Iranian Journal of Numerical Analysis and Optimization
Vol. 11, No. 2, (2021), pp 235-260
DOI:10.22067/ijnao.2021.67760.1002

Research Article

Trainable fourth-order partial differential equations for image noise removal

N. Khoeiniha*, S.M. Hosseini and R. Davoudi

Abstract

Image processing by partial differential equations (PDEs) has been an active topic in the area of image denoising, which is an important task in computer vision. In PDE-based methods for unprocessed image processing, the original image is considered as the initial value for the PDE and the solution of the equation is the outcome of the model. Despite the advantages of using PDEs in image processing, designing and modeling different equations for various types of applications have always been a challenging and interesting problem. In this article, we aim to tackle this problem by introducing a fourth-order equation with flexible and trainable coefficients, and with the help of an optimal control problem, the coefficients are determined; therefore the proposed model adapts itself to each particular application. At the final stage, the image enhancement is performed on the noisy test image and the performance of our proposed method is compared to other PDE-based models.

AMS subject classifications (2020): 35G25; 68U10; 90C90.

Keywords: Partial Differential Equations; Image Processing; Image Denoising; Optimal Control.

*Corresponding author

Received 14 December 2020; revised 8 February 2021; accepted 6 April 2021

Negin Khoeiniha
Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran. e-mail: Neginkhoiniha@modares.ac.ir

Mohammad Hosseini
Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran. e-mail: hossei_ m@modares.ac.ir

Ramtin Davoudi
Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran. e-mail: R.davoudi@modares.ac.ir

1 Introduction

Image denoising is an important preliminary step in many computer vision and image processing problems that can affect further processing, and has been an open research area for a long time [28, 31, 38]. By considering the image as a function I in $\mathbb{R}^2$, the aim of denoising an image is to extract the clear image I from the noisy image I_0 that has been degraded by the model $I_0 = I + n$, in which n is a noise function.

Various approaches for image denoising have been developed, such as methods based on nonlocal means filters [4, 18], wavelets [8, 25], Perona and Malik (PM) [29], block-matching, and three-dimensional filtering [6, 7, 26], sparse representation [46], adaptive image filtering [1, 2], bilateral filtering [35, 10], and so on.

In contrast to the discrete methods used in signal processing in the past, which were the basis of digital image processing, in the past decades, continuous methods based on partial differential equations (PDEs) have become an effective tool.

The first efforts of using PDEs in computer vision and image processing tasks date back to the 1960s [11, 15]; however, these methods did not attract much attention. In the 1980s, the importance of multi-scale descriptions of images has been recognized. The importance of scale-space filtering in images was introduced by Witkin [41] and then developed by Koenderink [17]. The idea of their approach is to generate a family of scaled images $I(x, y, t)$ by convolving the original image $I_0(x, y)$ with a Gaussian kernel $G(x, y, t)$ of variance t as follows:

$$I(x, y, t) = I_0(x, y) \otimes G(x, y, t).$$

The time variable t is said to be the scale parameter, and the higher the scale gets, it results in an image with reduced resolution; therefore, the noise vanishes.

It was pointed out by Koenderink [17] that, $I(x, y, t)$ is the fundamental solution to the heat conduction equation as below, with the original image $I_0(x, y)$ as the initial condition and the conductance coefficient c, a real number:

$$I_t = c(I_{xx} + I_{yy}).$$

Although the outcome of this equation is a less noisy image, the main disadvantage of this approach is the fact that while smoothing, it blurs important image features; in other words, the Gaussian scale-space does not respect the edges of objects in an image. Therefore, the scaled image loses important details after the process.

In 1992, a method introduced by Perona and Malik on anisotropic diffusion [29] drew great attention to PDE methods in image enhancement. In fact, they designed the following equation, known as PM second-order PDE, to achieve a good balance between noise removal and retaining the edges of objects in the image:

$$I_t = c(x, y, t)(I_{xx} + I_{yy}).$$

In the above PDE, the conductance coefficient function, $c(x, y, t)$, is chosen to decrease smoothing around the edges and reduce noise in other areas. Following their model, many equations have been introduced on this basis; see for instance [16, 39, 5].

Although the PM second-order PDE and its variances were successfully applied in image denoising problems; these methods cause blocky effects on the processed image. In [43], it is noted that the PM PDE is a second-order model, and this feature guarantees its ability to reconstruct images with discontinuities but is responsible for the blocky effect [45].

In recent years, many efforts have been made to resolve this problem in the PM method, such as introducing different coefficients, different equations for processing, raising the order of the equation, and so on. One of the innovative efforts to tackle the problems of second-order approaches was inspired by learning-based methods in machine learning [19]. Afterward, some second-order learning PDEs (L-PDEs) were introduced that adopt a technique called PDE-based optimal control [14] and use training images to learn the coefficients in the PDE, and once they are computed, then the PDE is obtained and can be applied to test images.

In learning PDEs proposed in [19, 22, 23], the coefficient functions are assumed to minimize a functional subject to some PDE constraints that are designed according to the problem.

In these classes of approaches, the structure of the problem is considered, and the learned coefficients will form the PDE according to the training images. Thus, the most effort in obtaining a PDE is to prepare some input/output training image pairs.

One of the other successful strategies in enhancing the performance of second-order PDEs for image denoising is to design higher-order equations in replace of quadratic ones [13, 21, 24]. In 2000, in a method proposed by You and Kaveh, known as YK fourth-order PDE [42], the following equation was introduced:

$$\frac{\partial I}{\partial t} = -\nabla^2 \left(c_1 \left(\left| \nabla^2 I \right| \right) \left| \nabla^2 I \right| \right), \tag{1}$$

where ∇^2 denotes the Laplacian operator and conductance coefficient $c_1(\cdot)$ in YK fourth-order PDE, is a function of the absolute value of the Laplacian of the image intensity, that is,

$$c_1 \left(\left| \nabla^2 I \right| \right) = \frac{1}{1 + \left(\left| \nabla^2 I \right| / k \right)^2}, \tag{2}$$

in which $k > 0$ is the Laplacian threshold. The YK fourth-order PDE attempts to remove noise and preserve edges by approximating an observed image with a piecewise planar image [42].

In recent years, more PDE-based models have been developed, improving the performance of the YK method. In 2016, Zhang and Ye [47] proposed the adaptive fourth-order PDE for noise removal

$$\frac{\partial u}{\partial t} = -div^2(h(\nabla u, D^2 u)\frac{D^2 u}{|D^2 u|}) + \mu div((1 - h(\nabla u, D^2 u))\frac{\nabla u}{|\nabla u|}),$$

where $h(\nabla u, D^2 u)$ is an edge detector chosen by

$$h(\nabla u, D^2 u) = \frac{1}{1 + k_1|\nabla(G_\sigma \otimes u)|^2 + k_2|D^2(G_\sigma \otimes u)|^2 + \gamma}, \tag{3}$$

in which $\gamma > 0$ is a small constant to guarantee $h < 1$, k_1 and k_2 are positive constants, and $\mu > 0$ and $\lambda > 0$ are parameters balancing the contribution of the three terms in the objective function.

Later in 2018, Siddig et al. [33] introduced the following equation:

$$\begin{cases} \frac{\partial u}{\partial t} + D^2_{ij}(\frac{\alpha(u) D^2_{ij} u}{|D^2_{ij} u|}) = 0, & (x,t) \in \Omega_T = (0,T) \times \Omega, \\ u(x,t) = 0, & (x,t) \in (0,T) \times \partial\Omega, \\ \frac{\partial u}{\partial \vec{n}} = 0, & (x,t) \in (0,T) \times \partial\Omega, \\ u(x,0) = u_0(x), & x \in \Omega, \end{cases}$$

where $\alpha(u) = \frac{1}{\sqrt{1 + |G_\sigma \otimes \nabla u|^2}}$ and $G_\sigma(x)$ is the Gaussian filter with σ as its parameter. In these models (Siddig's and Zhang's PDEs), they tend to overcome the disadvantages of YK and TV model while keeping the pros of PDE algorithms.

The theoretical analysis in [9, 12], shows the advantages of fourth-order equations over second-order ones. First, fourth-order diffusion reduces fluctuation faster than second-order ones. Second, since the second-order PDE will evolve toward a piecewise constant approximation in smooth regions, unlike second-order PDE, fourth-order PDE will evolve toward and settle down to a piecewise smooth image if the image support is infinite [44].

