \phi_\gamma(r)$.

Differentiate Φ_r with respect to ω and set to zero to obtain the optimality condition

$$\partial_\omega \Phi_r(\omega) = \frac{1}{2} r^2 + \psi'(\omega) = 0 \implies \psi'(\omega) = -\frac{1}{2} r^2.$$

To make this relation explicit, we use the slope-matching condition commonly used in half-quadratic constructions

$$\omega^*(r) = \frac{\phi'_\gamma(r)}{r} = \frac{2}{\gamma^2 + r^2}.$$

Invert this relation to express r^2 in terms of ω gives $r^2 = \frac{2}{\omega} - \gamma^2$. Substitute into $\psi'(\omega) = -\frac{1}{2} r^2$, and integrating with respect to ω yields

$$\psi(\omega) = -\ln \omega + \frac{\gamma^2}{2} \omega + C,$$

for an arbitrary integration constant C . Now, evaluate Φ_r at the stationary $\omega = \omega^*(r)$

$$\begin{aligned} \Phi_r(\omega^*) &= \frac{1}{2} \omega^* r^2 - \ln \omega^* + \frac{\gamma^2}{2} \omega^* + C \\ &= \frac{r^2}{\gamma^2 + r^2} - \ln\left(\frac{2}{\gamma^2 + r^2}\right) + \frac{\gamma^2}{\gamma^2 + r^2} + C \\ &= 1 - \ln 2 + \ln(\gamma^2 + r^2) + C. \end{aligned}$$

We want $\Phi_r(\omega^*) = \phi_\gamma(r) = \ln(\gamma^2 + r^2) - \ln \gamma^2$, which holds if and only if $C = \ln\left(\frac{2}{\gamma^2}\right) - 1$.

Because Φ_r is strictly convex in ω for $\omega > 0$ (its second derivative equals $\omega^{-2} > 0$ up to the constant term), the stationary point is indeed the global minimizer; hence,

$$\min_{\omega > 0} \Phi_r(\omega) = \Phi_r(\omega^*) = \phi_\gamma(r).$$

This proves the exact representation (14) with ψ given by (15). $\square$

The identity above allows us to rewrite the nonconvex u -subproblem (13) as the following jointly convex problem in (u, ω) :

$$\begin{aligned} \min_{\mathbf{U}, \omega > 0} \quad & \lambda \sum_{i,j} \left\{ \frac{1}{2} \omega_{i,j} (\mathbf{F}_{i,j} - \mathbf{U}_{i,j})^2 + \psi(\omega_{i,j}) \right\} \\ & + \frac{\beta}{2} \sum_{i,j} \|\mathbf{d}_{i,j}^k - \nabla^{\alpha,C} \mathbf{U}_{i,j} + \mathbf{b}_{i,j}^k\|_2^2. \end{aligned} \quad (17)$$

Alternating minimization in $\mathbf{U}$ and ω gives:

(a) **ω -step:** For fixed $\mathbf{U}^k$, minimize over $\omega_{i,j}$ independently using (16):

$$\omega_{i,j}^{k+1} = \frac{2}{\gamma^2 + (\mathbf{F}_{i,j} - \mathbf{U}_{i,j}^k)^2}. \quad (18)$$

(b) **u -step:** With ω^{k+1} fixed, minimize the quadratic functional

$$\begin{aligned} \mathbf{U}^{k+1} &= \arg \min_{\mathbf{U}} \frac{\lambda}{2} \sum_{i,j} \omega_{i,j}^{k+1} (\mathbf{U}_{i,j} - \mathbf{F}_{i,j})^2 \\ &+ \frac{\beta}{2} \sum_{i,j} \|\mathbf{d}_{i,j}^k - \nabla^{\alpha,C} \mathbf{U}_{i,j} + \mathbf{b}_{i,j}^k\|_2^2. \end{aligned}$$

This is a convex weighted least-squares problem that can be solved efficiently by conjugate gradient methods, using the Caputo operator and its adjoint for matrix–vector products:

$$(\lambda \mathbf{O}^k + \beta \operatorname{div}^\alpha \nabla^{\alpha,C}) \mathbf{U}^{k+1} = \lambda \mathbf{O}^k \mathbf{F} + \beta \operatorname{div}^\alpha (\mathbf{d}^k - \mathbf{b}^k), \quad (19)$$

where $\mathbf{O}^k = \operatorname{diag}(\omega_{i,j}^k)$, and $\operatorname{div}^\alpha$ denotes the discrete adjoint of the Caputo gradient operator.

3.2 d-subproblem

For each pixel, the **d**-update (11) is a 2D isotropic soft-thresholding

$$\begin{aligned} \mathbf{d}_{i,j}^{k+1} &= \\ \arg \min_{\mathbf{d}} \quad & \sum_{i,j} \left\{ \mathbf{w}_{i,j} |\mathbf{d}_{i,j}| + \frac{\beta}{2} \|\mathbf{d} - \nabla^{\alpha,C} \mathbf{U}_{i,j}^{k+1} + \mathbf{b}_{i,j}^k\|_2^2 \right\}. \end{aligned}$$

This is a closed form, and its solution is given explicitly by

$$\mathbf{d}_{i,j}^{k+1} = \max \left\{ 0, 1 - \frac{\mathbf{W}_{i,j}}{\beta \|\nabla^{\alpha,C} \mathbf{U}_{i,j}^{k+1} + \mathbf{b}_{i,j}^k\|_2} \right\} \left(\nabla^{\alpha,C} \mathbf{U}_{i,j}^{k+1} + \mathbf{b}_{i,j}^k \right), \quad (20)$$

with the convention $\frac{0}{0} = 0$. This formula is identical to weighted shrinkage used in TV but applied to the discrete Caputo gradient.

We then summarize the four updates $(\mathbf{U}, \mathbf{d}, \omega, \mathbf{b})$ in Algorithm 1.

Algorithm 1 SBsolver: Split–Bregman for Caputo weighted FOTV and Cauchy fidelity

Input: noisy image $\mathbf{F}$, weights $\mathbf{W}_{i,j}$, λ , penalty β , Caputo operator, tol Initialize $\mathbf{U}^0 = \mathbf{F}$, $\mathbf{d}^0 = \nabla^{\alpha,C} \mathbf{U}^0$, $\mathbf{b}^0 = 0$, λ, β, α and $k = 0$
1: **while** $\|\mathbf{U}^{k+1} - \mathbf{U}^k\| / \|\mathbf{U}^k\| > \text{tol}$ or $k < k_{\max}$ **do**
2: Compute (18) for $\omega_{i,j}^k$.
3: Solve (19) for $\mathbf{U}^{k+1}$.
4: Compute shrinkage-formula (20).
5: Compute the Bregman update $\mathbf{b}^{k+1} = \mathbf{b}^k + \nabla^{\alpha,C} \mathbf{U}^{k+1} - \mathbf{d}^{k+1}$
6: $k \leftarrow k + 1$
7: **end while**
Output: denoised image $\mathbf{U}^{k+1}$

Theorem 2 The sequence $\{(\mathbf{U}^k, \omega^k, \mathbf{d}^k, \mathbf{b}^k)\}$ generated from Algorithm 1 converges from any initialization to a stationary point of the augment Lagrangian

$$\begin{aligned} \mathcal{L}_\beta(\mathbf{U}, \mathbf{d}, \omega; \mathbf{b}) = & \lambda \sum_{i,j} \left(\frac{1}{2} \omega_{i,j} (\mathbf{F}_{i,j} - \mathbf{U}_{i,j})^2 + \psi(\omega_{i,j}) \right) \\ & + \sum_{i,j} \mathbf{W}_{i,j} |\mathbf{d}_{i,j}| + \frac{\beta}{2} \sum_{i,j} \|\mathbf{d}_{i,j} - \nabla^{\alpha,C} \mathbf{U}_{i,j} + \mathbf{b}_{i,j}\|_2^2. \end{aligned} \quad (21)$$

Remark 1 The proof follows the standard ADMM and split Bregman convergence framework, combining sufficient decrease arguments [2, 31, 32], nonconvex extensions [7, 33], and the Kurdyka–Łojasiewicz inequality [34, 35].

At iteration k , the variables ω^{k+1} , $\mathbf{U}^{k+1}$, and $\mathbf{d}^{k+1}$ are obtained by exact minimization of $\mathcal{L}_\beta$ with respect to each block variable while keeping the others fixed. Since each subproblem is convex, this yields the sufficient decrease in property

$$\mathcal{L}_\beta(\mathbf{U}^{k+1}, \mathbf{d}^{k+1}, \omega^{k+1}; \mathbf{b}^k) \leq \mathcal{L}_\beta(\mathbf{U}^k, \mathbf{d}^k, \omega^k; \mathbf{b}^k),$$

and thus, the sequence $\{\mathcal{L}_\beta(\mathbf{U}^k, \mathbf{d}^k, \omega^k; \mathbf{b}^k)\}$ is non-increasing.

Moreover, since $\omega^k > 0$, all terms of $\mathcal{L}_\beta$ are bounded from below, and hence, $\inf \mathcal{L}_\beta > -\infty$. In addition, $\mathcal{L}_\beta$

is coercive with respect to $(\mathbf{U}, \mathbf{d})$. Therefore, the sequence $\{(\mathbf{U}^k, \mathbf{d}^k, \omega^k, \mathbf{b}^k)\}$ remains in a compact sublevel set and is bounded [34]. Consequently, it admits accumulation points.

Let $(\mathbf{U}^*, \mathbf{d}^*, \omega^*, \mathbf{b}^*)$ be one such accumulation point. By the exact optimality of each subproblem, $(\mathbf{U}^*, \mathbf{d}^*, \omega^*, \mathbf{b}^*)$ satisfies the first-order optimality conditions of the augmented Lagrangian and is therefore a critical point of $\mathcal{L}_\beta$ [35]. By consistency of the splitting constraints, $\mathbf{U}^*$ is a stationary point of the original energy functional (6).

Finally, $\mathcal{L}_\beta$ is a finite sum of polynomial and quadratic functions, an ℓ^1 -norm, and a real-analytic function defined on $\omega > 0$. Since the discrete Caputo fractional operator $\nabla^{\alpha,C}$ is linear, $\mathcal{L}_\beta$ is a proper lower semi-continuous definable function and satisfies the Kurdyka–Łojasiewicz property [34, 35]. The classical KL convergence theorem then ensures that the whole sequence $\{(\mathbf{U}^k, \mathbf{d}^k, \omega^k, \mathbf{b}^k)\}$ converges to $(\mathbf{U}^*, \mathbf{d}^*, \omega^*, \mathbf{b}^*)$.

4 Numerical implementation

4.1 Discretization

For the discretization, we employ a matrix-free implementation following Algorithm 1. Using the backward quadrature form of the Caputo derivative in (4)–(5), the discrete operators can be expressed in matrix form as

$$D_x^{\alpha,C} \mathbf{U} = \frac{h^{-\alpha}}{\Gamma(2-\alpha)} \mathbf{A}_\alpha \mathbf{U}, \quad D_y^{\alpha,C} \mathbf{U} = \frac{h^{-\alpha}}{\Gamma(2-\alpha)} \mathbf{U} \mathbf{B}_\alpha^T,$$

where $\mathbf{A}_\alpha \in \mathbb{R}^{N \times N}$ and $\mathbf{B}_\alpha \in \mathbb{R}^{M \times M}$ are lower-triangular Toeplitz matrices whose entries are generated by the one-dimensional fractional weights $a_\ell^{(\alpha)} = (\ell+1)^{1-\alpha} - \ell^{1-\alpha}$. Let $\mathbf{U} \in \mathbb{R}^{N \times M}$ be the image and $\mathbf{u} = \text{vec}(\mathbf{U}) \in \mathbb{R}^{NM}$ its column-wise vectorization. In vectorized form, the two-dimensional fractional gradient yields the stacked discrete gradient operator

$$\nabla^{\alpha,C} \mathbf{u} = \frac{h^{-\alpha}}{\Gamma(2-\alpha)} \begin{bmatrix} (\mathbf{I}_M \otimes \mathbf{A}_\alpha) \\ (\mathbf{B}_\alpha \otimes \mathbf{I}_N) \end{bmatrix} \mathbf{u} \in \mathbb{R}^{2NM}.$$

The discrete adjoint operator is defined by the transposes of the discrete $D_x^{\alpha,C}$ and $D_y^{\alpha,C}$. Denote the stacked field $\mathbf{p} = [\mathbf{p}_1; \mathbf{p}_2] \in \mathbb{R}^{2NM}$ where $\mathbf{p}_1, \mathbf{p}_2$ correspond to the two directional components. Then, the adjoint operator $(\nabla^{\alpha,C})^T : \mathbb{R}^{2NM} \rightarrow \mathbb{R}^{NM}$ writes

$$(\nabla^{\alpha,C})^T \mathbf{p} = \frac{h^{-\alpha}}{\Gamma(2-\alpha)} (\mathbf{I}_M \otimes \mathbf{A}_\alpha^T) \mathbf{p}_1 + (\mathbf{B}_\alpha^T \otimes \mathbf{I}_N) \mathbf{p}_2.$$

We denote the discrete divergence by $\text{div}^\alpha(\mathbf{p}) := -(\nabla^{\alpha,C})^T \mathbf{p}$. Using the matrix notation above, the U -step linear system

(19) becomes

$$\begin{aligned} & (\lambda \mathbf{O}^k + \beta (\nabla^{\alpha, C})^T \nabla^{\alpha, C}) \mathbf{U}^{k+1} \\ & = \lambda \mathbf{O}^k \mathbf{F} + \beta (\nabla^{\alpha, C})^T (\mathbf{d}^k - \mathbf{b}^k), \end{aligned}$$

To solve the resulting linear systems, we employ a preconditioned conjugate gradient (PCG) method in a matrix-free manner. In our implementation, no explicit preconditioner is introduced, which corresponds to using the identity operator as a preconditioner. This choice is motivated by the structure of the system operator

$$\mathcal{A}^k := \lambda \mathbf{O}^k + \beta (\nabla^{\alpha, C})^T \nabla^{\alpha, C},$$

which is symmetric and positive definite. The diagonal matrix $\mathbf{O}^k = \text{diag}(\omega_{i,j}^k)$, arising from the iteratively reweighted least-squares scheme associated with the Cauchy data-fidelity term, acts as an intrinsic diagonal scaling that improves the conditioning of the system by down-weighting pixels affected by strong outliers. Moreover, the quadratic penalty term $\beta (\nabla^{\alpha, C})^T \nabla^{\alpha, C}$ introduced by the split Bregman framework further stabilizes the spectrum of the linear operator. From a numerical standpoint, the fractional operator $(\nabla^{\alpha, C})^T \nabla^{\alpha, C}$ is applied in a matrix-free fashion using the Toeplitz structures $\mathbf{A}_\alpha$ and $\mathbf{B}_\alpha$ through successive matrix-vector products. Constructing an explicit preconditioner involving these operators would require assembling dense fractional matrices or designing problem-specific approximations for nonlocal convolutions, leading to significantly increased memory usage and computational cost. Such approaches would be incompatible with the matrix-free strategy adopted in this work. In practice, the combination of the diagonal weight matrix $\mathbf{O}^k$ and the penalty parameter β provides an effective implicit preconditioning, resulting in fast and stable convergence of the PCG solver.

For full reproducibility, the discretization routines used in this work are provided in our public GitHub repository: <https://github.com/braiekhamdi/weightedfotv/tree/main>.

The discretization of the Caputo fractional-order gradient introduces intrinsic nonlocality, since the fractional operator at a given pixel depends on image values over an extended neighborhood. For solving the U -subproblem in (19), reflective (Neumann-type) boundary conditions are adopted and implemented via symmetric padding of the image domain, namely

$$\begin{aligned} \mathbf{U}(0, j) &= \mathbf{U}(2, j), & \mathbf{U}(N+1, j) &= \mathbf{U}(N-1, j), \\ \mathbf{U}(i, 0) &= \mathbf{U}(i, 2), & \mathbf{U}(i, M+1) &= \mathbf{U}(i, M-1). \end{aligned}$$

Such boundary conditions are commonly used in image denoising problems, as they provide a natural continuation of image intensities and avoid artificial edge artifacts introduced

by zero-padding or Dirichlet boundary conditions, which are known to adversely affect fractional operators near the boundaries [36]. The use of reflective boundary conditions is consistent with the convolution-type structure induced by the backward quadrature discretization of the Caputo derivative and preserves the symmetry of the discrete operator $(\nabla^{\alpha, C})^T \nabla^{\alpha, C}$ [37, 38]. The resulting linear system in the u -subproblem consists of a diagonal data-fidelity term $\lambda \mathbf{O}^k$ and a fractional diffusion term $\beta (\nabla^{\alpha, C})^T \nabla^{\alpha, C}$ and is therefore symmetric positive definite. This structure ensures stable numerical behavior and allows efficient solution via conjugate gradient methods in a matrix-free framework [39]. Moreover, a full discrete error analysis for Caputo fractional derivatives with reflective boundary conditions has been shown to achieve stable and accurate solutions in fractional diffusion problems [37, 38, 40, 41].

4.2 Bayesian optimization

The performance of the proposed weighted FOTV model depends critically on the regularization parameters λ , β , and the fractional order $\alpha \in [1, 2]$. A suboptimal choice may lead to either oversmoothing or insufficient noise removal. To achieve stable and near-optimal restoration quality, we adopt a data-driven Bayesian optimization (BO) strategy to automatically tune (λ, β, α) .

BO is a global black-box optimization framework designed for objectives that are expensive to evaluate and lack closed-form gradients [22, 23]. Compared to manual tuning or grid search, BO offers a statistically principled approach with superior sample efficiency. It adapts the parameters to both the noise characteristics and image content, achieving a consistent balance between edge preservation and smooth denoising.

Let $\Theta = (\lambda, \beta, \alpha)$ denote the vector of tunable parameters, and let $J_{\text{obj}}(\Theta)$ denote the quantitative performance measure of the model. When the ground-truth image $\mathbf{u}_{\text{true}}$ is available, we define

$$J_{\text{obj}}(\Theta) = -\text{PSNR}(\mathbf{U}_\Theta, \mathbf{U}_{\text{true}}),$$

where $\mathbf{U}_\Theta$ is the restored image obtained with parameters Θ , and PSNR is the peak signal-to-noise ratio,

$$\text{PSNR}(\mathbf{U}_\Theta, \mathbf{U}_{\text{true}}) = 10 \log_{10} \left(\frac{L^2}{\|\mathbf{U}_\Theta - \mathbf{U}_{\text{true}}\|_2^2 / (NM)} \right),$$

with L denoting the image dynamic range. The goal is to minimize $J_{\text{obj}}(\Theta)$ over a bounded search box $\Theta \in [\Theta_{\min}, \Theta_{\max}]$.

At each iteration, the BO algorithm maintains a probabilistic surrogate model of the objective function using a Gaussian

process (GP) prior,

$$J_{\text{obj}}(\Theta) \sim \mathcal{GP}(m(\Theta), \kappa(\Theta, \Theta')),$$

where $m(\Theta)$ and $\kappa(\Theta, \Theta')$ denote the mean and covariance functions. The mean function is chosen constant, $m(\Theta) = m_0$. The covariance is modeled with a Matern-5/2 kernel

$$\kappa(\Theta, \Theta') = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{\ell} + \frac{5r^2}{3\ell^2} \right) \exp\left(-\frac{\sqrt{5}r}{\ell}\right) + \sigma_n^2 \delta_{\Theta, \Theta'}, \quad (22)$$

where $r = \|\Theta - \Theta'\|_2$, σ_f^2 is the signal variance, ℓ the characteristic length scale, σ_n^2 the observation noise variance, and $\delta_{\Theta, \Theta'}$ the Kronecker delta. The hyperparameters $(m_0, \sigma_f^2, \ell, \sigma_n^2)$ are learned by maximizing the GP marginal likelihood. Given observed evaluations $\{(\Theta^i, J_{\text{obj}}^i)\}_{i=1}^{t-1}$, the GP yields a posterior predictive distribution characterized by a mean $\mu_t(\Theta)$ and standard deviation $\sigma_t(\Theta) > 0$. Let J_t^* denote the best (lowest) objective value observed so far.

