SIAM J. Imaging Sciences
Vol. 14, No. 3, pp. 1078–1113

A Weighted Difference of Anisotropic and Isotropic Total Variation for Relaxed Mumford–Shah Color and Multiphase Image Segmentation*

Kevin Bui[†], Fredrick Park[‡], Yifei Lou[§], and Jack Xin[†]

Abstract. In a class of piecewise-constant image segmentation models, we propose to incorporate a weighted difference of anisotropic and isotropic total variation (AITV) to regularize the partition boundaries in an image. In particular, we replace the total variation regularization in the Chan–Vese segmentation model and a fuzzy region competition model by the proposed AITV. To deal with the nonconvex nature of AITV, we apply the difference-of-convex algorithm (DCA), in which the subproblems can be minimized by the primal-dual hybrid gradient method with linesearch. The convergence of the DCA scheme is analyzed. In addition, a generalization to color image segmentation is discussed. In the numerical experiments, we compare the proposed models with the classic convex approaches and the two-stage segmentation methods (smoothing and then thresholding) on various images, showing that our models are effective in image segmentation and robust with respect to impulsive noises.

Key words. (multiphase) image segmentation, alternating minimization, total variation, difference-of-convex algorithm, primal-dual algorithms

AMS subject classifications. 49M20, 65D18, 65K10, 68U10, 90C90

DOI. 10.1137/20M1337041

---

1. Introduction.

Image segmentation is an important problem in computer vision, where the goal is to partition a given image into salient regions that usually represent specific objects of interest. Each partitioned region has uniform characteristics such as edges, intensities, colors, and textures. Mathematically, given an image $f : \Omega \to \mathbb{R}$, where the image domain Ω is a bounded and open subset of $\mathbb{R}^2$, the aim is to partition Ω into N predetermined number of regions $\{\Omega_i\}_{i=1}^N$ such that $\Omega_i \cap \Omega_j = \emptyset$ for each $i \neq j$ and $\Omega = \bigcup_{i=1}^N \Omega_i$.

In the past two decades, image segmentation has been studied extensively using variational methods and partial differential equations as common and popular methodologies. One class of models, such as the snake model and geodesic contour model, uses edge-detection functions and evolves the curves toward sharp gradients [8, 17, 33, 34]. However, these models are sensitive to noise. As an alternative, region-based models that incorporate region and boundary information are robust to noise. One of the most fundamental region-based models

---

[*] Received by the editors May 11, 2020; accepted for publication (in revised form) March 31, 2021; published electronically July 28, 2021.

https://doi.org/10.1137/20M1337041

Funding: The work of the authors was partially supported by the National Science Foundation grants IIS-1632935, DMS-1846690, DMS-1854434, and DMS-1952644.

[†] Department of Mathematics, University of California at Irvine, Irvine, CA 92697 USA (kevinb3@uci.edu, jxin@math.uci.edu).

[‡] Department of Mathematics & Computer Science, Whittier College, Whittier, CA 90602 USA (fpark@whittier.edu).

[§] Department of Mathematical Science, University of Texas at Dallas, Richardson, TX 75080 USA (yifei.lou@utdallas.edu).

is the Mumford–Shah (MS) model [52], which approximates an image using piecewise-smooth functions. The MS model is formulated as

$$\min_{g,\Gamma} \lambda \int_{\Omega} (f(x,y) - g(x,y))^2 \, dx \, dy + \mu \int_{\Omega \setminus \Gamma} |\nabla g(x,y)|^2 \, dx \, dy + |\Gamma| , \tag{1.1}$$

where $g : \Omega \to \mathbb{R}$ is a smooth approximation of the given image f, $\Gamma = \bigcup_{i=1}^{N} \partial \Omega_i$ is the union of the boundaries of the regions Ω_i, $|\Gamma|$ denotes the arc length of Γ, and λ, μ are positive parameters. Unfortunately, solving (1.1) is extremely complex and difficult because it requires discretizing the unknown set of edges.

Instead of piecewise-smooth functions, the Chan–Vese (CV) model [14] approximates f by piecewise-constant functions with two constant values c_1, c_2 for the regions inside and outside Γ. The CV model is expressed as

$$\min_{c_1,c_2,\Gamma} \lambda \int_{\text{inside}(\Gamma)} (f(x,y) - c_1)^2 \, dx \, dy + \lambda \int_{\text{outside}(\Gamma)} (f(x,y) - c_2)^2 \, dx \, dy \; + \; |\Gamma|. \tag{1.2}$$

Note that CV simultaneously optimizes c_1 and c_2 together with Γ. It is popular to minimize (1.2) via the level-set method [54]. Let $\phi : \Omega \to \mathbb{R}$ be a Lipschitz function such that $\Gamma = \{(x,y) \in \Omega : \phi(x,y) = 0\}$ and

$$\text{inside}(\Gamma) = \{(x,y) \in \Omega : \phi(x,y) > 0\}, \; \text{outside}(\Gamma) = \{(x,y) \in \Omega : \phi(x,y) < 0\}.$$

We denote the Heaviside function

$$H(\phi(x,y)) = \begin{cases} 1 & \text{if } \phi(x,y) \geq 0, \\ 0 & \text{if } \phi(x,y) < 0. \end{cases}$$

The level-set reformulation of (1.2) is

$$\begin{aligned} \min_{\phi,c_1,c_2} \lambda \int_{\Omega} & (f(x,y) - c_1)^2 H(\phi(x,y)) + (f(x,y) - c_2)^2 (1 - H(\phi(x,y))) \, dx \, dy \\ & + \int_{\Omega} |\nabla H(\phi(x,y))| \, dx \, dy. \end{aligned} \tag{1.3}$$

A numerical scheme for (1.3) requires solving the Euler–Lagrange equation for ϕ, followed by updating c_1, c_2 as average intensities inside and outside of Γ, respectively; please see [14, 25] for details. Lie, Lysaker, and Tai [43] introduced a binary level-set formulations of the MS model. Esedoglu and Tsai [22] later developed a more efficient algorithm using the Merriman–Bence–Osher scheme [51]. Chan, Esedoglu, and Nikolova [12] proposed a convex relaxation of the CV model, formulated as

$$\begin{aligned} \min_{u(x,y) \in [0,1], c_1, c_2} \lambda \int_{\Omega} & (f(x,y) - c_1)^2 u(x,y) + (f(x,y) - c_2)^2 (1 - u(x,y)) \, dx \, dy \\ & + \int_{\Omega} |\nabla u(x,y)| \, dx \, dy. \end{aligned} \tag{1.4}$$

The segmented regions can be defined by thresholding u as follows:

$$\text{inside}(\Gamma) = \{(x,y) \in \Omega : u(x,y) > \tau\}, \quad \text{outside}(\Gamma) = \{(x,y) \in \Omega : u(x,y) \leq \tau\}$$

with a chosen constant $\tau \in [0,1]$. Since the objective function in (1.4) is convex with respect to u, it can be minimized using popular convex optimization algorithms, such as split Bregman [27], alternating direction method of multipliers (ADMM) [3, 24], and primal-dual hybrid gradient (PDHG) [11, 23]. As a result, (1.4) inspired various segmentation models [1, 10, 32, 39, 58, 73, 74, 75] that can be solved by convex optimization.

In (1.4), the total variation (TV) term $\|\nabla u\|_1 = \int_\Omega |\nabla u(x,y)|\, dx\, dy$ approximates the length of the curves that partition the segmented regions. Furthermore, it is the tightest convex relaxation of the jump term $\|\nabla u\|_0$, which counts the number of jump discontinuities. When u is piecewise constant, $\|\nabla u\|_0$ is exactly the total arc length of the curves [63]. Unfortunately, minimizing $\|\nabla u\|_0$ is an NP-hard combinatorial problem, and it is often replaced by $\|\nabla u\|_1$ that is algorithmically and theoretically easier to work with. Numerically, $\|\nabla u\|_1$ can be approximated isotropically [60] or anisotropically [16, 21]:

$$J_{\text{iso}}(u) = \int_\Omega \sqrt{|D_x u(x,y)|^2 + |D_y u(x,y)|^2}\, dx\, dy, \tag{1.5}$$

$$J_{\text{ani}}(u) = \int_\Omega |D_x u(x,y)| + |D_y u(x,y)|\, dx\, dy, \tag{1.6}$$

where D_x and D_y denote the horizontal and vertical partial derivative operators, respectively.

In order to better approximate $\|\nabla u\|_0$, we consider the weighted anisotropic-isotropic TV (AITV),

$$\begin{aligned} &J_{\text{ani}}(u) - \alpha J_{\text{iso}}(u) \\ &= \int_\Omega |D_x u(x,y)| + |D_y u(x,y)| - \alpha \sqrt{|D_x u(x,y)|^2 + |D_y u(x,y)|^2}\, dx\, dy \end{aligned} \tag{1.7}$$

with $\alpha \in [0,1]$. The AITV term was inspired by recent successes of $L_1 - L_2$ minimization [20, 45, 46, 47, 71, 72] in compressed sensing. Compared with L_1, L_p for $p \in (0,1)$ [15, 36, 70], and L_0 [65], the $L_1 - L_2$ penalty was shown to have the best performance in recovering sparse solutions when the sensing matrix is highly coherent or violates the restricted isometry property [7]. Figure 1 compares L_0, L_1, and $L_1 - \alpha L_2$ by their contour lines in two dimensions. We observe that as α increases, the contour lines of $L_1 - \alpha L_2$ are bending more inward and closer to the ones of L_0. This phenomenon illustrates that $L_1 - \alpha L_2$ can encourage sparsity, and the constant α acts like a parameter controlling to what extent. By applying $L_1 - \alpha L_2$ on the gradient, Lou et al. [48] proposed AITV with a difference-of-convex algorithm (DCA) [38, 56, 57] for image denoising, deconvolution, and MRI reconstruction. Later, Li et al. [42] demonstrated the robustness of AITV with respect to impulsive noise corruption of the data. Both works [42, 48] showed that AITV preserves sharper image edges than the anisotropic TV. Moreover, AITV is preferred over the isotropic TV that tends to blur oblique edges [2, 18].

As edges are defined by gradient vectors, it is expected that AITV ($L_1 - \alpha L_2$) should produce sparser gradients and maintain sharper edges compared to TV (L_1). A preliminary work

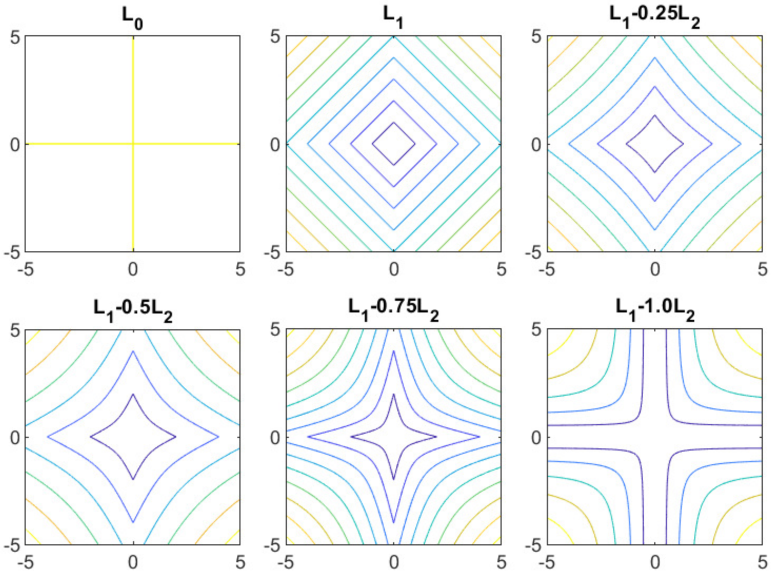


Figure 1. *Contour lines of* $\|x\|_0$ *(*L_0*) and* $\|x\|_1 - \alpha\|x\|_2$ *(*$L_1 - \alpha L_2$*), where* $x \in \mathbb{R}^2$ *and* $\alpha \in \{0, 0.25, 0.5, 0.75, 1.0\}$*. As* α *increases, the contour lines of* $L_1 - \alpha L_2$ *are closer to the ones of* L_0*.*

that replaced $\|\nabla u\|_1$ by AITV in (1.4) was conducted by Park, Lou, and Xin [55], showing better segmentation results than TV. However, this approach was limited to predetermined values of c_1/c_2, grayscale images, and two-phase segmentation (rather than multiphase).

The CV model can be extended to vector-valued images [13] and to multiphase segmentation [4, 66]. The vector-valued extension is straightforward, i.e., replacing f with a vector-valued input $\mathbf{f} : \Omega \to \mathbb{R}^C$ and replacing c_1, c_2 with vector-valued constants $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^C$, where C is the number of channels in an image. The multiphase CV model relies on $\log_2(N)$ level-set functions to partition Ω into N regions $\{\Omega_i\}_{i=1}^N$ and, hence, most CV-based multiphase segmentation methods are limited to a power-of-two number of regions so that $\log_2(N)$ is an integer. There are two approaches that can deal with an arbitrary number of regions. One approach represents each region by a single level-set function [61], which unfortunately causes vacuums and overlapping regions to appear. The other approach defines regions by membership functions, referred to as fuzzy region (FR) competition [40].

In this paper, we propose to incorporate the AITV term into both CV and FR models together with an extension to color image segmentation. To solve these models, we develop an alternating minimization framework that involves DCA and PDHG with linesearch (PDHGLS) [49]. We provide convergence analysis of the proposed algorithms. Experimentally, we compare the proposed models with the classic convex approaches and other segmentation methods to showcase the effectiveness and robustness of the AITV penalty. The major contributions of this work are threefold:

- We study the AITV regularization comprehensively in image segmentation, including grayscale/color image and multiphase segmentation.

- We propose an efficient algorithm that combines DCA and PDHGLS with guaranteed convergence. To the best of our knowledge, this paper pioneers the implementation of PDHGLS in image segmentation.
- We conduct extensive experiments to demonstrate the effect of the constant α in AITV on the segmentation performance and the robustness to impulsive noise. We compare the results with the two-stage segmentation methods.

The paper is organized as follows. Section 2 describes notations that will be used throughout the paper. In section 3, we introduce the AITV extension of the CV model, which can be solved by DCA with convergence analysis. In section 4, we incorporate AITV into the FR model [40] for multiphase segmentation with an algorithm similar to the CV model. In section 5, we extend both CV and FR models to color image segmentation. Numerical results are shown in section 6. Lastly, conclusions and future works are given in section 7.

2. Notations.

For simplicity, we adopt the discrete notations for images and related models. The space $\mathbb{R}^n$ is equipped with the standard inner product $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ and standard Euclidean norm $\|x\|_2 = \sqrt{\langle x, x\rangle}$ for $x, y \in \mathbb{R}^n$.

Without loss of generality, an image is represented as an $m \times n$ matrix, i.e. the image domain is $\Omega = \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$. We denote $X \coloneqq \mathbb{R}^{m\times n}$ and the all-ones matrix in X as $\mathbb{1}$. The vector space X is equipped with following inner product and norm:

$$\langle u, v\rangle_X = \sum_{i=1}^m \sum_{j=1}^n u_{i,j} v_{i,j}, \quad \|u\|_X = \sqrt{\sum_{i=1}^m \sum_{j=1}^n u_{i,j}^2} \qquad \forall u, v \in X.$$

We denote D_x, D_y by the horizontal and vertical partial derivative operators, respectively, i.e.,

$$(D_x u)_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } 1 \leq j \leq n-1, \\ u_{i,1} - u_{i,n} & \text{if } j = n, \end{cases}$$

$$(D_y u)_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } 1 \leq i \leq m-1, \\ u_{1,j} - u_{m,j} & \text{if } i = m. \end{cases}$$

Let $Y \coloneqq X \times X$. Then the discrete gradient operator $D : X \to Y$ is defined as

$$(Du)_{i,j} = ((D_x u)_{i,j}, (D_y u)_{i,j}) \in Y.$$

For any $p = (p_x, p_y), q = (q_x, q_y) \in Y$, the inner product on Y is defined by

$$\langle p, q\rangle_Y = \langle p_x, q_x\rangle_X + \langle p_y, q_y\rangle_X,$$

and the norms on Y are

$$\|p\|_Y = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |(p_x)_{i,j}|^2 + |(p_y)_{i,j}|^2}, \qquad \|p\|_1 = \sum_{i=1}^m \sum_{j=1}^n \left(|(p_x)_{i,j}| + |(p_y)_{i,j}|\right),$$

$$\|p\|_{2,1} = \sum_{i=1}^m \sum_{j=1}^n \sqrt{|(p_x)_{i,j}|^2 + |(p_y)_{i,j}|^2} = \sum_{i=1}^m \sum_{j=1}^n \|((p_x)_{i,j}, (p_y)_{i,j})\|_2.$$

We use a bold letter to denote a three-dimensional tensor, e.g., $\mathbf{u} = (u_1, u_2, \ldots, u_N) \in X^N$. We further denote $\mathbf{u}_{<k} \coloneqq (u_1, \ldots, u_{k-1})$ and $\mathbf{u}_{>k} \coloneqq (u_{k+1}, \ldots, u_N)$ for $1 \leq k \leq N$. The notations $\mathbf{u}_{\leq k}$ and $\mathbf{u}_{\geq k}$ are defined similarly by including u_k. Note that $\mathbf{u}_{<1}$ and $\mathbf{u}_{>N}$ are null or empty variables.