In this article, motivated by the previous efforts on solving second-order equations' problems, we propose a fourth-order PDE with trainable coefficients for image denoising to enhance the performance of the second-order L-PDEs. To mention two advantages of this PDE, one should note the advantage of using fourth-order PDEs, and also the proposed PDE uses image differential invariants up to fourth-order, which can detect more image features compared to those of second-order PDEs. The results obtained from our scheme are compared to the state-of-the-art previous models, such as second-order learning-based PDEs [22], YK fourth-order PDE [42], and a more recent fourth-order PDEs, Siddig [33], and Zhang's [47], on this topic. The experimental results show that, in terms of subjective and objective measures, the proposed model can outperform the pre-existing methods.

The rest of this article is organized as follows. Section 2 presents the trainable fourth-order PDE with a detailed description of the solution of the optimal control problem. Experimental results are presented in Section 3, and the article is concluded in Section 4.

2 Proposed model

In this section, first, the formulation of the fourth-order PDE in our model for noise reduction is presented and then an optimal control problem will be introduced to assist in completing the PDE. Finally, the optimality conditions will be investigated for the recommended model, which helps us to extract enough information to solve the problem.

The framework of this article is based on the fourth-order noise removal PDE models, and inspired by second-order learning PDEs mentioned in the previous section we will consider flexible trainable coefficients for the proposed PDE.

In this work, the typical notations in image processing and optimal control literature are used. We denote f as the input image and I as the desired output image. Moreover, Ω is an open bounded region of $\mathbb{R}^2$, $\partial\Omega$ is its boundary, the spatial variable (x, y) belongs to Ω, $t \in (0, T_f)$ is the temporal variable, $\Omega \times (0, T_f)$ is named Q, and Γ is its boundary. We then define the set η as bellow and $|p|$ is the length of string p in which $p \in \eta$:

$$\eta = \{0, x, y, xy, xx, yy, xxx, xxy, xyy, yyy, xxxx, xxxy, xxyy, xyyy, yyyy\}$$

Changing the viewpoints of an object causes a geometric deformation, which can be modeled by a two-dimensional Euclidean, similarity, affine, or projective transformation group. Many methods have been designed to correctly recognize planar objects, by extracting image invariant features for the action of various two-dimensional transformation groups, and differential invariants are one of them. We can express image differential invariants as the functions of image partial derivatives, and they are widely used to describe local image structures. Therefore, to construct our PDE, we used a set of differential invariants up to fourth-order. We will use a set of differential invariants that are more wildly used.

Thus, the main reason for using these types of partial derivatives to form our equation is to create a model that can cover more image features compared to second-order ones and therefore be more sensitive to edges in the image. Furthermore, these differential invariants stay unchanged under various two-dimensional transformations like rotation and translation [27]. Consider the fourth-order PDE below:

$$\begin{cases} \frac{\partial I}{\partial t} = \mathbf{u}(t)^T \mathrm{diff}(I), & in \quad Q, \\ I(x,y,t) = 0, & on \quad \Gamma, \\ I(x,y,0) = f, & in \quad \Omega, \end{cases} \tag{4}$$

in which $\mathbf{u}(t) = \begin{bmatrix} u_0(t) \cdots u_6(t) \end{bmatrix}^T$ is the coefficient function and $\mathrm{diff}(\cdot)$ includes the differential parts defined as

$$\begin{aligned} &\mathrm{diff}_1(I) = I, \\ &\mathrm{diff}_2(I) = {I_y}^2 + {I_x}^2, \\ &\mathrm{diff}_3(I) = I_{xx}I_{yy} - {I_{xy}}^2, \\ &\mathrm{diff}_4(I) = I_{xx}{I_y}^2 - 2I_{xy}I_xI_y + I_{yy}{I_x}^2, \\ &\mathrm{diff}_5(I) = I_{xx}I_xI_y - I_{xy}{I_y}^2 - I_{xy}{I_x}^2 + I_{yy}I_xI_y, \\ &\mathrm{diff}_6(I) = I_{xxx}{I_x}^3 + 3I_{xxy}I_y{I_x}^2 + 3I_{xyy}{I_y}^2I_x + I_{yyy}{I_y}^3, \\ &\mathrm{diff}_7(I) = I_{xxxx}{I_x}^4 + 4I_{xyyy}{I_y}^3I_x + 6I_{xxyy}{I_x}^2{I_y}^2 + 4I_{xxxy}I_y{I_x}^3 + I_{yyyy}{I_y}^4. \end{aligned} \tag{5}$$

The initial value for this equation is f, the noisy image, and the output $I(\cdot)$ is the desired clear image. To obtain a workable equation, we need to determine the coefficient function $\mathbf{u}(t)$, which is used to control the evolution of I. We aim to attain a flexible system that can train the coefficients; then the function $\mathbf{u}(t)$ would vary in different problems and be adapted according to the given training images.

As mentioned previously, by using differential invariants in our model, the PDE system (4) is rotationally invariant and the coefficient functions must be independent of (x,y) so that (4) is shift-invariant as well.

Proposition 1. Considering equation (4), the control functions $\{u_i\}_{i=0}^{6}$ must be independent of (x,y).

Proof. Considering

$$D(I,\mathbf{u}) := \mathbf{u}(t)^T \mathrm{diff}(I), \tag{6}$$

we then rewrite $D(I,\mathbf{u}) = \tilde{D}(I,x,y,t)$. Therefore, it is enough to prove the independence of $\tilde{D}$ of (x,y).

Due to the translation invariance of (4), when we change $f(x,y)$ to $f(x-x',y-y')$, $I(x,y)$ changes to $I(x-x',y-y')$ and we have

$$\frac{\partial I(x-x',y-y')}{\partial t} = \tilde{D}(I(x-x',y-y'),x,y,t).$$

Then, by the conversion $x - x' = x$ and $y - y' = y$, we get

$$\frac{\partial I(x,y)}{\partial t} = D(I(x,y),x+x',y+y',t).$$

On the other hand,

$$\frac{\partial I(x,y)}{\partial t} = D(I(x,y),x,y,t),$$

$$\Rightarrow \;\; D(I,x+x',y+y',t) = D(I,x,y,t) \quad \text{for all } (x',y').$$

Hence, D is independent of (x,y), which indicates the independence of the function $\mathbf{u}$. □

2.1 Extracting the coefficient function

Attempting to solve the problem (4), we need to determine the coefficients $\mathbf{u}(t)$, which is done by an optimal control problem. Therefore we recollect some necessary definitions in this field. We begin by assuming $I \in H_0^1(\Omega)$, defined to be the closure of the infinitely differentiable functions compactly supported in Ω in $H_1(\Omega)$.

Definition 1. Considering the optimal control problem

$$\begin{aligned} &\min J(I,u) \\ s.t.\;& T(I,u) = 0 \\ & u \in U_{ad} \end{aligned} \tag{7}$$

where $J : H_0^1(\Omega) \times L^2(\Omega) \to \mathbb{R}$ is called the objective function of the optimal control problem and $U_{ad} \in L^2(\Omega)$ is the admissible set, a nonempty set in $\mathbb{R}^m$.

A vector $\bar{u} \in U_{ad}$ is called an optimal control for (7), if $J(\bar{I},\bar{u}) \leq J(I,u)$ for each $u \in U_{ad}$ and $\bar{I}$ is called the optimal state associated with $\bar{u}$.

Definition 2 ([36])**.** Let U and V be normed vector spaces, let $U_{ad} \subseteq U$, and let $f : U_{ad} \to V$, and consider some $u \in U_{ad}$. Let $h \in U$. If $u + th \in U_{ad}$ for sufficiently small $t > 0$, and the limit

$$\delta f(u,h) = \lim_{t \to 0} \frac{f(u+th) - f(u)}{t},$$

exists in V, then $\delta f(u,h)$ is called the directional derivative of F at u in the direction h. If the directional derivative $\delta f(u,h)$ exists for each h, then the map

$$\delta f(u,\cdot) : U \to V, h \mapsto \delta f(u,h),$$

is called the first variation of f at u. In addition, if the first variation constitutes a bound linear operator, then it is called the Gateaux derivative of f.