To determine the next evaluation point, BO maximizes an acquisition function, which quantifies the expected improvement $\text{EI}(\Theta)$ over the current best value. The expected improvement acquisition is given in closed form as

$$\text{EI}_t(\Theta) = (J_t^* - \mu_t(\Theta)) \Phi\left(\frac{J_t^* - \mu_t(\Theta)}{\sigma_t(\Theta)}\right) + \sigma_t(\Theta) \varphi\left(\frac{J_t^* - \mu_t(\Theta)}{\sigma_t(\Theta)}\right), \quad (23)$$

where Φ and φ denote the CDF and PDF of the standard normal distribution. The next evaluation point is then obtained by

$$\Theta^t = \underset{\Theta \in [\Theta_{\min}, \Theta_{\max}]}{\operatorname{argmax}} \text{EI}_t(\Theta). \quad (24)$$

Equation (23) balances exploration (large σ_t) and exploitation (small μ_t). The process iteratively refines the surrogate model until convergence or a stopping criterion is met.

In practice, the optimization is performed using MATLAB's `bayesopt` function with a Matern-5/2 kernel and the expected improvement acquisition (Eq. (23)). In Algorithm 2, we describe the BO steps for parameter tuning of the SB solver using weighted fractional TV. It is important to note that each evaluation of the objective function within the BO loop involves a full run of the split Bregman solver with dynamically updated weights $w^k(x)$, so that BO implicitly optimizes parameters in the presence of the proposed spatially adaptive weighting mechanism.

In this algorithm, we set $T_{\max} = 20$ iterations. Parameter bounds are set to $\lambda \in [10^{-2}, 5]$, $\beta \in [10^{-2}, 5]$, and $\alpha \in [1, 2]$. The ranges for λ and β are chosen to ensure a flexible yet stable balance between the data-fidelity term and the regularization components across different noise levels.

Algorithm 2 Bayesian optimization for parameter tuning of the fractional SBSolver.

Require: Noisy image $\mathbf{F}$, reference image $\mathbf{U}_{\text{true}}$, parameter domain $\Theta = [\Theta_{\min}, \Theta_{\max}]$, maximum evaluations $T_{\max}$.

Initialization

- 1: Select an initial design $\{\Theta^{(i)}\}_{i=1}^{n_0}$ using Latin Hypercube sampling.
- 2: **for** $i = 1, \dots, n_0$ **do**
- 3: Compute $u^{(i)} \leftarrow \text{SBSolver}(\mathbf{F}; \Theta^{(i)})$
- 4: Evaluate objective

$$J_{\text{obj}}^{(i)} \leftarrow -\text{PSNR}(\mathbf{U}^{(i)}, \mathbf{U}_{\text{true}}).$$

5: **end for**

6: Fit a Gaussian process surrogate

$$J_{\text{obj}}(\Theta) \sim \mathcal{GP}(m(\Theta), \kappa(\Theta, \Theta')).$$

Bayesian Optimization Loop

- 7: **for** $t = n_0 + 1, \dots, T_{\max}$ **do**
- 8: Compute GP posterior: $\mu_t(\Theta)$ and $\sigma_t^2(\Theta)$.
- 9: Let $J_t^* = \min\{J_{\text{obj}}^{(1)}, \dots, J_{\text{obj}}^{(t-1)}\}$.
- 10: Determine next candidate $\Theta^{(t)}$ by maximizing (24).
- 11: Run the fractional SB solver:

$$\mathbf{U}^{(t)} \leftarrow \text{SBSolver}(\mathbf{F}; \Theta^{(t)}).$$

12: Evaluate: $J_{\text{obj}}^{(t)} \leftarrow -\text{PSNR}(\mathbf{U}^{(t)}, \mathbf{U}_{\text{true}})$.

13: Update the GP with the new pair $(\Theta^{(t)}, J_{\text{obj}}^{(t)})$.

14: **end for**

$$\text{Output: } \Theta^* = \arg \min_{1 \leq i \leq T_{\max}} J_{\text{obj}}^{(i)}.$$

The search range of the fractional order $\alpha \in [1, 2]$ is motivated by both theoretical and practical considerations. When $\alpha = 1$, the fractional-order total variation model reduces to the classical first-order TV, which provides strong edge preservation but is prone to staircasing artifacts. Increasing α introduces a higher degree of smoothness and implicitly captures longer-range interactions, thereby reducing staircasing and improving texture representation. Values of α below 1 tend to introduce excessive nonlocal smoothing and may lead to loss of fine structures, whereas values significantly larger than 2 can oversmooth edges and increase numerical instability in the discretized fractional operator. Therefore, the interval $[1, 2]$ provides a meaningful compromise between edge preservation and texture smoothing and has been widely adopted in fractional-order image regularization studies [8, 16].

Once the optimal triplet Θ^* is identified, the solver is executed with these parameters for the final restoration.

5 Experiments

To validate the proposed weighted FOTV model, we perform experiments on synthetic and standard test images. The observed image f is generated by corrupting a clean

reference image u_{true} with additive Cauchy noise of scale parameter $\gamma > 0$. Let η be a random variable distributed according to the standard Cauchy law with density $p(\eta) = \frac{1}{\pi} \frac{\gamma}{\eta^2 + \gamma^2}$. In practice, a noisy image can be simulated by $f(x) = u_{\text{true}}(x) + \eta(x)$ where $\eta(x) = \gamma \tan(\pi(\xi(x) - \frac{1}{2}))$, where $\xi(x) \sim \mathcal{U}(0, 1)$ is uniformly distributed.

To quantitatively assess the quality of the restored image u_{rec} , we compute the PSNR between the reconstructed image u_{rec} and the ground truth u_{true} , we use also the structural similarity index (SSIM), and the relative error (ReErr) with respect to the ground truth

$$\text{SSIM}(u_{\text{rec}}, u_{\text{true}}) = \frac{(2\mu_u \mu_f + c_1)(2\sigma_{uf} + c_2)}{(\mu_u^2 + \mu_f^2 + c_1)(\sigma_u^2 + \sigma_f^2 + c_2)},$$

$$\text{ReErr}(u_{\text{rec}}, u_{\text{true}}) = \frac{\|u_{\text{rec}} - u_{\text{true}}\|_2}{\|u_{\text{true}}\|_2},$$

where $\mu_u, \mu_f, \sigma_u, \sigma_f, \sigma_{uf}$ denote the empirical means, standard deviations, and cross-covariance of u_{rec} and u_{true} . The iterative procedure is terminated once the relative error satisfies $\text{ReErr}(X^k, X^{k-1}) < 10^{-3}$. Additionally, we stop when the evolution of image quality indices, SNR, SNR, SSIM, becomes numerically stable, meaning that successive iterations yield negligible variation in the metrics.

We also compare our model with four other competitive approaches widely used for image restoration under heavy-tailed noise. The first one is the classical total variation model (**TV**) $\mathcal{J}(u) = TV(u) + \lambda \mathcal{D}(u; f)$, proposed in [7]. The second is the penalized total variation model (**pen-TV**) $\mathcal{J}(u) = TV(u) + \lambda \mathcal{D}(u; f) + \mu \|u - u_0\|_2^2$, introduced in [6] to improve stability and alleviate the staircase effect. u_0 is the filtered image f by the median filter. The third baseline is the second-order total generalized variation model (**TGV**) $\mathcal{J}(u) = TGV_{\mu_1, \mu_2}^2(u) + \lambda \mathcal{D}(u; f)$, recently adapted for Cauchy noise removal in [42]. The fourth model is the group-based low-rank prior model (**GBLR**), which is included as a representative strong nonlocal low-rank denoising method widely used in recent years. Its energy functional is given by $\mathcal{J}(u) = \|u\|_* + \lambda \mathcal{D}(u; f)$, where $\|X\|_* = \sum_{r=1}^{\min(m,n)} \sigma_r(X)$ is the nuclear norm of a matrix $X \in \mathbb{R}^{n \times m}$ and σ_r is the r th largest singular value. This model was proposed in [43]. We also include the fractional-order total variation model (**FOTV**), corresponding to our proposed model (2) with a uniform weight function $w(x) = 1$. For a fair comparison, all models are implemented using the same SB optimization framework, initialized with identical settings and using the penalty parameter β . The regularization parameters $\{\lambda, \beta, \mu, \mu_1, \mu_2\}$ of the compared models, as well as (λ, β, α) in our method, are tuned using the BO strategy. We also use eleven ground-truth images to compare the models. The images are shown in Fig. 1.

The adaptive weight $w(x)$ in (3) aims to preserve edges while enforcing stronger regularization in smooth regions. Following edge-preserving diffusion principles, the weight map w is updated iteratively after each u -update according to the current restored image:

$$w^{k+1}(x) = \exp(-|\nabla(G_\sigma * u^{k+1}(x))|).$$

With this choice, it adapts the regularization strength spatially and across iterations.

5.1 Grayscale images

We present quantitative and qualitative evaluations of the proposed weighted FOTV model on grayscale images corrupted by Cauchy noise. The parameters λ, β , and α were tuned from Algorithm 2 to achieve the best trade-off between noise suppression and detail preservation.

Figures 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 illustrate the performance of the competing models on several grayscale test images. Table 1 reports the corresponding quantitative measures (SNR, PSNR, SSIM) together with the number of SB iterations needed for each method. Figure 2 presents the reconstructions models for the ‘Clock’ image in Fig. 1(a), including zoomed patches for detailed comparison.

Visual inspection of the results shows a clear pattern: Classical TV removes a large portion of the Cauchy outliers but introduces piecewise-constant artifacts and blunted thin structures. The pen-TV variant reduces some staircasing at the cost of mild additional smoothing, while TGV better preserves smooth transitions but occasionally attenuates small scale ridges. The GBLR model, which exploits non-local self-similarity through low-rank group representations, produces competitive results and effectively suppresses noise in homogeneous regions. Nevertheless, it may oversmooth fine textures and thin structures, particularly in areas with strong local contrast. The FOTV model recovers medium-range oscillatory structures more faithfully than first-order TV; however, without spatial weighting it can still underperform near high-contrast edges. The proposed weighted FOTV consistently returns the most faithful texture and edge reconstruction, retaining thin circular contours and local contrast while suppressing extreme outliers. These qualitative observations are supported by the numbers in Table 1: The weighted FOTV achieves the highest PSNR and SSIM in most test cases, indicating that the combined effect of fractional smoothing and adaptive weighting produces both visually pleasing and objectively better restorations 3.

Figure 4 illustrates standard BO diagnostics produced by Algorithm 2 for the clock image. Each panel provides a different view of the optimizer’s behavior and of the computational cost of searching the parameter space Θ . The top-left panel shows the Gaussian-process surrogate posterior mean of the

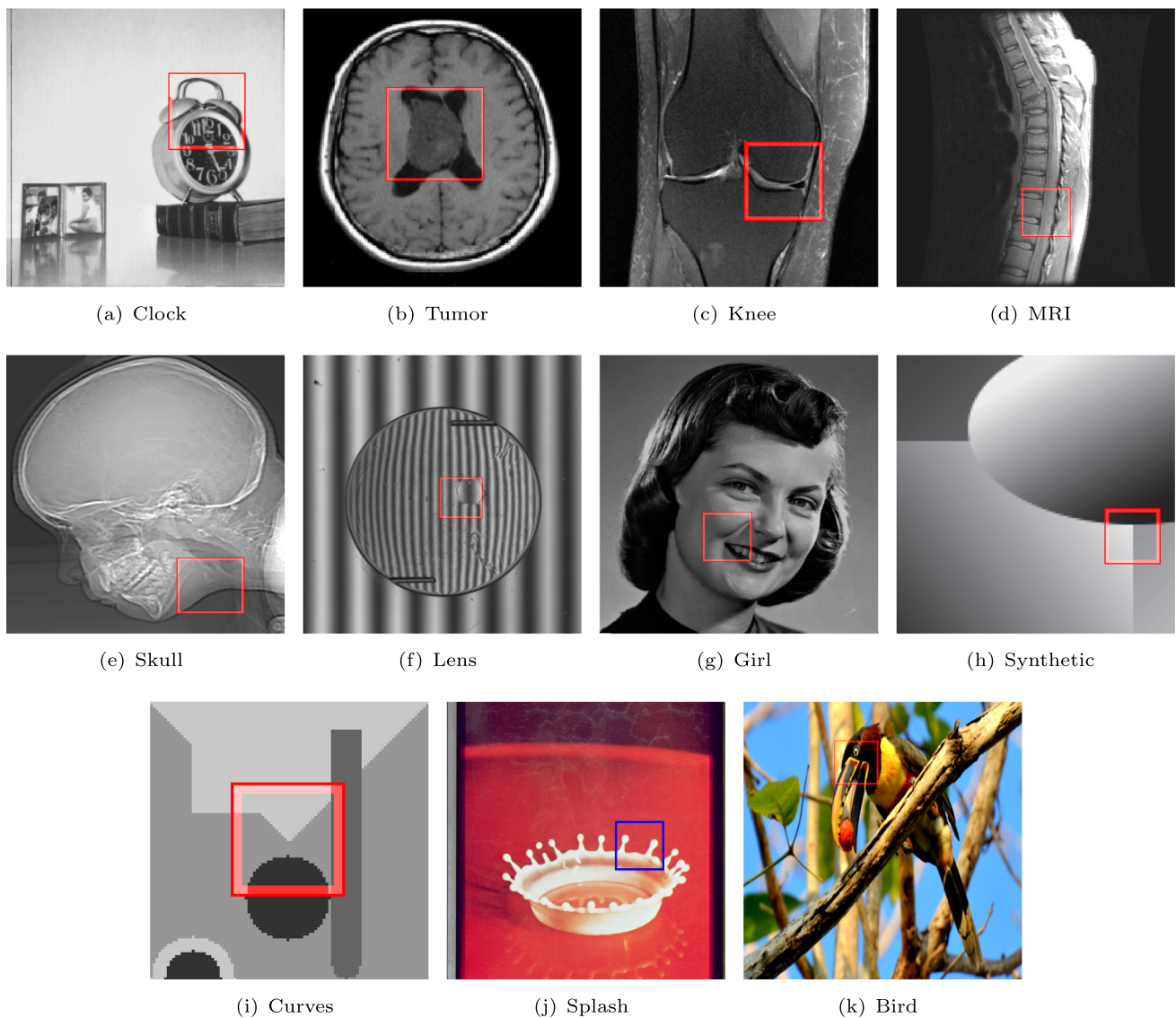
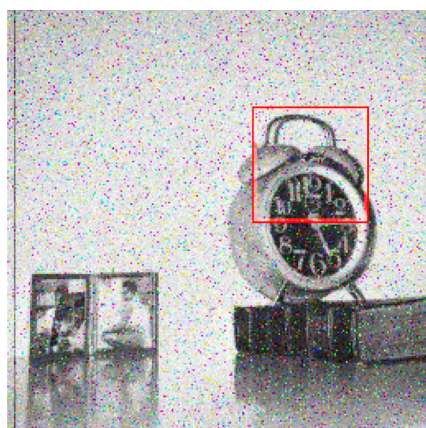


Fig. 1 The ground-truth images intended to use in this work

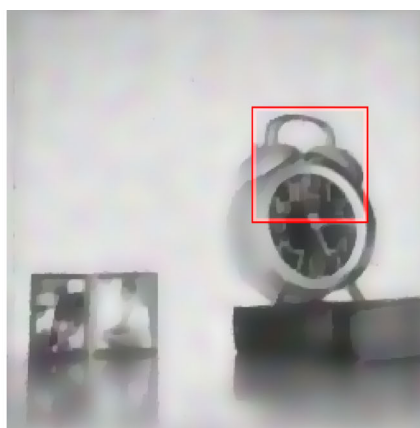
objective function $J_{\text{obj}}(\Theta)$, together with observed evaluations (red dots) and the next query point (black marker) selected by the expected improvement criterion. The learned surface reveals a well-defined low-objective basin, and the concentration of samples in this region indicates a transition from exploration to exploitation and convergence toward consistent parameter values. The top-right panel reports the cumulative runtime as a function of the number of objective evaluations. The dominant contribution arises from objective evaluations (blue curve), each corresponding to a full split Bregman reconstruction, while the overhead associated with surrogate modeling and acquisition optimization (red curve) remains comparatively small. This confirms that the computational cost is primarily driven by the number of solver runs rather than by Bayesian optimization itself. The bottom-

left panel plots the minimum observed objective value and the surrogate-estimated minimum versus the number of evaluations. A sharp decrease within the first few iterations demonstrates that Bayesian optimization rapidly identifies high-quality parameter configurations, after which the objective stabilizes. The close agreement between observed and estimated minima indicates a reliable surrogate model near the optimum. Finally, the bottom-right panel shows the per-evaluation runtime, which remains approximately constant between 1.3 and 1.6 s, further emphasizing the importance of minimizing the number of objective evaluations.