3. Anisotropic-isotropic CV model.

Let $f \in X$ be an observed image. Suppose the image domain Ω has $N = 2^M$ nonoverlapping regions, i.e., $\Omega = \bigcup_{i=1}^N \Omega_i$ and $\Omega_i \cap \Omega_j = \emptyset$ for each $i \neq j$. Let $\mathbf{u} = (u_1, \ldots, u_M) \in X^M$ and $\mathbf{c} = (c_1, \ldots, c_N) \in \mathbb{R}^N$. We propose an AITV-regularized CV (AICV) model for multiphase segmentation as follows:

$$\min_{\substack{\mathbf{u}\in\mathcal{B}\\ \mathbf{c}\in\mathbb{R}^N}} \sum_{k=1}^{M} (\|Du_k\|_1 - \alpha\|Du_k\|_{2,1}) + \lambda \sum_{\ell=1}^{N} \langle f_\ell(\mathbf{c}), R_\ell(\mathbf{u})\rangle_X, \tag{3.1}$$

where $\mathcal{B} = \left\{\mathbf{u} \in X^M : (u_k)_{i,j} \in \{0,1\} \forall i,j,k\right\}$, $f_\ell(\mathbf{c}) = (f - c_\ell \mathbb{1})^2$ with square defined elementwise, and $R_\ell(\mathbf{u})$ is a function of $\mathbf{u}$ related to the region Ω_ℓ such that

$$R_\ell(\mathbf{u})_{i,j} = \begin{cases} 1 & \text{if } (i,j) \in \Omega_\ell, \\ 0 & \text{if } (i,j) \notin \Omega_\ell \end{cases}$$

with $\sum_{\ell=1}^N R_\ell(\mathbf{u}) = \mathbb{1}$. Specifically when $N = 2$ ($M = 1$), we have $R_1(\mathbf{u}) = u_1$ and $R_2(\mathbf{u}) = \mathbb{1} - u_1$. When $N = 4$ ($M = 2$), we have

$$\begin{aligned} R_1(\mathbf{u})_{i,j} &= (u_1)_{i,j}(u_2)_{i,j}, & R_2(\mathbf{u})_{i,j} &= (u_1)_{i,j}[1 - (u_2)_{i,j}], \\ R_3(\mathbf{u})_{i,j} &= [1 - (u_1)_{i,j}](u_2)_{i,j}, & R_4(\mathbf{u})_{i,j} &= [1 - (u_1)_{i,j}][1 - (u_2)_{i,j}]. \end{aligned}$$

When $N = 8$ ($M = 3$), we have

$$\begin{aligned} R_1(\mathbf{u})_{i,j} &= (u_1)_{i,j}(u_2)_{i,j}(u_3)_{i,j}, & R_2(\mathbf{u})_{i,j} &= (u_1)_{i,j}(u_2)_{i,j}[1 - (u_3)_{i,j}], \\ R_3(\mathbf{u})_{i,j} &= (u_1)_{i,j}[1 - (u_2)_{i,j}](u_3)_{i,j}, & R_4(\mathbf{u})_{i,j} &= (u_1)_{i,j}[1 - (u_2)_{i,j}][1 - (u_3)_{i,j}], \\ R_5(\mathbf{u})_{i,j} &= [1 - (u_1)_{i,j}](u_2)_{i,j}(u_3)_{i,j}, & R_6(\mathbf{u})_{i,j} &= [1 - (u_1)_{i,j}](u_2)_{i,j}[1 - (u_3)_{i,j}], \\ R_7(\mathbf{u})_{i,j} &= [1 - (u_1)_{i,j}][1 - (u_2)_{i,j}](u_3)_{i,j}, & R_8(\mathbf{u})_{i,j} &= [1 - (u_1)_{i,j}][1 - (u_2)_{i,j}][1 - (u_3)_{i,j}]. \end{aligned}$$

For $N = 2^M$ with $M \geq 4$, R_ℓ depends on ℓ's binary representation to decide whether to include u_k or $\mathbb{1} - u_k$ as a factor in R_ℓ.

Due to the binary constraint set $\mathcal{B}$, (3.1) is a nonconvex optimization problem, thus numerically difficult to solve. We relax the binary constraint $\{0,1\}$ by a $[0,1]$ box constraint, which in turn has $R_\ell(\mathbf{u})_{i,j} \in [0,1]$. In particular, we rewrite (3.1) as an unconstrained formulation by introducing the indicator function

$$\chi_U(u) = \begin{cases} 0 & \text{if } u_{i,j} \in [0,1] \text{ for all } i,j, \\ +\infty & \text{otherwise.} \end{cases}$$

Hence, a relaxed model of (3.1) can be expressed as

$$\min_{\substack{\mathbf{u}\in X^M\\ \mathbf{c}\in\mathbb{R}^N}} \tilde{F}(\mathbf{u},\mathbf{c}) \coloneqq \sum_{k=1}^{M} \Big(\|Du_k\|_1 - \alpha\|Du_k\|_{2,1} + \chi_U(u_k)\Big) + \lambda \sum_{\ell=1}^{N} \langle f_\ell(\mathbf{c}), R_\ell(\mathbf{u})\rangle_X. \tag{3.2}$$

3.1. Numerical algorithm. We propose an alternating minimization algorithm to find a solution of (3.2) with the following framework:

$$\mathbf{u}^{t+1} \in \arg\min_{\mathbf{u}} \tilde{F}(\mathbf{u}, \mathbf{c}^t), \tag{3.3}$$

$$\mathbf{c}^{t+1} \in \arg\min_{\mathbf{c}} \tilde{F}(\mathbf{u}^{t+1}, \mathbf{c}), \tag{3.4}$$

where t counts the (outer) iterations. Below, we discuss how to solve each subproblem.

We start with the **c**-subproblem (3.4), as it is simpler than the other. Notice that we can solve c_ℓ separately for each $\ell = 1, \dots, N$, i.e.,

$$c_\ell^{t+1} \in \arg\min_{c_\ell} \lambda \langle f_\ell(\mathbf{c}), R_\ell(\mathbf{u}^{t+1})\rangle_X = \arg\min_{c_\ell} \lambda \sum_{i=1}^{m}\sum_{j=1}^{n} (f_{i,j} - c_\ell)^2 R_\ell(\mathbf{u}^{t+1})_{i,j}. \tag{3.5}$$

If $\sum_{i=1}^m \sum_{j=1}^n R_\ell(\mathbf{u}^{t+1})_{i,j} \neq 0$, we differentiate the objective function in (3.5) with respect to c_ℓ, set the derivative equal to zero, and solve for c_ℓ; otherwise, since the objective function does not depend on c_ℓ, the solution can take on any value, so we set the solution to 0 as a default. In summary, there is a closed-form solution to (3.5) for updating c_ℓ^{t+1}, i.e.,

$$c_\ell^{t+1} = \begin{cases} \dfrac{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{i,j} R_\ell(\mathbf{u}^{t+1})_{i,j}}{\sum_{i=1}^{m}\sum_{j=1}^{n} R_\ell(\mathbf{u}^{t+1})_{i,j}} & \text{if } \sum_{i=1}^{m}\sum_{j=1}^{n} R_\ell(\mathbf{u}^{t+1})_{i,j} \neq 0, \\ 0 & \text{if } \sum_{i=1}^{m}\sum_{j=1}^{n} R_\ell(\mathbf{u}^{t+1})_{i,j} = 0. \end{cases} \tag{3.6}$$

The formula (3.6) implies that c_ℓ^{t+1} is the mean intensity value of the region $\Omega_\ell \subset \Omega$ at the $(t+1)$th iteration.

The **u**-subproblem (3.3) is separable with respect to each k, i.e.,

$$u_k^{t+1} \in \arg\min_{u_k} \|Du_k\|_1 - \alpha \|Du_k\|_{2,1} + \chi_U(u_k) + \lambda \langle r_k(\mathbf{c}^t, \mathbf{u}_{<k}^{t+1}, \mathbf{u}_{>k}^t), u_k\rangle_X, \tag{3.7}$$

where $r_k(\mathbf{c}^t, \mathbf{u}_{<k}^{t+1}, \mathbf{u}_{>k}^t)$ is a multivariate polynomial of $(\mathbf{u}_{<k}^{t+1}, \mathbf{u}_{>k}^t)$ obtained by rewriting $\sum_{\ell=1}^N \langle f_\ell(\mathbf{c}), R_\ell(\mathbf{u})\rangle_X$ in (3.2) and getting the coefficients in front of u_k. Because a general form of r_k is complicated, we provide some specific examples in smaller dimensions. When $N = 2$ $(M = 1)$, we have $r_1(\mathbf{c}, \mathbf{u}_{<1}, \mathbf{u}_{>1})_{i,j} = (f_{i,j} - c_1)^2 - (f_{i,j} - c_2)^2$; when $N = 4$ $(M = 2)$, we have

$$\begin{aligned} r_1(\mathbf{c}, \mathbf{u}_{<1}, \mathbf{u}_{>1})_{i,j} &= \left[(f_{i,j} - c_1)^2 - (f_{i,j} - c_2)^2 - (f_{i,j} - c_3)^2 + (f_{i,j} - c_4)^2\right](u_2)_{i,j} \\ &\quad + (f_{i,j} - c_2)^2 - (f_{i,j} - c_4)^2, \\ r_2(\mathbf{c}, \mathbf{u}_{<2}, \mathbf{u}_{>2})_{i,j} &= \left[(f_{i,j} - c_1)^2 - (f_{i,j} - c_2)^2 - (f_{i,j} - c_3)^2 + (f_{i,j} - c_4)^2\right](u_1)_{i,j} \\ &\quad + (f_{i,j} - c_3)^2 - (f_{i,j} - c_4)^2. \end{aligned}$$

In order to minimize (3.7), we apply a descent algorithm called DCA [38, 56, 57] for solving a difference-of-convex (DC) optimization problem of the form $\min_{u\in X} g(u)-h(u)$, where g and h are proper, lower semicontinuous, and strongly convex functions. The algorithm consists of two steps per iteration with u^0 as initialization:

$$\begin{cases} v^t \in \partial h(u^t), \\ u^{t+1} \in \arg\min_{u\in X} g(u) - \langle v^t, u\rangle_X. \end{cases} \tag{3.8}$$

For each $k = 1,\dots,M$, we can express (3.7) as a DC function $g(u_k) - h(u_k)$ with

$$\begin{cases} g(u_k) = \|Du_k\|_1 + \chi_U(u_k) + \lambda\langle r_k(\mathbf{c}^t, \mathbf{u}^{t+1}_{<k}, \mathbf{u}^t_{>k}), u_k\rangle_X + c\|u_k\|_X^2, \\ h(u_k) = \alpha\|Du_k\|_{2,1} + c\|u_k\|_X^2, \end{cases} \tag{3.9}$$

where $c > 0$ enforces strong convexity on the functions g and h. Experimentally, c can be chosen arbitrarily small for better performance. We then compute the subgradient of $h(u)$, i.e.,

$$\alpha \frac{D_x^\top D_x u + D_y^\top D_y u}{\sqrt{|D_x u|^2 + |D_y u|^2}} + 2cu \in \partial h(u).$$

Therefore, the u-subproblem in (3.8) can be expressed as

$$\begin{aligned} u_k^{t+1} = \arg\min_{u_k} &\|Du_k\|_1 + \chi_U(u_k) + \lambda\langle r_k(\mathbf{c}^t, \mathbf{u}^{t+1}_{<k}, \mathbf{u}^t_{>k}), u_k\rangle_X + c\|u_k\|_X^2 \\ &- \alpha\langle Du_k, q_k^t\rangle_Y - 2c\langle u_k, u_k^t\rangle_X, \end{aligned} \tag{3.10}$$

where $q_k^t := ((q_x)_k^t, (q_y)_k^t) = (D_x u_k^t, D_y u_k^t)/\sqrt{|D_x u_k^t|^2 + |D_y u_k^t|^2}$. Note that we compute q_k^t elementwise and adopt the convention that if the denominator is zero at some point, the corresponding q_k^t value is set to zero, which aligns with the subgradient definition. To solve the convex problem (3.10), we apply the PDHG algorithm [11, 23, 77] since it was demonstrated in [11] that PDHG solves imaging models with the TV term [60] efficiently.

In general, the PDHG algorithm [11, 23, 77] targets at a saddle-point problem

$$\min_u \max_v \Psi(u) + \langle Au, v\rangle_Y - \Phi(v),$$

where Ψ, Φ are convex functions and A is a linear operator. The PDHG algorithm is outlined as

$$\begin{aligned} u^{\eta+1} &= (I + \tau\partial\Psi)^{-1}(u^\eta - \tau A^\top v^\eta), \\ \bar{u}^{\eta+1} &= u^{\eta+1} + \theta(u^{\eta+1} - u^\eta), \\ v^{\eta+1} &= (I + \sigma\partial\Phi)^{-1}(v^\eta + \sigma A\bar{u}^{\eta+1}) \end{aligned}$$

with $\tau, \sigma > 0, \theta \in [0,1]$. The inverse is defined by the proximal operator, i.e.,

$$(I + \tau\partial\Psi)^{-1}(z) = \min_u \left(\Psi(u) + \frac{\|u - z\|_X^2}{2\tau}\right),$$

and similarly for $(I + \sigma\partial\Phi)^{-1}$.

In order to apply PDHG for the u_k-problem in (3.10), we define its saddle-point formulation:

$$\begin{aligned}\min_{u_k} \max_{(p_x)_k,(p_y)_k} &\langle D_x u_k, (p_x)_k\rangle_X + \langle D_y u_k, (p_y)_k\rangle_X + \chi_U(u_k)\\ &+ \lambda\langle r_k(\mathbf{c}^t, \mathbf{u}^{t+1}_{<k}, \mathbf{u}^t_{>k}), u_k\rangle_X + c\|u_k\|_X^2 - \alpha\langle Du_k, q_k^t\rangle_Y - 2c\langle u_k, u_k^t\rangle_X\\ &- \chi_P((p_x)_k) - \chi_P((p_y)_k),\end{aligned} \tag{3.11}$$

where $(p_x)_k, (p_y)_k$ are dual variables of $D_x u_k, D_y u_k$, and $P = \{p : |p_{i,j}| \leq 1 \ \forall i,j\}$ is a convex set. Please refer to [9, 11] for the derivation of the saddle-point formulation in more detail. Then we have

$$\begin{aligned}\Psi_{k,t}(u_k) &= \chi_U(u_k) + \lambda\langle r_k(\mathbf{c}^t, \mathbf{u}^{t+1}_{<k}, \mathbf{u}^t_{>k}), u_k\rangle_X + c\|u_k\|_X^2\\ &\quad - \ \alpha\langle Du_k, q_k^t\rangle_Y - 2c\langle u_k, u_k^t\rangle_X,\\ Au_k &= (D_x u_k, D_y u_k),\\ \Phi((p_x)_k, (p_y)_k) &= \chi_P((p_x)_k) + \chi_P((p_y)_k).\end{aligned}$$

With the initial condition $u_k^{t,0} = u_k^t$, the u-subproblem can be computed as

$$\begin{aligned}u_k^{t,\eta+1} &= (I + \tau\partial\Psi_{k,t})^{-1}\left(u_k^{t,\eta} - \tau\left(D_x^\top(p_x)_k^\eta + D_y^\top(p_y)_k^\eta\right)\right)\\ &= \min_{0\leq(u_k)_{i,j}\leq 1}\left\{\lambda\langle r_k(\mathbf{c}^t, \mathbf{u}^{t+1}_{<k}, \mathbf{u}^t_{>k}), u_k\rangle_X + c\|u_k\|_X^2\right.\\ &\qquad - \alpha\langle Du_k, q_k^t\rangle_Y - 2c\langle u_k, u_k^t\rangle_X\\ &\qquad \left. + \frac{\|u_k - \left(u_k^{t,\eta} - \tau\left(D_x^\top(p_x)_k^\eta + D_y^\top(p_y)_k^\eta\right)\right)\|_X^2}{2\tau}\right\},\end{aligned} \tag{3.12}$$

where η indexes the inner iteration, as opposed to t for the outer iteration. To solve (3.12), we derive a closed-form solution that is similar to the one for the u-subproblem of (1.4) determined in [26]. In particular, we observe that the objective function in (3.12) is proper, continuous, and strongly convex with respect to u_k, so it has a unique minimizer. By ignoring the constraint and differentiating the objective function in (3.12) with respect to u_k, we obtain

$$\tilde{u}_k^{t,\eta+1} = \frac{2cu_k^t + \frac{1}{\tau}u_k^{t,\eta}}{2c + \frac{1}{\tau}} - \frac{\lambda r_k(\mathbf{c}^t, \mathbf{u}^{t+1}_{<k}, \mathbf{u}^t_{>k}) - \alpha D^\top q_k^t + (D_x^\top(p_x)_k^\eta + D_y^\top(p_y)_k^\eta)}{2c + \frac{1}{\tau}}.$$

If $(\tilde{u}_k^{t,\eta+1})_{i,j}$ lies in the interval $[0,1]$, then the (i,j)-entry of the unique minimizer also coincides with the minimizer of the constrained problem (3.12). If $(\tilde{u}_k^{t,\eta+1})_{i,j}$ is outside of the interval, then the (i,j)-entry of the unique minimizer lies at the interval endpoint closest to the unconstrained minimizer due to the quadratic objective function. As a result, we project $\tilde{u}_k^{t,\eta+1}$ onto $[0,1]$, leading to a closed-form solution for $u_k^{t,\eta+1}$:

$$u_k^{t,\eta+1} = \min\{\max\{\tilde{u}_k^{t,\eta+1}, 0\}, 1\}, \tag{3.13}$$

where min and max are executed elementwise.

It is straightforward to derive closed-form solutions for $(p_x)_k, (p_y)_k$ in (3.11) given by

$$\begin{aligned}(p_x)_k^{\eta+1} &= \mathrm{Proj}_P((p_x)_k^\eta + \sigma D_x \bar{u}_k^{t,\eta+1}),\\ (p_y)_k^{\eta+1} &= \mathrm{Proj}_P((p_y)_k^\eta + \sigma D_y \bar{u}_k^{t,\eta+1})\end{aligned} \tag{3.14}$$

with $\bar{u}_k^{t,\eta+1} = u_k^{t,\eta+1} + \theta(u_k^{t,\eta+1} - u_k^{t,\eta})$ and $\mathrm{Proj}_P(p) = \frac{p}{\max\{|p|,1\}}$. We see that (3.13) is projected gradient descent of the primal variable u with entrywise box constraint $[0,1]$, while (3.14) is projected gradient ascent of the dual variable (p_x, p_y) that is constrained to the set P. The update order between the primal variable $u_k^{t,\eta}$ and the dual variables $(p_x)_k^\eta, (p_y)_k^\eta$ does not matter for PDHG [11, 49]. To further improve the speed and solution quality of PDHG, we incorporate a linesearch technique [49] that starts with the primal variable, followed by the dual update. The PDHG algorithm with linesearch is referred to as PDHGLS. Both PDHG and PDHGLS provide a saddle-point solution $(u_k^*, (p_x)_k^*, (p_y)_k^*)$ for (3.11) upon convergence [11, 49]. Since (3.10) is convex, u_k^* is indeed its solution, independent of the choice between using PDHG or PDHGLS. We summarize the proposed DCA-PDHGLS algorithm to solve (3.2) in Algorithm 3.1.

3.2. Convergence analysis. We analyze the convergence of the sequence $\{(\mathbf{u}^t, \mathbf{c}^t)\}_{t=1}^\infty$ generated by (3.3) and (3.4), which are solved by (3.10) and (3.6), respectively. We establish in Lemma 3.1 that the sequence $\{\tilde{F}(\mathbf{u}^t, \mathbf{c}^t)\}_{t=1}^\infty$ decreases sufficiently, followed by the convergence result in Theorem 3.2.