Definition 3 ([36])**.** Let X and Y be Banach spaces and let $A : X \to Y$ be a bounded linear operator. The map

$$A^* : Y^* \to X^*, A^*(f) = foA,$$

is called the adjoint of A. For the optimal control problem (7), the adjoint equation is of the form

$$D_y T(I,u)^T \lambda = \nabla_y J(y,u), \tag{8}$$

and its solution λ is the adjoint state.

Definition 4 ([36])**.** The function $L(I,u,\lambda) : H_0^2(\Omega) \times L^2(\Omega) \times H_0^2(\Omega) \to \mathbb{R}$ defined as bellow is called the Lagrangian function for the problem (7):

$$L(I,u,\lambda) := J(I,u) - \langle T(I,u), \lambda \rangle \tag{9}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product defined on $H_0^2(\Omega)$.

Our motivation here is the optimal heating problem, which uses an optimal control problem to get the desired output from a heat equation. For instance, in a transient heating problem, the goal is to reach a certain temperature $\bar{y}$ by controlling the heating elements [30]. Likewise, our objective in this article is to approximate the target clean noise-free image $\tilde{I}$ by controlling the coefficient function. Therefore, a PDE-based optimal control problem will be presented. As the first step, we need to prepare some input/output training images $(f_k, \tilde{I}_k)$, where f_k is the noisy image and $\tilde{I}_k$ is the clear image. The final output of the equation needs to be close to the ground truth; thus the coefficient functions must minimize the following functional:

$$J(\{I_k\}_{k=1}^K, \mathbf{u}) = \frac{1}{2}\sum_{k=1}^{K} \int_\Omega \left(I_k(T_f) - \tilde{I}_k\right)^2 d\Omega + \frac{1}{2}\sum_{i=0}^{6} \alpha_i \int\limits_0^{T_f} u_i^2(t)dt, \tag{10}$$

where K is the number of the training images, $I_k(T_f)$ is the output image determined from (4) at time $t = T_f$ when the initial value is f_k, and α_i are positive weighting parameters [1]. The first term of the functional J requires the final output of our PDE to be close to the ground truth and the second term is for regularization so that this optimal control problem is well-posed and counteracts the tendency of the control to become locally bounded as J approaches its minimum [36].

We have the following optimal control problem with PDE constrains:

$$\min_{\mathbf{u}} J(\{I_k\}_{k=1}^K, \mathbf{u}); \quad s.t. \begin{cases} \frac{\partial I_k}{\partial t} = \mathbf{u}(t)^T \mathrm{diff}(I_k), & in\ Q, \\ I_k(x,y,t) = 0, & on\ \Gamma, \\ I_k(x,y,0) = f_k, & in\ \Omega. \end{cases} \tag{11}$$

In addition, we consider an inequality constraint $0 \leq \mathbf{u}(t)$ for the problem (11) to make sure that the coefficients are all positive. In the following, the

[1] in this article, we fix $\alpha_i = 10^{-7}$, $i = 0, \ldots, 6$.

necessary conditions for this problem (11) will be examined, so that we would be able to solve the problem by proper numerical methods.

Lemma 1. Suppose that $J(I,u) = \int_Q g(u,I)dQ$ is of class C^1, where g is a smooth function. Then considering the function $D(I, \mathbf{u})$ in (6) as

$$\frac{DJ}{Du} = D_I^*(I,u)(\lambda) + g_I^*(I,u)(T_f),$$

where D_I^* and g_I^* are the adjoint operators of D_I and g_I, respectively, $\frac{DJ}{Du}$ is the Gateaux derivative of J with respect to the control u, and λ is the adjoint state of (4).

Proof. We first define an operator ψ, mapping u to the solution of (4), and set $I = \psi(u)$. Now consider

$$d = \lim_{\varepsilon \to 0} \frac{\psi(u + \varepsilon.\delta u) - \psi(u)}{\varepsilon},$$

where δu is the perturbation of u.

We have

$$\begin{aligned}
(\frac{DJ}{Du}, \delta u)_Q &= \lim_{\varepsilon \to 0} \frac{J(I, u + \varepsilon.\delta u) - J(I,u)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{J(\psi(u + \varepsilon.\delta u), u + \varepsilon.\delta u) - J(\psi(u), u)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{J(I + \varepsilon.d + o(\varepsilon), u + \varepsilon.\delta u) - J(I,u)}{\varepsilon} \\
&= \int\limits_Q (g_I(u,I)(d) + g_I(u,I)(\delta u))dQ \\
&= (g_I(u,I)(d), 1)_Q + (g_I(u,I)(\delta u), 1)_Q \\
&= (g_I^*(u,I)(1), d)_Q + (g_I^*(u,I)(1), \delta u)_Q.
\end{aligned} \tag{12}$$

If we take λ as the adjoint state of (11), then from the PDE constrains of (11), we have

$$((\psi(u))_t, \lambda)_Q = (D(u, \psi(u)), \lambda)_Q \tag{13}$$

and

$$((\psi(u + \varepsilon.\delta u))_t, \lambda)_Q = (D(u + \varepsilon.\delta u, \psi(u + \varepsilon.\delta u)), \lambda)_Q. \tag{14}$$

Now by subtracting (14) with (13), we have

$$\begin{aligned}
&((\psi(u + \varepsilon.\delta u) - \psi(u))_t, \lambda)_Q \\
&= (D(u + \varepsilon.\delta u, \psi(u + \varepsilon.\delta u)) - D(u, \psi(u)), \lambda)_Q \\
&= (\varepsilon.D_u(u,I)(\delta u) + D_I(u,I)(\psi(u + \varepsilon.\delta u) - \psi(u), \lambda) + o(\varepsilon).
\end{aligned} \tag{15}$$

In this step, we divide both sides with ε, and let $\varepsilon \to 0$. Therefore

$$(d_t, \lambda)_Q = (D_u(u, I)(\delta u), \lambda)_Q + (D_I(u, I)(d), \lambda)_Q. \tag{16}$$

Integrating by parts implies

$$(d, \lambda)_\Omega|_0^T - (d, \lambda_t)_Q = (\delta u, D_u^*(u, I)(\lambda))_Q + (d, D_I^*(u, I)(\lambda)). \tag{17}$$

Now, since λ is the adjoint equation of the problem, we have

$$(d, \lambda)_\Omega|_0^T = 0, (g_I^*(u, I)(T), d)_Q = (-\lambda_t - D_I^*(u, I)(\lambda), d), \tag{18}$$

and combining it with (16) yields

$$(g_I^*(u, I)(T), d)_Q = (-\lambda_t - D_I^*(u, I)(\lambda), d) = (\delta u, D_u^*(u, I)(\lambda)). \tag{19}$$

Therefore we arrive at

$$\frac{DJ}{Du} = g_I^*(u, I)(T) + D_u^*(u, I)(\lambda).$$

□

Now, we investigate the conditions that the optimal vectors $\bar{u}$ and $\bar{I}$ must satisfy to be then determined, using numerical methods.

Theorem 1. Consider the optimal control problem (11), and suppose that

$$U_{ad} = \{\mathbf{u} \in L_2(\Omega); \mathbf{u}_a \leq \mathbf{u}\} \tag{20}$$

is the admissible control set, where $U_{ad} \subseteq O \subseteq L^2(\Omega)$, O is an open subset of $L^2(\Omega)$. In our case, $\mathbf{u}_a = 0$ and $J : H_0^1(\Omega) \times L_2(\Omega) \to \mathbb{R}$ is of class C^1. If $\bar{\mathbf{u}} \in U_{ad}$ is an optimal control for (11) in the sense of Definition 1 with corresponding state $\bar{I}$ and adjoint state $\bar{\lambda}$, then $\bar{\mathbf{u}} = (\bar{u}_i)_{0 \leq i \leq 6}$ satisfies for each $i \in \{0, \dots, 6\}$

$$\bar{u}_i = u_{a,i}, \quad \textit{where } L_u(\bar{I}, \bar{u}, \bar{\lambda}) > 0. \tag{21}$$

In addition, by introducing the extended Lagrange function

$$L(I, u, \lambda, \mu_a) = L(I, u, \lambda) + \langle u_a - u, \mu_a \rangle\,, \tag{22}$$

and letting

$$\mu_a := \max\{0, L_u(I, u, \lambda)\},$$

the 4-tuples $(\bar{I}, \bar{u}, \bar{\lambda}, \mu_a) \in (H_0^2(\Omega), L^2(\Omega), H_0^2(\Omega), L^2(\Omega))$ satisfies

$$\frac{\partial J}{\partial u_i} = \alpha_i u_i - \sum_{k=1}^{K} \int_\Omega \lambda_k \mathrm{diff}_i(I_k) d\Omega - \mu_a, \quad i = 0, \dots, 6. \tag{23}$$

Proof. First note that, (20) implies that U_{ad} is convex. We have the following optimality system which can be used to determine the optimal control $\bar{u}$ and $\bar{I}$:

$$\begin{array}{l} \nabla_\lambda L(\bar{y},\bar{u},\bar{\lambda}) = 0, \\ \nabla_I L(\bar{y},\bar{u},\bar{\lambda}) = 0, \\ \nabla_u L(\bar{y},\bar{u},\bar{\lambda}) = 0, \\ \left\langle L_u(\bar{I},\bar{u},\bar{\lambda}), u-\bar{u}\right\rangle \geq 0, \ \ for\ each\ u \in U_{ad}. \end{array} \tag{24}$$

Every solution to the general form of optimal control (7) must satisfy this system [36].