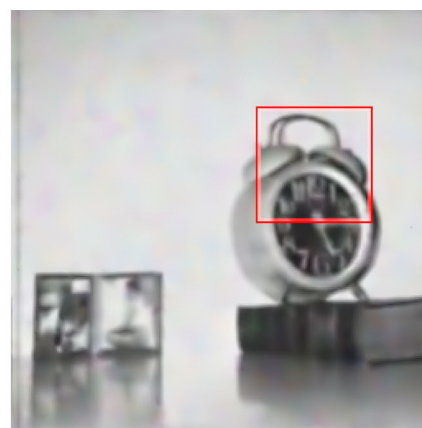
We present numerical experiments for medical images, including Brain Tumor CT (Fig. 5), Knee MRI (Fig. 8 and zoomed regions in Fig. 9), human thoracic spine MRI (Fig. 10), and Skull CT (Fig. 12). These modalities present



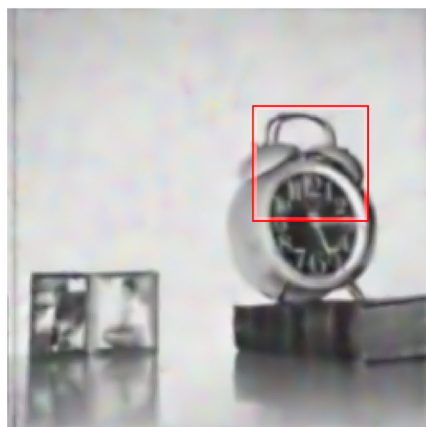
(a) Noisy image



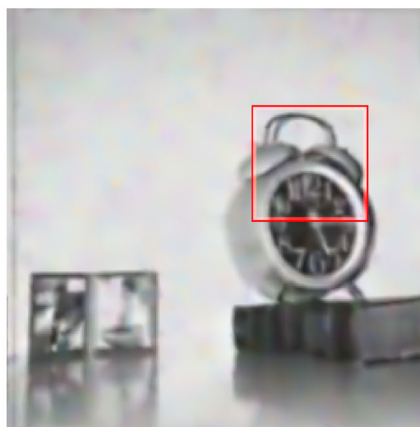
(b) TV



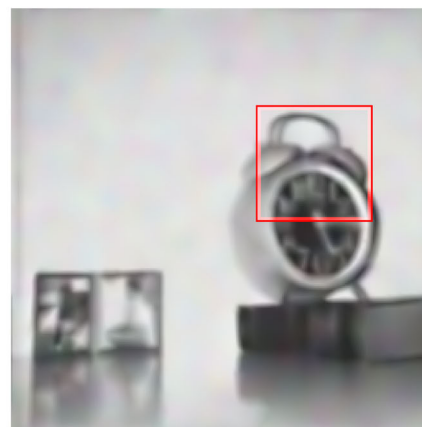
(c) pen-TV



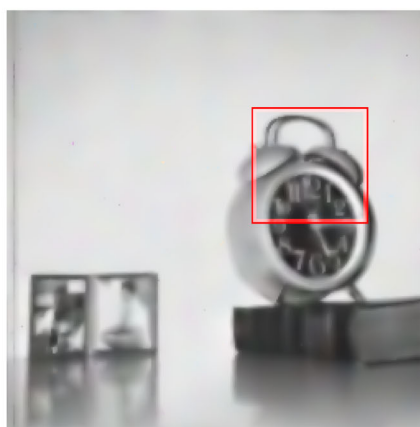
(d) TGV



(e) GBLR



(f) FOTV



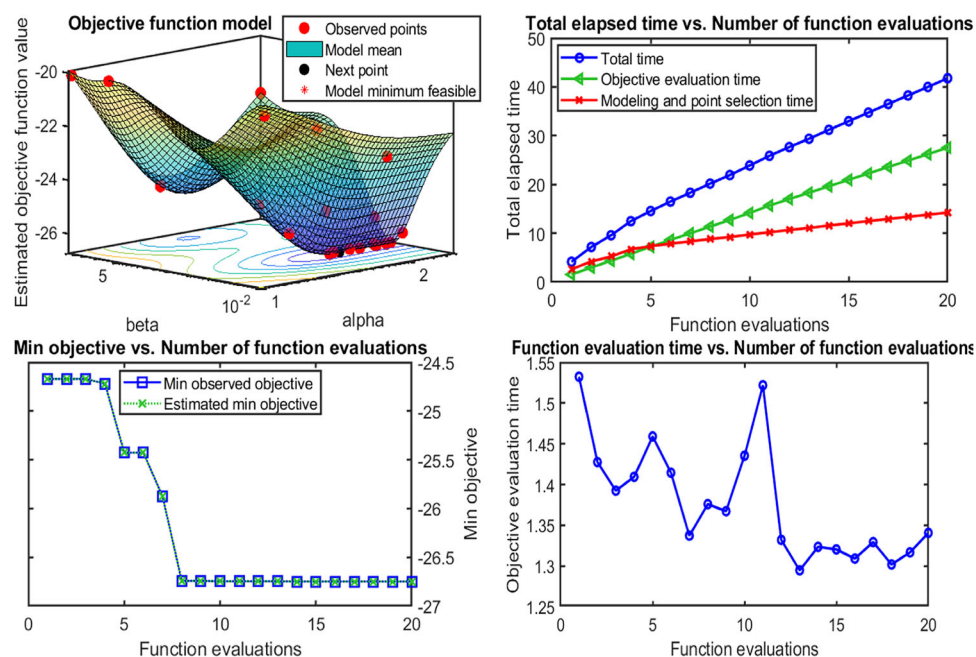
(g) Weighted FOTV

Fig. 2 From left to right: The noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for image in Fig. 1a



Fig. 3 From left to right: The zoomed regions of the noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 2

Fig. 4 From left to right: The tuned parameters using Algorithm 2 for the image in Fig. 2f



unique challenges, combining smooth anatomical regions with fine structural details that are clinically significant.

In the Brain Tumor CT case (Fig. 5 and zoomed regions in Fig. 6), all compared models exhibit varying degrees of noise suppression, yet tend to either oversmooth tumor boundaries or fail to fully preserve subtle parenchymal textures. The proposed weighted FOTV not only achieves superior noise removal but also enhances the visibility of tumor margins, as reflected in the adaptive weight map (Fig. 17(p)), which selectively reduces regularization near edges and pathological structures.

For the thoracic spine MRI (Figs. 10 and Fig. 11), the model demonstrates remarkable fidelity in preserving thin anatomical features such as vertebral bodies, intervertebral disks, and the spinal canal (Fig. 17(h)). While GBLR and TGV offer improved smoothness, they often attenuate these fine details. The weighted FOTV maintains high contrast in these regions without introducing staircase artifacts, as corroborated by the quantitative metrics in Table 1.

Similarly, in Knee MRI and Skull CT, the proposed approach consistently delivers the best balance between homogeneity in soft-tissue/bone regions and edge sharpness at structural interfaces. The adaptive weighting mechanism ensures that regions with high diagnostic value, such as cartilage interfaces in the knee or suture lines in the skull, are preserved, whereas more uniform areas are effectively smoothed.

Across all medical images, the weighted FOTV achieves the highest PSNR, SNR, and SSIM values in Table 1, while also requiring fewer SB iterations for convergence. This combination of quantitative superiority, visual fidelity, and

computational efficiency underscores the model's robustness and clinical applicability in restoring medical images corrupted by heavy-tailed Cauchy noise.

To further assess the robustness of the proposed Bayesian optimization strategy, we conducted a sensitivity analysis of the main regularization parameters $\Theta = \alpha, \beta, \lambda$ involved in the weighted FOTV model. Figure 7 illustrates the evolution of the reconstruction quality, measured by PSNR. The analysis is performed on the Tumor image shown in Fig. 1(b).

The PSNR curve with respect to the fractional order α shows a clear unimodal behavior, indicating that intermediate fractional orders provide a favorable balance between noise suppression and detail preservation. A similar trend is observed for the penalty parameter β , where excessive values degrade performance due to overly strong constraint enforcement. The regularization weight λ also exhibits a well-defined optimum, confirming its critical role in balancing data fidelity and fractional regularization.

Bayesian optimization automatically identifies the optimal parameter set $\alpha = 1.313$, $\beta = 0.09$, and $\lambda = 1.516$, which yields a maximum PSNR of 26.478 dB (see Table 2). These results demonstrate that the proposed optimization framework is not only effective in selecting high-performing parameters, but also robust, as small deviations around the optimal values lead to gradual rather than abrupt performance degradation. This sensitivity analysis confirms the stability of the proposed model and strengthens the validation of the Bayesian optimization procedure 13.

We also present, in Figs. 14, 15, 16, results on the synthetic and natural images and highlight the model's behavior on highly textured patterns and on complex natural content.

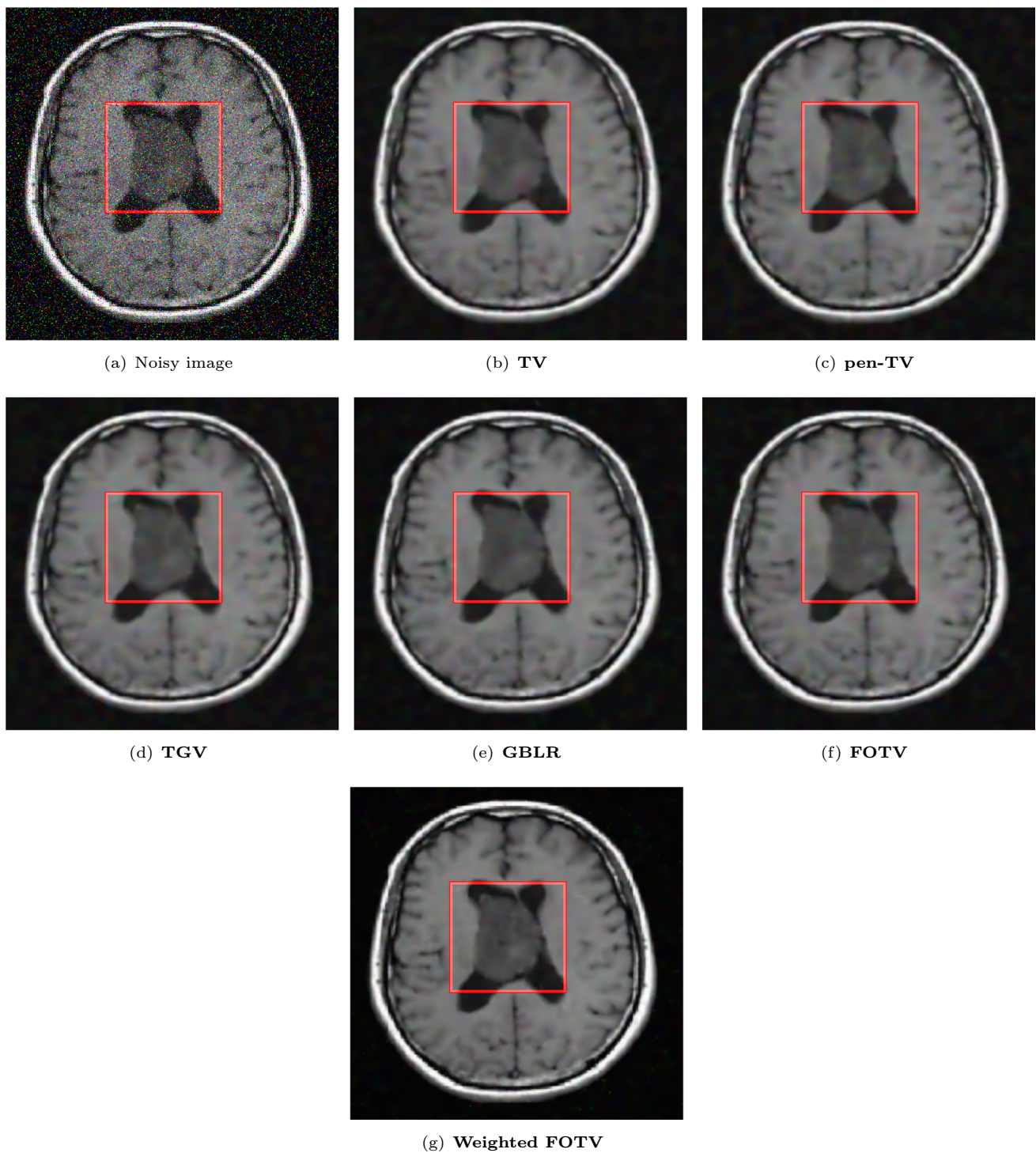


Fig. 5 From left to right: The noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 1b

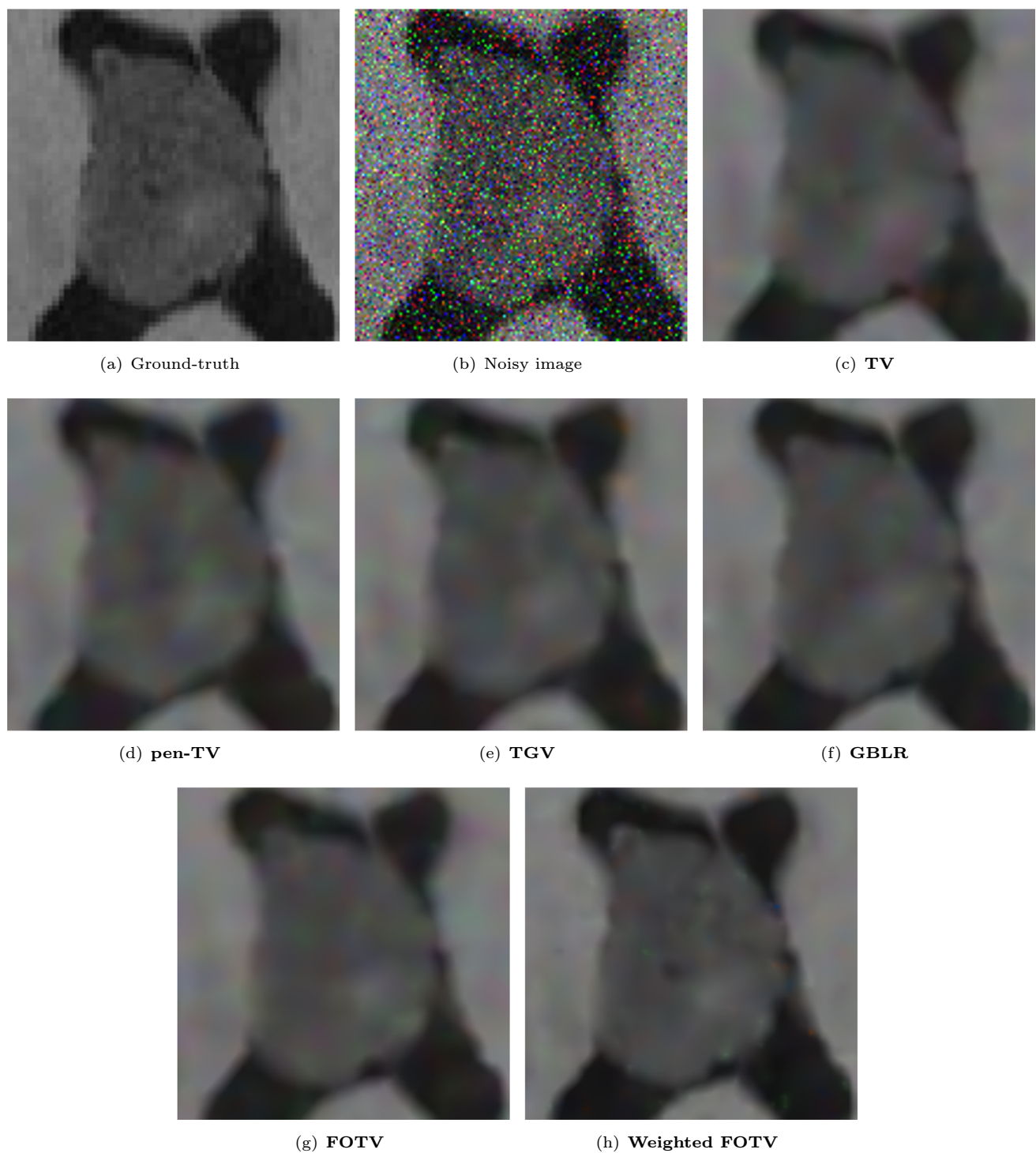


Fig. 6 From left to right: The zoomed regions for the noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 5

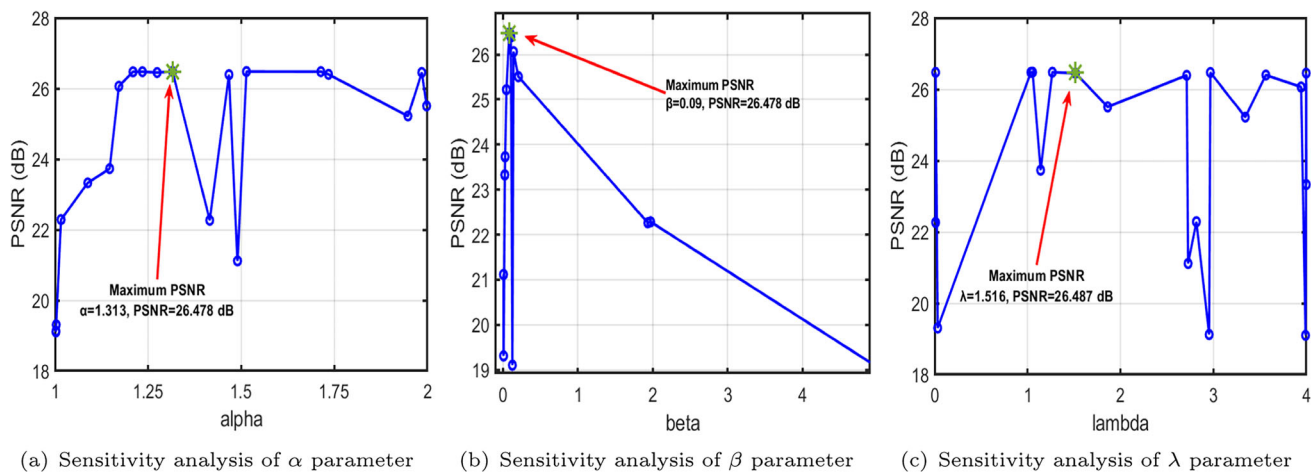


Fig. 7 Sensitivity analysis of the proposed weighted FOTV model parameters $\Theta = \{\alpha, \beta, \lambda\}$ using Bayesian optimization. The criterion function $|J| = \text{PSNR}$ is plotted as a function of each parameter. The optimal parameters identified by Bayesian optimization are $\alpha = 1.313$,

$\beta = 0.09$, and $\lambda = 1.502$, yielding a maximum PSNR of 26.478 dB for the ‘Tumor’ test image in Fig. 1b. The green markers indicate the optimal parameter values corresponding to the maximum PSNR

On the synthetic texture, the fractional operator’s capacity to model medium-range interactions reduces the staircasing and restores repetitive structures with higher fidelity than local TV. For the girl face image, which contains both smooth skin regions and high-frequency hair details, the adaptive weight prevents oversmoothing of perceptually important features: Edges and hair filaments are preserved, while homogeneous regions remain clean 15.

In Fig. 17, we present the evolution of the weight map w (Eq. (3)) across the SB iterations. The visualization clearly illustrates how the adaptive mechanism progressively refines the spatial distribution of the regularization strength. At early iterations, the weight map remains relatively uniform, reflecting the fact that the initial estimate has not yet revealed reliable edge and texture information. As the reconstruction improves, the weight function becomes increasingly structured: Lower weights concentrate around edges, high-frequency contours, and textured regions, whereas homogeneous areas are assigned larger weights. This behavior is consistent with the design of w , which is intended to attenuate regularization near image discontinuities while enforcing stronger smoothing in flat regions.

Table 2 summarizes the Bayesian optimization setup for the five models. For each model, the parameter vector Θ is tuned automatically by minimizing the objective function $J_{\text{obj}}(\Theta)$ (e.g., the negative PSNR). The optimization is performed over 20 BO iterations using a Matérn-5/2 kernel and the expected improvement acquisition function. Each evaluation involves running the SB solver for inner iterations. The optimal parameter Θ^* corresponds to the configuration yielding the minimal objective value J_{obj}^* , ensuring a balanced compromise between data fidelity and regularization.

Although Bayesian optimization introduces an additional overhead due to surrogate model construction and acquisition optimization, the dominant computational cost of the proposed framework arises from objective evaluations, each requiring a full split Bregman reconstruction. As illustrated in Figure 3 (top-right), the cumulative runtime is overwhelmingly driven by solver executions, while the overhead associated with Gaussian-process modeling remains negligible. Compared to simpler tuning strategies such as grid search, whose computational complexity grows exponentially with the number of parameters, Bayesian optimization drastically reduces the number of required objective evaluations by directing the search toward promising regions of the parameter space. For instance, a coarse grid search with only 10 candidate values per parameter would already require 10^3 reconstructions, whereas the proposed BO strategy identifies competitive parameter configurations within fewer than 20 evaluations (Table 2). Consequently, despite a modest per-iteration overhead, Bayesian optimization leads to substantial overall computational savings in practice.