Lemma 3.1. *Suppose $\alpha \in [0,1]$ and $\lambda > 0$. Let $\{(\mathbf{u}^t, \mathbf{c}^t)\}_{t=1}^\infty$ be a sequence such that $\mathbf{u}^t$ is generated by* (3.10) *and $\mathbf{c}^t$ is generated by* (3.6)*. Then we have*

$$\tilde{F}(\mathbf{u}^t, \mathbf{c}^t) - \tilde{F}(\mathbf{u}^{t+1}, \mathbf{c}^{t+1}) \geq 2c \sum_{k=1}^M \|u_k^t - u_k^{t+1}\|_X^2.$$

Proof. Since $\mathbf{c}^{t+1}$ satisfies (3.6), we have

$$\tilde{F}(\mathbf{u}^{t+1}, \mathbf{c}^{t+1}) \leq \tilde{F}(\mathbf{u}^{t+1}, \mathbf{c}^t). \tag{3.15}$$

Then we estimate

$$\begin{aligned}&\tilde{F}((\mathbf{u}_{\leq k-1}^{t+1}, \mathbf{u}_{\geq k}^t), \mathbf{c}^t) - \tilde{F}((\mathbf{u}_{\leq k}^{t+1}, \mathbf{u}_{\geq k+1}^t), \mathbf{c}^t)\\ =\ &\|Du_k^t\|_1 - \|Du_k^{t+1}\|_1 - \alpha(\|Du_k^t\|_{2,1} - \|Du_k^{t+1}\|_{2,1}) + \chi_U(u_k^t) - \chi_U(u_k^{t+1})\\ &+ \lambda \sum_{\ell=1}^N \langle f_\ell(\mathbf{c}), R_\ell(\mathbf{u}_{\leq k-1}^{t+1}, \mathbf{u}_{\geq k}^t) - R_\ell(\mathbf{u}_{\leq k}^{t+1}, \mathbf{u}_{\geq k+1}^t)\rangle_X.\end{aligned} \tag{3.16}$$

It follows from the first-order optimality condition of (3.10) at u_k^{t+1} that there exists $p_k^{t+1} \in \partial\left(\|Du_k^{t+1}\|_1 + \chi_U(u_k^{t+1})\right)$ such that

$$0 = p_k^{t+1} - \alpha D^\top q_k^t + 2c(u_k^{t+1} - u_k^t) + \lambda r_k(\mathbf{c}^t, \mathbf{u}_{<k}^{t+1}, \mathbf{u}_{>k}^t).$$

Algorithm 3.1 DCA-PDHGLS algorithm to solve (3.2)

```
Input:
    • Image $f$
    • model parameters $\alpha, \lambda > 0$
    • strong convexity parameter $c > 0$
    • PDHGLS initial step size $\tau_0 > 0$
    • PDHGLS primal-dual step size ratio $\beta > 0$
    • PDHGLS parameter $\delta \in (0, 1)$
    • PDHGLS step size multiplier $\mu \in (0, 1)$
 1: Set $u_k^0 = 1$ $(k = 1, \dots, M)$ for some region $\Sigma \subset \Omega$ and 0 elsewhere.
 2: Compute $\mathbf{c}^0 = (c_1^0, \dots, c_N^0)$ by (3.6).
 3: Set $t := 0$.
 4: while stopping criterion for DCA is not satisfied do
 5:     for $k = 1$ to $M$ do
 6:         Set $u_k^{t,0} := u_k^t$ and $(p_x)_k^0 = (p_y)_k^0 = 0$.
 7:         Compute $((q_x)_k^t, (q_y)_k^t) = (D_x u_k^t, D_y u_k^t)/\sqrt{|D_x u_k^t|^2 + |D_y u_k^t|^2}$.
 8:         Set $\theta_0 = 1$.
 9:         Set $\eta := 0$.
10:         while stopping criterion for PDHGLS is not satisfied do
11:             Compute $u_k^{t,\eta+1}$ by (3.13) with $\tau := \tau_\eta$.
12:             Set $\tau_{\eta+1} = \tau_\eta \sqrt{1 + \theta_\eta}$.
                Linesearch:
13:             Compute $\theta_{\eta+1} = \frac{\tau_{\eta+1}}{\tau_\eta}$ and $\sigma_{\eta+1} = \beta \tau_{\eta+1}$.
14:             Compute $\bar{u}_k^{t,\eta+1} = u_k^{t,\eta+1} + \theta_{\eta+1}(u_k^{t,\eta+1} - u_k^{t,\eta})$.
15:             Compute $p_k^{\eta+1} := ((p_x)_k^{\eta+1}, (p_y)_k^{\eta+1})$ by (3.14) with $\sigma := \sigma_{\eta+1}$.
16:             if $\sqrt{\beta}\tau_{\eta+1} \|(D_x^\top (p_x)_k^{\eta+1}, D_y^\top (p_y)_k^{\eta+1}) - (D_x^\top (p_x)_k^\eta, D_y^\top (p_y)_k^\eta)\|_Y \leq \delta \|p_k^{\eta+1} - p_k^\eta\|_Y$
                then
17:                 Set $\eta := \eta + 1$, and break linesearch
18:             else
19:                 Set $\tau_{\eta+1} := \mu \tau_{\eta+1}$ and go back to line 13.
20:             end if
                End of linesearch
21:         end while
22:         Set $u_k^{t+1} := u_k^{t,\eta}$.
23:     end for
24:     Compute $\mathbf{c}^{t+1}$ by (3.6).
25:     Set $t := t + 1$.
26: end while
Output: $(\mathbf{u}, \mathbf{c}) := (\mathbf{u}^t, \mathbf{c}^t)$.
```

Taking the inner product with $u_k^t - u_k^{t+1}$ and rearranging it, we obtain

$$\begin{aligned}&\lambda\langle r_k(\mathbf{c}^t,\mathbf{u}^{t+1}_{<k},\mathbf{u}^t_{>k}), u_k^t-u_k^{t+1}\rangle_X\\ &=-\langle p_k^{t+1}-\alpha D^\top q_k^t, u_k^t-u_k^{t+1}\rangle_X+2c\|u_k^{t+1}-u_k^t\|_X^2.\end{aligned}\tag{3.17}$$

The last term in (3.16) can be simplified to

$$\sum_{\ell=1}^{N}\langle f_\ell, R_\ell(\mathbf{u}^{t+1}_{\leq k-1},\mathbf{u}^t_{\geq k})-R_\ell(\mathbf{u}^{t+1}_{\leq k},\mathbf{u}^t_{\geq k+1})\rangle_X=\langle r_k(\mathbf{c}^t,\mathbf{u}^{t+1}_{<k},\mathbf{u}^t_{>k}),u_k^t-u_k^{t+1}\rangle_X,$$

as $R_\ell(\mathbf{u})$ consists of terms with at most one u_k, and the terms without u_k^t and u_k^{t+1} are canceled out. Together with (3.16) and (3.17), we get

$$\begin{aligned}&\tilde{F}((\mathbf{u}^{t+1}_{\leq k-1},\mathbf{u}^t_{\geq k}),\mathbf{c}^t)-\tilde{F}((\mathbf{u}^{t+1}_{\leq k},\mathbf{u}^t_{\geq k+1}),\mathbf{c}^t)\\
&=\|Du_k^t\|_1-\|Du_k^{t+1}\|_1-\alpha(\|Du_k^t\|_{2,1}-\|Du_k^{t+1}\|_{2,1})\\
&\quad+\chi_U(u_k^t)-\chi_U(u_k^{t+1})+\lambda\langle r_k(\mathbf{c}^t,\mathbf{u}^{t+1}_{<k},\mathbf{u}^t_{>k}),u_k^t-u_k^{t+1}\rangle_X\\
&=\|Du_k^t\|_1-\|Du_k^{t+1}\|_1-\alpha(\|Du_k^t\|_{2,1}-\|Du_k^{t+1}\|_{2,1})\\
&\quad+\chi_U(u_k^t)-\chi_U(u_k^{t+1})-\langle p_k^{t+1}-\alpha D^\top q_k^t,u_k^t-u_k^{t+1}\rangle_X+2c\|u_k^{t+1}-u_k^t\|_X^2\\
&=\left[\left(\|Du_k^t\|_1-\langle p_k^{t+1},u_k^t-u_k^{t+1}\rangle_X+\chi_U(u_k^t)\right)-\|Du_k^{t+1}\|_1-\chi_U(u_k^{t+1})\right]\\
&\quad+\alpha(\|Du_k^{t+1}\|_{2,1}-\langle D^\top q_k^t,u_k^t-u_k^{t+1}\rangle_X-\|Du_k^t\|_{2,1})+2c\|u_k^{t+1}-u_k^t\|_X^2.\end{aligned}\tag{3.18}$$

The definitions of convexity and subgradient yield that

$$\|Du_k^t\|_1+\chi_U(u_k^t)-\langle p_k^{t+1},u_k^t-u_k^{t+1}\rangle_X\geq\|Du_k^{t+1}\|_1+\chi_U(u_k^{t+1}),\tag{3.19}$$

$$\|Du_k^{t+1}\|_{2,1}-\langle D^\top q_k^t,u_k^{t+1}-u_k^t\rangle_X\geq\|Du_k^t\|_{2,1}.\tag{3.20}$$

Combining (3.18)–(3.20), we have

$$\tilde{F}((\mathbf{u}^{t+1}_{\leq k-1},\mathbf{u}^t_{\geq k}),\mathbf{c}^t)-\tilde{F}((\mathbf{u}^{t+1}_{\leq k},\mathbf{u}^t_{\geq k+1}),\mathbf{c}^t)\geq 2c\|u_k^{t+1}-u_k^t\|_X^2.$$

Summing over $k=1,\ldots,M$ leads to

$$\begin{aligned}\tilde{F}(\mathbf{u}^t,\mathbf{c}^t)-\tilde{F}(\mathbf{u}^{t+1},\mathbf{c}^t)&=\sum_{k=1}^{M}\tilde{F}((\mathbf{u}^{t+1}_{\leq k-1},\mathbf{u}^t_{\geq k}),\mathbf{c}^t)-\tilde{F}((\mathbf{u}^{t+1}_{\leq k},\mathbf{u}^t_{\geq k+1}),\mathbf{c}^t)\\
&\geq 2c\sum_{k=1}^{M}\|u_k^{t+1}-u_k^t\|_X^2.\end{aligned}\tag{3.21}$$

Therefore, (3.15) and (3.21) establish the desired result. ■

Theorem 3.2. *Suppose* $\alpha\in[0,1]$ *and* $\lambda>0$. *Let* $\{(\mathbf{u}^t,\mathbf{c}^t)\}_{t=1}^{\infty}$ *be a sequence such that* $\mathbf{u}^t$ *is generated by* (3.10) *and* $\mathbf{c}^t$ *is generated by* (3.6). *We have the following:*

(a) $\{(\mathbf{u}^t,\mathbf{c}^t)\}_{t=1}^{\infty}$ *is bounded.*

(b) *For* $k=1,\ldots,M$, *we have* $\|u_k^{t+1}-u_k^t\|_X\to 0$ *as* $t\to\infty$.

(c) *The sequence* $\{(\mathbf{u}^t, \mathbf{c}^t)\}_{t=1}^{\infty}$ *has a limit point* $(\mathbf{u}^*, \mathbf{c}^*)$ *satisfying*

$$\mathbf{0} \in \partial\|Du_k^*\|_1 - \alpha\partial\|Du_k^*\|_{2,1} + \partial\chi_U(u_k^*) + \lambda r_k(\mathbf{c}^*, \mathbf{u}_{<k}^*, \mathbf{u}_{>k}^*) \tag{3.22}$$

for $k = 1, \dots, M$, *and*

$$0 \in \frac{\partial \tilde{F}(\mathbf{u}^*, \mathbf{c}^*)}{\partial c_\ell}, \quad \ell = 1, \dots, N. \tag{3.23}$$

Proof. (a) As each entry of u_k^t is bounded by $[0,1]$ for $k = 1, \dots, M$, $\{\mathbf{u}^t\}_{t=1}^{\infty}$ is a bounded sequence. It further follows from (3.6) that $0 \leq |c_\ell^{t+1}| \leq \max_{i,j} |f_{i,j}|$. Therefore, $\{\mathbf{c}^t\}_{t=1}^{\infty}$ is also bounded, and, altogether, so is the sequence $\{(\mathbf{u}^t, \mathbf{c}^t)\}_{t=1}^{\infty}$.

(b) Since $\alpha\|Du_k\|_{2,1} \leq \|Du_k\|_1$ for $\alpha \in [0,1]$, we have

$$\tilde{F}(\mathbf{u}, \mathbf{c}) \geq \sum_{k=1}^{M} \chi_U(u_k) + \lambda \sum_{\ell=1}^{N} \langle f_\ell, R_\ell(\mathbf{u})\rangle_X \geq 0, \tag{3.24}$$

which implies that $\tilde{F}(\mathbf{u}, \mathbf{c})$ is lower bounded. As it is also decreasing by Lemma 3.1, the sequence $\{\tilde{F}(\mathbf{u}^t, \mathbf{c}^t)\}_{t=1}^{\infty}$ converges. By a telescope summation of (3.21), we obtain

$$\tilde{F}(\mathbf{u}^1, \mathbf{c}^1) - \lim_{t\to\infty} \tilde{F}(\mathbf{u}^t, \mathbf{c}^t) \geq 2c \sum_{t=1}^{\infty}\sum_{k=1}^{M} \|u_k^t - u_k^{t+1}\|_X^2 = 2c \sum_{k=1}^{M}\sum_{t=1}^{\infty} \|u_k^t - u_k^{t+1}\|_X^2.$$

Therefore, $\sum_{t=1}^{\infty} \|u_k^t - u_k^{t+1}\|_X^2 < \infty$, leading to $\lim_{t\to\infty} \|u_k^t - u_k^{t+1}\|_X^2 = 0$ for $k = 1, \dots, M$.

(c) By the Bolzano–Weierstrass theorem, the bounded sequence $\{(\mathbf{u}^t, \mathbf{c}^t)\}_{t=1}^{\infty}$ has a convergent subsequence $\{(\mathbf{u}^{t_L}, \mathbf{c}^{t_L})\}_{L=1}^{\infty}$ such that $\lim_{L\to\infty} (\mathbf{u}^{t_L}, \mathbf{c}^{t_L}) = (\mathbf{u}^*, \mathbf{c}^*)$. By (b), $\lim_{L\to\infty} u_k^{t_L+1} - u_k^{t_L} = 0$. As $\lim_{L\to\infty} u_k^{t_L+1} = \lim_{L\to\infty} u_k^{t_L} = u_k^*$, we have $\lim_{L\to\infty} \mathbf{u}^{t_L+1} = \mathbf{u}^*$. Since $\mathbf{u}^{t_L}$ is generated by (3.10), all of its entries are bounded by $[0,1]$; otherwise, the objective function would be at $+\infty$. Hence, $\chi_U(u_k^{t_L}) = 0$ and similarly $\chi_U(u_k^{t_L+1}) = 0$ for all L, from which follows that $\chi_U(u_k^*) = 0$. In short, we have

$$\lim_{L\to\infty} \chi_U(u_k^{t_L}) = \chi_U(u_k^*) \quad \text{for } k = 1, \dots, M. \tag{3.25}$$

Now we establish (3.23) by showing that $\tilde{F}(\mathbf{u}^*, \mathbf{c}^*) \leq \tilde{F}(\mathbf{u}^*, \mathbf{c})$ for all $\mathbf{c} \in \mathbb{R}^n$. On one hand, we have

$$\begin{aligned}
&\lim_{L\to\infty} \tilde{F}(\mathbf{u}^{t_L}, \mathbf{c}^{t_L}) \\
&= \lim_{L\to\infty} \left[\sum_{k=1}^{M} \left(\|Du_k^{t_L}\|_1 - \alpha\|Du_k^{t_L}\|_{2,1} + \chi_U(u_k^{t_L})\right) + \lambda \sum_{\ell=1}^{N} \langle f_\ell(\mathbf{c}^{t_L}), R_\ell(\mathbf{u}^{t_L})\rangle_X\right] \\
&= \sum_{k=1}^{M} \lim_{L\to\infty} \left(\|Du_k^{t_L}\|_1 - \alpha\|Du_k^{t_L}\|_{2,1} + \chi_U(u_k^{t_L})\right) + \lambda \sum_{\ell=1}^{N} \lim_{L\to\infty} \langle f_\ell(\mathbf{c}^{t_L}), R_\ell(\mathbf{u}^{t_L})\rangle_X \\
&= \sum_{k=1}^{M} (\|Du_k^*\|_1 - \alpha\|Du_k^*\|_{2,1} + \chi_U(u_k^*)) + \lambda \sum_{\ell=1}^{N} \langle f_\ell(\mathbf{c}^*), R_\ell(\mathbf{u}^*)\rangle_X = \tilde{F}(\mathbf{u}^*, \mathbf{c}^*).
\end{aligned} \tag{3.26}$$

We can take the limit as all the terms of $\tilde{F}$ except for χ_U are continuous with respect to $(\mathbf{u}, \mathbf{c})$. On the other hand, we have

$$\begin{aligned}
&\lim_{L\to\infty} \tilde{F}(\mathbf{u}^{t_L}, \mathbf{c}) \\
&= \lim_{L\to\infty} \left[\sum_{k=1}^{M} \left(\|Du_k^{t_L}\|_1 - \alpha \|Du_k^{t_L}\|_{2,1} + \chi_U(u_k^{t_L}) \right) + \lambda \sum_{\ell=1}^{N} \langle f_\ell(\mathbf{c}), R_\ell(\mathbf{u}^{t_L}) \rangle_X \right] \\
&= \sum_{k=1}^{M} \lim_{L\to\infty} \left(\|Du_k^{t_L}\|_1 - \alpha \|Du_k\|_{2,1} + \chi_U(u_k^{t_L}) \right) + \lambda \sum_{\ell=1}^{N} \lim_{L\to\infty} \langle f_\ell(\mathbf{c}), R_\ell(\mathbf{u}^{t_L}) \rangle_X \\
&= \sum_{k=1}^{M} (\|Du_k^*\|_1 - \alpha \|Du_k^*\|_{2,1} + \chi_U(u_k^*)) + \lambda \sum_{\ell=1}^{N} \langle f_\ell(\mathbf{c}), R_\ell(\mathbf{u}^*) \rangle_X = \tilde{F}(\mathbf{u}^*, \mathbf{c}).
\end{aligned} \tag{3.27}$$

It follows from (3.4) that for all $L \in \mathbb{N}$, we have

$$\tilde{F}(\mathbf{u}^{t_L}, \mathbf{c}^{t_L}) \leq \tilde{F}(\mathbf{u}^{t_L}, \mathbf{c}) \quad \forall\, \mathbf{c} \in \mathbb{R}^N. \tag{3.28}$$

Combined with (3.26)–(3.27),

$$\tilde{F}(\mathbf{u}^*, \mathbf{c}^*) = \lim_{L\to\infty} \tilde{F}(\mathbf{u}^{t_L}, \mathbf{c}^{t_L}) \leq \lim_{L\to\infty} \tilde{F}(\mathbf{u}^{t_L}, \mathbf{c}) = \tilde{F}(\mathbf{u}^*, \mathbf{c}) \quad \forall\, \mathbf{c} \in \mathbb{R}^N$$

or, equivalently, $\tilde{F}(\mathbf{u}^*, \mathbf{c}^*) = \inf_{\mathbf{c}\in\mathbb{R}^N} \tilde{F}(\mathbf{u}^*, \mathbf{c})$. The minimization with respect to $\mathbf{c}$ can be expressed elementwise for each c_ℓ, leading to the optimality condition of (3.23).