Slightly rearranging the fourth optimal condition, we get

$$\left\langle L_u(\bar{I},\bar{u},\bar{\lambda}),\bar{u}\right\rangle_{L^2(\Omega)} \leq \left\langle L_u(\bar{I},\bar{u},\bar{\lambda}),u\right\rangle_{L^2(\Omega)}.$$

It implies that $\bar{u}$ is the solution to the optimal problem

$$\min_{u\in U_{ad}} \left\langle L_u(\bar{I},\bar{u},\bar{\lambda}),u\right\rangle_{L^2(\Omega)} = \min \sum_{i=0}^{6} L_u(\bar{I},\bar{u}_i,\bar{\lambda})u_i. \tag{25}$$

The components u_i can be varied independently because of the form of U_{ad}, such that the sum in (25) attains its min if and only if each summand is minimal. Therefore,

$$L_u(\bar{I},\bar{u}_i,\bar{\lambda})\bar{u}_i = \min_{u\in U_{ad}} L_u(\bar{I},\bar{u}_i,\bar{\lambda})u_i. \tag{26}$$

Now, (21) is a direct result of (26).

Moreover, according to the definition of $L(I,u,\lambda,\mu_a)$ in (22), we have

$$\begin{array}{l} \nabla_\lambda L(I,u,\lambda,\mu_a) = \nabla_\lambda L(\bar{I},\bar{u},\bar{\lambda}), \\ \nabla_I L(I,u,\lambda,\mu_a) = \nabla_I L(\bar{I},\bar{u},\bar{\lambda}), \\ \nabla_u L(I,u,\lambda,\mu_a) = \nabla_u L(\bar{I},\bar{u},\bar{\lambda}) - \mu_a. \end{array} \tag{27}$$

Therefore, we continue the proof with the old form of Lagrangian for simplicity. By Definition 4, the Lagrangian function for the optimal control problem (11) is as follows:

$$L(\{I_k\}_{k=1}^K,\mathbf{u}) = J(\{I_k\}_{k=1}^K,\mathbf{u}) + \sum_{k=1}^{K}\int_Q \lambda_k((I_k)_t - D(I_k,\mathbf{u}))dQ, \tag{28}$$

where the multiplier λ_k is the adjoint state. It can be seen that the first optimality condition $\nabla_\lambda L(\bar{y},\bar{u},\bar{\lambda}) = 0$ is the PDE constraint in (11).

To find the adjoint state λ, we first perturb $D(I_k)$ with respect to I:

$$\begin{aligned}&D(I_k+\varepsilon.\delta I_k,\mathbf{u})-D(I_k,\mathbf{u})\\&=\varepsilon.\Big(\frac{\partial D}{\partial (I_k)_x}\frac{\partial(\delta I_k)}{\partial x}+\cdots+\frac{\partial D}{\partial (I_k)_{yyyy}}\frac{\partial^4(\delta I_k)}{\partial y^4}\Big)+O(\varepsilon)\\&=\varepsilon.\sum_{p\in\eta}\sigma_p(I_k)\frac{\partial^{|p|}(\delta I_k)}{\partial p}+O(\varepsilon),\end{aligned}\tag{29}$$

in which

$$\sigma_p(I_k)=\frac{\partial D(I_k,\mathbf{u})}{\partial I_k}=\mathbf{u}^T(t).\frac{\partial(\mathrm{diff}(I_k))}{\partial (I_k)_p}=\sum_{i=0}^{6}u_i\frac{\partial \mathrm{diff}_i(I_k)}{\partial (I_k)_p}.\tag{30}$$

Besides, for having $L_{I_k}(I,u,\lambda)$, we perturb I_k in L:

$$\begin{aligned}L_{I_k}(I,u,\lambda)=&\lim_{\varepsilon\to 0}\frac{L(\dots,I_k+\varepsilon.\delta I_k,\dots)-L(\dots,I_k,\dots)}{\varepsilon}\\=&\lim_{\varepsilon\to 0}\big(\frac{1}{2.\varepsilon}\int_\Omega((I_k+\varepsilon.\delta I_k)(x,y,T_f)-\tilde{I}_k(x,y))^2d\Omega\\&-\frac{1}{2}\int_\Omega(I_k(x,y,T_f)-\tilde{I}_k(x,y))^2d\Omega\\&+\int_Q\lambda_k[((I_k+\varepsilon.\delta I_k)_t-D(I_k+\varepsilon.\delta I_k))-((I_k)_t-D(I_k))]dQ\big)\\=&\int_\Omega(I_k(x,y,T_f)-\tilde{I}_k(x,y))\delta I_k(x,y,T_f)d\Omega\\&+\int_Q\lambda_k(\delta I_k(x,y,T_f))_t dQ-\int_Q\lambda_k\sum_{p\in\eta}\sigma_p\frac{\partial^{|p|}(\delta I_k)}{\partial p}dQ+O(\varepsilon),\end{aligned}$$

and δI_k should satisfy

$$\begin{aligned}\delta I_k&=0\ \ on\ \Gamma,\\\delta I_k&=0\ \ in\ \Omega.\end{aligned}$$

Due to the boundary and initial conditions of I_k, if we assume that $\lambda_k=0$ on Γ, then upon integration by parts (with respect to t in $(I_k)_t$ and spatial variables in $\partial^{|p|}(\delta I_k)$), we have

$$\begin{aligned}L_{I_k}(I,u,\lambda)=&\int_\Omega(I_k(x,y,T_f)-\tilde{I}_k(x,y))\delta I_k(x,y,T_f)d\Omega+\int_\Omega(\lambda_k.\delta I_k)(x,y,T_f)d\Omega\\&-\int_Q(\lambda_k)_t\delta I_k dQ-\int_Q\sum_{p\in\eta}(-1)^{|p|}\frac{\partial^{|p|}(\sigma_p(I_k)\lambda_k)}{\partial p}.\delta I_k dQ\end{aligned}$$

$$= \int\limits_{Q} [(\lambda_k + I_k - \tilde{I}_k)\delta(t - T_f) - (\lambda_k)_t - \sum_{p\in\eta} (-1)^{|p|} \frac{\partial^{|p|}(\sigma_p \lambda_k)}{\partial p}]\delta I_k dQ + O(\varepsilon),$$

where $\delta(\cdot)$ is the Dirac function. Finally, by considering the second optimality condition $L_u(I, u, \lambda) = 0$, our argument yields the following system as the adjoint equation for λ_k:

$$\begin{cases} \frac{\partial \lambda_k}{\partial t} + \sum\limits_{p\in\eta} (-1)^{|p|}(\sigma_p \lambda_k)_p = 0, & in\ Q, \\ \lambda_k = 0, & on\ \Gamma, \\ \lambda_k(x, y, T_f) = \tilde{I}_k - I_k(T_f), & in\ \Omega, \end{cases} \tag{31}$$

for $k = 1, \ldots, K$.