To provide further insight into the robustness of the proposed Bayesian-optimized model, we analyze its behavior under different noise levels and image characteristics. Figures 18 and 19 present visual denoising results for three increasing noise levels, namely $\gamma \in 0.04, 0.08, 0.12$, on the grayscale image in Fig. 1(i). These experiments allow us to evaluate the stability of the model parameters selected by Bayesian optimization when the noise intensity changes. From the visual results in Fig. 18, it can be observed that classical TV-based models tend to oversmooth the image as the noise level increases, leading to loss of fine structures and contrast. In contrast, the proposed weighted fractional-order

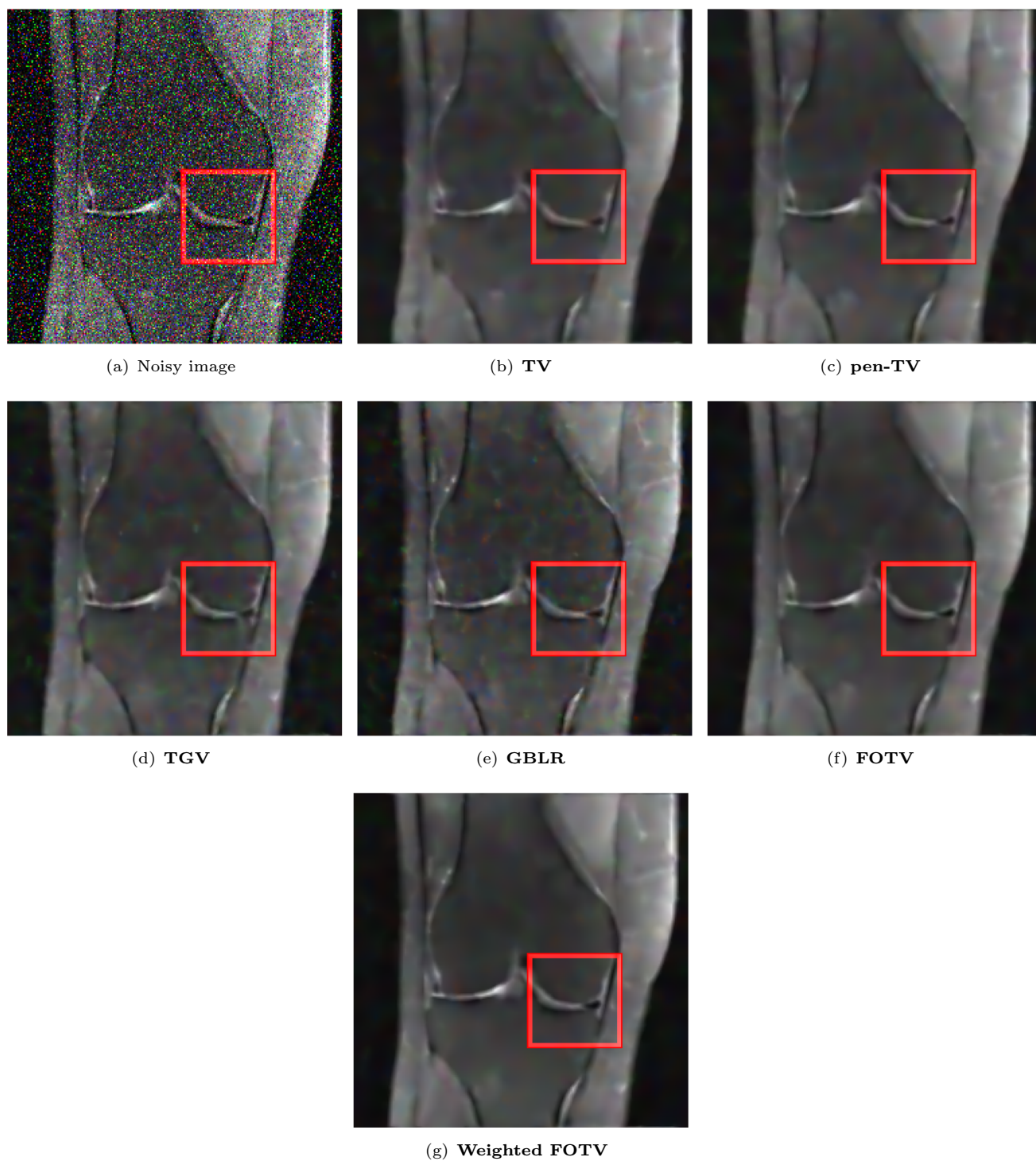


Fig. 8 From left to right: The zoomed regions for the noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 1c

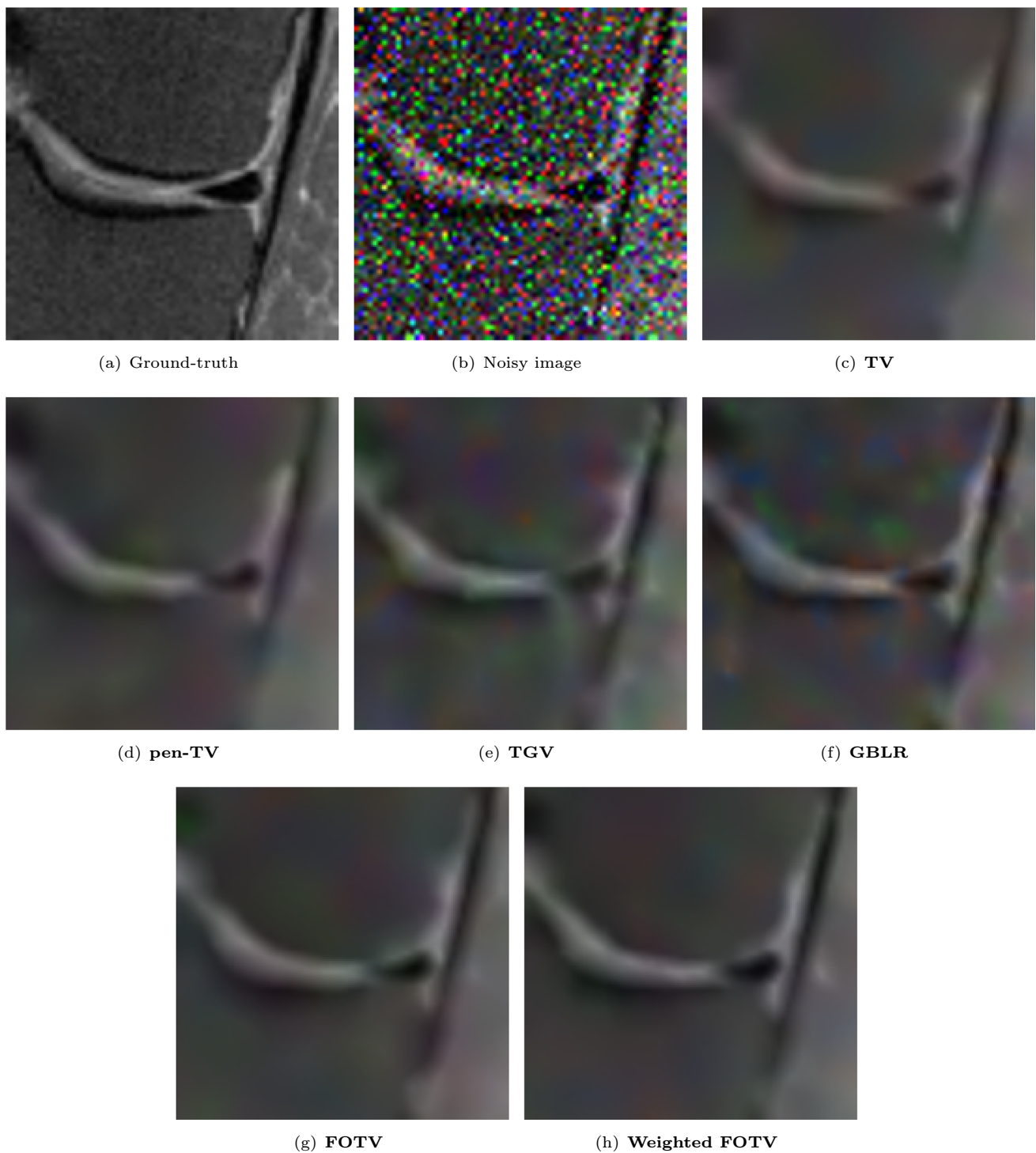


Fig. 9 From left to right: The zoomed regions for the noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 8

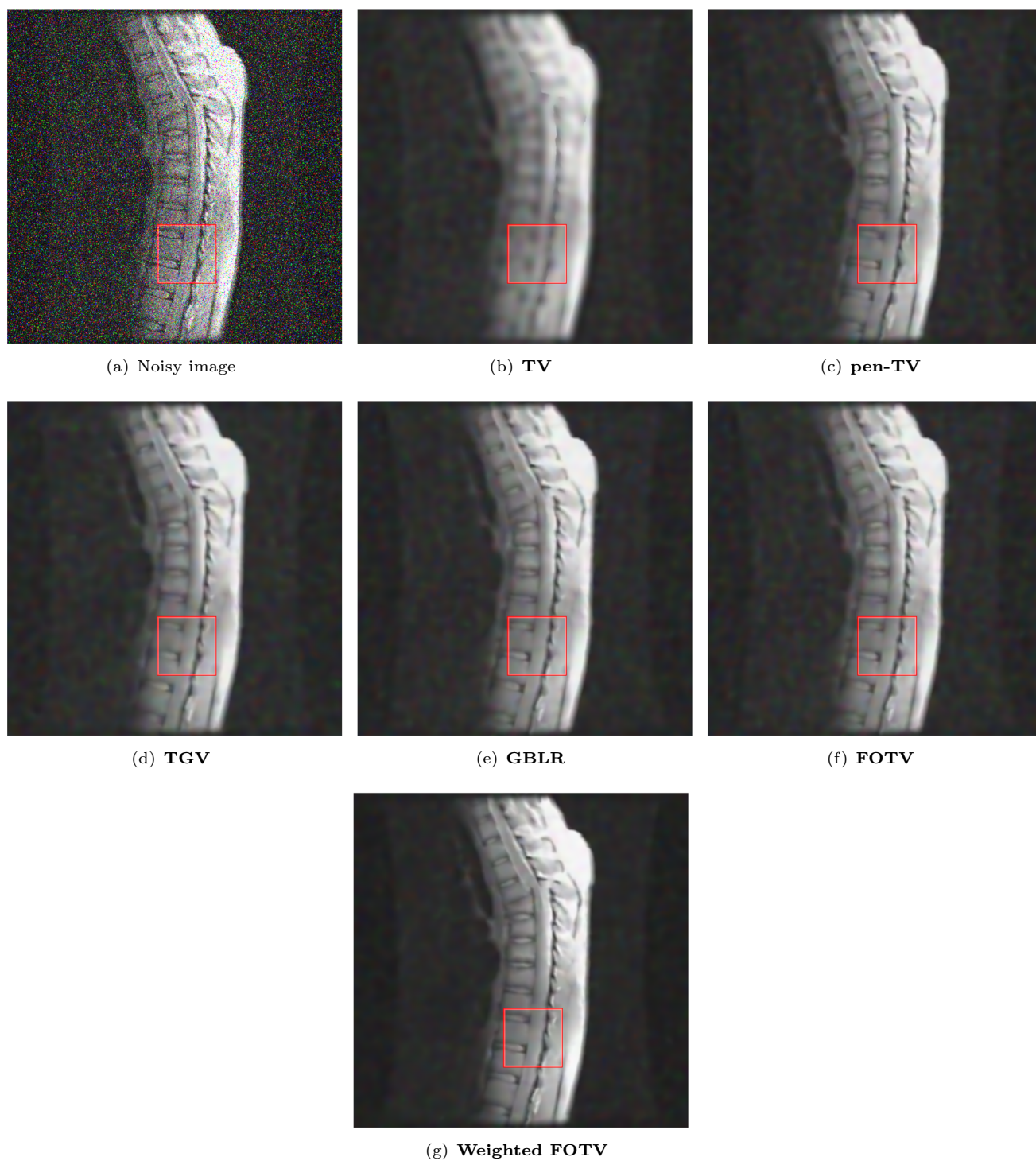


Fig. 10 From left to right: The noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 1d

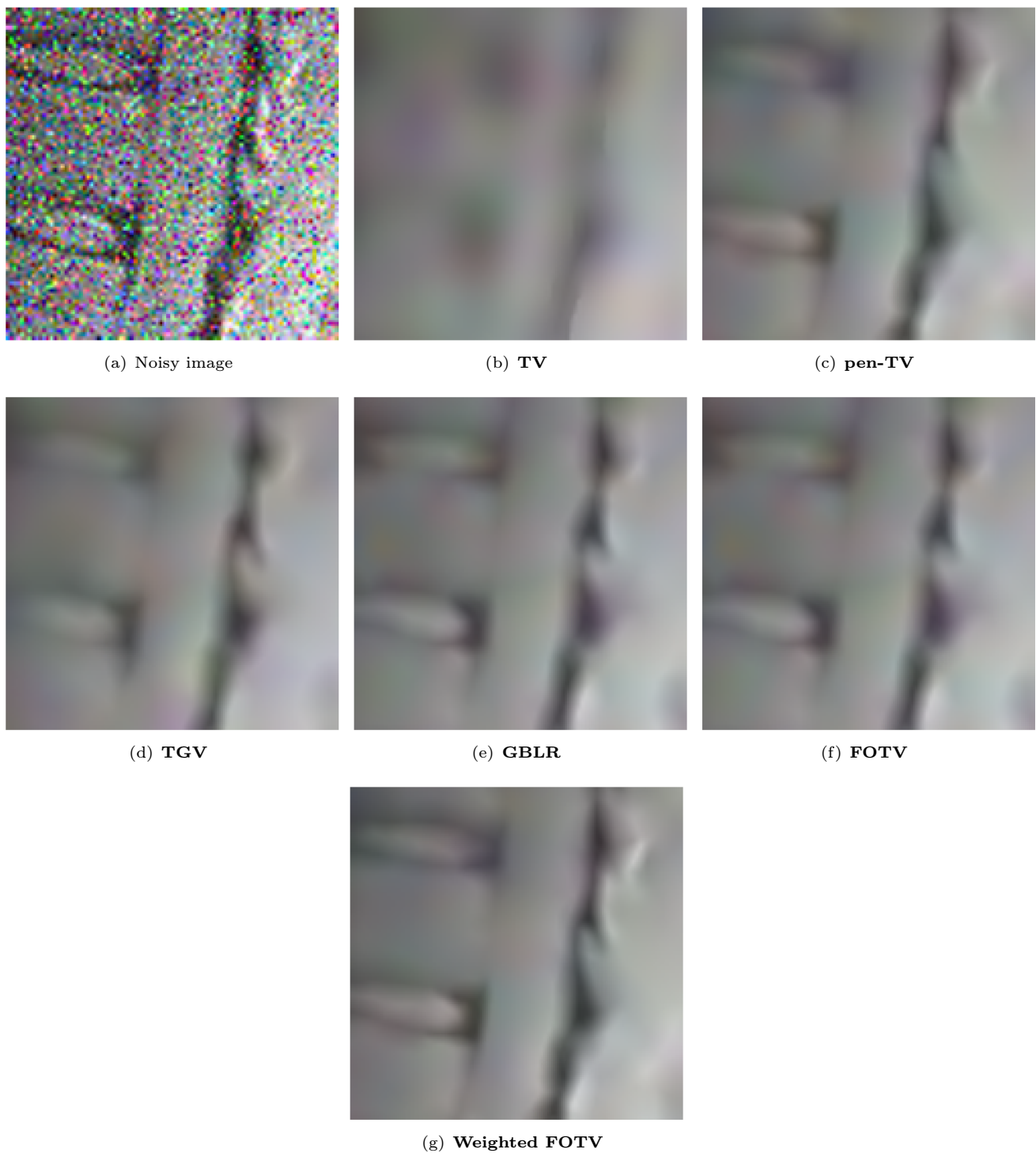


Fig. 11 From left to right: The zoomed region for the noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 10

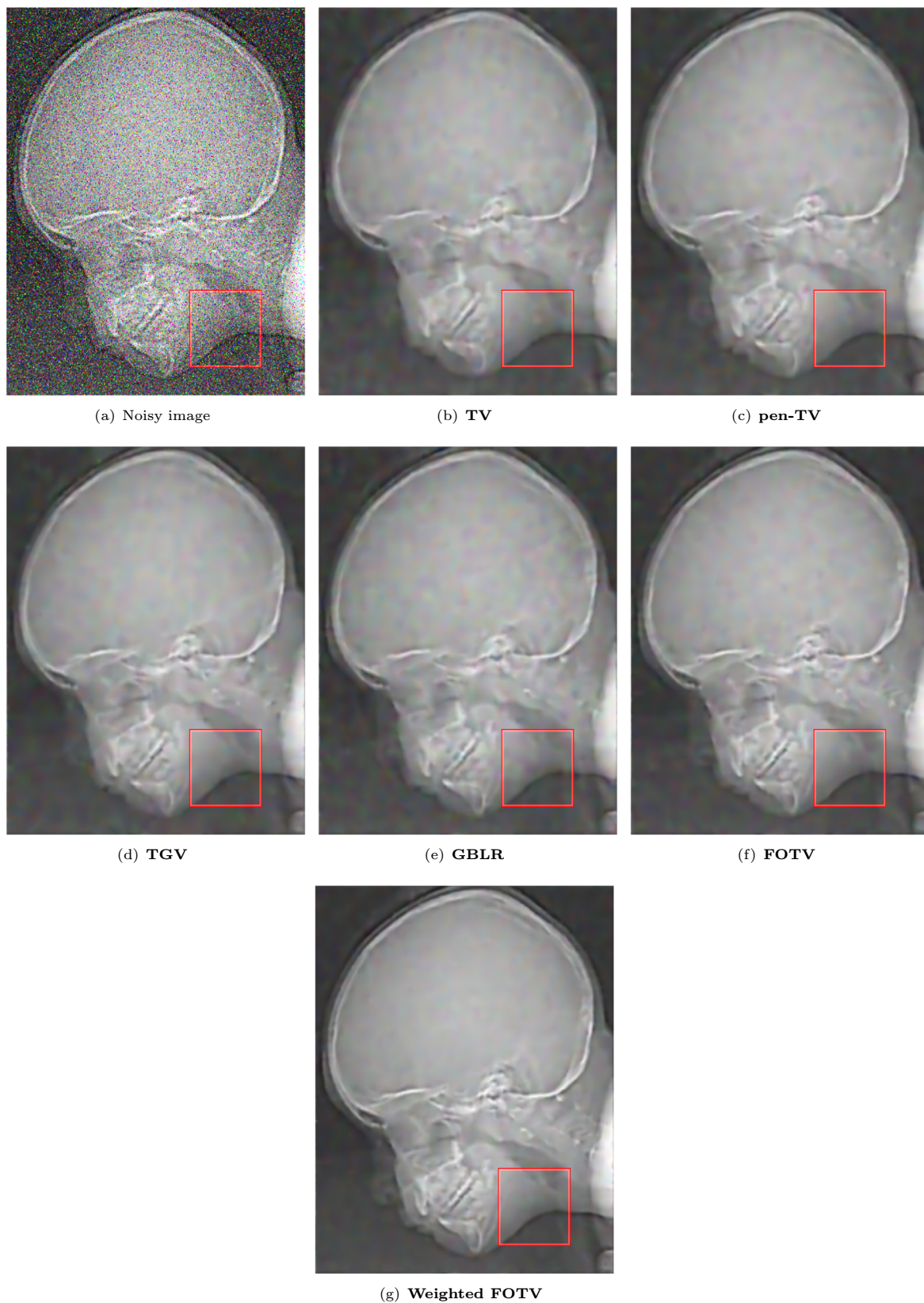


Fig. 12 From left to right: The noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for image in Fig. 1e