For the rest of the proof, we establish (3.22). For each $k = 1, \ldots, M$, the optimality condition at the $(t_L + 1)$th step of (3.10) is

$$\begin{aligned}
\mathbf{0} \in\ & \partial(\|Du_k^{t_L+1}\|_1 + \chi_U(u_k^{t_L+1})) + \lambda r_k(\mathbf{c}^{t_L}, \mathbf{u}_{<k}^{t_L+1}, \mathbf{u}_{>k}^{t_L}) + 2c(u_k^{t_L+1} - u_k^{t_L}) \\
& - \alpha D^\top q_k^{t_L}.
\end{aligned} \tag{3.29}$$

Denote $s_k^L \coloneqq -\lambda r_k(\mathbf{c}^{t_L}, \mathbf{u}_{<k}^{t_L+1}, \mathbf{u}_k^{t_L}) - 2c(u_k^{t_L+1} - u_k^{t_L}) + \alpha D^\top q_k^{t_L}$. Then (3.29) implies that

$$s_k^L \in \partial(\|Du_k^{t_L+1}\|_1 + \chi_U(u_k^{t_L+1})). \tag{3.30}$$

Since $r_k(\mathbf{c}, \mathbf{u}_{<k}, \mathbf{u}_{>k})$ is continuous in $(\mathbf{c}, \mathbf{u}_{<k}, \mathbf{u}_{>k})$, we have

$$\lim_{L\to\infty} r_k(\mathbf{c}^{t_L}, \mathbf{u}_{<k}^{t_L+1}, \mathbf{u}_{>k}^{t_L}) = r_k(\mathbf{c}^*, \mathbf{u}_{<k}^*, \mathbf{u}_{>k}^*).$$

To compute the limit of $D^\top q_k^{t_L}$, we recall the multivariate subgradient of $\partial\|Du_k\|_{2,1} = \prod_{(i,j)} \partial\|(Du_k)_{i,j}\|_2$, where

$$\partial\|(x_1, x_2)\|_2 = \begin{cases} \left\{ \frac{(x_1, x_2)}{\sqrt{x_1^2 + x_2^2}} \right\} & \text{if } (x_1, x_2) \neq (0,0) \in \mathbb{R}^2, \\ \{(y_1, y_2) \in \mathbb{R}^2 : y_1^2 + y_2^2 \leq 1\} & \text{if } (x_1, x_2) = (0,0). \end{cases}$$

Let $((v_{x,k}^*)_{i,j}, (v_{y,k}^*)_{i,j}) \coloneqq ((D_x u_k^*)_{i,j}, (D_y u_k^*)_{i,j})$ be the discrete gradient of u_k^* at entry (i,j) for $k = 1, \dots, M$, which satisfies

$$\begin{aligned}\partial \|(v_{x,k}^*)_{i,j}, (v_{y,k}^*)_{i,j}\|_2 \\ &= \begin{cases} \left\{ \frac{((v_{x,k}^*)_{i,j}, (v_{y,k}^*)_{i,j})}{\sqrt{|(v_{x,k}^*)_{i,j}|^2 + |(v_{y,k}^*)_{i,j}|^2}} \right\} & \text{if } ((v_{x,k}^*)_{i,j}, (v_{y,k}^*)_{i,j}) \neq (0,0), \\ \{(y_1, y_2) \in \mathbb{R}^2 : y_1^2 + y_2^2 \leq 1\} & \text{if } ((v_{x,k}^*)_{i,j}, (v_{y,k}^*)_{i,j}) = (0,0). \end{cases}\end{aligned}$$

Note that we define $q_k^{t_L}$ in the following way:

$$(q_k^{t_L})_{i,j} = \begin{cases} \frac{((D_x u_k^{t_L})_{i,j}, (D_y u_k^{t_L})_{i,j})}{\sqrt{|(D_x u_k^{t_L})_{i,j}|^2 + |(D_y u_k^{t_L})_{i,j}|^2}} & \text{if } ((D_x u_k^{t_L})_{i,j}, (D_y u_k^{t_L})_{i,j}) \neq (0,0), \\ (0,0) & \text{if } ((D_x u_k^{t_L})_{i,j}, (D_y u_k^{t_L})_{i,j}) = (0,0). \end{cases} \tag{3.31}$$

Denote $q_k^* \coloneqq \lim_{L \to \infty} q_k^{t_L}$. Therefore, by (3.31), when $((v_x^*)_{i,j}, (v_y^*)_{i,j}) \neq (0,0)$, we have

$$(q_k^*)_{i,j} = \lim_{L \to \infty} (q_k^{t_L})_{i,j} = \frac{((v_{x,k}^*)_{i,j}, (v_{y,k}^*)_{i,j})}{\sqrt{|(v_{x,k}^*)_{i,j}|^2 + |(v_{y,k}^*)_{i,j}|^2}} \in \partial \|((v_{x,k}^*)_{i,j}, (v_{y,k}^*)_{i,j})\|_2$$

and when $((v_x^*)_{i,j}, (v_y^*)_{i,j}) = (0,0)$, we have

$$(q_k^{t_L})_{i,j} \in \{(y_1, y_2) \in \mathbb{R}^2 : y_1^2 + y_2^2 \leq 1\} \subseteq \partial \|((v_{x,k}^*)_{i,j}, (v_{y,k}^*)_{i,j})\|_2,$$

for all $L \in \mathbb{N}$ so that taking the limit $L \to \infty$ yields $(q_k^*)_{i,j} \in \partial \|((v_{x,k}^*)_{i,j}, (v_{y,k}^*)_{i,j})\|_2$. By the chain rule of the subgradient (Corollary 16 in [28]), we have

$$\partial \|(Du_k^*)_{i,j}\|_2 = D^\top \partial \|((v_{x,k}^*)_{i,j}, (v_{y,k}^*)_{i,j})\|_2.$$

Since $D^\top$ is a linear operator (thus continuous), we get

$$\lim_{L \to \infty} D^\top q_k^{t_L} = D^\top q_k^* \in \partial \|Du_k^*\|_{2,1}. \tag{3.32}$$

In short, we obtain that $s_k^* \coloneqq \lim_{L \to \infty} s_k^L = -\lambda r_k(\mathbf{c}^*, \mathbf{u}_{<k}^*, \mathbf{u}_{>k}^*) + \alpha D^\top q_k^*$.

It further follows from (3.30) and the subgradient definition that

$$\begin{aligned} \|Du_k\|_1 + \chi_U(u_k) &\geq \|Du_k^{t_L+1}\|_1 + \chi_U(u_k^{t_L+1}) + \langle s_k^L, u_k - u_k^{t_L+1} \rangle \\ &= \|Du_k^{t_L+1}\|_1 + \langle s_k^L, u_k - u_k^{t_L+1} \rangle \end{aligned} \tag{3.33}$$

for all $u_k \in X$ and $L \in \mathbb{N}$. By continuity, we obtain

$$\begin{aligned} \|Du_k\|_1 + \chi_U(u_k) &\geq \lim_{L \to \infty} \left(\|Du_k^{t_L+1}\|_1 + \langle s_k^L, u_k - u_k^{t_L+1} \rangle \right) \\ &= \|Du_k^*\|_1 + \langle s_k^*, u_k - u_k^* \rangle = \|Du_k^*\|_1 + \chi_U(u_k^*) + \langle s_k^*, u_k - u_k^* \rangle, \end{aligned}$$

where the last equality is due to $\chi_U(u_k^*) = 0$. Since both $\|Du\|_1$ and $\chi_U(u)$ are convex, $s_k^* \in \partial(\|Du_k^*\|_1 + \chi_U(u_k^*)) = \partial\|Du_k^*\|_1 + \partial\chi_U(u_k^*)$. Therefore, we have

$$\begin{aligned}\mathbf{0} &\in \partial\|Du_k^*\|_1 + \partial\chi_U(u_k^*) + \lambda r_k(\mathbf{c}^*, \mathbf{u}^*_{<k}, \mathbf{u}^*_{>k}) - \alpha D^\top q_k^* \\ &\subseteq \partial\|Du_k^*\|_1 - \alpha\partial\|Du_k^*\|_{2,1} + \partial\chi_U(u_k^*) + \lambda r_k(\mathbf{c}^*, \mathbf{u}^*_{<k}, \mathbf{u}^*_{>k}).\end{aligned}$$

This concludes the proof. ■

Remark 3.3. The limit point $(\mathbf{u}^*, \mathbf{c}^*)$ is not guaranteed to be a global optimal solution for (3.2) because the objective function is nonconvex, and $(\mathbf{u}^*, \mathbf{c}^*)$ may not even satisfy a first-order optimality condition $\mathbf{0} \in \partial_{(\mathbf{u},\mathbf{c})}\tilde{F}(\mathbf{u}^*, \mathbf{c}^*)$. However, according to Theorem 3.2(c), each coordinate u_k^* or c_ℓ^* satisfies its respective first-order optimality condition, since $(\mathbf{u}^*, \mathbf{c}^*) = (u_1^*, \dots, u_M^*, c_1^*, \dots, c_N^*)$. In convex optimization, if g is convex, a point x^* is a critical point if $0 \in \partial g(x^*)$. (3.23) establishes c_ℓ^* to be a critical point of the function convex in c_ℓ,

$$\sum_{i=1}^m \sum_{j=1}^n (f_{i,j} - c_\ell)^2 R_\ell(\mathbf{u})_{i,j},$$

which is derived from (3.2) when minimizing for c_ℓ. In DC optimization, a point x^* is a critical point of DC function $g-h$ if $0 \in \partial g(u^*) - \partial h(u^*)$ [38]. However, this optimality condition is not as strong as the optimality condition $0 \in \partial(g-h)(u^*)$ because $\partial(g-h)(u^*) \subset \partial g(u^*) - \partial h(u^*)$ in terms of either the Clarke subdifferential or the Fréchet subdifferential [38]. (3.22) establishes u_k^* to be a DC critical point of the DC function

$$\underbrace{\|Du_k\|_1 + \chi_U(u_k) + \lambda\langle r_k(\mathbf{c}, \mathbf{u}_{<k}, \mathbf{u}_{>k}), u_k\rangle_X}_{g(u_k)} - \underbrace{\alpha\|Du_k\|_{2,1}}_{h(u_k)},$$

which is derived from (3.2) when minimizing for u_k.

4. Fuzzy extension of the AICV model.

One limitation of the CV models is that they are only applicable for image segmentation that has specifically a power-of-two number (i.e., 2^M) of regions. To generalize to an arbitrary number of regions N, we associate each region Ω_ℓ with a membership function u_ℓ for $\ell = 1, \dots, N$. A membership function u_ℓ represents a region Ω_ℓ in the following way:

$$(u_\ell)_{i,j} = \begin{cases} 1 & \text{if } (i,j) \in \Omega_\ell, \\ 0 & \text{if } (i,j) \notin \Omega_\ell. \end{cases}$$

To avoid overlap between the u_ℓ's, we enforce the constraint $\sum_{\ell=1}^N u_\ell = \mathbb{1}$, but we relax it with a quadratic penalty to make the model numerically tractable. As such, we propose an AITV extension to the FR model, referred to as AIFR,

$$\begin{aligned}\min_{\substack{\mathbf{u}\in X^N\\ \mathbf{c}\in\mathbb{R}^N}} \hat{F}(\mathbf{u}, \mathbf{c}) &:= \sum_{\ell=1}^N (\|Du_\ell\|_1 - \alpha\|Du_\ell\|_{2,1} + \chi_U(u_\ell)) + \lambda\sum_{\ell=1}^N \langle f_\ell(\mathbf{c}), u_\ell\rangle_X \\ &\quad + \frac{\nu}{2}\left\|\sum_{\ell=1}^N u_\ell - \mathbb{1}\right\|_X^2 \end{aligned} \tag{4.1}$$

with $\nu > 0$. Similarly to (3.3)–(3.4), we adopt the alternating minimization framework to solve (4.1), i.e.,

$$\mathbf{u}^{t+1} \in \arg\min_{\mathbf{u}} \hat{F}(\mathbf{u}, \mathbf{c}^t), \tag{4.2}$$

$$\mathbf{c}^{t+1} \in \arg\min_{\mathbf{c}} \hat{F}(\mathbf{u}^{t+1}, \mathbf{c}). \tag{4.3}$$

The $\mathbf{c}$-subproblem (4.3) has a closed-form solution for $\ell = 1, \dots, N$:

$$c_\ell^{t+1} = \begin{cases} \dfrac{\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n} f_{i,j}(u_\ell^{t+1})_{i,j}}{\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n}(u_\ell^{t+1})_{i,j}} & \text{if } \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n}(u_\ell^{t+1})_{i,j} \neq 0, \\ 0 & \text{if } \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n}(u_\ell^{t+1})_{i,j} = 0. \end{cases} \tag{4.4}$$

For (4.2), we can find u_ℓ^{t+1} coordinatewise with respect to ℓ by solving

$$\begin{aligned} u_\ell^{t+1} \in \arg\min_{u_\ell} & \|Du_\ell\|_1 - \alpha\|Du_\ell\|_{2,1} + \chi_U(u_\ell) + \lambda\langle f_\ell(\mathbf{c}), u_\ell\rangle_X \\ & + \frac{\nu}{2}\left\| \sum_{j<\ell} u_j^{t+1} + u_\ell + \sum_{j>\ell} u_\ell^t - \mathbb{1} \right\|_X^2 . \end{aligned} \tag{4.5}$$

Applying DCA (3.8) to solve for (4.5) gives

$$\begin{aligned} u_\ell^{t+1} = \arg\min_{u_\ell} & \|Du_\ell\|_1 + \chi_U(u_\ell) + \lambda\langle f_\ell(\mathbf{c}), u_\ell\rangle_X \\ & + \frac{\nu}{2}\left\| \sum_{j<\ell} u_j^{t+1} + u_\ell + \sum_{j>\ell} u_\ell^t - \mathbb{1} \right\|_X^2 + c\|u_\ell\|_X^2 \\ & - \alpha\langle Du_\ell, q_\ell^t\rangle_Y - 2c\langle u_\ell, u_\ell^t\rangle_X, \end{aligned} \tag{4.6}$$

where $q_\ell^t := ((q_x)_\ell^t, (q_y)_\ell^t) = (D_x u_\ell^t, D_y u_\ell^t)/\sqrt{|D_x u_\ell^t|^2 + |D_y u_\ell^t|^2}$ if the denominator is not zero. Similarly to (3.10), we apply PDHGLS to find u_ℓ^{t+1} in (4.6) with the following iteration:

$$\begin{aligned} u_\ell^{t,\eta+1} = \min\Bigg\{ \max\Bigg\{ & \frac{2cu_\ell^t + \frac{1}{\tau}u_\ell^{t,\eta} + \nu\left(\mathbb{1} - \sum_{j<\ell} u_j^{t+1} - \sum_{j>\ell} u_\ell^t\right)}{2c + \frac{1}{\tau} + \nu} \\ & - \frac{\lambda f_\ell(\mathbf{c}) - \alpha D^\top q_\ell^t + (D_x^\top (p_x)_\ell^\eta + D_y^\top (p_y)_\ell^\eta)}{2c + \frac{1}{\tau} + \nu}, 0 \Bigg\}, 1 \Bigg\}, \end{aligned} \tag{4.7}$$

$$\bar{u}_\ell^{t,\eta+1} = u_\ell^{t,\eta+1} + \theta(u_\ell^{t,\eta+1} - u_\ell^{t,\eta}), \tag{4.8}$$

$$(p_x)_\ell^{\eta+1} = \mathrm{Proj}_P((p_x)_\ell^\eta + \sigma D_x \bar{u}_\ell^{t,\eta+1}), \tag{4.9}$$

$$(p_y)_\ell^{\eta+1} = \mathrm{Proj}_P((p_y)_\ell^\eta + \sigma D_y \bar{u}_\ell^{t,\eta+1}) \tag{4.10}$$

for $u_\ell^{t,0} = u_\ell^t$ and $\tau, \sigma > 0$, $\theta \in [0,1]$. The proposed algorithm is referred to as DCA-PDHGLS, summarized in Algorithm 4.1. Convergence analysis of the sequence $\{(\mathbf{u}^t, \mathbf{c}^t)\}_{t=1}^{\infty}$ generated

Algorithm 4.1 DCA-PDHGLS algorithm to solve (4.1)

```
Input:
    • Image $f$
    • model parameters $\alpha, \lambda > 0$
    • strong convexity parameter $c > 0$
    • quadratic penalty parameter $\nu > 0$
    • PDHGLS initial step size $\tau_0 > 0$
    • PDHGLS primal-dual step size ratio $\beta > 0$
    • PDHGLS parameter $\delta \in (0,1)$
    • PDHGLS step size multiplier $\mu \in (0,1)$
 1: Set $u_\ell^0 = 1$ $(\ell = 1, \dots, N)$ for some region $\Sigma \subset \Omega$ and 0 elsewhere.
 2: Compute $\mathbf{c}^0 = (c_1^0, \dots, c_N^0)$ by (4.4).
 3: Set $t := 0$.
 4: while stopping criterion for DCA is not satisfied do
 5:     for $\ell = 1$ to $M$ do
 6:         Set $u_\ell^{t,0} := u_\ell^t$ and $(p_x)_\ell^0 = (p_y)_\ell^0 = 0$.
 7:         Compute $((q_x)_\ell^t, (q_y)_\ell^t) = (D_x u_\ell^t, D_y u_\ell^t) / \sqrt{|D_x u_\ell^t|^2 + |D_y u_\ell^t|^2}$.
 8:         Set $\theta_0 = 1$.
 9:         Set $\eta := 0$.
10:         while stopping criterion for PDHGLS is not satisfied do
11:             Compute $u_\ell^{t,\eta+1}$ by (4.7) with $\tau := \tau_\eta$.
12:             Set $\tau_{\eta+1} = \tau_\eta \sqrt{1 + \theta_\eta}$.
                Linesearch:
13:             Compute $\theta_{\eta+1} = \frac{\tau_{\eta+1}}{\tau_\eta}$ and $\sigma_{\eta+1} = \beta \tau_{\eta+1}$.
14:             Compute $\bar{u}_\ell^{t,\eta+1} = u_\ell^{t,\eta+1} + \theta_{\eta+1}(u_\ell^{t,\eta+1} - u_\ell^{t,\eta})$.
15:             Compute $p_\ell^{\eta+1} := ((p_x)_\ell^{\eta+1}, (p_y)_\ell^{\eta+1})$ by (4.8)–(4.10) with $\sigma := \sigma_{\eta+1}$.
16:             if $\sqrt{\beta}\tau_{\eta+1} \|(D_x^\top (p_x)_\ell^{\eta+1}, D_y^\top (p_y)_\ell^{\eta+1}) - (D_x^\top (p_x)_\ell^\eta, D_y^\top (p_y)_\ell^\eta)\|_Y \leq \delta \|p_\ell^{\eta+1} - p_\ell^\eta\|_Y$
                then
17:                 Set $\eta := \eta + 1$, and break linesearch
18:             else
19:                 Set $\tau_{\eta+1} := \mu \tau_{\eta+1}$ and go back to line 13.
20:             end if
                End of linesearch
21:         end while
22:         Set $u_\ell^{t+1} := u_\ell^{t,\eta}$.
23:     end for
24:     Compute $\mathbf{c}^{t+1}$ by (4.4).
25:     Set $t := t + 1$.
26: end while
Output: $(\mathbf{u}, \mathbf{c}) := (\mathbf{u}^t, \mathbf{c}^t)$.
```

by (4.6) and (4.4) can be established similarly to the one in section 3.2. Hence, we have the following theorem, but for the sake of brevity, the proof is omitted.