Now by perturbing u_i in L, we get $L_{u_i}(I, u, \lambda)$

$$\begin{aligned} L_{u_i}(I, u, \lambda) =& \lim_{\varepsilon\to 0} \frac{L(\ldots, u_i + \varepsilon.\delta u_i, \ldots) - L(\ldots, u_i, \ldots)}{\varepsilon} \\ =& \lim_{\varepsilon\to 0} \Big(\frac{1}{2.\varepsilon} \alpha_i \int_0^T (u_i + \varepsilon.\delta u_i)^2(t)dt - \frac{1}{2}\alpha_i \int_0^T u_i^2(t)dt \\ & - \sum_{k=1}^{K} \int_Q \lambda_k((u_i + \varepsilon.\delta u_i)(t) - u_i^t(t)) diff(I) dQ \Big) \\ =& \int\limits_0^T (\alpha_i u_i.\delta u_i)(t)dt - \int_0^T (\sum_{k=1}^{K} \int_\Omega \lambda_k diff(I_k) d\Omega) \delta u_i dt. \end{aligned}$$

Therefore, we get

$$L_{u_i}(I, u, \lambda) = \alpha_i u_i - \sum_{k=1}^{K} \int_\Omega \lambda_k \mathrm{diff}_i(I_k) d\Omega, \quad i = 0, \ldots, 6,$$

where the function λ_k is the solution to (31). Now we get back to the Lagrangian (22) of our model, we have

$$L_{u_i}(I, u, \lambda, \mu_a) = \alpha_i u_i - \sum_{k=1}^{K} \int_\Omega \lambda_k \mathrm{diff}_i(I_k) d\Omega - \mu_{a,i}, \quad i = 0, \ldots, 6,$$

From Lemma 1, we have $\frac{DJ}{Du} = L_u(I, u, \lambda)$. Therefore we arrive at

$$\frac{DJ}{Du_i} = \alpha_i u_i - \sum_{k=1}^{K} \int_\Omega \lambda_k \mathrm{diff}_i(I_k) d\Omega - \mu_{a,i}, \quad i = 0, \ldots, 6. \tag{32}$$

□

In summary, we have the Gateaux derivative of the object function J with respect to the control u,

$$\frac{DJ}{Du_i} = \alpha_i u_i - \sum_{k=1}^{K} \int_{\Omega} \lambda_k \mathrm{diff}_i(I_k) d\Omega - \mu_{a,i}, \quad i = 0, \ldots, 6, \tag{33}$$

where the adjoint state λ is the solution to the PDE bellow:

$$\begin{cases} \frac{\partial \lambda_k}{\partial t} + \sum\limits_{p \in \eta} (-1)^{|p|} (\sigma_p \lambda_k)_p = 0, & in\ Q, \\ \lambda_k = 0, & on\ \Gamma, \\ \lambda_k(x, y, T_f) = \tilde{I}_k - I_k(T_f), & in\ \Omega. \end{cases} \tag{34}$$

For more details and a more mathematically precise explanation, the interested reader is referred to [36, 20]. Now that we have extracted enough information from the optimality conditions, we can use numerical methods to determine $\bar{\mathbf{u}}$ for (11).

Now, we need to find the initial $\mathbf{u}(t)$ as the starter of the numerical method. At each time step, $\frac{\partial I_k(t)}{\partial t}$ is expected to be $\frac{\tilde{I}_k - I_k(t)}{T_f - t}$ so that I_k tends to the expected output $\tilde{I}_k$. On the other hand, we have $\frac{\partial I_k(t)}{\partial t} = L(I_k, \mathbf{u})$. Next, we aim to find $\mathbf{u}(t)$ to minimize

$$\sum_{k=1}^{K} \int_{\Omega} \left(L(I_k, \mathbf{u}) - \frac{\partial I_k(t)}{\partial t}\right)^2 d\Omega = \sum_{k=1}^{K} \int_{\Omega} \left[p_k(t)^T \mathbf{u}(t) - d_k(t)\right]^2 d\Omega.$$

Therefore, the initial $\mathbf{u}(t)$ can be obtained by solving the following system:

$$P(t)\mathbf{u}(t) = \mathbf{d}(t), \tag{35}$$

where we have $P(t) = \sum\limits_{k=1}^{K} \int_{\Omega} p_k(t) p_k(t)^T d\Omega$ and $\mathbf{d}(t) = \sum\limits_{k=1}^{K} \int_{\Omega} p_k(t) d_k(t)^T d\Omega$, in which $p_k(t) = \mathrm{diff}(I_k)$ and $d_k(t) = \frac{\tilde{I}_k - I_k(t)}{T_f - t}$.

Finally, with the help of the above $\mathbf{u(t)}$ and by considering the Gateaux derivative (33), we can the perform Conjugate Gradient (CG) method [32, 34] to solve the optimal control problem (11).

Our PDE framework is summarized in Algorithm 1. Once the coefficients are learned, equation (4) is completed and can be applied on test images, where the noisy image f is the input as the initial value and the output $I(T_f)$ is the desired denoised image.

Algorithm 1 The framework to learn PDEs for image restoration

```
Input: Training image pairs $(f_k, \tilde{I}_k)$, $k = 1, \dots, K$ , $T_f$.
1: Initialize $\mathbf{u}(t)$, $t \in [0, T_f)$, by solving (35).
2: while not converged do
3:     Compute $\frac{\partial J}{\partial u_i}$, $i = 0, \dots, 6$ using (33).
4:     Perform conjugate gradient method using $\frac{\partial J}{\partial u_i}$ [32, 34].
5: end while
6: return Coefficient functions $\mathbf{u}(t)$, $t \in [0, T_f)$.
7: Solve (4) using the result from previous step.
```

3 Experimental Results

In this section, the application of our proposed PDE for image denoising is demonstrated. The experiments are performed on gray-scale images. We compare the results obtained from our PDE with the evolutionary previous works, PM second-order PDE [29], YK fourth-order PDE [42], second-order L-PDE [22], Zhang [47], and Siddig's [33] fourth-order PDE approaches for image denoising. For selecting the parameters in PM and YK method, we chose the values that produce the most appealing results and the best result is chosen as the output of the method. For all experiments, the grid sizes $h = 1$ and $T_f = 5$ are considered. All our experiments are performed in MATLAB R2018b on an Intel(R) Core(TM) i5, 2.40 GHz, 8 GB RAM laptop.

For a quantitative comparison between the previous methods and our scheme, the peak signal-to-noise ratio (PSNR)[44] and structural similarity index (SSIM)[37] of the processed images are considered. It should be noted that PSNR is most easily defined via the mean squared error (MSE). Given a noise-free $m \times n$ monochrome image I and its noisy approximation f, MSE is defined as:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (I(i,j) - f(i,j))^2.$$

Then PSNR is obtained as follows:

$$\begin{aligned} PSNR &= 10.\log_{10}(\frac{\max^2 I}{\sqrt{MSE}}) \\ &= 20.\log_{10}(\max I) - 10.\log_{10}(MSE). \end{aligned} \tag{36}$$

Here, $\max I$ is the maximum possible pixel value of the image. It can easily be seen that, the closer an image is to its ground truth, the higher its PSNR is. Given similar assumptions, SSIM is defined as

$$SSIM = \frac{(2\bar{I}\bar{f} + C_1)(2\sigma_{If} + C_2)}{(\mu_I^2 + \mu_f^2 + C_1)(\sigma_I^2 + \sigma_f^2 + C_2)},$$

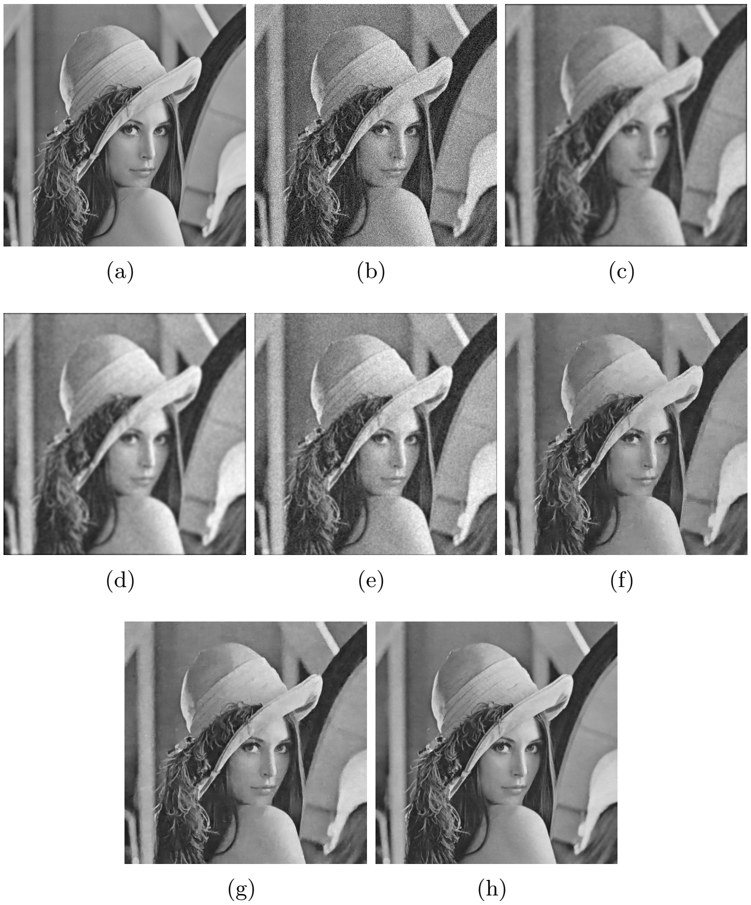
in which $\bar{I}$ and $\bar{f}$ are the mean values of I and f, σ_I and σ_f are the standard deviation of them, σ_I^2 and σ_f^2 are their variances, and σ_{If} is the variance of I and f. For stabilizing the division with the weak denominator, $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$ are two variables and L is the dynamic range of pixel values, $K_1 = 0.01$ and $K_2 = 0$. Like PSNR, the closer an image is to its ground truth, the closer the SSIM will be to one.