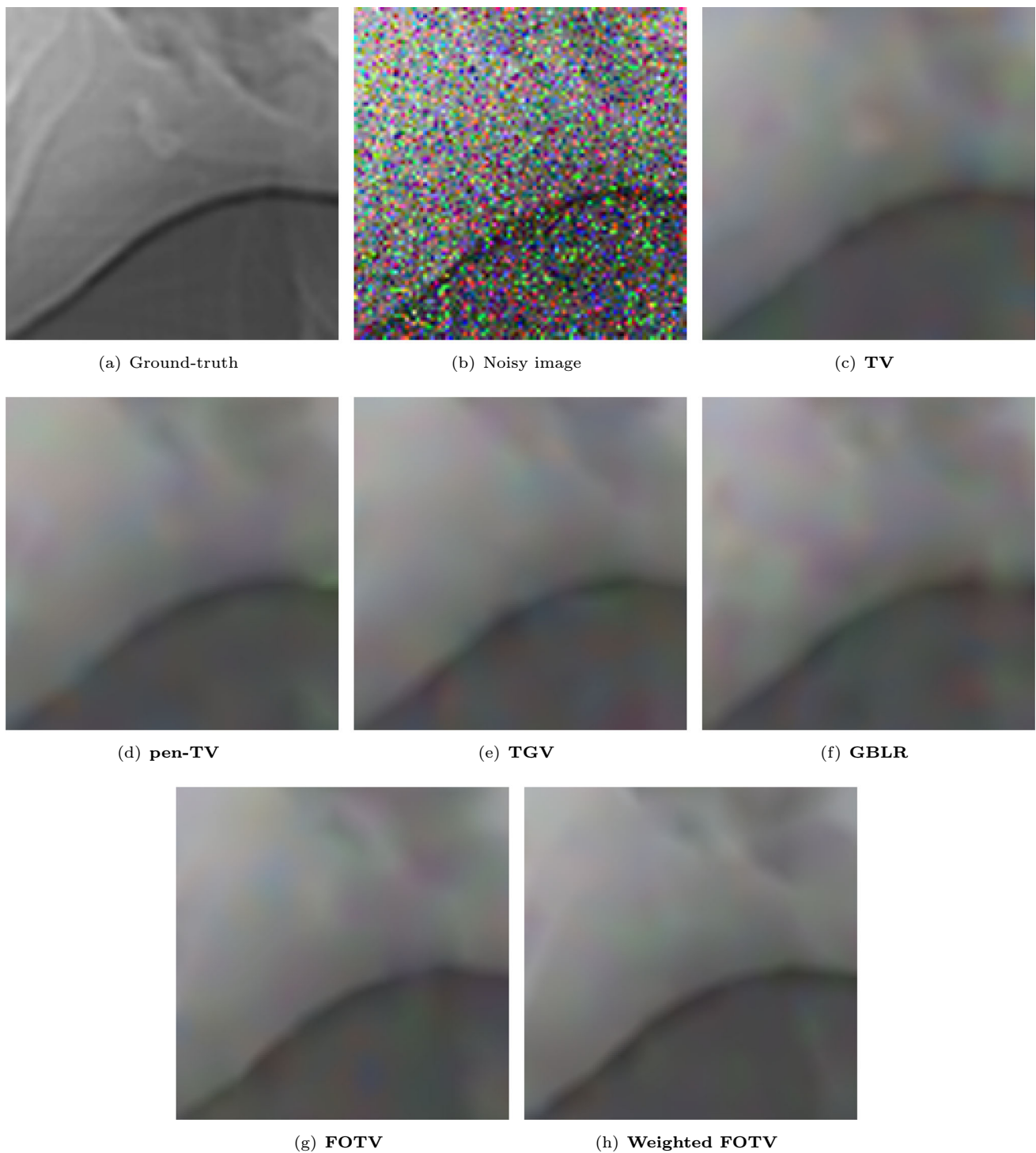


Fig. 13 From left to right: The zoomed regions for the noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 12

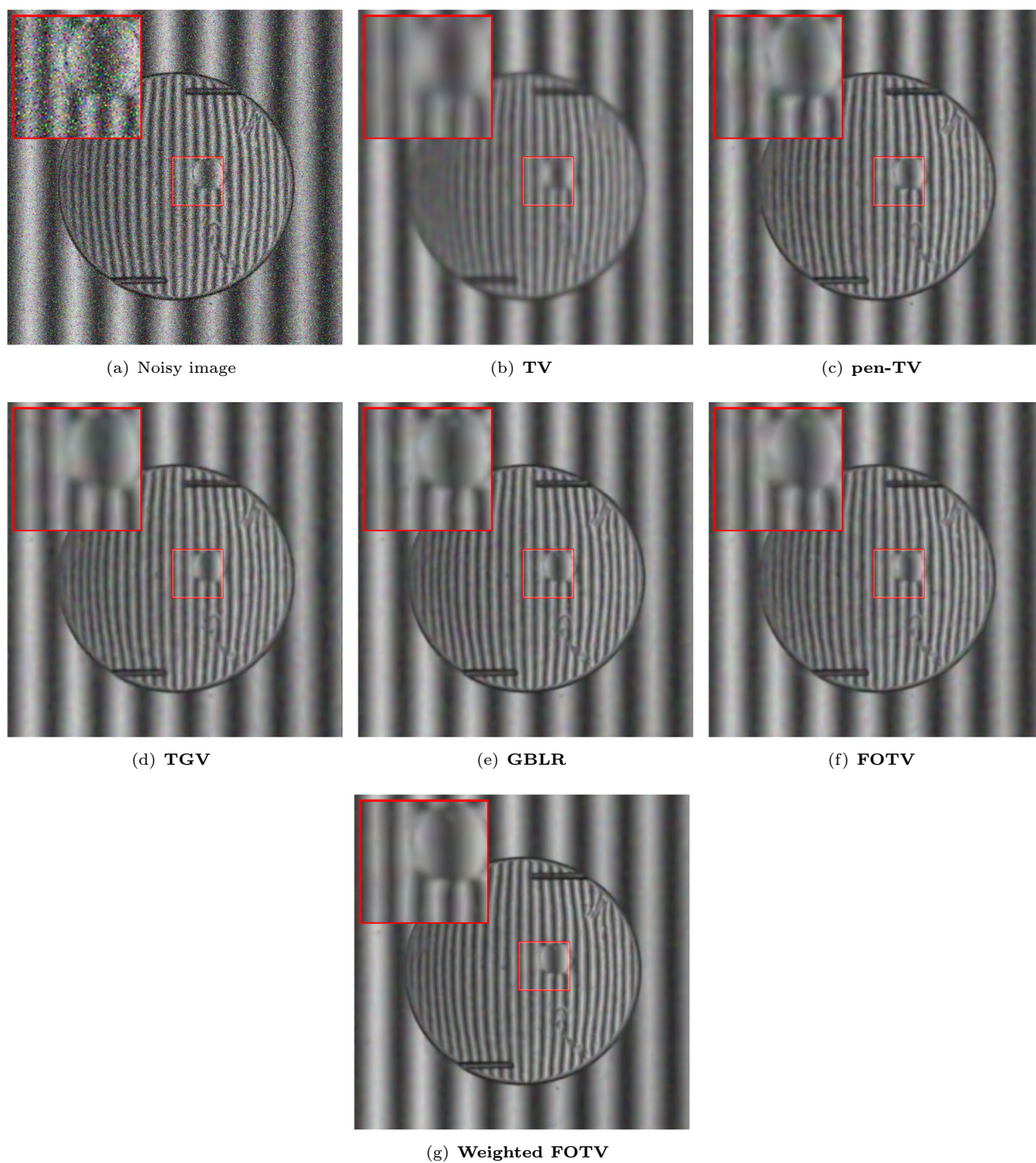


Fig. 14 From left to right: The noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for image in Fig. 1(f)

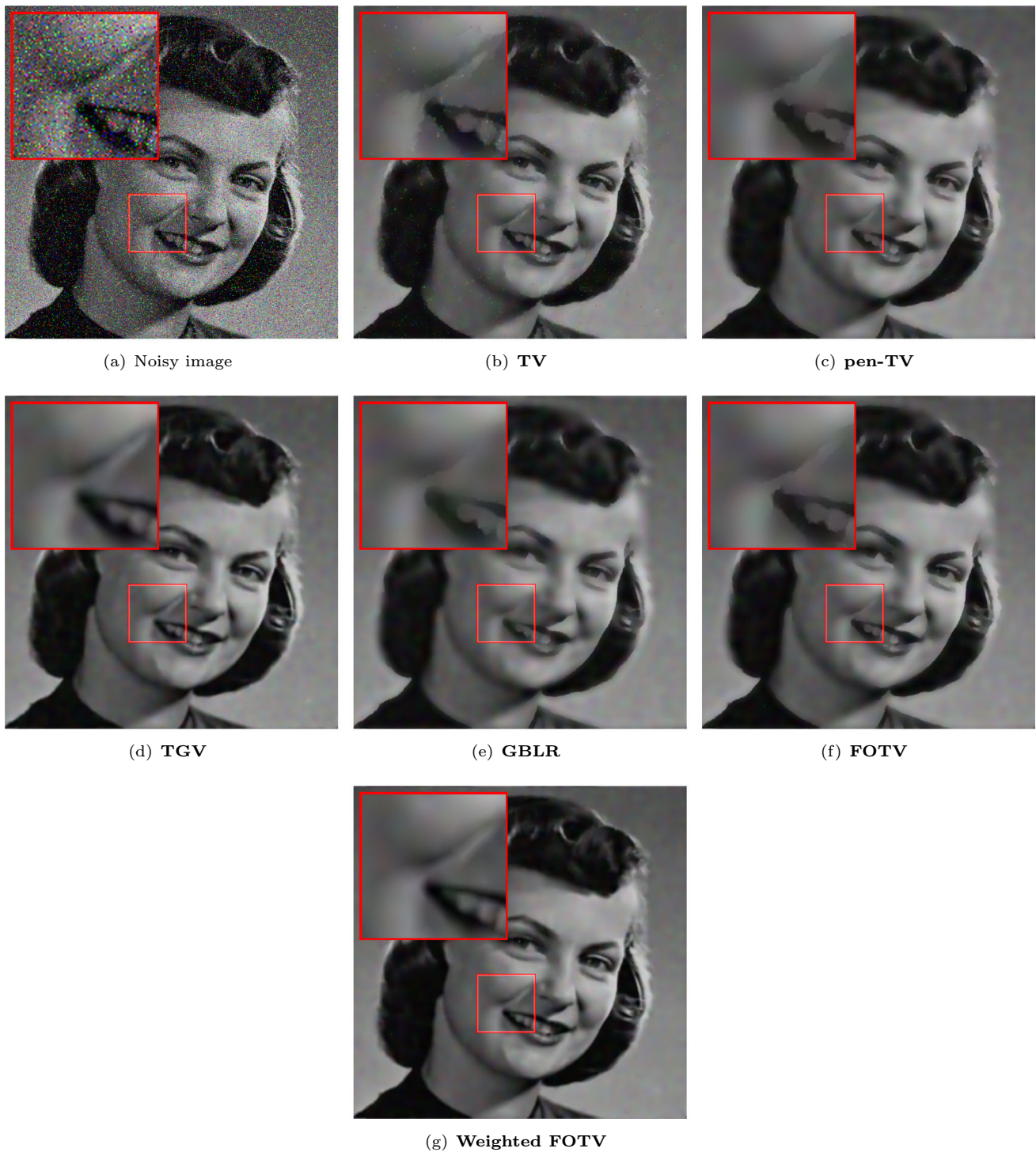


Fig. 15 From left to right: The noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 1(g)

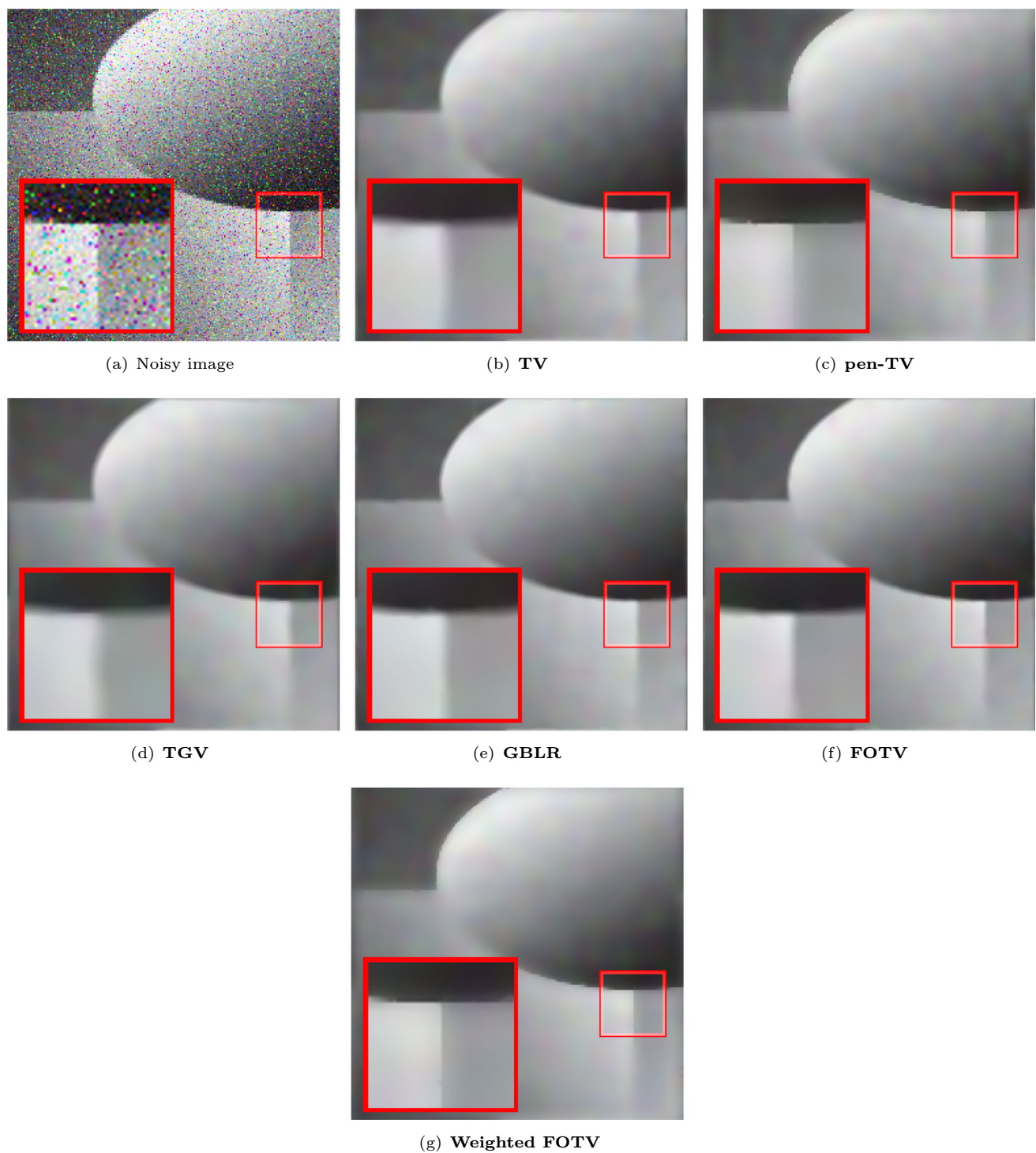


Fig. 16 From left to right: The noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 1h

Table 1 Quantitative comparison of denoising performance on the medical and natural grayscale images from Fig. 1(a)–1(h). Results are shown for noisy images and the denoising models at specific noise lev-els (γ) as function as SNR (dB), PSNR (dB), SSIM metrics. Best results are in **bold**, and second-best are underlined

Image	Metrics	Noisy	Models					FOTV $^\alpha_w$
			TV	pen-TV	TGV	GBLR	FOTV $^\alpha$	
Clock (256×256) $\gamma = 0.05$	SNR	12.887	22.372	24.039	24.00	23.542	<u>24.367</u>	25.129
	PSNR	15.2841	24.926	26.550	26.532	26.104	<u>26.878</u>	27.595
	SSIM	0.170	0.782	0.822	0.812	0.805	<u>0.824</u>	0.843
	Iterations	–	180	160	<u>140</u>	150	<u>140</u>	110
Tumor (400×420) $\gamma = 0.05$	SNR	7.117	15.300	15.265	15.635	<u>16.704</u>	16.136	17.729
	PSNR	15.262	24.424	24.392	24.767	<u>25.852</u>	25.286	26.884
	SSIM	0.150	0.742	0.739	0.748	<u>0.777</u>	0.759	0.808
	Iterations	–	140	140	<u>120</u>	100	140	100
MRI (512×512) $\gamma = 0.07$	SNR	6.176	15.899	16.736	16.786	<u>17.631</u>	16.817	20.710
	PSNR	13.944	24.879	25.68	25.738	<u>26.593</u>	25.762	29.701
	SSIM	0.070	0.741	0.763	0.768	<u>0.785</u>	0.764	0.840
	Iterations	–	200	180	<u>140</u>	130	150	130
Knee (249×249) $\gamma = 0.07$	SNR	5.139	13.883	14.299	14.610	15.283	<u>15.672</u>	16.451
	PSNR	13.963	24.316	24.799	25.096	25.830	<u>26.295</u>	27.141
	SSIM	0.108	0.790	0.802	0.801	0.814	<u>0.851</u>	0.882
	Iterations	–	170	160	130	160	130	130
Lens (590×614) $\gamma = 0.05$	SNR	9.367	21.350	22.697	24.110	<u>25.497</u>	24.138	27.123
	PSNR	15.316	27.893	29.226	30.621	<u>32.003</u>	30.644	33.622
	SSIM	0.12943	0.865	0.891	0.900	<u>0.918</u>	0.902	0.936
	Iterations	–	240	220	200	190	<u>150</u>	130
Skull (374×452) $\gamma = 0.07$	SNR	9.159	21.805	22.661	22.908	22.510	<u>23.761</u>	26.490
	PSNR	13.998	27.331	28.049	28.294	27.914	<u>29.131</u>	32.820
	SSIM	0.077	0.726	0.754	0.759	0.740	<u>0.784</u>	0.874
	Iterations	–	160	140	160	140	<u>120</u>	110
Girl (512×512) $\gamma = 0.04$	SNR	8.589	19.128	19.061	<u>20.686</u>	18.394	19.453	21.778
	PSNR	16.172	27.379	27.346	<u>28.959</u>	26.689	27.749	30.064
	SSIM	0.126	0.666	0.717	<u>0.749</u>	0.698	0.726	0.782
	Iterations	–	170	190	130	190	<u>110</u>	100
Synthetic (253×253) $\gamma = 0.04$	SNR	11.526	25.811	26.774	25.841	<u>27.695</u>	27.099	28.418
	PSNR	16.143	30.755	31.710	30.781	<u>32.604</u>	32.014	33.310
	SSIM	0.066	0.886	0.905	0.890	0.914	0.918	<u>0.916</u>
	Iterations	–	120	120	90	120	<u>100</u>	90

total variation model (FOTV $^\alpha_w$) preserves edges and structural details more effectively across all noise levels. A closer inspection of local image details is provided in Fig. 19. The zoomed regions highlight the ability of FOTV $^\alpha_w$ to maintain texture continuity and sharp edges while suppressing noise-related artifacts. This behavior remains consistent from low- to high-noise regimes, indicating that the optimized parameters are not overly sensitive to noise variations.