Theorem 4.1. *Suppose* $\alpha \in [0,1]$ *and* $\lambda > 0$. *Let* $\{(\mathbf{u}^t, \mathbf{c}^t)\}_{t=1}^{\infty}$ *be a sequence such that* $\mathbf{u}^t$ *is generated by* (4.6) *and* $\mathbf{c}^t$ *is generated by* (4.4). *We have the following:*

(a) $\{(\mathbf{u}^t, \mathbf{c}^t)\}_{t=1}^{\infty}$ *is bounded.*
(b) *For* $\ell = 1, \dots, N$, *we have* $\|u_\ell^{t+1} - u_\ell^t\|_X \to 0$ *as* $t \to \infty$.
(c) *The sequence* $\{(\mathbf{u}^t, \mathbf{c}^t)\}_{t=1}^{\infty}$ *has a limit point* $(\mathbf{u}^*, \mathbf{c}^*)$ *satisfying*

$$\mathbf{0} \in \partial \|Du_\ell^*\|_1 - \alpha \partial \|Du_\ell^*\|_{2,1} + \partial \chi_U(u_\ell^*) + \lambda f_\ell(\mathbf{c}^*) + \nu \left(\sum_{j=1}^N u_j^* - \mathbb{1} \right), \tag{4.11}$$

$$0 \in \frac{\partial \hat{F}(\mathbf{u}^*, \mathbf{c}^*)}{\partial c_\ell} \quad \forall\, \ell = 1, \dots, N. \tag{4.12}$$

5. Extension to color images.

Both AICV (3.2) and AIFR (4.1) models can be extended to color image segmentation. Let $\mathbf{f} = (f_r, f_g, f_b) \in X^3$ be a color image and $(c_{\ell,r}, c_{\ell,g}, c_{\ell,b}) \in \mathbb{R}^3$ for $\ell = 1, \dots, N$. By replacing $f_\ell(\mathbf{c})$ with

$$\mathbf{f}_\ell(\mathbf{c}_r, \mathbf{c}_g, \mathbf{c}_b) = \sum_{\iota \in \{r,g,b\}} (f_\iota - c_{\ell,\iota} \mathbb{1})^2,$$

where $\mathbf{c}_\iota = (c_{1,\iota}, \dots, c_{N,\iota})$ for $\iota \in \{r, g, b\}$, the AICV model for color segmentation is

$$\min_{\substack{\mathbf{u} \in X^M \\ \mathbf{c}_r, \mathbf{c}_g, \mathbf{c}_b \in \mathbb{R}^N}} \sum_{k=1}^M (\|Du_k\|_1 - \alpha \|Du_k\|_{2,1} + \chi_U(u_k)) + \lambda \sum_{\ell=1}^N \langle \mathbf{f}_\ell(\mathbf{c}_r, \mathbf{c}_g, \mathbf{c}_b), R_\ell(\mathbf{u}) \rangle_X. \tag{5.1}$$

Similarly, the AIFR model for color segmentation is

$$\begin{aligned} \min_{\substack{\mathbf{u} \in X^N \\ \mathbf{c}_r, \mathbf{c}_g, \mathbf{c}_b \in \mathbb{R}^N}} & \sum_{\ell=1}^N (\|Du_\ell\|_1 - \alpha \|Du_\ell\|_{2,1} + \chi_U(u_\ell)) + \lambda \sum_{\ell=1}^N \langle \mathbf{f}_\ell(\mathbf{c}_r, \mathbf{c}_g, \mathbf{c}_b), u_\ell \rangle_X \\ & + \frac{\nu}{2} \left\| \sum_{\ell=1}^N u_\ell - \mathbb{1} \right\|_X^2 . \end{aligned} \tag{5.2}$$

For (5.1) and (5.2), their respective update formulas for $\mathbf{c}_\iota$ with $\iota \in \{r, g, b\}$ are

$$c_{\ell,\iota} = \begin{cases} \dfrac{\sum_{i=1}^m \sum_{j=1}^n (f_\iota)_{i,j} R_\ell(\mathbf{u})_{i,j}}{\sum_{i=1}^m \sum_{j=1}^n R_\ell(\mathbf{u})_{i,j}} & \text{if } \sum_{i=1}^m \sum_{j=1}^n R_\ell(\mathbf{u})_{i,j} \neq 0, \\ 0 & \text{if } \sum_{i=1}^m \sum_{j=1}^n R_\ell(\mathbf{u})_{i,j} = 0 \end{cases} \tag{5.3}$$

and

$$c_{\ell,\iota} = \begin{cases} \dfrac{\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n}(f_\iota)_{i,j}(u_\ell)_{i,j}}{\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n}(u_\ell)_{i,j}} & \text{if } \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n}(u_\ell)_{i,j} \neq 0, \\ 0 & \text{if } \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n}(u_\ell)_{i,j} = 0. \end{cases} \tag{5.4}$$

The update formulas for $\mathbf{u}$ are similar to their grayscale counterparts since only f_ℓ needs to be replaced with $\mathbf{f}_\ell$. Hence, their algorithms are straightforward to derive, thus omitted.

6. Numerical results.

In this section, we present extensive experiments on various synthetic and real images to demonstrate the effectiveness of AITV in image segmentation. In particular, we compare the AICV and AIFR models for $\alpha \in \{0, 0.25, 0.5, 0.75, 1.0\}$ with the two-stage segmentation methods that use $L_1 + L_2^2$ [5, 6], L_0 [62, 69], and real-time MS (R_{MS}) [64] penalties. When $\alpha = 0$, the AICV model reduces to the original CV (L_1 CV) model [13, 14], while the AIFR model becomes the FR competition (L_1 FR) model [40]. The two-stage segmentation methods find a smooth approximation $\bar{f}$ of the underlying image f with certain regularization, followed by k-means clustering on $\bar{f}$ to obtain the segmentation result. Specifically, Cai et al. [5, 6] proposed an $L_1 + L_2^2$ regularization problem[1]

$$\min_u \lambda\|f - u\|_X^2 + \gamma\|Du\|_Y^2 + \|Du\|_{2,1}. \tag{6.1}$$

Throughout our numerical experiments, we set $\gamma = 1$, which is suggested in [5, 6]. The L_0-regularized model [62, 69] is given by

$$\min_u \lambda\|f - u\|_X^2 + \|D_x u\|_0 + \|D_y u\|_0, \tag{6.2}$$

where $\|\cdot\|_0$ counts the number of nonzero entries of the matrix. The model in (6.2) can be solved in two different ways. One is by alternating minimization with half-quadratic splitting [69].[2] Another approach [62] incorporates weights for a better isotropic discretizatation than the original L_0 model, followed by ADMM.[3] The R_{MS} model [64] replaces the L_0 norm in (6.2) by $R_{MS}(u) = \sum_{i=1}^m \sum_{j=1}^n \min\{\gamma u_{i,j}, 1\}$, thus leading to

$$\min_u \lambda\|f - u\|_X^2 + R_{MS}(D_x u) + R_{MS}(D_y u). \tag{6.3}$$

In our numerical experiments, we consider the piecewise-constant limit case, where $\gamma \to \infty$. Its implementation is described in [64, Algorithm 1]. We refer to the models (6.1), (6.2), and (6.3) as $L_1 + L_2^2$, L_0, and R_{MS}, respectively.

---

[1] Code is available at https://xiaohaocai.netlify.app/download/.

[2] Code is available at http://www.cse.cuhk.edu.hk/~leojia/projects/L0smoothing/.

[3] Code is available at https://github.com/mstorath/Pottslab.

For the proposed Algorithms 3.1 and 4.1, we set $c = 10^{-8}$, $\tau_0 = 1/8$, $\beta = 1.0$, $\delta = 0.9999$, and $\mu = 7.5 \times 10^{-5}$, as suggested in [48, 49]. The parameter λ depends on the image, which will be specified for each testing case. For the stopping criteria, we use the relative error

$$\text{relerr}(u, v) = \frac{\|u - v\|_X}{\max\{\|u\|_X, \|v\|_X, \epsilon\}}, \tag{6.4}$$

where ϵ is the machine's precision. Following [48], we choose the stopping criterion for the inner PDHGLS algorithm as $\text{relerr}(u^{t,\eta+1}, u^{t,\eta}) < 10^{-6}$. As for the outer iterations, DCA minimization terminates when $\text{relerr}(u^{t+1}, u^t) < 10^{-6}$ and $\text{relerr}(u^{t+1}, u^t) < 10^{-4}$ for two-phase and four-phase AICV models, respectively. For the AIFR models, we use the same stopping criterion in [41] for the outer iterations, i.e., when all the relative errors of the membership functions are less than 10^{-4}. We further adjust the maximum number of outer/inner iterations for multiple channels and multiphase segmentation, which are selected empirically for each image.

We shall apply postprocessing to define the segmented regions. In particular, we convert the results of Algorithm 3.1 to a binary output by setting any pixel values greater than or equal to 0.5 to 1, and 0 otherwise. For the results from Algorithm 4.1, we set a pixel value $(u_\ell)_{i,j}$ to 1 if it is the maximum among all the membership functions $\{u_k\}_{k=1}^N$ at pixel (i, j), and 0 otherwise. For a grayscale image f, we define its reconstructed image

$$\tilde{f} = \sum_{k=1}^{N} c_k \mathbb{1}_{\tilde{\Omega}_k}, \tag{6.5}$$

where $\{c_k\}_{k=1}^N$ and $\{\tilde{\Omega}_k\}_{k=1}^N$ are sets of constants and regions obtained by a segmentation algorithm, respectively, and $\mathbb{1}_{\tilde{\Omega}_k}$ is a binary image corresponding to the region $\tilde{\Omega}_k$. The matrix $\mathbb{1}_{\tilde{\Omega}_k}$ is obtained by thresholding for Algorithms (3.1) and (4.1) or by k-means clustering for the two-stage segmentation framework. Specifically for Algorithms 3.1 and 4.1, the constants $\{c_k\}_{k=1}^N$ are the final outputs of (3.6) and (4.4), respectively. For the two-stage segmentation framework, we compute a smoothed image of f by one of the models (6.1)–(6.3), thus getting $\bar{f}$, and define the constants in (6.5) by

$$c_k = \frac{\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n} \bar{f}_{i,j} (\mathbb{1}_{\tilde{\Omega}_k})_{i,j}}{\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n} (\mathbb{1}_{\tilde{\Omega}_k})_{i,j}}, \quad k = 1, \dots, N. \tag{6.6}$$

As k-means clustering applied to $\bar{f}$ does not produce an empty cluster, the denominator of (6.6) is nonzero. Similarly, the color image $\mathbf{f}$ is approximated by $\tilde{\mathbf{f}} = (\tilde{f}_r, \tilde{f}_g, \tilde{f}_b)$ given by

$$\tilde{f}_\iota = \sum_{k=1}^{N} c_{k,\iota} \mathbb{1}_{\tilde{\Omega}_k} \text{ for } \iota \in \{r, g, b\}, \tag{6.7}$$

where $\{c_{k,\iota}\}_{k=1}^N$ is a set of constants for channel ι. For the color versions of Algorithms 3.1 and 4.1, the constants are obtained by (5.3) and (5.4), respectively. For the color version of

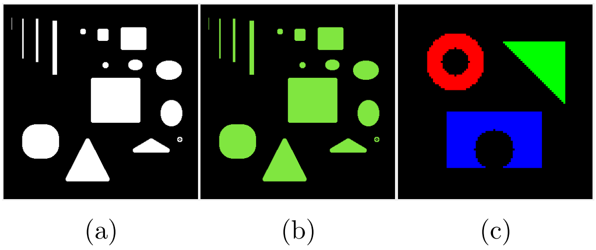


Figure 2. *Synthetic images for image segmentation.* (a) *Grayscale image for two-phase segmentation. Size:* 385×385. (b) *Color image for two-phase segmentation. Size:* 385×385. (c) *Color image for four-phase segmentation. Size:* 100×100.

the two-stage segmentation framework, the constants are computed by (6.6) applied to each channel of the smoothed image $\bar{\mathbf{f}} = (\bar{f}_r, \bar{f}_g, \bar{f}_b)$.

All the algorithms are coded in MATLAB R2019a and all the computations are performed on a Dell laptop with a 1.80 GHz Intel Core i7-8565U processor and 16.0 GB of RAM. The codes are available at https://github.com/kbui1993/L1mL2Segmentation.

6.1. Synthetic images.

We apply various segmentation algorithms on the synthetic images presented in Figure 2. We scale the intensity values of all the images to be $[0,1]$ to ease the parameter tuning. To demonstrate the robustness of the algorithms with respect to noises, we contaminate the original images with either salt-and-pepper impulsive noise (SPIN) or random-valued impulsive noise (RVIN). To evaluate the model performance, we compute the DICE index [19] between the segmentation result and the ground truth. The metric is defined by

$$\text{DICE} = 2\frac{\#\{A(i) \cap A'(i)\}}{\#\{A(i)\} + \#\{A'(i)\}},$$

where $A(i)$ is the set of pixels with label i in the ground-truth image f or $\mathbf{f}$, $A'(i)$ is the set of pixels with label i in the segmented image $\tilde{f}$ or $\tilde{\mathbf{f}}$, and $\#\{A\}$ refers to the number of pixels in the set A. If the DICE index equals 1, it means the perfect alignment of the segmentation result to the ground truth. For two-phase segmentation, we compute the DICE index only for the object of interest, not the background. For multiphase segmentation, we compute the mean of the DICE indices across the regions, including the background.

For the two-phase AICV model, the initialization u_1^0 in Algorithm 3.1 is a binary step function that represents a circle of radius 10 in the center of the image (i.e., taking the value 1 if inside the circle and 0 elsewhere). Since the binary step function forms two regions in an image, it can be used as initialization for the two-phase AIFR model, i.e., u_1^0 and $u_2^0 = \mathbb{1} - u_1^0$ for Algorithm 4.1. The initialization for the four-phase segmentation requires two step functions, which are set to be two circles of radius 30 shifted by 5 pixels to the right of the image center and another by 5 pixels to the left. The circle functions are used here for simplicity. Contours of the initialization are marked as colored circles in the noisy images.

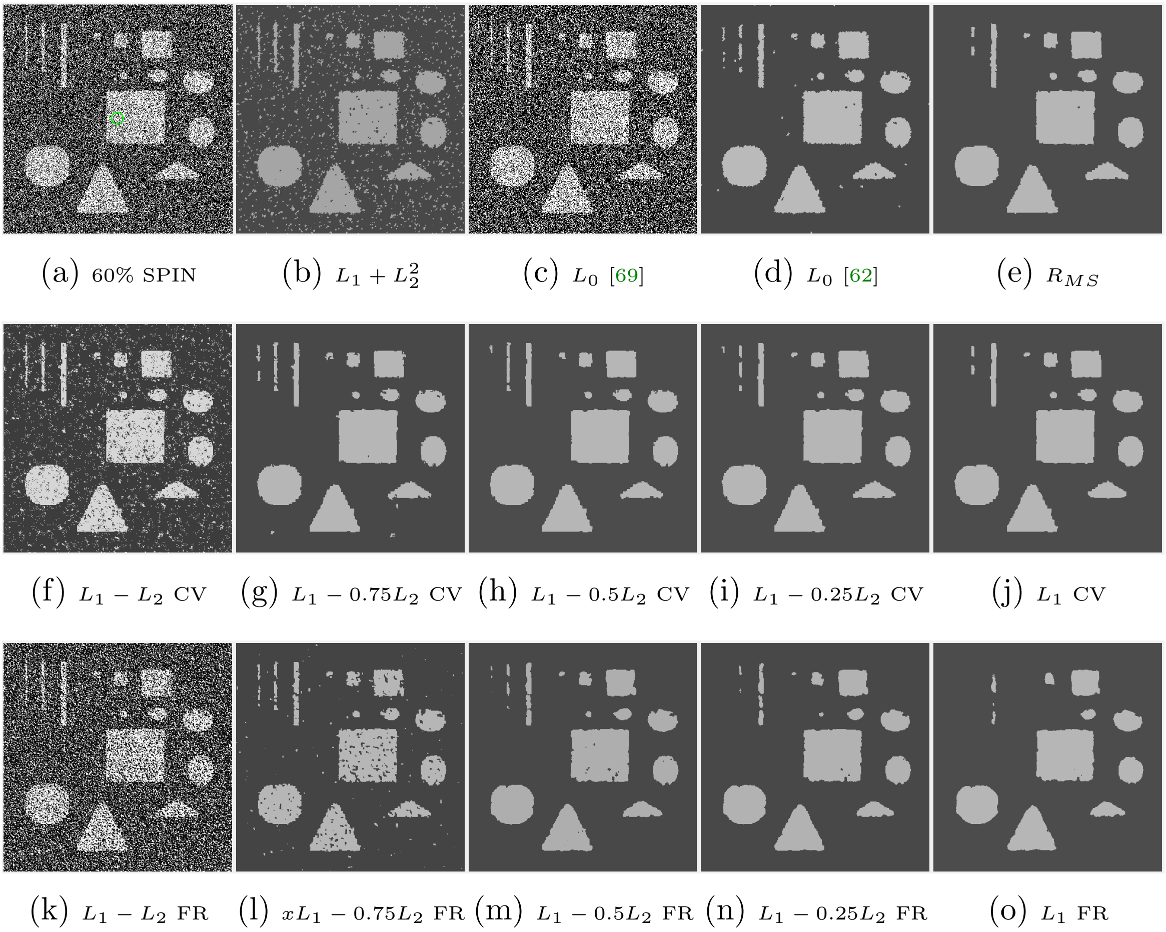
For Figure 2(a), we set $\lambda = 2$ for all methods, except for L_0 [69] in which $\lambda = 50$. For the AIFR models, we set $\nu = 10$. The maximum number of inner iterations for the AITV models

Table 1

DICE indices of various segmentation models applied to Figure 2(a) corrupted with different levels of impulsive noise.