For choosing images in our experiment, first, we used a group of images containing 50 clear images and their noisy version, with additive white zero mean, Gaussian white noise with the variance of 0.01, in various sizes, categorized in ten different classes including animals, sea, personal picture, city view, nature images, and so on. These images are gathered from SIPI image database [40], Berkeley image database [3], and Matlab's image library. Four pictures in each class were used for training the PDE, and one was used as the test image and the same test image was used as the initial image for the other denoising methods.

In Figures 1 and 2, the performance of our proposed method and the other five denoising models are shown. First, "Lena" picture with size 512 × 512, Figure 1(a) is the original clean image, Figure 1(b) is a noisy version of it with additive white noise followed by Figures 1(c-g) showing the denoised images by PM, YK, 2nd L-PDE, Zhang's's PDE, and Siddig's model, respectively, and Figure 1(h) is the denoised image with our proposed method. It can be seen that compared to the other five methods, our fourth-order PDE demonstrates better results in preserving edges while decreasing the noise of the image. In Figure 2, the "City" image with size 850 × 480 is displayed followed by Figure 2(b), its noisy version, and on Figures 2(c-g), the enhanced form of it by the Pm, YK, 2nd order L-PDE, Zhang's model, and Siddig's approach is shown, and in Figure 2(h), the result of our denoising model is illustrated. It is vivid that the recommended PDE reduces noise without blurring the image features or leaving any speckles.

The PSNR and SSIM values collected from denoising ten images with additive white Gaussian noise with all the mentioned models are presented in Table 1. Comparing the values obtained from discussed methods in each image shows that our proposed scheme has the best performance among the other methods, which is indicated by the higher PSNR and SSIM values.

In addition, to better investigate the performance of our suggested method on more complex noises, the second group consisting of 50 images, in five separate classes with unknown noise, meaning a combination of zero-mean Gaussian white noise, Poisson noise, and the salt & pepper noise ($d = 0.1$) is added to the images, such as aerial and texture images are chosen. This group of images is also gathered from SIPI and Berkeley image databases and pictures are taken by a Canon EOS 750D camera. Table 2 represents the results of performing all the methods on our five image classes, which can be seen that our suggested model results in better outcomes. In training the L-PDEs in [22] and our model for denoising the images with unknown noise, we used more training images to learn the coefficients due to the more complex noise in each image. In Figures 3 and 4, the noisy test image“Aerial” and “Tile” are shown with their denoised versions and it is observed that our model is superior in decreasing noise while preserving the image features.


(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 1: Comparison of proposed scheme with five methods on “Lena” image with added white Gaussian noise. (a) Original image; (b) Noisy image; (c) PM second-order PDE; (d) YK fourth-order PDE; (e) second-order L-PDE; (f) Zhang’s fourth-order PDE; (g) Siddig’s model; (h) proposed fourth-order PDE

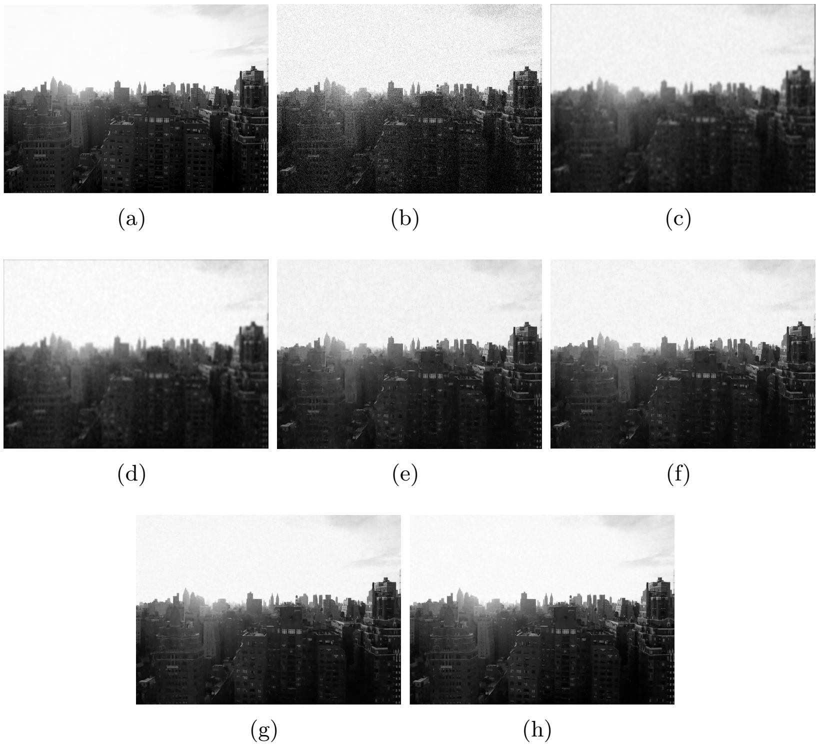


Figure 2: Comparison of proposed scheme with five methods on “City” image with added white Gaussian noise. (a) Original image; (b) Noisy image; (c) PM second-order PDE; (d) YK fourth-order PDE; (e) second-order L-PDE; (f) Zhang’s fourth-order PDE; (g) Siddig’s model; (h) proposed fourth-order PDE

In Figure 3, a noisy aerial image with unknown noise is shown and its denoised forms are Figures 3(b-g). Even with the complex unknown noise, the performance of our proposed quadratic model is better than the other competitors. In Figure 4, a noisy tile image is presented. As expected from the PSNR and SSIM values in Table 2, the recommended learning-based PDE still yields better results in comparison to the other contestants.

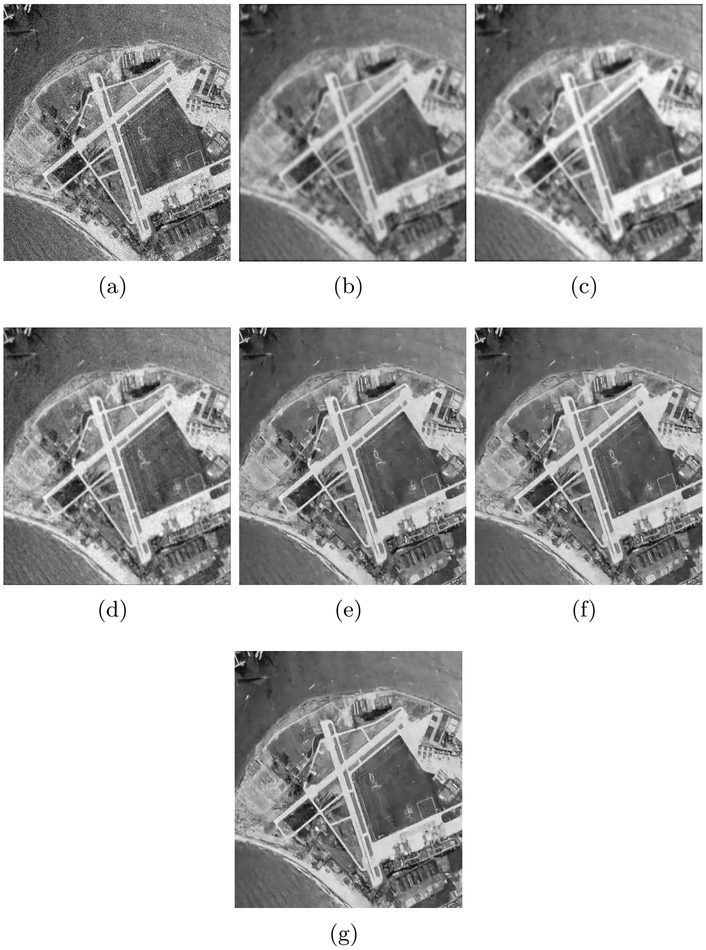
Finally, we compare the average PSNR values resulted from all the models, in Figure 5. Comparing the average values obtained from each method on both image groups with Gaussian and unknown noise, we can see that the average values gained from our proposed PDE are larger than the other approaches on both types of noises, which indicates that our scheme outperforms other models in both image categories.