Quantitative results reported in Table 3 further support these observations. For all three noise levels, the proposed model achieves the highest PSNR, SNR, and SSIM values compared to competing methods. The performance gap

becomes more pronounced as the noise level increases, which suggests that the model is particularly robust in challenging high-noise scenarios. Figure 20 summarizes these trends by plotting the metrics as functions of the noise level, clearly showing stable and superior performance of FOTV $^\alpha_w$.

5.2 Color images

In this subsection, we generalize the proposed weighted FOTV model to color images. Let $\mathbf{u} : \Omega \rightarrow \mathbb{R}^{Ch}$ denote a color image with Ch channels, for RGB images, $Ch = 3$. We write $\mathbf{u} = (u_1, \dots, u_{Ch})$, where u_c is the c -th chan-

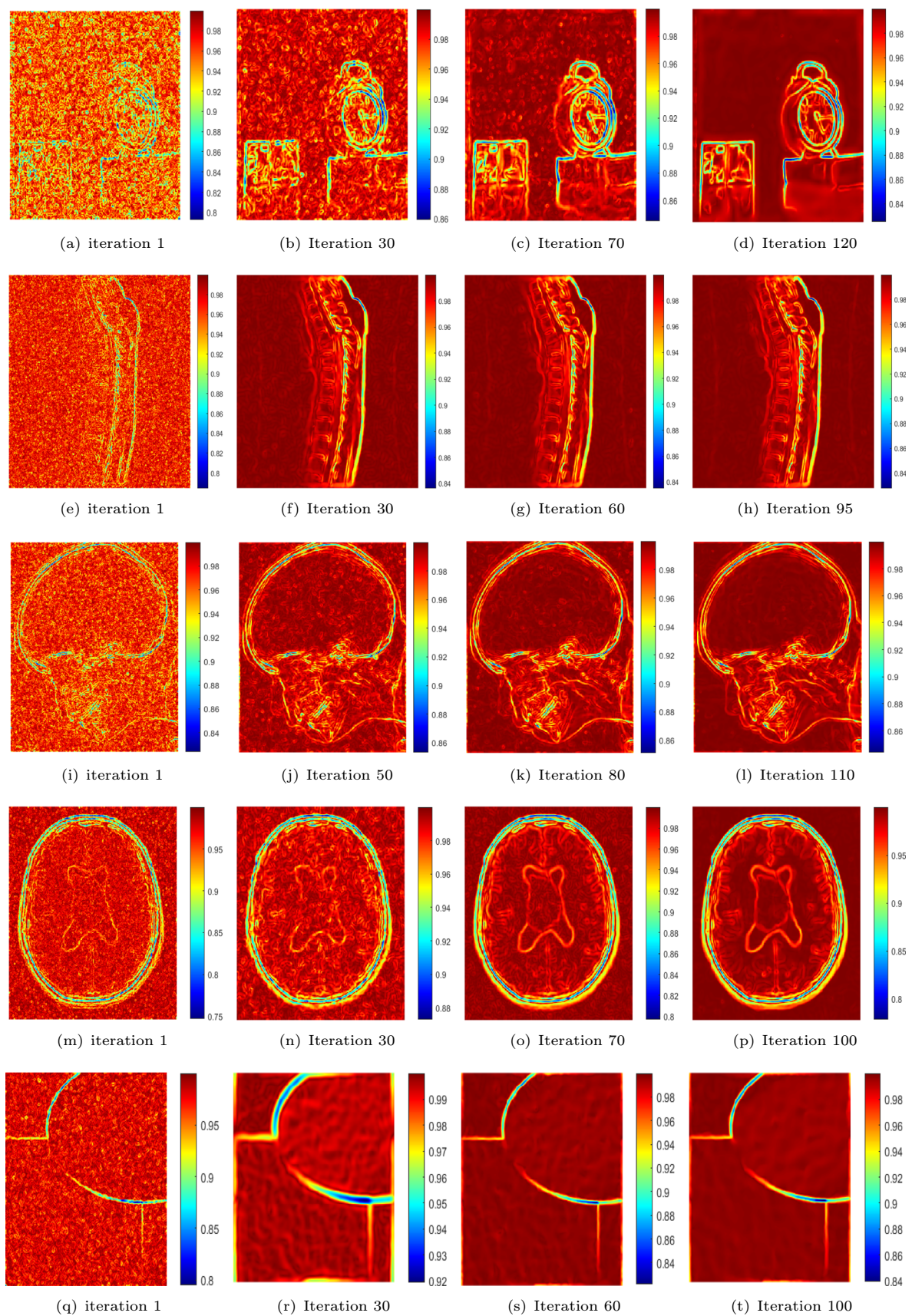


Fig. 17 From left to right: The weighted function w at different iterations

Table 2 Parameter tuning using Algorithm 2 for the five competing models. Each model is optimized over its corresponding parameter set Θ , within the given search ranges. Each model was tuned for 20 outer BO iterations, each invoking an inner SB iterations. The best parameter configuration Θ^* is selected according to the minimal objective value J_{obj}^*

Image	Model	Parameters	Best Θ^*	$ J_{\text{obj}}^* $	Iter. (BO)	Inner Iter. (SB)	Time (s)
Clock	TV	(λ, β)	(1, 0.15)	24.258	17	45	36
	pen-TV	(λ, β, μ)	(0.70, 0.09, 2.38)	25.131	15	40	67
	TGV	$(\lambda, \beta, \mu_1, \mu_2)$	(1.5, 0.08, 1.02, 0.09)	25.800	18	35	91
	GBLR	(λ, β)	(0.82, 0.11)	23.110	17	40	100
	FOTV	(λ, β, α)	(0.45, 0.02, 1.87)	<u>26.241</u>	12	30	41
	Proposed	(λ, β, α)	(0.50, 0.25, 1.64)	26.936	11	30	42
Tumor	TV	(λ, β)	(0.77, 0.12)	24.125	18	45	53
	pen-TV	(λ, β, μ)	(0.82, 0.11, 0.08)	24.126	16	40	59
	TGV	$(\lambda, \beta, \mu_1, \mu_2)$	(1.26, 0.07, 1.14, 0.12)	24.502	15	40	88
	GBLR	(λ, β)	(1.34, 0.11)	<u>25.131</u>	15	30	120
	FOTV	(λ, β, α)	(1.17, 0.07, 1.42)	24.949	15	30	46
	Proposed	(λ, β, α)	(1.51, 0.09, 1.31)	26.478	14	30	49
MRI	TV	(λ, β)	(0.50, 0.11)	24.555	17	30	74
	pen-TV	(λ, β, μ)	(0.68, 0.16, 0.04)	24.780	18	30	79
	TGV	$(\lambda, \beta, \mu_1, \mu_2)$	(0.79, 0.16, 0.08, 0.03)	24.640	18	30	86
	GBLR	(λ, β)	(0.43, 0.13)	<u>26.148</u>	19	30	130
	FOTV	(λ, β, α)	(1.23, 0.23, 1.61)	25.375	18	30	88
	Proposed	(λ, β, α)	(1.42, 0.19, 1.70)	28.865	16	30	91
Lens	TV	(λ, β)	(0.83, 0.14)	27.884	18	40	240
	pen-TV	(λ, β, μ)	(0.69, 0.08, 0.02)	28.772	18	40	305
	TGV	$(\lambda, \beta, \mu_1, \mu_2)$	(1.05, 0.20, 0.13, 0.05)	29.681	17	30	460
	GBLR	(λ, β)	(1.03)	<u>31.009</u>	18	30	490
	FOTV	(λ, β, α)	(1.08, 0.11, 1.50)	30.164	13	30	300
	Proposed	(λ, β, α)	(1.50, 0.09, 1.60)	32.615	11	30	303
Skull	TV	(λ, β)	(0.50, 0.12)	26.758	18	30	83
	pen-TV	(λ, β, μ)	(1, 0.19, 0.07)	27.290	18	30	110
	TGV	$(\lambda, \beta, \mu_1, \mu_2)$	(0.73, 0.11, 0.14, 0.09)	28.112	16	30	186
	GBLR	(λ, β)	(1.12, 0.22)	27.401	3	30	196
	FOTV	(λ, β, α)	(0.65, 0.20, 1.45)	<u>28.789</u>	13	30	115
	Proposed	(λ, β, α)	(0.60, 0.25, 1.68)	31.553	13	30	115
Synthetic	TV	(λ, β)	(0.79, 0.16)	29.988	17	35	30
	pen-TV	(λ, β, μ)	(0.70, 0.20, 0.03)	30.333	16	35	34
	TGV	$(\lambda, \beta, \mu_1, \mu_2)$	(0.74, 0.22, 0.15, 0.10)	29.963	16	30	56
	GBLR	(λ, β)	(0.83, 0.11)	31.646	18	30	65
	FOTV	(λ, β, α)	(0.65, 0.10, 1.47)	<u>30.647</u>	11	30	36
	Proposed	(λ, β, α)	(0.50, 0.07, 1.34)	32.212	9	30	33

nel. The observed noisy image is $\mathbf{f} : \Omega \rightarrow \mathbb{R}^{Ch}$ with the same structure. A natural extension of the model to the multi-channel case is to define a joint fractional gradient that couples both spatial directions and color channels. For the Caputo fractional directional derivatives, we define the vectorial fractional gradient

$$\nabla^{\alpha, C} \mathbf{u}(x) := (\nabla^{\alpha, C} u_1(x), \dots, \nabla^{\alpha, C} u_{Ch}(x)) \in \mathbb{R}^{2Ch},$$

where each $\nabla^{\alpha, C} u_c = ({}^C D_x^\alpha u_c, {}^C D_y^\alpha u_c)$.

The variational model for color images reads

$$\begin{aligned} \min_{\mathbf{u}: \Omega \rightarrow \mathbb{R}^{Ch}} \lambda \int_{\Omega} \sum_{c=1}^{Ch} \log \left(1 + \frac{(f_c(x) - u_c(x))^2}{\gamma^2} \right) dx \\ + \int_{\Omega} w(x) \sqrt{\sum_{c=1}^{Ch} |\nabla^{\alpha, C} u_c(x)|^2} dx. \end{aligned}$$

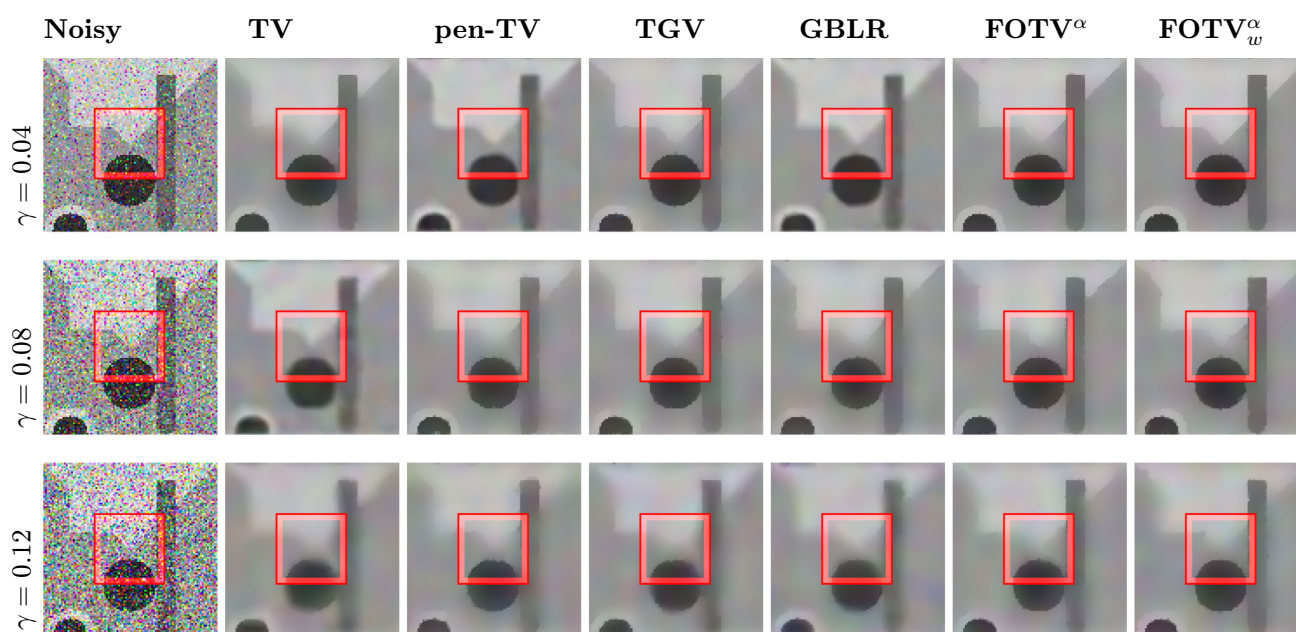


Fig. 18 Visual comparison of denoising results under different noise levels. Rows correspond to increasing noise levels $\gamma \in \{0.04, 0.08, 0.12\}$, and columns show the noisy image and the restored results obtained by TV, pen-TV, TGV, GBLR, FOTV $^\alpha$, and the proposed FOTV $^\alpha_w$ model

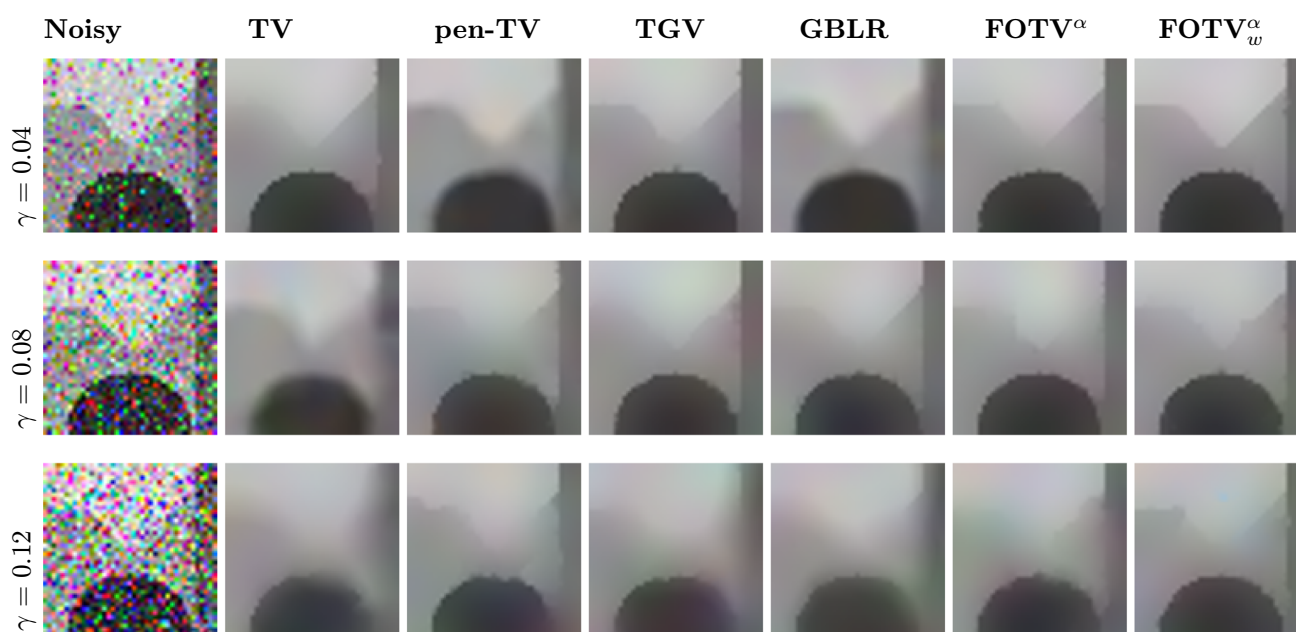


Fig. 19 Zoomed regions from Fig. 18. Detailed visual comparison showing fine details preservation across noise levels $\gamma \in \{0.04, 0.08, 0.12\}$ and different denoising models

where the second term is an isotropic vectorial weighted FOTV, coupling the spatial derivatives of all channels [17, 44, 45]. The stacked fractional gradient operator acting on all channels can be written as

$$\nabla^{\alpha, C} \mathbf{u} = \frac{h^{-\alpha}}{\Gamma(2-\alpha)} \begin{bmatrix} \mathbf{I}_{Ch} \otimes (\mathbf{I}_M \otimes \mathbf{A}_\alpha) \\ \mathbf{I}_{Ch} \otimes (\mathbf{B}_\alpha \otimes \mathbf{I}_N) \end{bmatrix} \mathbf{u} \in \mathbb{R}^{2ChNM},$$

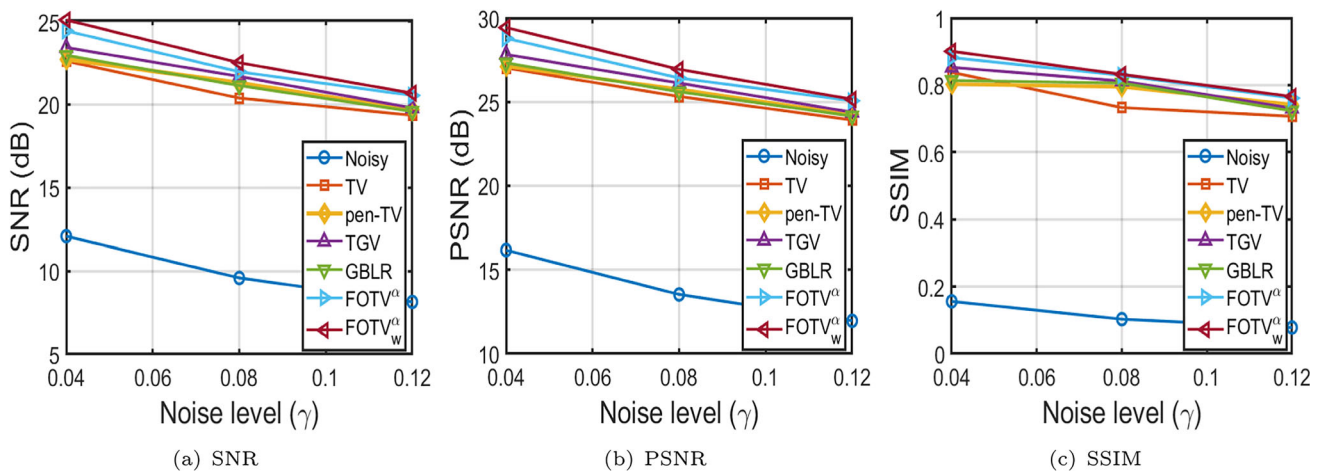
where $\mathbf{u}$ stacks all channel components column-wise.