Salt & pepper (%)	0	10	20	30	40	50	60	70
$L_1 - L_2$ CV	**1**	0.9977	0.9932	0.9854	0.9594	0.9062	0.8138	0.7643
$L_1 - 0.75L_2$CV	**1**	0.9978	0.9929	0.9853	0.9795	0.9727	0.9678	0.9550
$L_1 - 0.5L_2$ CV	**1**	0.9975	0.9941	0.9893	0.9850	0.9801	**0.9726**	**0.9554**
$L_1 - 0.25L_2$ CV	**1**	0.9974	0.9954	0.9910	0.9870	**0.9823**	0.9711	0.9483
L_1 CV	**1**	**0.9981**	0.9960	0.9922	0.9877	0.9802	0.9681	0.9338
$L_1 - L_2$ FR	**1**	0.8753	0.7719	0.6833	0.6129	0.5425	0.4702	0.4138
$L_1 - 0.75L_2$ FR	**1**	0.9896	0.9841	0.9693	0.9585	0.9437	0.9183	0.7775
$L_1 - 0.5L_2$ FR	0.9998	0.9978	0.9956	0.9923	**0.9879**	0.9788	0.9495	0.7760
$L_1 - 0.25L_2$ FR	0.9995	0.9979	**0.9961**	**0.9925**	0.9865	0.9737	0.9347	0.6883
L_1 FR	0.9992	0.9978	0.9949	0.9877	0.9812	0.9663	0.8990	0.5053
$L_1 + L_2^2$	0.9996	0.9961	0.9925	0.9857	0.9733	0.9328	0.8375	0.6840
L_0 [69]	**1**	0.8731	0.7666	0.6736	0.5943	0.5226	0.4601	0.4035
L_0 [62]	0.9995	0.9944	0.9874	0.9792	0.9738	0.9690	0.9605	0.9474
R_{MS}	0.9995	0.9969	0.9947	0.9887	0.9851	0.9784	0.9670	0.9312
Random valued (%)	0	10	20	30	40	50	60	70
$L_1 - L_2$ CV	**1**	0.9986	0.9957	0.9909	0.9846	0.9739	0.9534	0.9542
$L_1 - 0.75L_2$ CV	**1**	0.9988	0.9971	0.9948	0.9926	0.9894	**0.9840**	**0.9712**
$L_1 - 0.5L_2$ CV	**1**	0.9989	**0.9973**	0.9958	0.9930	**0.9899**	0.9816	0.9614
$L_1 - 0.25L_2$ CV	**1**	**0.9990**	0.9971	0.9957	**0.9935**	0.9898	0.9808	0.9560
L_1 CV	**1**	0.9984	0.9972	**0.9959**	0.9928	0.9863	0.9700	0.9332
$L_1 - L_2$ FR	**1**	0.9505	0.9053	0.8578	0.8015	0.7369	0.6478	0.5662
$L_1 - 0.75L_2$ FR	**1**	0.9987	0.9971	0.9945	0.9913	0.9879	0.9715	0.5364
$L_1 - 0.5L_2$ FR	0.9998	0.9984	0.9972	0.9955	0.9921	0.9833	0.9538	0.3540
$L_1 - 0.25L_2$ FR	0.9995	0.9983	0.9972	0.9940	0.9880	0.9763	0.9299	0.5984
L_1 FR	0.9992	0.9983	0.9970	0.9925	0.9833	0.9643	0.8800	0.4503
$L_1 + L_2^2$	0.9996	0.9980	0.9960	0.9937	0.9903	0.9858	0.9776	0.9668
L_0 [69]	**1**	0.8753	0.7697	0.6768	0.5981	0.5247	0.4627	0.4054
L_0 [62]	0.9995	0.9966	0.9933	0.9904	0.9874	0.9810	0.9688	0.9462
R_{MS}	0.9995	0.9983	0.9971	0.9954	0.9932	0.9850	0.9731	0.9361

is 300, while the maximum number of outer iterations is 20 for AICV and 40 for AIFR. Table 1 records the DICE indices of the segmentation results for varying levels of both SPIN and RVIN from 0% to 70%. When the noise level is at least 50%, both $L_1 - 0.5L_2$ and $L_1 - 0.25L_2$ CV models outperform L_1 CV. For AIFR, $L_1 - 0.5L_2$ and $L_1 - 0.25L_2$ outperform L_1 across all levels of SPIN corruption. In addition, $L_1 - L_2$ FR is less robust than other values of α when the noise level increases. Most of the best results in the cases of intermediate to high RVIN noise levels are attained by the proposed models. Figures 3–4 display the segmentation results of Figure 2(a) corrupted with 60% SPIN and 60% RVIN, respectively. (We note that the contrast


(a) 60% SPIN (b) $L_1 + L_2^2$ (c) L_0 [69] (d) L_0 [62] (e) R_{MS}

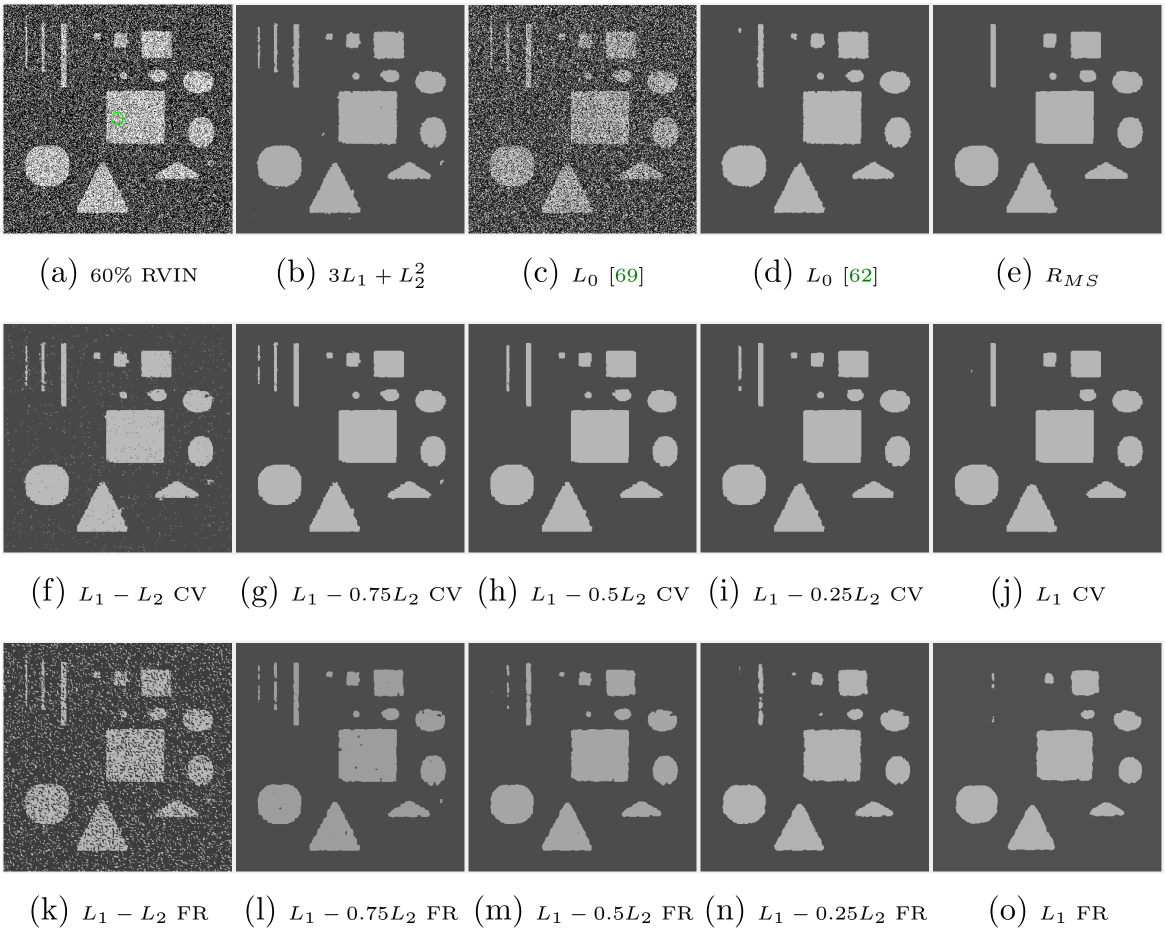
(f) $L_1 - L_2$ CV (g) $L_1 - 0.75L_2$ CV (h) $L_1 - 0.5L_2$ CV (i) $L_1 - 0.25L_2$ CV (j) L_1 CV

(k) $L_1 - L_2$ FR (l) $xL_1 - 0.75L_2$ FR (m) $L_1 - 0.5L_2$ FR (n) $L_1 - 0.25L_2$ FR (o) L_1 FR

Figure 3. *Reconstruction results on Figure 2(a) corrupted with 60% SPIN.*

of the reconstructed images is different from Figure 2(a) because the impulsive noise in the corrupted image skews the values of the constants $\{c_k\}_{k=1}^N$ computed by the segmentation algorithms. This phenomenon repeats for Figures 2(b)–2(c).) As α decreases in both the AICV and AIFR models, the results become less noisy, but they have fewer segmented regions. Therefore, $\alpha = 0.5$ yields the best compromise in the case of SPIN. For RVIN, the AICV and AIFR results are not as noisy as in the case of SPIN and, hence, $\alpha = 0.75$ is the best for RVIN. The two-stage methods generally produce noisy results in the presence of SPIN and RVIN.

Figure 2(b) is a color version of Figure 2(a). We corrupt the image by 0% to 50% SPIN/RVIN for each color channel. When a color image is corrupted with noise, one channel might be noisier than the others. In addition, image structures may vary with color channels, thus making the color extension of finding a balanced segmentation across all the color channels more challenging than for grayscale images. For Figure 2(b), we set $\lambda = 0.5$ for all methods, except for L_0 [69] in which $\lambda = 50$. For the AIFR models, we set $\nu = 2.5$. The maximum number of inner/outer iterations are the same as the case for Figure 2(a). The


(a) 60% RVIN (b) $3L_1 + L_2^2$ (c) L_0 [69] (d) L_0 [62] (e) R_{MS}

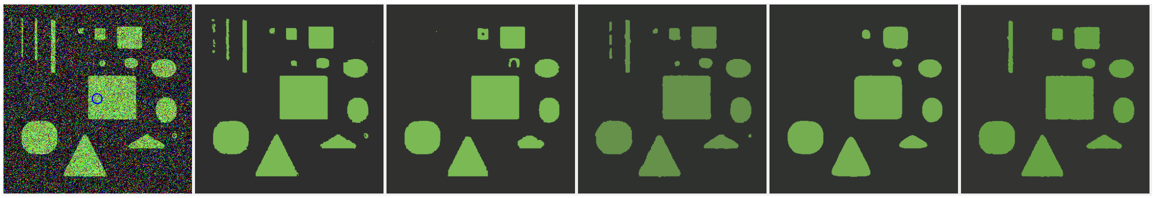
(f) $L_1 - L_2$ CV (g) $L_1 - 0.75L_2$ CV (h) $L_1 - 0.5L_2$ CV (i) $L_1 - 0.25L_2$ CV (j) L_1 CV

(k) $L_1 - L_2$ FR (l) $L_1 - 0.75L_2$ FR (m) $L_1 - 0.5L_2$ FR (n) $L_1 - 0.25L_2$ FR (o) L_1 FR

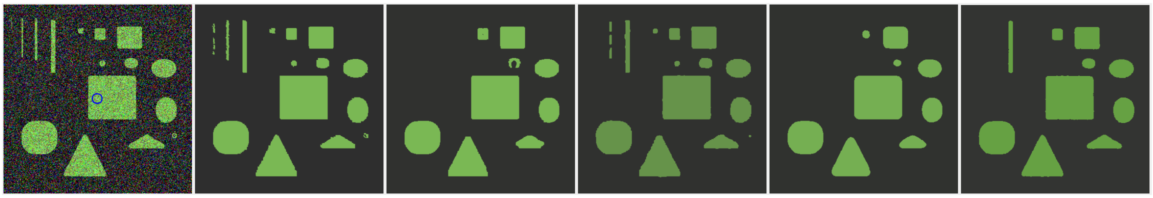
Figure 4. *Reconstruction results on Figure* 2(a) *corrupted with* 60% *RVIN.*

DICE indices of the segmentation results are reported in Table 2, which shows that $L_1 - L_2$ CV generally yields the best results and AIFR is slightly worse than its AICV counterpart but better than L_1 FR. Figure 5 presents the comparison results of AICV (with optimal α), L_1 CV, AIFR (with optimal α), L_1 FR, and $L_1 + L_2^2$ for 40% SPIN and 40% RVIN, showing that AICV and AIFR segment more salient regions than their L_1 counterparts and $L_1 + L_1^2$.

Figure 2(c) is a color image for multiphase segmentation. We set $\lambda = 2.25$ for all methods, except for L_0 [69] in which $\lambda = 50$. For the AIFR models, we set $\nu = 5$. The maximum number of inner iterations for the AITV models is 1000, while the maximum number of outer iterations is 40 for AICV and 160 for AIFR. Table 3 presents the DICE indices of the segmentation results under 0% to 40% SPIN/RVIN contamination for each color channel. For SPIN, $L_1 - 0.25L_2$ FR is comparable to L_1 FR and outperforms it when the noise level is 40%. For RVIN, $L_1 - 0.5L_2$ and $L_1 - 0.25L_2$ FR give the best results in general. We also observe that the smaller α is, the more robust AICV/AIFR are with respect to impulsive noise. The visual results 4 are presented in Figure 6 for 40% SPIN/RVIN, clearly showing that AIFR provides

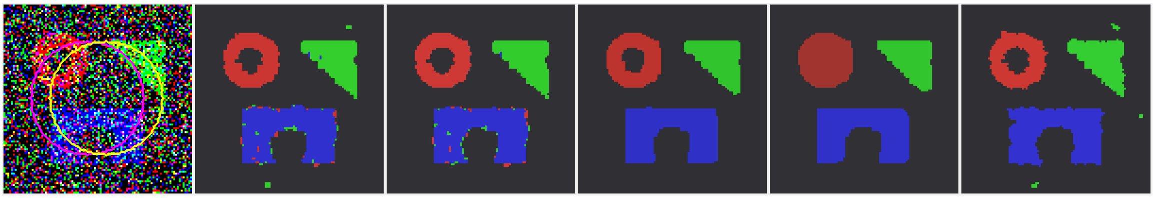


(a) 40% SPIN (b) $L_1 - L_2$ CV (c) L_1 CV (d) $L_1 - 0.75L_2$ FR (e) L_1 FR (f) $L_1 + L_2^2$

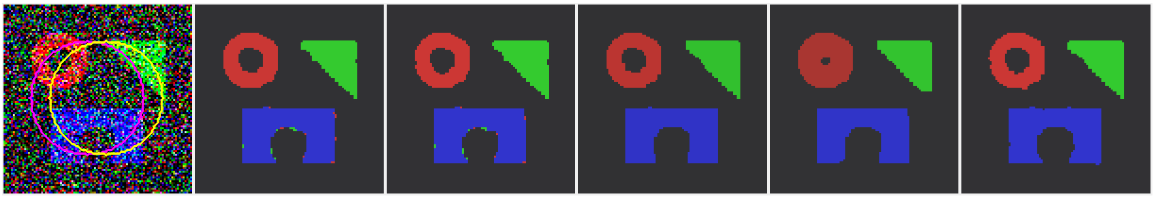


(g) 40% RVIN (h) $L_1 - L_2$ CV (i) L_1 CV (j) $L_1 - 0.75L_2$ FR (k) L_1 FR (l) $L_1 + L_2^2$

Figure 5. *Reconstruction results on Figure 2(b) corrupted with 40% SPIN (top) and 40% RVIN (bottom).*



(a) 40% SPIN (b) $L_1 - 0.25L_2$ CV (c) L_1 CV (d) $L_1 - 0.25L_2$ FR (e) L_1 FR (f) R_{MS}



(g) 40% RVIN (h) $L_1 - 0.25L_2$ CV (i) L_1 CV (j) $L_1 - 0.5L_2$ FR (k) L_1 FR (l) R_{MS}

Figure 6. *Reconstruction results on Figure 2(c) corrupted with 40% SPIN (top) and 40% RVIN (bottom).*

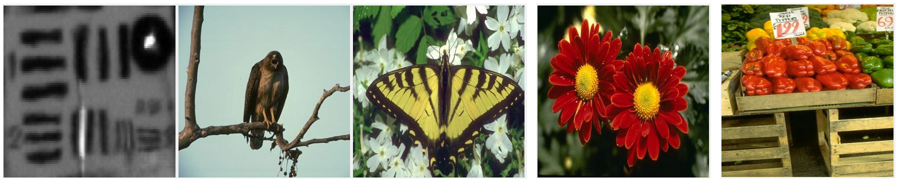
the best segmentation. AICV and L_1 CV contain noise along the edges of the blue region, L_1 FR oversegments the red region, and R_{MS} appears slightly worse than AIFR.