Table 1: PSNR and SSIM results for ten test images with additive white Gaussian noise.

	PM	YK	2nd L-PDE	Zhang	Siddig	**Proposed**
Lena	19.4755	20.8096	23.8938	27.7531	28.5738	**30.4873**
	0.7527	0.8031	0.8093	0.8069	0.8421	**0.9184**
City	20.3853	20.7052	24.2705	28.2597	29.6089	**29.1105**
	0.6908	0.7288	0.8041	0.8472	0.8133	**0.8563**
Text	18.0646	21.9834	23.9050	28.0671	28.5636	**31.2598**
	0.7093	0.7313	0.8527	0.8530	0.8600	**0.8697**
Woman	19.8594	20.7346	23.0504	29.4682	30.1833	**30.4666**
	0.6551	0.6995	0.8211	0.8873	0.8237	**0.9030**
Eiffel	19.1024	21.4389	24.1924	28.0046	30.1453	**30.2713**
	0.6694	0.6890	0.7806	0.8074	0.8497	**0.8901**
Sea	20.2217	21.0962	23.9878	27.2673	27.5078	**31.0694**
	0.7589	0.7854	0.8030	0.8364	0.7890	**0.8751**
Nightcity	19.8372	21.9698	25.2569	28.5490	28.6472	**29.3441**
	0.7048	0.7589	0.8497	0.7573	0.8628	**0.8638**
Nature	20.2255	22.3523	24.6548	29.7069	29.6567	**30.4155**
	0.6508	0.7001	0.7945	0.7886	0.8349	**0.9153**
Flowers	19.3536	20.9230	25.2612	28.9910	29.0531	**29.9513**
	0.6461	0.6934	0.7914	0.8199	0.8276	**0.8785**
Trees	20.3819	22.1760	25.8949	28.6375	28.8983	**30.5744**
	0.6835	0.7412	0.7364	0.8536	0.8619	**0.9143**

Table 2: PSNR and SSIM results for five test images with real unknown noise.

	PM	YK	2nd L-PDE	Zhang	Siddig	**Proposed**
Truck	16.1163	17.2961	18.6548	19.9679	21.0791	**22.5288**
	0.7001	0.7058	0.7413	0.7538	0.8103	**0.8526**
Aerial	15.9888	18.1222	18.9908	19.5223	20.9317	**22.9611**
	0.6846	0.7140	0.7851	0.7987	0.8159	**0.8794**
Personal	16.6055	18.3664	19.0258	20.6046	21.6548	**23.7669**
	0.6422	0.6988	0.7423	0.7642	0.7941	**0.8682**
Texture	15.4097	17.0267	18.9516	19.5583	20.9513	**22.0304**
	0.7085	0.7631	0.7970	0.8155	0.8289	**0.8437**
Tile	16.2849	17.9806	18.9823	20.7140	21.1675	**23.6656**
	0.6827	0.7125	0.7681	0.8155	0.8468	**0.8549**



(a) (b) (c)

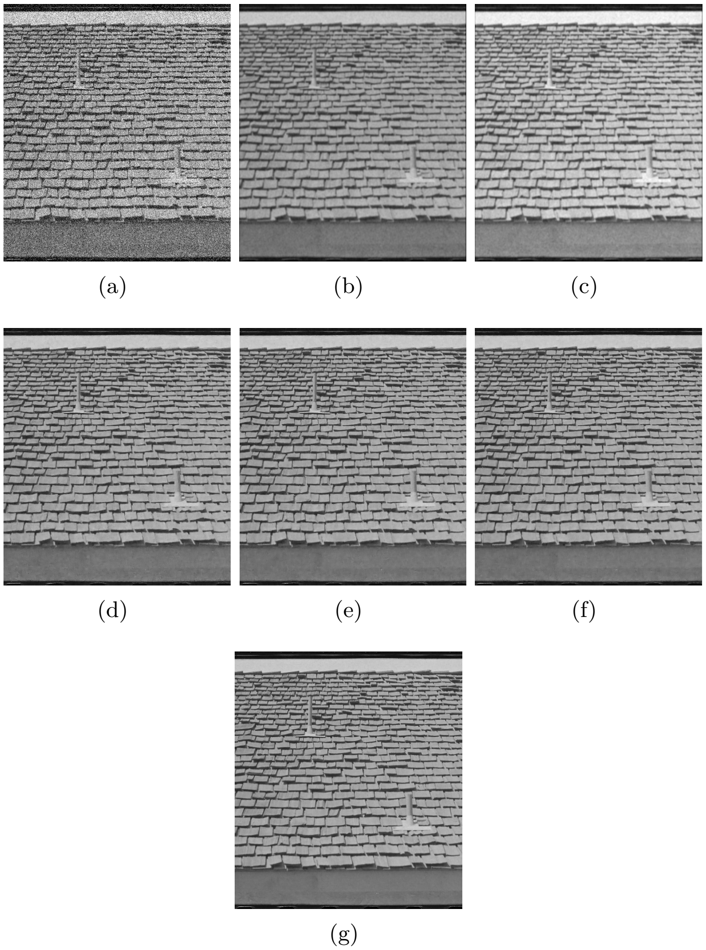
(d) (e) (f)

(g)

Figure 3: Comparison of proposed scheme with five methods on “Aerial” image with real unknown noise. (a) Noisy image; (b) PM second-order PDE; (c) YK fourth-order PDE; (d) second-order L-PDE; (e) Zhang’s fourth-order PDE; (f) Siddig’s model; (g) proposed fourth-order PDE

4 Conclusion

In this article, we proposed a fourth-order PDE for image denoising with trainable coefficients asserted by an optimal control problem. The experimental results show improvements in performance in comparison to previous approaches in the field of PDE-based methods for image denoising. Compared to previous methods, our scheme results in better outcomes, clearer image with higher PSNR and SSIM. For future studies, we want to strengthen our approach in the following areas. First, design a more general equation to adapt better to different problems, more precisely, having an optimal control problem with controls depending on spatial variables as well. Second, to introduce a region-based PDE with trainable coefficients, which leads to having two controls for each region. Finally, finding a more suitable numerical method for solving the optimal control problem to have a faster and more accurate numerical result.


(a) (b) (c)

(d) (e) (f)

(g)

Figure 4: Comparison of proposed scheme with five methods on "Tile" image with real unknown noise. (a) Noisy image; (b) PM second-order PDE; (c) YK fourth-order PDE; (d) second-order L-PDE; (e) Zhang's fourth-order PDE; (f) Siddig's model; (g) proposed fourth-order PDE

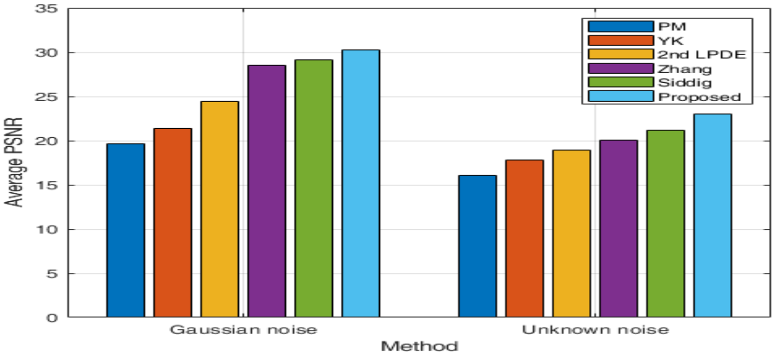


Figure 5: Comparison of average PSNR values in four methods

Acknowledgements

Authors are grateful to there anonymous referees and editor for their constructive comments.