The discrete color model becomes

$$\min_{\mathbf{u} \in \mathbb{R}^{N \times M \times Ch}} \lambda \sum_{c=1}^{Ch} \sum_{i=1}^N \sum_{j=1}^M \log \left(1 + \frac{(\mathbf{F}_{i,j,c} - \mathbf{U}_{i,j,c})^2}{\gamma^2} \right) + \sum_{i=1}^N \sum_{j=1}^M \mathbf{w}_{i,j,c} \sqrt{\sum_{c=1}^{Ch} \|\nabla^{\alpha, C} U_{i,j,c}\|_2^2}.$$

Table 3 Comparison of denoising results on grayscale image (in Fig. 18) under varying noise levels γ . The best results are in **bold**, and the second-best are underlined

γ	Metrics	Noisy	Models					
			TV	pen-TV	TGV	GBLR	FOTV ^{α}	FOTV ^{α} _w
$\gamma = 0.04$	SNR	12.081	22.606	22.732	23.425	22.954	<u>24.428</u>	25.083
	PSNR	16.168	27.021	27.127	27.830	27.322	<u>28.794</u>	29.432
	SSIM	0.156	0.838	0.803	0.853	0.814	<u>0.882</u>	0.901
$\gamma = 0.08$	SNR	9.586	20.386	21.296	21.687	21.138	<u>21.967</u>	22.510
	PSNR	13.534	25.320	25.750	26.122	25.606	<u>26.424</u>	26.940
	SSIM	0.103	0.733	0.795	0.812	0.806	<u>0.829</u>	0.832
$\gamma = 0.12$	SNR	8.143	19.349	19.654	19.792	19.581	<u>20.542</u>	20.680
	PSNR	11.961	23.922	24.232	24.379	24.152	<u>25.063</u>	25.154
	SSIM	0.078	0.707	0.742	0.731	0.723	<u>0.761</u>	0.766

**Fig. 20** Comparison of denoising performance across noise levels: **a** SNR, **b** PSNR, and **c** SSIM metrics versus noise level γ for all models in Table 3

Notice that the resulting optimization problem remains separable in the **U**- and **d**-subproblems. The SB iterations are extended to the multi-channel case by promoting the auxiliary variables $\mathbf{d}_{i,j} \in \mathbb{R}^{2Ch}$ and $\mathbf{b}_{i,j} \in \mathbb{R}^{2Ch}$ at each pixel. The **d**-update becomes an isotropic vector shrinkage across both spatial directions and color channels [46], which promotes simultaneous sparsity of fractional gradients across all RGB channels.

We now evaluate the proposed weighted FOTV model on RGB images. In contrast to grayscale denoising, the color case requires treating the three channels jointly to avoid color shifts and spatial misalignment of edges.

Figures 21 and 23 show denoising results for the Splash and Bird images heavily degraded by Cauchy noise. Across both images, TV and penalized-TV remove noise but tend to produce desaturation, texture loss, and visible color flattening, especially in smooth regions. TGV alleviates some staircasing but remains sensitive to impulsive outliers, leading to residual blotches in high-frequency regions. The GBLR model, which groups similar patches into low-rank representations, demonstrates effective noise suppression in

homogeneous areas but can introduce blurring across color edges and subtle chromatic textures, particularly in regions with fine structural variations. The standard FOTV model provides improved texture recovery thanks to the nonlocal character of fractional derivatives, yet it still applies a uniform degree of regularization across the image. In contrast, the proposed weighted FOTV significantly enhances visual quality in both examples: Edges remain crisp and well aligned across channels, chromatic structures are preserved without false color bleeding, and fine textures are reconstructed more faithfully. This improvement is particularly noticeable in the feathers of the Bird image and the smooth color transitions in Splash, where our adaptive regularizer suppresses strong outliers while maintaining color consistency.

To better understand the mechanism of the model, Fig. 25 illustrates the evolution of the adaptive weight w over several SB iterations. Initially, the weight map is highly perturbed because it is computed from the noisy gradients. As the iterations proceed, impulsive noise is progressively attenuated and the weight adapts to the emerging geo-

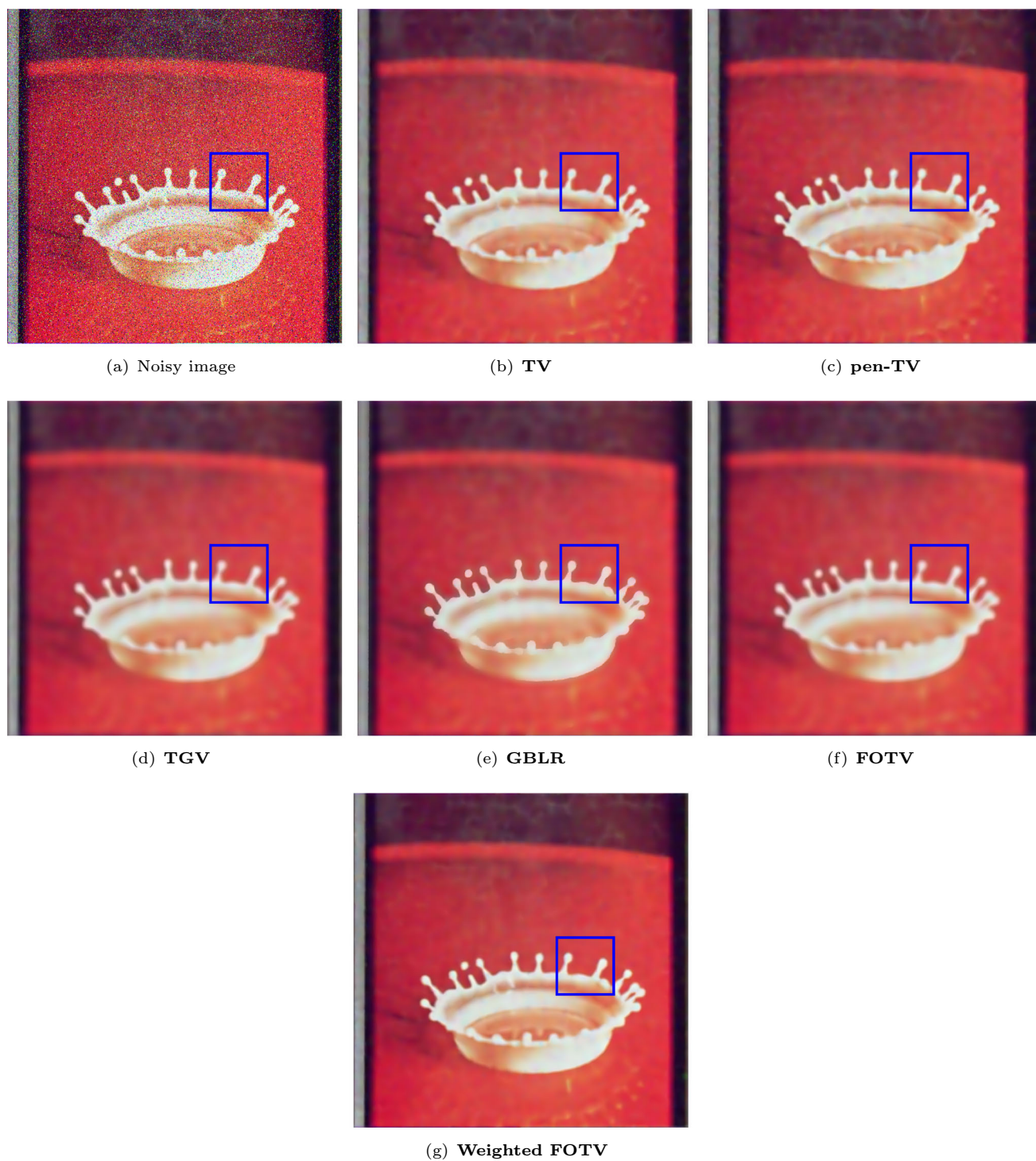


Fig. 21 From left to right: The noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 1j

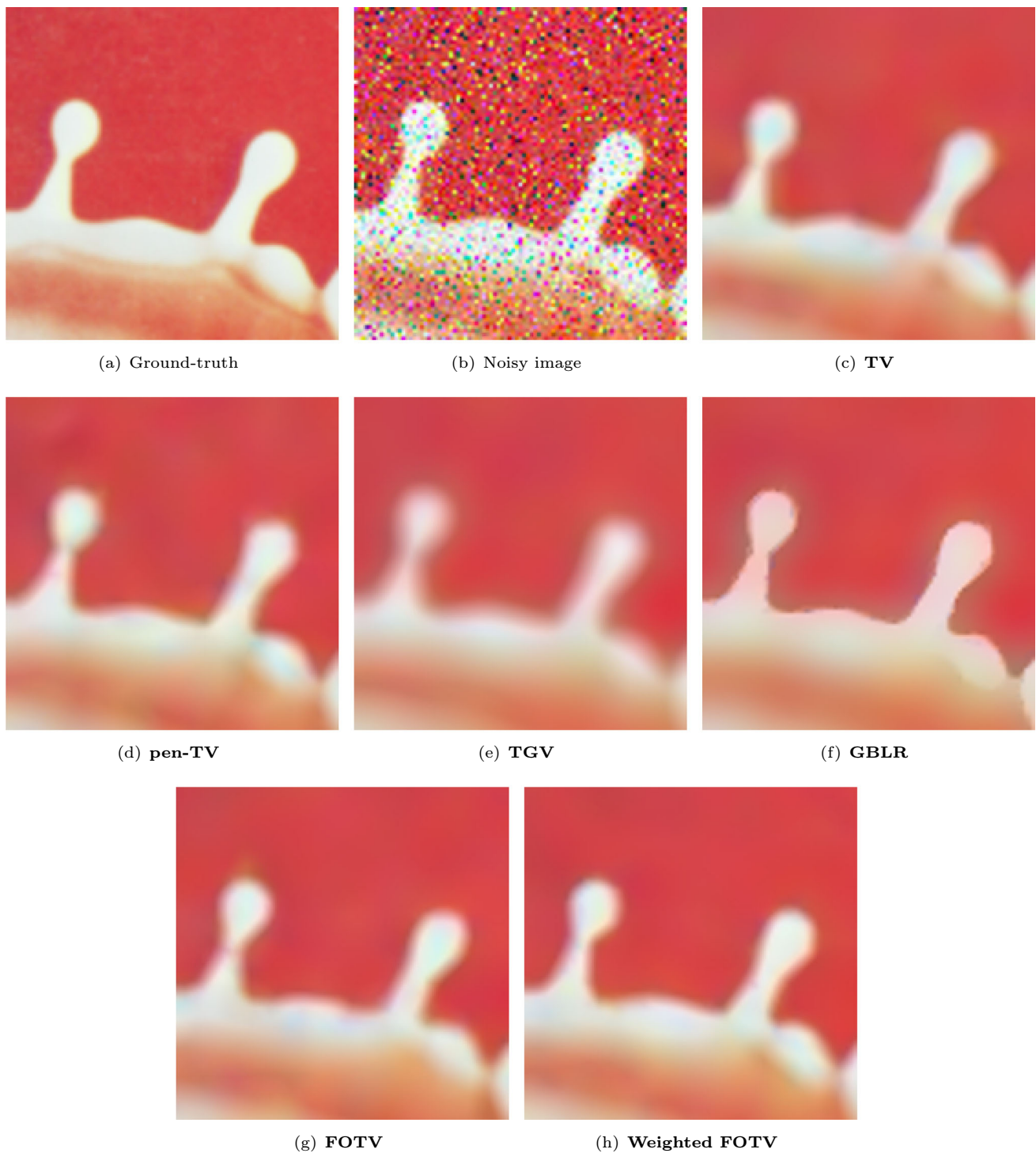


Fig. 22 From left to right: The zoomed regions of the noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 21



Fig. 23 From left to right: The noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for image in Fig. 1(k)

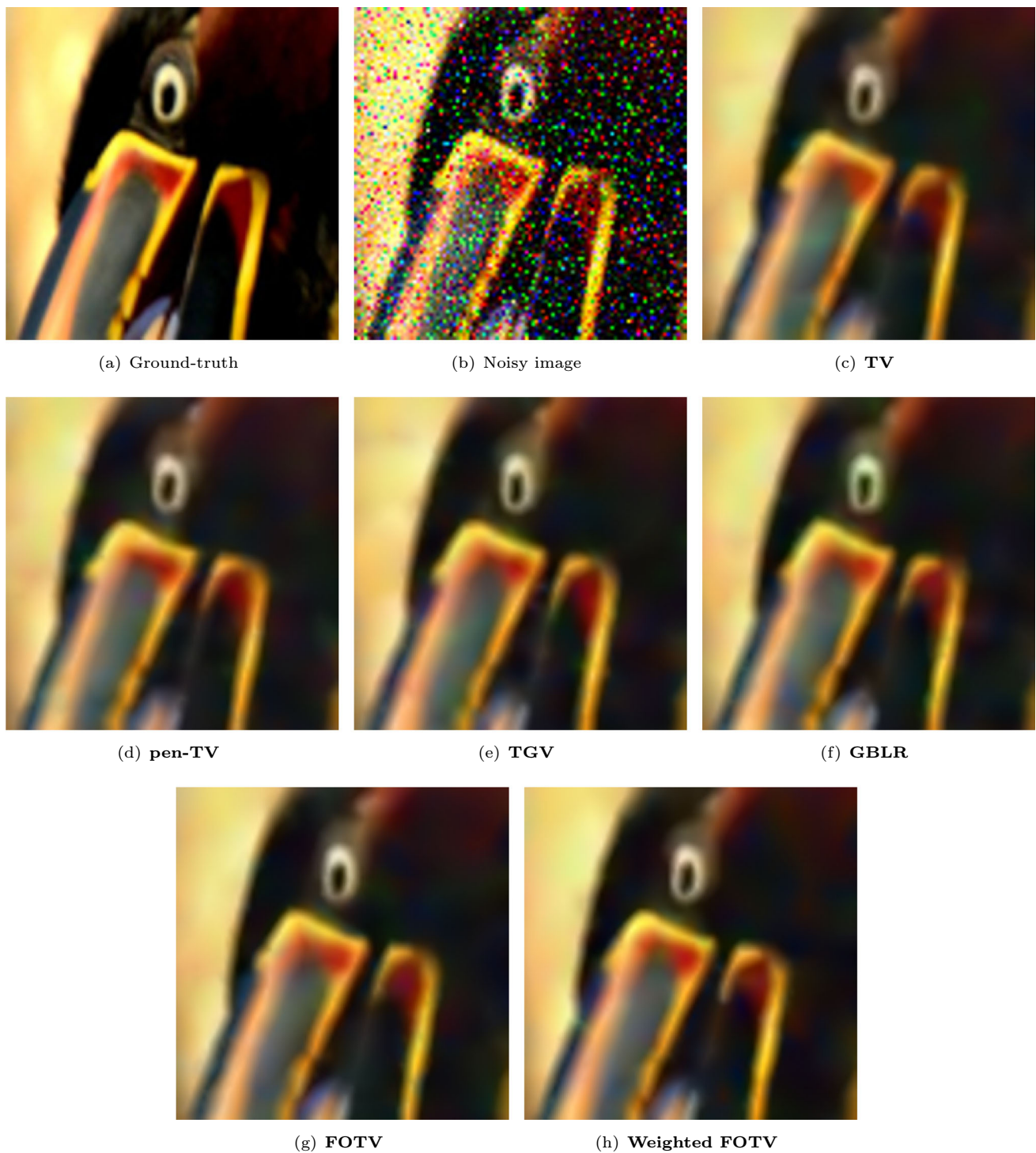


Fig. 24 From left to right: The noisy image degraded by Cauchy noise and the restored images by TV, pen-TV, TGV, GBLR, FOTV, and weighted FOTV models, respectively, for images in Fig. 23

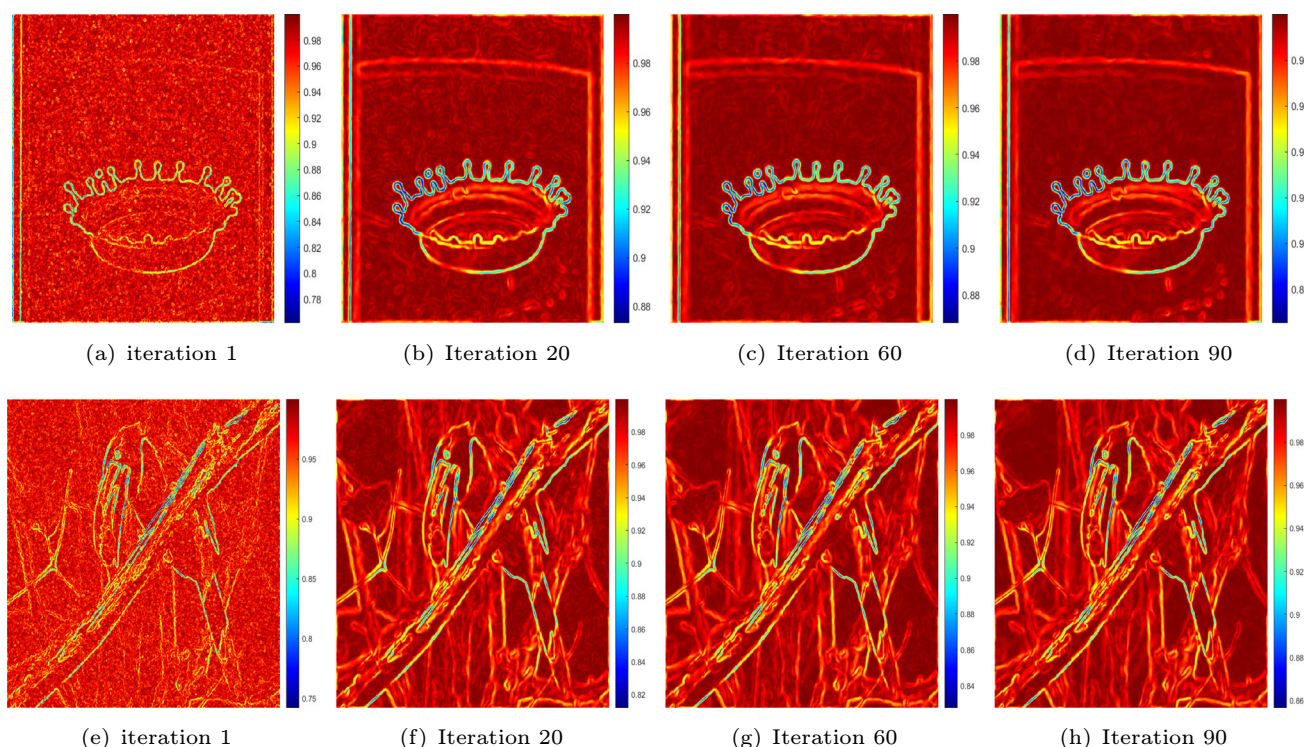


Fig. 25 From left to right: The weighted function w at different iterations.

Table 4 Table for comparison of results for color images in Figs. 21 and 23

Images	Metrics	Noisy	Models				FOTV $^\alpha$	FOTV $^\alpha_w$
			TV	pen-TV	TGV	GBLR		
Splash (512×512×3) $\gamma = 0.05$	SNR	9.826	21.111	21.206	19.497	20.852	<u>21.645</u>	22.832
	PSNR	15.258	27.035	27.126	25.439	26.776	<u>27.551</u>	28.697
	SSIM	0.619	0.949	0.949	0.931	0.944	<u>0.953</u>	0.961
	Iterations	–	150	140	<u>120</u>	150	<u>120</u>	90
Bird (640×640×3) $\gamma = 0.07$	SNR	9.221	18.608	18.554	19.267	19.079	<u>19.284</u>	20.070
	PSNR	13.927	23.877	24.506	23.828	24.327	<u>24.522</u>	25.269
	SSIM	0.506	0.873	0.873	<u>0.882</u>	0.880	<u>0.882</u>	0.892
	Iterations	–	150	180	130	120	<u>115</u>	90

metric structures of the underlying image. High weights concentrate around salient boundaries, indicating regions where stronger regularization is required to stabilize large fractional gradients, while the interior of smooth and homogeneous regions is assigned smaller weights. The final weight map is well aligned with the true edges and color transitions of the restored image, demonstrating the effectiveness of the adaptive mechanism in guiding the fractional regularization.