Overall, the proposed AICV/AIFR methods are robust against impulsive noise, unlike the two-stage methods. For the three synthetic images, AICV and AIFR with appropriately chosen α outperform their L_1 counterparts under a high level of impulsive noise. Unfortunately, there is no optimal choice of α that works for all images, as demonstrated by our experiments. For example, $\alpha = 1.0$ yields the highest DICE indices for Figure 2(b) according to Table 2, but it does not perform as well for Figure 2(a) according to Table 1.

Table 2

DICE indices of various segmentation models applied to Figure 2(b) corrupted with different levels of impulsive noise.

Salt & pepper (%)	0	10	20	30	40	50
$L_1 - L_2$ CV	**1**	**0.9979**	0.9952	**0.9920**	**0.9867**	**0.9775**
$L_1 - 0.75L_2$ CV	0.9994	0.9978	**0.9957**	0.9896	0.9856	0.9737
$L_1 - 0.5L_2$ CV	0.9992	0.9970	0.9910	0.9889	0.9826	0.9512
$L_1 - 0.25L_2$ CV	0.9982	0.9924	0.9904	0.9829	0.9726	0.9308
L_1 CV	0.9938	0.9918	0.9808	0.9755	0.9457	0.9109
$L_1 - L_2$ FR	0.9977	0.9960	0.9931	0.9685	0.8187	0.7273
$L_1 - 0.75L_2$ FR	0.9979	0.9955	0.9920	0.9873	0.9795	0.9626
$L_1 - 0.5L_2$ FR	0.993	0.9908	0.9802	0.9720	0.9635	0.9409
$L_1 - 0.25L_2$ FR	0.9818	0.9786	0.9690	0.9462	0.9441	0.9195
L_1 FR	0.9774	0.9705	0.9524	0.9383	0.9301	0.8906
$L_1 + L_2^2$	0.9931	0.9907	0.9874	0.9794	0.9726	0.9686
L_0 [69]	**1**	0.8734	0.7687	0.6745	0.5945	0.4307
L_0 [62]	0.9939	0.9904	0.9823	0.9762	0.9543	0.9266
R_{MS}	0.9853	0.9801	0.9676	0.9444	0.9116	0.8225
Random valued (%)	0	10	20	30	40	50
$L_1 - L_2$ CV	**1**	**0.9987**	**0.9966**	**0.9932**	**0.9887**	**0.9826**
$L_1 - 0.75L_2$ CV	0.9994	0.9983	0.9960	0.9915	0.9877	0.9759
$L_1 - 0.5L_2$ CV	0.9992	0.9975	0.9916	0.9899	0.9815	0.9535
$L_1 - 0.25L_2$ CV	0.9982	0.9928	0.9913	0.9784	0.9748	0.9344
L_1 CV	0.9938	0.9920	0.9798	0.9773	0.9493	0.9145
$L_1 - L_2$ FR	0.9977	0.9965	0.9943	0.9902	0.9071	0.7154
$L_1 - 0.75L_2$ FR	0.9979	0.9960	0.9921	0.9879	0.9815	0.9520
$L_1 - 0.5L_2$ FR	0.993	0.9907	0.9797	0.9742	0.9644	0.9526
$L_1 - 0.25L_2$ FR	0.9818	0.9781	0.9702	0.9620	0.9534	0.9161
L_1 FR	0.9774	0.9656	0.9533	0.9519	0.9316	0.8770
$L_1 + L_2^2$	0.9931	0.9912	0.9877	0.9812	0.9755	0.9726
L_0 [69]	**1**	0.9032	0.7991	0.6972	0.6089	0.5312
L_0 [62]	0.9939	0.9852	0.9846	0.9786	0.9573	0.9298
R_{MS}	0.9853	0.9797	0.9782	0.9465	0.9074	0.8260



(a) (b) (c) (d) (e)

Figure 7. *Real images for image segmentation.* (a) *Close-up of a target board in a video. Size:* 89 × 121. (b) *Image of a hawk. Size:* 318 × 370. (c) *Image of a butterfly. Size:* 321 × 481. (d) *Image of a flower. Size:* 321 × 481. (e) *Image of peppers. Size:* 481 × 321.

6.2. Real images. We apply the proposed methods and the two-stage methods on real images (all rescaled to $[0, 1]$ for the pixel values) shown in Figure 7 without additive noise. Figure 7(a) is provided in [44] while Figures 7(b)–7(e) are provided by the Berkeley Segmentation

Table 3

DICE indices of various segmentation models applied to Figure 2(c) corrupted with different levels of impulsive noise.

Salt & pepper (%)	0	10	20	30	40
$L_1 - L_2$ CV	0.9990	0.9762	0.9524	0.9245	0.8548
$L_1 - 0.75L_2$ CV	0.9992	0.9763	0.9649	0.9288	0.8978
$L_1 - 0.5L_2$ CV	0.9992	0.9789	0.9704	0.9509	0.9292
$L_1 - 0.25L_2$ CV	0.9994	0.9852	0.9686	0.9608	0.9448
L_1 CV	0.9987	0.9832	0.9788	0.9597	0.9496
$L_1 - L_2$ FR	0.9994	0.7869	0.6566	0.5424	0.4552
$L_1 - 0.75L_2$ FR	0.9994	0.9328	0.8736	0.8058	0.6541
$L_1 - 0.5L_2$ FR	0.9980	0.9905	0.9847	0.9720	0.8976
$L_1 - 0.25L_2$ FR	0.9976	0.9921	0.9863	0.9801	**0.9753**
L_1 FR	0.9976	**0.9924**	**0.9869**	**0.9804**	0.9474
$L_1 + L_2^2$	0.9984	0.9904	0.9691	0.8984	0.7562
L_0 [69]	**1**	0.7611	0.6284	0.5134	0.4225
L_0 [62]	0.9997	0.9245	0.7977	0.6536	0.4884
R_{MS}	**1**	0.9900	0.9771	0.9649	0.9575
Random valued (%)	0	10	20	30	40
$L_1 - L_2$ CV	0.9990	0.9895	0.9757	0.9594	0.9261
$L_1 - 0.75L_2$ CV	0.9992	0.9910	0.9831	0.9755	0.9664
$L_1 - 0.5L_2$ CV	0.9992	0.9934	0.9875	0.9797	0.9737
$L_1 - 0.25L_2$ CV	0.9994	0.9934	0.9876	0.9798	0.9771
L_1 CV	0.9987	0.9941	0.9884	0.9789	0.9761
$L_1 - L_2$ FR	0.9994	0.8841	0.7118	0.6604	0.5972
$L_1 - 0.75L_2$ FR	0.9994	0.9916	0.9875	0.9353	0.8790
$L_1 - 0.5L_2$ FR	0.998	0.9947	**0.9912**	**0.9851**	**0.9833**
$L_1 - 0.25L_2$ FR	0.9976	0.9942	**0.9912**	0.9849	0.9821
L_1 FR	0.9976	0.9921	0.9892	**0.9851**	0.9553
$L_1 + L_2^2$	0.9984	0.9949	0.9857	0.9803	0.9705
L_0 [69]	**1**	0.7744	0.6932	0.5302	0.4478
L_0 [62]	0.9997	0.9828	0.9614	0.9482	0.9311
R_{MS}	**1**	**0.9953**	0.9900	0.9849	0.9831

Dataset and Benchmark [50]. Specifically, Figures 7(a) and 7(b) are for two-phase segmentation, Figure 7(c) is for four-phase segmentation, and Figures 7(d) and 7(e) are for five-phase and seven-phase segmentation, respectively. We set the maximum number of inner iterations for CV/FR methods as 300, and the maximum number of outer iterations for CV as 20. The maximum outer iteration number of the FR methods depends on images, which is set to 40 for Figures 7(a)–7(b), 80 for Figure 7(c), and 160 for Figures 7(d)–7(e). Following the work of [31], we compute the peak signal-to-noise ratio (PSNR) between the reconstructed image $\tilde{\mathbf{f}}$ derived by (6.7) and the original image $\mathbf{f}$. PSNR is defined by $10\log_{10}\frac{3mn}{\sum_{\iota\in\{r,g,b\}}\|\tilde{f}_\iota - f_\iota\|_X^2}$, and it quantitatively measures the quality of the segmentation results for real color images without ground truth. The PSNR values are recorded in Table 4. As the CV methods are inapplicable to non-power-of-2 segmentation examples, we indicate by NA ("not applicable") their results on Figures 7(d)–7(e) in Table 4.

Table 4

PSNR values of segmentation methods applied to real color images. NA stands for "not applicable."

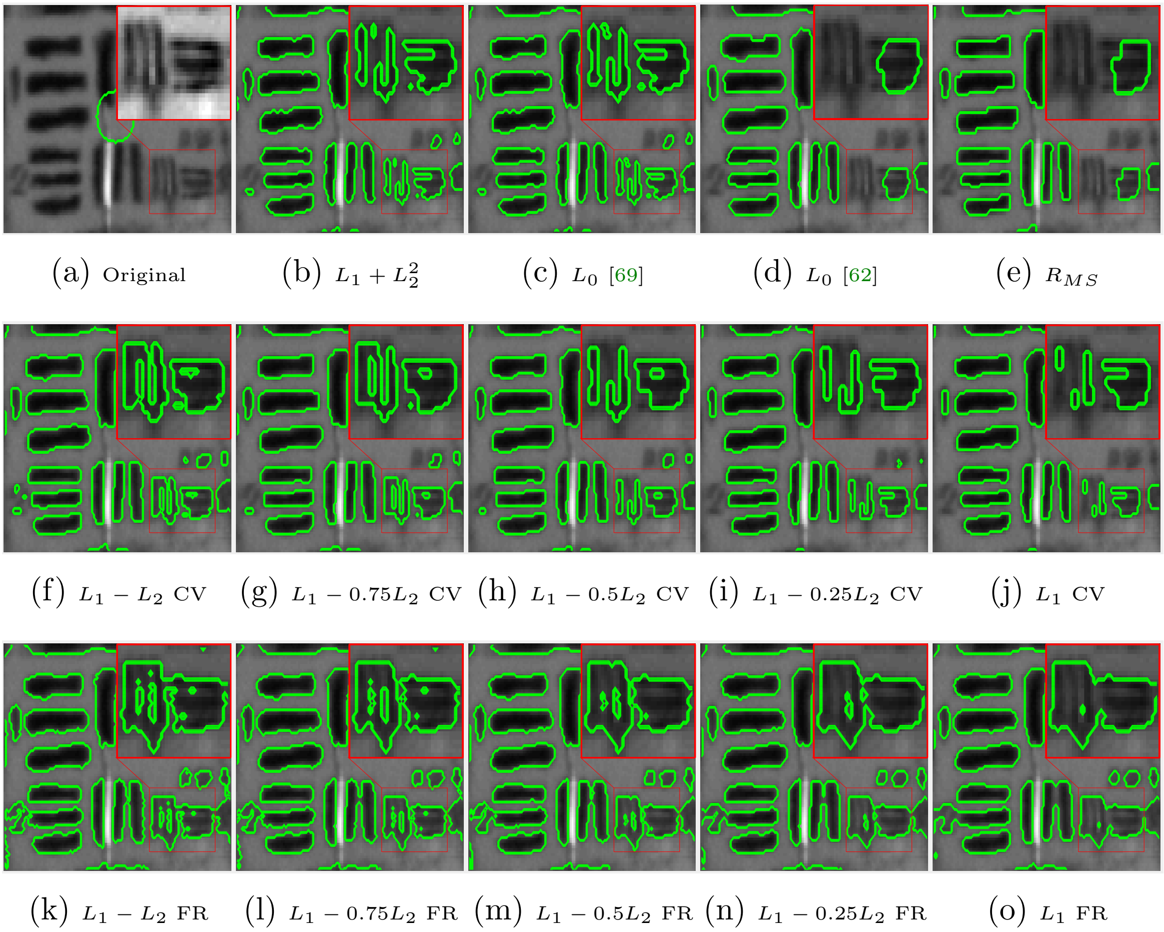
	Figure 7(b)	Figure 7(c)	Figure 7(d)	Figure 7(e)
$L_1 - L_2$ CV	23.3949	21.9000	NA	NA
$L_1 - 0.75L_2$ CV	23.3933	21.9001	NA	NA
$L_1 - 0.5L_2$ CV	**23.4001**	21.8976	NA	NA
$L_1 - 0.25L_2$ CV	23.3913	21.8985	NA	NA
L_1 CV	23.3690	21.8977	NA	NA
$L_1 - L_2$ FR	23.4223	22.2574	21.8283	22.2597
$L_1 - 0.75L_2$ FR	23.4014	**22.2578**	21.8383	22.4880
$L_1 - 0.5L_2$ FR	23.3814	22.2576	**21.8418**	**22.4901**
$L_1 - 0.25L_2$ FR	23.3523	22.2575	**21.8418**	22.4672
L_1 FR	23.3173	22.2570	21.8409	21.9482
$L_1 + L_2^2$	23.2601	21.6077	21.1802	21.0277
L_0 [69]	23.2419	22.2570	21.7914	22.0361
L_0 [62]	23.1985	17.7573	21.8129	21.9703
R_{MS}	23.0865	17.7140	21.7832	22.0904

For Figure 7(a), we set $\lambda = 100$ for all methods, except for L_0 [69] in which $\lambda = 10000$. For all FR methods, we set $\nu = 35$. The initialization for the CV and FR methods is a step function of a circle in the image center with radius 10. The segmentation results of these competing methods are displayed in Figure 8, each equipped with a zoomed-in region of the bottom right of the image. We observe that as α decreases, the CV methods segment lesser regions, while the FR methods identify lesser gaps. The results of the two-stage methods are not as detailed as the results provided by $L_1 - L_2$ CV and FR.

For Figure 7(b), we set $\lambda = 50$ for L_0 [69], $\lambda = 10$ for the other methods, and $\nu = 10.0$ for the FR methods. The initialization for the CV and FR methods is the same as Figure 7(a). Quantitative comparison of these methods is listed in Table 4, showing that the AICV and AIFR methods outperform their L_1 counterparts. The visual results in Figure 9 demonstrate that AICV and AIFR can segment finer details, especially on the branch on the left side of the image and on the hawk, than their L_1 counterparts, which thereby explains their higher PSNR values.

For Figure 7(c), we set $\lambda = 1000$ for all methods and $\nu = 650$ for the FR methods. Initialization for the CV methods are two step functions of circles both with radius 10, one shifted 5 pixels to the left of the image center and the other shifted 5 pixels to the right. For the FR methods, the initialization of the membership functions are uniformly distributed in $[0, 1]$ and then normalized. Figure 10 compares the AIFR and AICV methods (using the optimal α value that corresponds to the highest PSNR in Table 4) with their L_1 counterparts. As PSNR values are all similar, we do not observe much visual differences between the images in Figure 10.

For Figure 7(d), we set $\lambda = 650$ for all methods, except L_0 [69] in which $\lambda = 1000$. For the FR methods, we set $\nu = 1050$. For Figure 7(e), we set $\lambda = 500$ for all methods and $\nu = 400$ for the FR methods. Initialization of the membership functions for the FR methods is the same as for Figure 7(c). The segmentation results of the FR methods and the two-stage methods are shown in Figures 11 and 12. In Figure 11, the results of the FR methods have better


(a) Original (b) $L_1 + L_2^2$ (c) L_0 [69] (d) L_0 [62] (e) R_{MS}

(f) $L_1 - L_2$ CV (g) $L_1 - 0.75L_2$ CV (h) $L_1 - 0.5L_2$ CV (i) $L_1 - 0.25L_2$ CV (j) L_1 CV

(k) $L_1 - L_2$ FR (l) $L_1 - 0.75L_2$ FR (m) $L_1 - 0.5L_2$ FR (n) $L_1 - 0.25L_2$ FR (o) L_1 FR

Figure 8. *Segmentation results on Figure 7(a). (The images may need to be zoomed in on a PDF reader to see the differences.)*

contrast than the result of $L_1 + L_2^2$ and thus they look more similar to the original image. In Figure 12, $L_1 - L_2$ FR, L_1 FR, and L_0 are unable to identify the yellow/orange peppers behind the red peppers, which explains their lower PSNR values. Although the results of the AIFR methods for $\alpha = 0.25, 0.5, 0.75$ appear similar to $L_1 + L_2^2$ and R_{MS}, $L_1 - 0.5L_2$ attains the best segmentation based on its PSNR value.

Last, we report the computational times of the segmentation methods in Table 5. Admittedly, the proposed methods are slower compared to other segmentation methods. Besides, our computational times largely depend on the image size, the number of channels, and the number of u_k's needed to segment. The acceleration of the proposed scheme will be left for future investigation.

In summary, given particular choices of α, the AITV models outperform their L_1 counterparts and the two-stage methods. For Figure 7(a), larger values of α provide better segmentation results, but this may not be the case for other images. Thus, the optimal α value in an AITV model varies for an individual image. In addition, although the AITV methods tend

(a) Original (b) $L_1 - 0.5L_2$ CV (c) L_1 CV (d) $L_1 - 0.75L_2$ FR (e) L_1 FR

Figure 9. *Reconstruction results on Figure* 7(b).