References

1. Awate, S.P., and Whitaker, R.T. *Unsupervised, information-theoretic, adaptive image filtering for image restoration.* IEEE Transactions on pattern analysis and machine intelligence 28, 3 (2006), 364–376.

2. Barash, D. *Fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation.* IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 6 (2002), 844–847.

3. Besser, H. *Visual access to visual images: the UC berkeley image database project.* Library Trends. 38, 1990.

4. Buades, A., Coll, B., and Morel, J.-M. *A review of image denoising algorithms, with a new one.*Multiscale Model. Simul. 4 (2005), 490–530.

5. Chen, Q., Montesinos, P., Sun, Q.S., and Xia, D.S. *Ramp preserving Perona–Malik model.* Signal Processing, 90 (6) (2010), 1963–1975.

6. Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. *Image denoising with block-matching and 3d filtering.* In Image Processing: Algorithms

and Systems, Neural Networks, and Machine Learning (2006), vol. 6064, International Society for Optics and Photonics, p. 606414.

7. Danielyan, A., Katkovnik, V., and Egiazarian, K. *BM3D frames and variational image deblurring.* IEEE Trans. Image Process. 21(4) (2012), 1715–1728.

8. Dautov, Ç.P., and Özerdem, M.S. *Wavelet transform and signal denoising using wavelet method.* 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2018, pp. 1-4.

9. Didas, S., Weickert, J., and Burgeth, B. *Properties of higher order nonlinear diffusion filtering.* J. Math. Imaging Vision 35 (3) (2009), 208–226.

10. Elad, M. *On the origin of the bilateral filter and ways to improve it.*IEEE Trans. Image Process. 11 (10) (2002), 1141–1151.

11. Gabor, D. *Information theory in electron microscopy.* Lab Invest. 14 (1965), 801–807.

12. Greer, J.B., and Bertozzi, A.L. *Traveling wave solutions of fourth order PDEs for image processing.* SIAM J. Math. Anal. 36 (1) (2004), 38–68.

13. Hajiaboli, M.R. *An anisotropic fourth-order diffusion filter for image noise removal.* Int. J. Comput. Vis. 92 (2) (2011), 177–191.

14. Hinze, M., Pinnau, R., Ulbrich, M., and Ulbrich, S. *Optimization with PDE constraints, Mathematical Modelling: Theory and Applications.* 23. Springer, New York, 2009.

15. Jain, A.K. *Partial differential equations and finite-difference methods in image processing, part 1: Image representation.* J. Optim. Theory Appl. 23 (1) (1977), 65–91.

16. Kichenassamy, S. *The Perona-Malik paradox.* SIAM J. Appl. Math. 57 (5) (1997), 1328–1342.

17. Koenderink, J.J. *The structure of images.* Biol. Cybernet. 50 (5) (1984), 363–370.

18. Li, M. *An improved non-local filter for image denoising.* 1–4.

19. Lin, Z., Zhang, W., and Tang, X. *Learning partial differential equations for computer vision.* Peking Univ., Chin. Univ. of Hong Kong 2008.

20. Lions, J.L. *Optimal control of systems governed by partial differential equations* Translated from the French by S. K. Mitter. Die Grundlehren der mathematischen Wissenschaften, Band 170 Springer-Verlag, New York-Berlin 1971.

21. Liu, Q., Yao, Z., and Ke, Y. *Entropy solutions for a fourth-order nonlinear degenerate problem for noise removal.* Nonlinear Anal. 67 (6) (2007), 1908–1918.

22. Liu, R., Lin, Z., Zhang, W., and Su, Z. *Learning PDEs for image restoration via optimal control.* Daniilidis K., Maragos P., Paragios N. (eds) Computer Vision – ECCV 2010. ECCV 2010. 115–128, Lecture Notes in Computer Science, vol 6311. Springer, Berlin, Heidelberg.

23. Liu, R., Zhong, G., Cao, J., Lin, Z., Shan, S., and Luo, Z. *Learning to diffuse: A new perspective to design pdes for visual analysis.* IEEE transactions on pattern analysis and machine intelligence 38 (12) (2016), 2457–2471.

24. Lysaker, M., Lundervold, A., and Tai, X.-C. *Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time.* IEEE Transactions on image processing 12 (12) (2003), 1579–1590.

25. Malfait, M., and Roose, D. *Wavelet-based image denoising using a Markov random field a priori model.* IEEE Transactions on image processing 6 (4) (1997), 549–565.

26. Milanfar, P. *A tour of modern image filtering: New insights and methods, both practical and theoretical.* IEEE signal processing magazine 30 (1) (2012), 106–128.

27. Mo, H., and Li, H. *Image differential invariants.* arXiv:1911.05327 (2019).

28. Own, C., Tsai, H., Yu, P., and Lee, Y. *Adaptive type-2 fuzzy median filter design for removal of impulse noise.* NSIP 2005. Abstracts. IEEE-Eurasip Nonlinear Signal and Image Processing, 2005., Sapporo, Japan, 2005, Imaging Sci. J. 54 (1) (2006), 3–18.

29. Perona, P., and Malik, J. *Scale-space and edge detection using anisotropic diffusion.* IEEE Transactions on pattern analysis and machine intelligence 12 (7) (1990), 629–639.

30. Philip, P. *Optimal control of partial differential equations.* Lecture Notes, Ludwig-Maximilians-Universität, Germany (2009).

31. Rudin, L.I., Osher, S., and Fatemi, E. *Nonlinear total variation based noise removal algorithms.* Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991). Phys. D 60 (1-4) (1992), 259–268.

32. Shewchuk, J.R. *An introduction to the conjugate gradient method without the agonizing pain.* School of computer science. Carnegie Mellon University, Pittsburgh, PA 15213 (1994), 10.

33. Siddig, A., Guo, Z., Zhou, Z., and Wu, B. *An image denoising model based on a fourth-order nonlinear partial differential equation.* Comput. Math. Appl. 76 (5) (2018), 1056–1074.

34. Stoer, J., and Bulirsch, R. *Introduction to numerical analysis*, vol. 12. Springer Science & Business Media, 2013.

35. Tomasi, C., and Manduchi, R. *Bilateral filtering for gray and color images.* Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), Bombay, India, 1998, 839–846.

36. Tröltzsch, F. *Optimal control of partial differential equations: theory, methods, and applications*, Translated from the 2005 German original by Jürgen Sprekels. Graduate Studies in Mathematics, 112. American Mathematical Society, Providence, RI, 2010.

37. Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P. *Image quality assessment: from error visibility to structural similarity.* IEEE transactions on image processing 13 (4) (2004), 600–612.

38. Wang, Y., Chen, W., Zhou, S., Yu, T., and Zhang, Y. *Mtv: modified total variation model for image noise removal.* IEE Electronics Letters, 47 (10) (2011), 592–594.

39. Wang, Y., Guo, J., Chen, W., and Zhang, W. *Image denoising using modified Perona–Malik model based on directional Laplacian.* Signal Processing 93 (9) (2013), 2548–2558.

40. Weber, A.G. *The USC-SIPI image database version 5.* USC-SIPI Report 315, 1 (1997).

41. Witkin, A.P. *Scale-space filtering.* Readings in Computer Vision (1987) 329–332.

42. You, Y.-L., and Kaveh, M. *Fourth-order partial differential equations for noise removal.* IEEE Transactions on Image Processing 9 (10) (2000), 1723–1730.

43. You, Y.-L., Xu, W., Tannenbaum, A., and Kaveh, M. *Behavioral analysis of anisotropic diffusion in image processing.* IEEE Transactions on Image Processing 5 (11) (1996), 1539–1553.

44. Zeng, W., and Lu, X. *A robust variational approach to super- resolution with nonlocal tv regularisation term.* Imaging Sci. J. 61 (2) (2013), 268–278.

45. Zeng, W., Lu, X., and Tan, X. *A local structural adaptive partial differential equation for image denoising.* Multimed. Tools Appl. 74 (3) (2015), 743–757.

46. Zhang, M., and Desrosiers, C. *Image denoising based on sparse representation and gradient histogram.* IET Image Processing 11 (1) (2016), 54–63.

47. Zhang, X., and Ye, W. *An adaptive fourth-order partial differential equation for image denoising.* Comput. Math. Appl. 74 (10) (2017), 2529–2545.