Table 4 provides a quantitative comparison using SNR, PSNR, and SSIM. For both test images, the proposed weighted FOTV achieves the highest values across all three

metrics, confirming the qualitative observations. Compared to standard FOTV, the adaptive model yields an average improvement of 1.0–4.0 dB in PSNR and similar gains in SNR. The method also requires fewer SB iterations to converge, thanks to the structural guidance provided by the adaptive weight. These results show that the interaction between joint fractional derivatives and spatially adaptive weighting produces a robust color-denoising framework that is simultaneously edge-preserving, texture-aware, and resilient to heavy-tailed noise Figs. 22, 24.

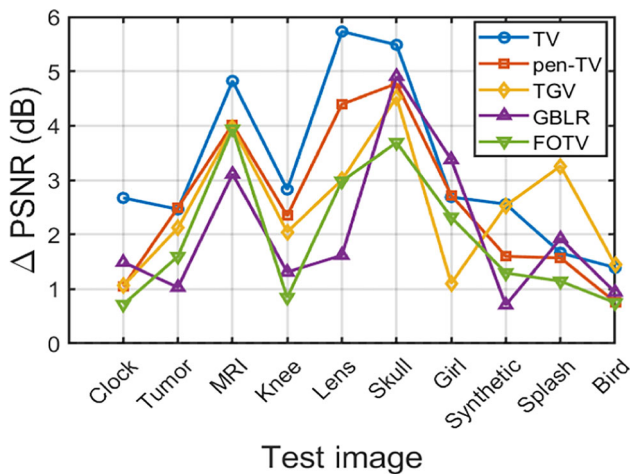


Fig. 26 Paired PSNR differences: $FOTV_w$ versus competing methods in Table 1 and Table 4

5.3 Statistical analysis improvements

To evaluate whether the observed performance differences in Tables 1 and 4 reflect consistent improvements rather than random variation across test images, we conducted a statistical significance analysis on the per-image results obtained by the competing models. Since each denoising method was applied to the same set of test images (the eight grayscale and the two color images in Tables 1 and 4), the resulting performance measurements form matched pairs, making paired hypothesis testing appropriate [47, 48]. In this context, standard statistical techniques such as the paired Student's t test and its nonparametric counterpart, the Wilcoxon signed-rank test, are widely used for paired comparisons when evaluating image quality metrics across methods [49, 50].

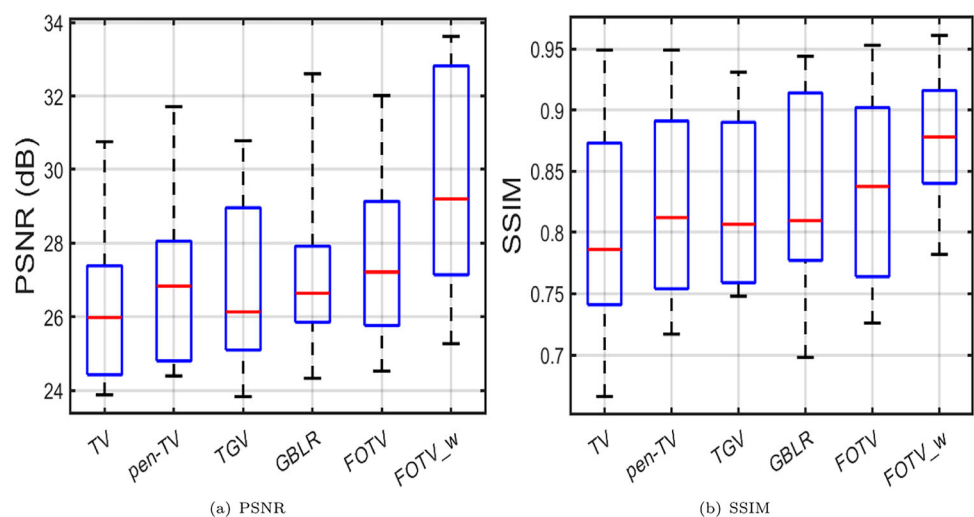
Let X_i and Y_i denote the PSNR values obtained on image i by two competing models, for $i = 1, \dots, n$. Under the null hypothesis that there is no systematic difference in PSNR

between the two models, the distribution of the paired differences $D_i = X_i - Y_i$ is symmetric about zero. The Wilcoxon signed-rank test is a nonparametric rank test well suited for this setting because it does not require the assumption of normality for the D_i and provides robust inference with moderate sample sizes [47, 51]. The paired t test can also be applied when normality of the differences is plausible; however, the nonparametric signed-rank test was adopted here as the preferred primary test due to its broader applicability and common use in image processing studies [47].

We applied the Wilcoxon signed-rank test to compare the proposed $FOTV_w^\alpha$ model against TV, pen-TV, TGV, and the unweighted $FOTV^\alpha$ model using PSNR, SNR, and SSIM as evaluation metrics. Statistical significance was assessed at the 5% level. For all pairwise comparisons and all three metrics, the resulting p values were equal 0.0020 for the Wilcoxon signed-rank test and it is between 0.0001 and 0.001, indicating that the improvements achieved by the proposed method are statistically significant. To further illustrate the paired nature of the statistical tests, Fig. 26 depicts the per-image PSNR differences between $FOTV_w^\alpha$ model and each competing model. All differences are positive across images, visually confirming the systematic improvement quantified by the Wilcoxon signed-rank tests.

To complement the hypothesis testing, Fig. 27 reports box-plots of PSNR, SNR, and SSIM values across all test images. Figure 27 illustrates the distribution of the three metrics values across all test images for the competing methods. The proposed $FOTV_w^\alpha$ model achieves the highest median of PSNR, SNR, and SSIM among all approaches and exhibits a relatively compact interquartile range compared to the baseline methods. This indicates that the proposed method not only improves average reconstruction quality but also delivers more stable performance across different images. In contrast, classical TV-based methods show lower medians and larger dispersion, reflecting greater sensitivity to image

Fig. 27 PSNR and SSIM distribution across grayscale and color test images in Table 1 and Table 4



content and noise characteristics. These results confirm that the quantitative gains reported in Tables 1 and 4 are statistically robust and consistent across images.

6 Conclusions

In this work, we introduced a new variational image restoration model that combines a Caputo weighted FOTV regularizer with an edge-adaptive weighting strategy. The proposed weighted FOTV formulation preserves the desirable nonlocal smoothing effects of fractional derivatives while dynamically adjusting the regularization strength according to local image structures. This mechanism allows the model to suppress Cauchy noise efficiently while retaining fine geometric details and sharp edges. A SB scheme was developed to solve the resulting nonsmooth optimization problem. A Bayesian optimization procedure was further employed to identify the optimal fractional and weighting parameters, providing a principled and reproducible parameter-selection strategy. Extensive experiments on grayscale and color images demonstrate the effectiveness of the approach. The weighted FOTV model systematically achieves higher SNR, PSNR, and SSIM scores than classical TV, penalized TV, TGV, and standard FOTV. Several research directions naturally emerge from this study. A first extension concerns multi-phase regularization schemes. We plan to combine the proposed weighted FOTV with a second regularizer such as TGV or penalized TV may lead to a two-phase or double-layer model capable of simultaneously capturing piecewise smooth regions and fine-scale textures. A second direction is the design of multi-objective Bayesian optimization, where a vector-valued objective that jointly maximizes PSNR and SSIM, to enable a more perceptually consistent selection of optimal parameters. Another important direction concerns unsupervised parameter selection in situations where no ground-truth image is available.

Acknowledgements The author would like to thank the anonymous reviewers for their careful reading of the manuscript and for their constructive comments and suggestions. Their insightful feedback has significantly improved the clarity, rigor, and overall quality of this work.

Author Contributions H.B. wrote the main manuscript text, prepared all figures, and reviewed the manuscript.

Funding Not applicable.

Data Availability Not applicable.

Declarations

Conflict of interest The author states that the publishing of this paper does not include any conflict of interest.

Ethics approval and consent to participate Not applicable.

References

1. Getreuer, P.: Rudin-Osher-Fatemi total variation denoising using split Bregman. *Image Processing On Line* **2**, 74–95 (2012). <https://doi.org/10.5201/ipol.2012.g-tvd>
2. Goldstein, T., Osher, S.: The split Bregman method for L1-regularized problems. *SIAM J. Imag. Sci.* **2**(2), 323–343 (2009). <https://doi.org/10.1137/080725891>
3. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60**(1–4), 259–268 (1992). [https://doi.org/10.1016/0167-2789\(92\)90242-F](https://doi.org/10.1016/0167-2789(92)90242-F)
4. Hamza, A.B., Krim, H.: A variational approach to maximum a posteriori estimation for image denoising. In: *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 19–34. Springer (2001). <https://doi.org/10.1007/3-540-44745-82>
5. Hamza, A.B., Krim, H., Unal, G.B.: Unifying probabilistic and variational estimation. *IEEE Signal Process. Mag.* **19**(5), 37–47 (2002). <https://doi.org/10.1109/MSP.2002.1028351>
6. Sciacchitano, F., Dong, Y., Zeng, T.: Variational approach for restoring blurred images with Cauchy noise. *SIAM J. Imag. Sci.* **8**(3), 1894–1922 (2015). <https://doi.org/10.1137/140997816>
7. Mei, J.-J., Dong, Y., Huang, T.-Z., Yin, W.: Cauchy noise removal by nonconvex ADMM with convergence guarantees. *J. Sci. Comput.* **74**(2), 743–766 (2018). <https://doi.org/10.1007/s10915-017-0460-5>
8. Zhang, J., Chen, K.: A total fractional-order variation model for image restoration with nonhomogeneous boundary conditions and its numerical solution. *SIAM J. Imag. Sci.* **8**(4), 2487–2518 (2015). <https://doi.org/10.1137/14097121X>
9. Ren, Z., He, C., Zhang, Q.: Fractional order total variation regularization for image super-resolution. *Signal Process.* **93**(9), 2408–2421 (2013). <https://doi.org/10.1016/j.sigpro.2013.02.015>
10. Gilboa, G., Osher, S.: Nonlocal linear image regularization and supervised segmentation. *Multiscale Model. Simul.* **6**(2), 595–630 (2007). <https://doi.org/10.1137/060669358>
11. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. *Multiscale Model. Simul.* **7**(3), 1005–1028 (2009). <https://doi.org/10.1137/070698592>
12. Peyré, G., Bougleux, S., Cohen, L.: Non-local regularization of inverse problems. In: *European Conference on Computer Vision*, pp. 57–68. Springer (2008). https://doi.org/10.1007/978-3-540-88690-7_5
13. Kindermann, S., Osher, S., Jones, P.W.: Deblurring and denoising of images by nonlocal functionals. *Multiscale Model. Simul.* **4**(4), 1091–1115 (2005). <https://doi.org/10.1137/050622249>
14. Li, M., Bi, S., Cai, G.: An adaptive fractional-order regularization primal-dual image denoising algorithm based on non-convex function. *Appl. Math. Model.* **131**, 67–83 (2024). <https://doi.org/10.1016/j.apm.2024.04.001>
15. Xie, D., Wang, X., Wang, C., Yuan, K., Wei, X., Liu, X., Huang, T.: A fractional-order total variation regularization-based method for recovering Geiger-mode avalanche photodiode light detection and ranging depth images. *Fractal Fract.* **7**(6), 445 (2023). <https://doi.org/10.3390/fractalfract7060445>
16. Yao, W., Shen, J., Guo, Z., Sun, J., Wu, B.: A total fractional-order variation model for image super-resolution and its SAV algorithm. *J. Sci. Comput.* **82**(3), 81 (2020). <https://doi.org/10.1007/s10915-020-01185-1>
17. Zhang, J., Kong, X., Zhang, J., Yang, F., Chen, K.: Vectorial fractional-order regularizer-based diffeomorphic image registration model and its numerical algorithm. *J. Math. Imag. Vision* **67**(4), 38 (2025). <https://doi.org/10.1007/s10851-025-01254-w>
18. Diwakar, M., Singh, P., Garg, D.: Edge-guided filtering based CT image denoising using fractional order total variation. *Biomed. Sig-*

- nal Process. Control **92**, 106072 (2024). <https://doi.org/10.1016/j.bspc.2024.106072>
19. Bhutto, J.A., Khan, A., Rahman, Z.: Image restoration with fractional-order total variation regularization and group sparsity. *Mathematics* **11**(15), 3302 (2023). <https://doi.org/10.3390/math11153302>
 20. Cai, J.-F., Osher, S., Shen, Z.: Split Bregman methods and frame based image restoration. *Multiscale Model. Simul.* **8**(2), 337–369 (2010). <https://doi.org/10.1137/090753504>
 21. Houichet, H., Theljani, A., Moakher, M., Rjaibi, B.: A nonstandard higher-order variational model to speckle noise removal and thin-structure detection. *J. Math. Study* **52**(4), 394–424 (2019). <https://doi.org/10.4208/jms.v52n4.19.03>
 22. Snoek, J., Larochelle, H., Adams, R.P.: Practical Bayesian optimization of machine learning algorithms. In: *Advances in neural information processing systems*, p. 25. (2012)
 23. Frazier, P.I.: Bayesian optimization. In: *Recent Advances in Optimization and Modeling of Contemporary Problems*, pp. 255–278. *Inform*s (2018). <https://doi.org/10.1287/educ.2018.0188>
 24. Jung, M., Kang, M.: Image restoration under Cauchy noise with sparse representation prior and total generalized variation. *J. Comput. Math.* **39**(1), 81–107 (2021). <https://doi.org/10.4208/jcm.1907-m2018-0234>
 25. Podlubny, I.: *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*, vol. 198, Elsevier (1998)
 26. Comi, G.E., Stefani, G.: A distributional approach to fractional Sobolev spaces and fractional variation: Existence of blow-up. *J. Funct. Anal.* **277**(10), 3373–3435 (2019). <https://doi.org/10.1016/j.jfa.2019.03.011>
 27. Li, C., Zeng, F.: *Numerical Methods for Fractional Calculus*, CRC Press (2015)
 28. Geman, D., Yang, C.: Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Image Process.* **4**(7), 932–946 (1995). <https://doi.org/10.1109/83.392335>
 29. Hunter, D.R., Lange, K.: A tutorial on MM algorithms. *Am. Stat.* **58**(1), 30–37 (2004). <https://doi.org/10.1198/0003130042836>
 30. Nikolova, M., Ng, M.K.: Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM J. Sci. Comput.* **27**(3), 937–966 (2005). <https://doi.org/10.1137/030600862>
 31. Shi, B., Pang, Z.-F., Wu, J.: Alternating split bregman method for the bilaterally constrained image deblurring problem. *Appl. Math. Comput.* **250**, 402–414 (2015). <https://doi.org/10.1016/j.amc.2014.11.004>
 32. Liu, X., Huang, L.: Split bregman iteration algorithm for total bounded variation regularization based image deblurring. *J. Math. Anal. Appl.* **372**(2), 486–495 (2010). <https://doi.org/10.1016/j.jmaa.2010.07.013>
 33. Wang, Y., Yin, W., Zeng, J.: Global convergence of ADMM in nonconvex nonsmooth optimization. *J. Sci. Comput.* **78**(1), 29–63 (2019). <https://doi.org/10.1007/s10915-018-0757-z>
 34. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.* **35**(2), 438–457 (2010). <https://doi.org/10.1287/moor.1100.0449>
 35. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* **146**(1), 459–494 (2014). <https://doi.org/10.1007/s10107-013-0701-9>
 36. Kumar, A., Ahmad, M.O., Swamy, M.: A framework for image denoising using first and second order fractional overlapping group sparsity (hf-olgs) regularizer. *IEEE Access* **7**, 26200–26217 (2019). <https://doi.org/10.1109/ACCESS.2019.2901691>
 37. Zhang, Y.-N., Sun, Z.-Z., Liao, H.-I.: Finite difference methods for the time fractional diffusion equation on non-uniform meshes. *Journal of Computational Physics* **265**, 195–210 (2014). <https://doi.org/10.1016/j.jcp.2014.02.008>
 38. Baeumer, B., Kovács, M., Meerschaert, M.M., Sankaranarayanan, H.: Boundary conditions for fractional diffusion. *J. Comput. Appl. Math.* **336**, 408–424 (2018). <https://doi.org/10.1016/j.cam.2017.12.053>
 39. Benzi, M., Golub, G.H., Liesen, J.: Numerical solution of saddle point problems. *Acta Numer* **14**, 1–137 (2005). <https://doi.org/10.1017/S0962492904000212>
 40. Deng, W.: Finite element method for the space and time fractional Fokker-Planck equation. *SIAM J. Numer. Anal.* **47**(1), 204–226 (2009). <https://doi.org/10.1137/080714130>
 41. Kelly, J.F., Sankaranarayanan, H., Meerschaert, M.M.: Boundary conditions for two-sided fractional diffusion. *J. Comput. Phys.* **376**, 1089–1107 (2019). <https://doi.org/10.1016/j.jcp.2018.10.010>
 42. Bai, L.: A new approach for Cauchy noise removal. *AIMS Mathemat.* **6**(9), 10296–10312 (2021). <https://doi.org/10.3934/math.2021596>
 43. Ding, M., Huang, T.-Z., Ma, T.-H., Zhao, X.-L., Yang, J.-H.: Cauchy noise removal using group-based low-rank prior. *Appl. Math. Comput.* **372**, 124971 (2020). <https://doi.org/10.1016/j.amc.2019.124971>
 44. Blomgren, P., Chan, T.F.: Color TV: total variation methods for restoration of vector-valued images. *IEEE Trans. Image Process.* **7**(3), 304–309 (1998). <https://doi.org/10.1109/83.661180>
 45. Wang, W., Xia, X.-G., Zhang, S., He, C., Chen, L.: Vector total fractional-order variation and its applications for color image denoising and decomposition. *Appl. Math. Model.* **72**, 155–175 (2019). <https://doi.org/10.1016/j.apm.2019.03.010>
 46. Wu, C., Tai, X.-C.: Augmented lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. *SIAM J. Imag. Sci.* **3**(3), 300–339 (2010). <https://doi.org/10.1137/090767558>
 47. Wilcoxon, F.: *Individual Comparisons by Ranking Methods*. Presented at the (1992). <https://doi.org/10.1007/978-1-4612-4380-916>
 48. Conover, W.J.: *Practical Nonparametric Statistics*. John Wiley & Sons (1999)
 49. Kwong, J.C., Palomo, J.M., Landers, M.A., Figueroa, A., Hans, M.G.: Image quality produced by different cone-beam computed tomography settings. *Am. J. Orthod. Dentofac. Orthop.* **133**(2), 317–327 (2008). <https://doi.org/10.1016/j.ajodo.2007.02.053>
 50. Angelone, F., Franco, A., Ponsiglione, A.M., Ricciardi, C., Belfiore, M.P., Gatta, G., Grassi, R., Sansone, M., Amato, F.: Assessment of an unsupervised denoising approach based on Noise2Void in digital mammography. *Sci. Rep.* **15**(1), 35712 (2025). <https://doi.org/10.1038/s41598-025-19499-1>
 51. Hsu, H., Lachenbruch, P.A.: *Paired t Test*. John Wiley & Sons, Ltd (2014). doi: <https://doi.org/10.1002/9781118445112.stat05929>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Hamdi Braiek received the Ph.D. degree in Applied Mathematics from the National Engineering School of Tunis in 2019. He is currently an Associate Professor at ESPRIT School of Engineering, Tunisia. He also holds a master's degree in business analytics and data science. His research interests include image processing, deep learning, machine learning, variational methods, partial differential equations, optimization, and inverse problems. His recent research explores variable expo-

nent models, machine learning techniques, and physics informed neural networks for scientific and engineering applications.