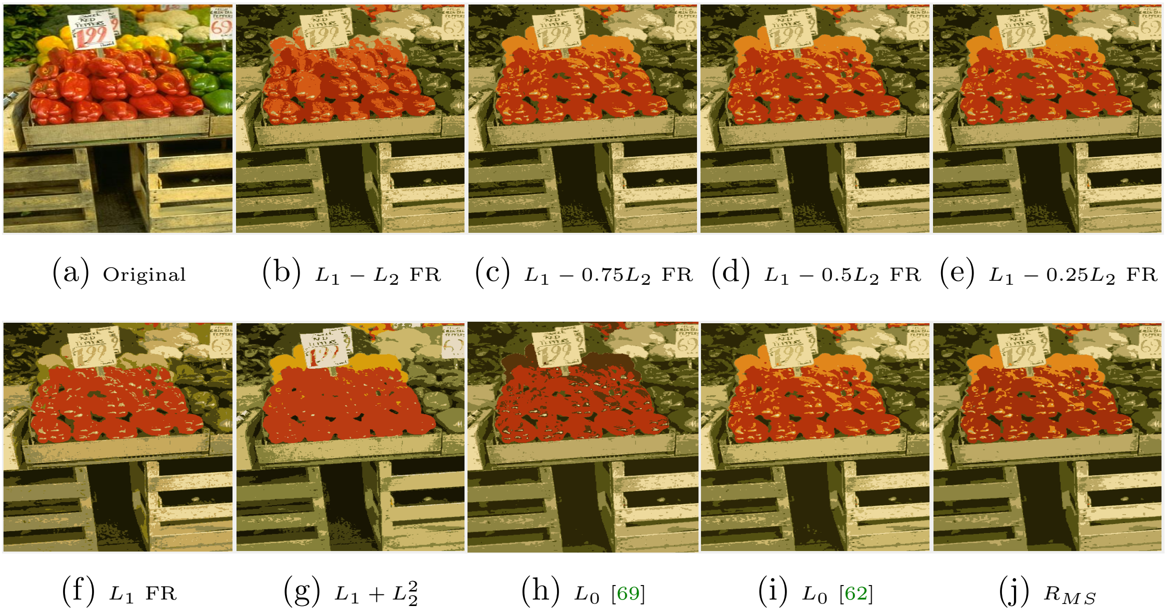
(a) Original (b) $L_1 - 0.75L_2$ CV (c) L_1 CV (d) $L_1 - 0.75L_2$ FR (e) L_1 FR

Figure 10. *Reconstruction results on Figure* 7(c).

(a) Original (b) $L_1 - L_2$ FR (c) $L_1 - 0.75L_2$ FR (d) $L_1 - 0.5L_2$ FR (e) $L_1 - 0.25L_2$ FR

(f) L_1 FR (g) $L_1 + L_2^2$ (h) L_0 [69] (i) L_0 [62] (j) R_{MS}

Figure 11. *Reconstruction results on Figure* 7(d).


(a) Original (b) $L_1 - L_2$ FR (c) $L_1 - 0.75L_2$ FR (d) $L_1 - 0.5L_2$ FR (e) $L_1 - 0.25L_2$ FR

(f) L_1 FR (g) $L_1 + L_2^2$ (h) L_0 [69] (i) L_0 [62] (j) R_{MS}

Figure 12. *Reconstruction results on Figure 7(e).*

Table 5

Computational time (seconds) of segmentation methods applied to real color images. NA stands for "not applicable."

	Figure 7(a)	Figure 7(b)	Figure 7(c)	Figure 7(d)	Figure 7(e)
$L_1 - L_2$ CV	2.06	16.09	49.27	NA	NA
$L_1 - 0.75L_2$ CV	1.86	15.91	55.91	NA	NA
$L_1 - 0.5L_2$ CV	2.08	15.89	70.68	NA	NA
$L_1 - 0.25L_2$ CV	2.17	16.09	71.23	NA	NA
L_1 CV	1.78	16.23	54.94	NA	NA
$L_1 - L_2$ FR	2.51	43.65	66.27	191.30	212.28
$L_1 - 0.75L_2$ FR	1.91	46.26	64.98	185.26	233.79
$L_1 - 0.5L_2$ FR	1.23	15.29	68.3	175.67	263.52
$L_1 - 0.25L_2$ FR	0.92	13.18	69.49	182.08	227.62
L_1 FR	0.72	13.18	69.49	182.08	227.62
$L_1 + L_2^2$	0.24	1.8	**1.2**	1.75	2.48
L_0 [69]	**0.15**	**0.92**	1.71	**1.6**	**1.97**
L_0 [62]	0.17	2.96	3.06	3.05	4.26
R_{MS}	0.61	6.60	17.71	17.24	20.10

to be slower than the two-stage methods, they are consistently more accurate based on their PSNR values. This observation is apparent in Figures 7(c)–7(e), the most complex images tested in this section.

7. Conclusions and future works.

In this paper, we proposed AICV and AIFR models for piecewise-constant segmentation that can deal with both grayscale and color images. We developed alternating minimization algorithms utilizing DCA and PDHGLS to efficiently solve the models. Convergence analyses were provided to demonstrate that the objective functions were monotonically decreasing and to validate the efficacy of the algorithms. Numerical results illustrated that the AICV/AIFR models outperform their anisotropic counterparts on various images in a robust manner. The segmentation results are comparable and sometimes better than those of the two-stage segmentation methods.

In the future, we will consider the application of the weighted anisotropic-isotropic penalty to other types of segmentation approaches, such as piecewise-smooth formulations [31, 37], the Potts models [58, 63, 68], the FR model [41], and deep learning techniques [29, 30, 35]. Since the two-stage methods are generally faster to run than our methods, we will leave the acceleration as a future work. Another future direction involves segmenting blurry images by combining our proposed models with some deblurring techniques. The numerical experiments demonstrated that there is no optimal, universal α for all images, which motivates us to develop an automatic method to select α for any given image in the future. As AICV/AIFR models indicate the success of using nonconvex penalty terms in image processing, we aim at other nonconvex penalties, such as transformed L_1 [53, 76] and L_1/L_2 [59, 67], for image segmentation and other imaging problems including denoising and deblurring.

Acknowledgment. We would like to thank the anonymous referees for their useful suggestions and feedback, which significantly improved the presentation of the paper.

REFERENCES

[1] E. Bae, J. Yuan, and X.-C. Tai, *Global minimization for continuous multiphase partitioning problems using a dual approach*, Int. J. Comput. Vis., 92 (2011), pp. 112–129.

[2] H. Birkholz, *A unifying approach to isotropic and anisotropic total variation denoising models*, J. Comput. Appl. Math., 235 (2011), pp. 2502–2514.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122.

[4] E. S. Brown, T. F. Chan, and X. Bresson, *Completely convex formulation of the Chan–Vese image segmentation model*, Int. J. Comput. Vis., 98 (2012), pp. 103–121.

[5] X. Cai, R. Chan, M. Nikolova, and T. Zeng, *A three-stage approach for segmenting degraded color images: Smoothing, lifting and thresholding (SLaT)*, J. Sci. Comput., 72 (2017), pp. 1313–1332.

[6] X. Cai, R. Chan, and T. Zeng, *A two-stage image segmentation method using a convex variant of the Mumford–Shah model and thresholding*, SIAM J. Imaging Sci., 6 (2013), pp. 368–390.

[7] E. J. Candes, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.

[8] V. Caselles, R. Kimmel, and G. Sapiro, *Geodesic active contours*, Int. J. Comput. Vis., 22 (1997), pp. 61–79.

[9] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, *An introduction to total variation for image analysis*, Theor. Found. Numer. Methods Sparse Recov., 9 (2010), pp. 263–340.

[10] A. Chambolle, D. Cremers, and T. Pock, *A convex approach to minimal partitions*, SIAM J. Imaging Sci., 5 (2012), pp. 1113–1158.

[11] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[12] T. F. Chan, S. Esedoglu, and M. Nikolova, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM J. Appl. Math., 66 (2006), pp. 1632–1648.

[13] T. F. Chan, B. Y. Sandberg, and L. A. Vese, *Active contours without edges for vector-valued images*, J. Vis. Commun. Image Represent., 11 (2000), pp. 130–141.

[14] T. F. Chan and L. A. Vese, *Active contours without edges*, IEEE Trans. Image Proces., 10 (2001), pp. 266–277.

[15] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett., 14 (2007), pp. 707–710.

[16] R. Choksi, Y. G. Gennip, and A. Oberman, *Anisotropic total variation regularized L^1 approximation and denoising/deblurring of* 2D *bar codes*, Inverse Probl. Imaging, 5 (2011), pp. 591–617.

[17] L. D. Cohen, *On active contour models and balloons*, CVGIP Image Underst., 53 (1991), pp. 211–218.

[18] L. Condat, *Discrete total variation: New definition and minimization*, SIAM J. Imaging Sci., 10 (2017), pp. 1258–1290.

[19] L. R. Dice, *Measures of the amount of ecologic association between species*, Ecology, 26 (1945), pp. 297–302.

[20] L. Ding and W. Han, *$\alpha\ell_1 - \beta\ell_2$ regularization for sparse recovery*, Inverse Problems, 35 (2019), 125009.

[21] S. Esedoglu and S. J. Osher, *Decomposition of images by the anisotropic Rudin–Osher–Fatemi model*, Comm. Pure Appl. Math., 57 (2004), pp. 1609–1626.

[22] S. Esedoglu and Y.-H. R. Tsai, *Threshold dynamics for the piecewise constant Mumford–Shah functional*, J. Comput. Phys., 211 (2006), pp. 367–384.

[23] E. Esser, X. Zhang, and T. F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM J. Imaging Sci., 3 (2010), pp. 1015–1046.

[24] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrargian Methods, Stud. Math. Appl., 15, Elsevier, Amsterdam, 1983, pp. 299–331.

[25] P. Getreuer, *Chan–Vese segmentation*, IPOL Image Process. On Line, 2 (2012), pp. 214–224.

[26] T. Goldstein, X. Bresson, and S. Osher, *Geometric applications of the split Bregman method: Segmentation and surface reconstruction*, J. Sci. Comput., 45 (2010), pp. 272–293.

[27] T. Goldstein and S. Osher, *The split Bregman method for $L1$-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.

[28] A. Hantoute, M. A. López, and C. Zălinescu, *Subdifferential calculus rules in convex analysis: A unifying approach via pointwise supremum functions*, SIAM J. Optim., 19 (2008), pp. 863–882.

[29] F. Jia, J. Liu, and X.-C. Tai, *A regularized convolutional neural network for semantic image segmentation*, Anal. Appl., 19 (2021), pp. 147–165.

[30] F. Jia, X.-C. Tai, and J. Liu, *Nonlocal regularized CNN for image segmentation*, Inverse Prob. Imaging, 14 (2020), pp. 891–911.

[31] M. Jung, *Piecewise-smooth image segmentation models with L^1 data-fidelity terms*, J. Sci. Comput., 70 (2017), pp. 1229–1261.

[32] M. Jung, M. Kang, and M. Kang, *Variational image segmentation models involving non-smooth data-fidelity terms*, J. Sci. Comput., 59 (2014), pp. 277–308.

[33] M. Kass, A. Witkin, and D. Terzopoulos, *Snakes: Active contour models*, Int. J. Comput. Vis., 1 (1988), pp. 321–331.

[34] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, *Gradient flows and geometric active contour models*, in Proceedings of IEEE International Conference on Computer Vision, IEEE Computer Society, Los Alamitos, CA 1995, pp. 810–815.

[35] B. Kim and J. C. Ye, *Mumford–Shah loss functional for image segmentation with deep learning*, IEEE Trans. Image Process., 29 (2019), pp. 1856–1866.

[36] M.-J. Lai, Y. Xu, and W. Yin, *Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization*, SIAM J. Numer. Anal., 51 (2013), pp. 927–957.

[37] T. M. Le and L. A. Vese, *Additive and mutiplicative piecewise-smooth segmentation models in a functional minimization approach*, Contemp. Math., 445 (2007), pp. 207–224.

[38] H. A. Le Thi and T. P. Dinh, *DC programming and DCA: thirty years of developments*, Math. Program., 169 (2018), pp. 5–68.

[39] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr, *Convex multi-class image labeling by simplex-constrained total variation*, in International Conference on Scale Space and Variational Methods in Computer Vision, Springer, Berlin, 2009, pp. 150–162.

[40] F. Li, M. K. Ng, T. Y. Zeng, and C. Shen, *A multiphase image segmentation method based on fuzzy region competition*, SIAM J. Imaging Sci., 3 (2010), pp. 277–299.

[41] F. Li, S. Osher, J. Qin, and M. Yan, *A multiphase image segmentation based on fuzzy membership functions and L1-norm fidelity*, J. Sci. Comput., 69 (2016), pp. 82–106.

[42] P. Li, W. Chen, H. Ge, and M. K. Ng, *$\ell_1 - \alpha\ell_2$ minimization methods for signal and image reconstruction with impulsive noise removal*, Inverse Problems, 36 (2020), 055009.

[43] J. Lie, M. Lysaker, and X.-C. Tai, *A binary level set model and some applications to Mumford–Shah image segmentation*, IEEE Trans. Image Process., 15 (2006), pp. 1171–1181.

[44] Y. Lou, S. H. Kang, S. Soatto, and A. L. Bertozzi, *Video stabilization of atmospheric turbulence distortion*, Inverse Probl. Imaging, 7 (2013), pp. 839–861.

[45] Y. Lou and M. Yan, *Fast $L_1 - L_2$ minimization via a proximal operator*, J. Sci. Comput., 74 (2018), pp. 767–785.

[46] Y. Lou, P. Yin, Q. He, and J. Xin, *Computing sparse representation in a highly coherent dictionary based on difference of L_1 and L_2*, J. Sci. Comput., 64 (2015), pp. 178–196.

[47] Y. Lou, P. Yin, and J. Xin, *Point source super-resolution via non-convex L_1 based methods*, J. Sci. Comput., 68 (2016), pp. 1082–1100.

[48] Y. Lou, T. Zeng, S. Osher, and J. Xin, *A weighted difference of anisotropic and isotropic total variation model for image processing*, SIAM J. Imaging Sci., 8 (2015), pp. 1798–1823.

[49] Y. Malitsky and T. Pock, *A first-order primal-dual algorithm with linesearch*, SIAM J. Optim., 28 (2018), pp. 411–432.

[50] D. Martin, C. Fowlkes, D. Tal, and J. Malik, *A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics*, in Proceedings of the Eighth IEEE International Conference on Computer Vision, Vol. 2, IEEE, IEEE Computer Society, Los Alamitos, CA, 2001, pp. 416–423.

[51] B. Merriman, J. K. Bence, and S. J. Osher, *Motion of multiple junctions: A level set approach*, J. Comput. Phys., 112 (1994), pp. 334–363.

[52] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math., 42 (1989), pp. 577–685.

[53] M. Nikolova, *Local strong homogeneity of a regularized estimator*, SIAM J. Appl. Math., 61 (2000), pp. 633–658.

[54] S. Osher and J. A. Sethian, *Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations*, J. Comput. Phys., 79 (1988), pp. 12–49.

[55] F. Park, Y. Lou, and J. Xin, *A weighted difference of anisotropic and isotropic total variation for relaxed Mumford–Shah image segmentation*, in 2016 IEEE International Conference on Image Processing (ICIP), IEEE, IEEE, Piscataway, NJ, 2016, pp. 4314–4318.

[56] T. Pham Dinh and H.A. Le Thi, *Convex analysis approach to DC programming: Theory, algorithms and applications*, Acta Math. Vietnam, 22 (1997), pp. 289–355.

[57] T. Pham Dinh and H.A. Le Thi, *A dc optimization algorithm for solving the trust-region subproblem*, SIAM J. Optim., 8 (1998), pp. 476–505.

[58] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers, *A convex formulation of continuous multi-label problems*, in European Conference on Computer Vision, Springer, Berlin, 2008, pp. 792–805.

[59] Y. Rahimi, C. Wang, H. Dong, and Y. Lou, *A scale-invariant approach for sparse signal recovery*, SIAM J. Sci. Comput., 41 (2019), pp. A3649–A3672.

[60] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.

[61] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia, *A level set model for image classification*, Int. J. Comput. Vis., 40 (2000), pp. 187–197.

[62] M. Storath and A. Weinmann, *Fast partitioning of vector-valued images*, SIAM J. Imaging Sci., 7 (2014), pp. 1826–1852.

[63] M. Storath, A. Weinmann, J. Frikel, and M. Unser, *Joint image reconstruction and segmentation using the Potts model*, Inverse Problems, 31 (2015), 025003.

[64] E. Strekalovskiy and D. Cremers, *Real-time minimization of the piecewise smooth Mumford-Shah functional*, in European Conference on Computer Vision, Springer, Cham, Switzerland, 2014, pp. 127–141.

[65] J. A. Tropp, *Greed is good: Algorithmic results for sparse approximation*, IEEE Trans. Inform. Theory, 50 (2004), pp. 2231–2242.

[66] L. A. Vese and T. F. Chan, *A multiphase level set framework for image segmentation using the Mumford and Shah model*, Int. J. Comput. Vis., 50 (2002), pp. 271–293.

[67] C. Wang, M. Yan, Y. Rahimi, and Y. Lou, *Accelerated schemes for the L_1/L_2 minimization*, IEEE Trans. Signal Process., 68 (2020), pp. 2660–2669.

[68] K. Wei, K. Yin, X.-C. Tai, and T. F. Chan, *New region force for variational models in image segmentation and high dimensional data clustering*, Ann. Math. Sci. Appl., 3 (2018), pp. 255–286.

[69] L. Xu, C. Lu, Y. Xu, and J. Jia, *Image smoothing via L_0 gradient minimization*, ACM Trans. Graph., 30 (2011), pp. 1–12.

[70] Z. Xu, X. Chang, F. Xu, and H. Zhang, *$L_{1/2}$ regularization: A thresholding representation theory and a fast solver*, IEEE Trans. Neural Netw. Learn. Syst., 23 (2012), pp. 1013–1027.

[71] P. Yin, E. Esser, and J. Xin, *Ratio and difference of ℓ_1 and ℓ_2 norms and sparse representation with coherent dictionaries*, Comm. Inform. Systems, 14 (2014), pp. 87–109.

[72] P. Yin, Y. Lou, Q. He, and J. Xin, *Minimization of ℓ_{1-2} for compressed sensing*, SIAM J. Sci. Comput., 37 (2015), pp. A536–A563.

[73] J. Yuan, E. Bae, X.-C. Tai, and Y. Boykov, *A continuous max-flow approach to Potts model*, in European Conference on Computer Vision, Springer, Berlin, 2010, pp. 379–392.

[74] J. Yuan, K. Yin, Y.-G. Bai, X.-C. Feng, and X.-C. Tai, *Bregman-proximal augmented Lagrangian approach to multiphase image segmentation*, in International Conference on Scale Space and Variational Methods in Computer Vision, Springer, Cham, Switzerland, 2017, pp. 524–534.

[75] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer, *Fast global labeling for real-time stereo using multiple plane sweeps.*, in VMV, IOS Press, Amsterdam, 2008, pp. 243–252.

[76] S. Zhang and J. Xin, *Minimization of transformed L_1 penalty: Theory, difference of convex function algorithm, and robust application in compressed sensing*, Math. Program., 169 (2018), pp. 307–336.

[77] M. Zhu and T. Chan, *An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration*, UCLA CAM Report, 34, Los Angeles, 2008.