科学写作（Scientific Writing）

董彬
北京国际数学研究中心
北京大学
http://bicmr.pku.edu.cn/~dongbin

论文内容

- 题目（Title）
- 摘要（Abstract）
- 绪论（Introduction）
- 背景/预备知识（Background/Preliminaries）
- 模型/方法/理论（Method/Theory）
- 数值结果和对结果的讨论（Results and Discussions）
- 结论（Conclusions）
- 致谢（Acknowledgements）

Image restoration: Total variation, wavelet frames, and beyond

Authors: Jian-Feng Cai, Bin Dong, Stanley Osher and Zuowei Shen
Journal: J. Amer. Math. Soc. **25** (2012), 1033-1089
MSC (2010): Primary 35A15, 41A25, 42C40, 45Q05, 65K10, 68U10
DOI: https://doi.org/10.1090/S0894-0347-2012-00740-1
Published electronically: May 17, 2012
MathSciNet review: 2947945

论文整体逻辑构架

Introduction

From the beginning of science, visual observations have been playing important roles.

Image processing has wide applications

Image restoration is crucial

We provide a unified theory

Prelim & Motivation

Methods

Models

Algorithms

Theorems

Proofs

Discussions

Unified theory established

Impact on theory and computations

Inspirations on modeling for image restoration

Possible extensions to other image processing tasks

Implications for applications in science and/or engineering

Modified from "*the paper hourglass: keep the scales in logical order*" by D. Hinsbergen (PGP, UiO)

一些误解

- "I cannot write a clear paper because my English is poor." It has nothing to do with language. It's often the problem of expressing yourself clearly.
- "When I prove the theorem my advisor gave me, or have some excellent numerical results, the work is mostly done." It's probably only 20% done.
- Other general tips:
  - Do not trust the so-called recipe/routine for writing.
  - Use plain English, avoid fancy words.
  - Be illustrative, avoid long sentences.

论文内容

- 题目（Title）
- 摘要（Abstract）
- 绪论（Introduction）
- 背景/预备知识（Background/Preliminaries）
- 模型/方法/理论（Method/Theory）
- 数值结果和对结果的讨论（Results and Discussions）
- 结论（Conclusions）
- 致谢（Acknowledgements）

标题（Title）

- The title is the shortest summary of a paper
- Make titles as short as possible
- Make titles as precise as possible
- Make them conclusive instead of descriptive
- Make titles proper English

Tips: refer to the titles of research articles in the same field

标题（Title）

- Examples
  - Titles such as “Studies on … ”, “Investigations on …” are too vague and contain no solid information.
  - Some good titles:
    - Hilbert’s tenth problem is unsolvable
    - A proof of Minkowski’s inequality for convex curves

标题（Title）

- Examples
  - Titles that are too grand:
    - Computational methods for Euler equations
    - Numerical analysis for fluid mechanics
  - Some good titles:
    - An efficient numerical method for solving a class of delay differential equations
    - A new numerical method for solving convex nonsmooth optimization problems

标题（TITLE）

- Examples (my own)
  - Level set based surface capturing in 3D medical images (not a good summary, not proper English)
  - Distillation ≈ Early Stopping? Harvesting Dark Knowledge Utilizing Anisotropic Information Retrieval for Overparameterized Neural Network (too long, not proper English)
  - PDE-Net: Learning PDEs from Data (good)
  - Sparsifying the Fisher Linear Discriminant by Rotation (good)

标题（TITLE）

- Examples (in machine learning)
  - We used Neural Networks to Detect Clickbaits: You won't believe what happened Next! (arXiv:1612.01340) (not a good summary, too descriptive)
  - BERT has a Mouth, and It Must Speak: BERT as a Markov Random Field Language Model (arXiv:1902.04094) (-_-!!)

标题（TITLE）

- Examples (meanwhile in computer vision)
  - Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval. O. Chum et al. CVPR 2007
  - Total Recall II: Query Expansion Revisited. O. Chum, A. Mikulik, M.Perdoch, and J. Matas. CVPR 2011
  - The Devil is in the Details: an Evaluation of Recent Feature Encoding Methods. K. Chatfield, V. Lempitsky, A. Vedaldi, A. Zisserman. BMVC 2011
  - People Watching: Human Actions as a Cue for Single View Geometry. D. Fouhey et al. ECCV 2012
  - Using the Forest to See the Trees: Exploiting Context for Visual Object Detection and Localization. Torralba, Murphy, and Freeman. CACM 2009
  - Thinking Inside the Box: Using Appearance Models and Context Based on Room Geometry. V. Hedau, D. Hoiem, and D. Forsyth. ECCV 2010
  - Putting Objects in Perspective, by D. Hoiem, A. Efros, and M. Hebert, CVPR 2006
  - What, Where and Who? Classifying Events by Scene and Object Recognition, L.-J. Li and L. Fei-Fei, ICCV 2007
  - Hedging Your Bets: Optimizing Accuracy-Specificity Trade-offs in Large Scale Visual Recognition. J. Deng, J. Krause, A. Berg, L. Fei-Fei. CVPR 2012
  - Some Objects are More Equal Than Others: Measuring and Predicting Importance, M. Spain and P. Perona. ECCV 2008
  - Reading Between the Lines: Object Localization Using Implicit Cues from Image Tags. S. J. Hwang and K.Grauman. CVPR 2010

论文内容

- 题目（Title）
- 摘要（Abstract）
- 绪论（Introduction）
- 背景/预备知识（Background/Preliminaries）
- 模型/方法/理论（Method/Theory）
- 数值结果和对结果的讨论（Results and Discussions）
- 结论（Conclusions）
- 致谢（Acknowledgements）

摘要（Abstract）

- Problem – Action – Result.
- Give readers headlines, not endless summaries.
- If you can give results in numbers, do it.

First few lines: what's the big problem?

Last few lines: what's the breakthrough?

摘要（Abstract）

- Examples

Abstract. Brain aneurysm rupture has been reported to be directly related to the size of aneurysms. The current method used to determine aneurysm size is to manually measure the width of the neck and height of the dome on a computer screen. Because aneurysms usually have complicated shapes, using the size of the aneurysm neck and dome may not be accurate and may overlook important geometrical information. In this paper we present a level set based illusory surface algorithm to capture the aneurysms from the vascular tree. Since the aneurysms are described by level set functions, not only the volume but also the curvature of aneurysms can be computed for medical studies. Experiments and comparisons with models used for capturing illusory contours in 2D images are performed. This includes applications to clinical image data demonstrating the procedure of accurately capturing a middle cerebral artery aneurysm.

Problem: didn’t mention the highlights of the proposed methodology

摘要（ABSTRACT）

- Examples

The wavelet frame based approach is one of the most successful mathematical tools for image restoration due to its good capability of sparsely approximating piecewise smooth functions such as images. In this paper, we introduce a new edge driven wavelet frame based model for image restoration which treats an image as an approximation of piecewise smooth function. By the term "edge driven", we mean the alternative update of the image and its discontinuities, as previous models in [13, 46]. The major difference between our model and the previous models is that our model uses ℓ_1 norm of wavelet frame coefficients instead of ℓ_2 norm to promote the smoothness through a careful choice of wavelet frame systems and parameters. Since ℓ_1 norm does not smear the discontinuities out, our model is less sensitive to the estimation of discontinuities compared to the previous ℓ_2 norm based models. In addition, our model provides a regularization on the discontinuity regions by relaxing the binary image of discontinuity regions into the image taking the intermediate values. This enables to provide an asymptotic analysis of the edge driven model as the resolution goes to infinity, which establishes a rigorous connection between the discrete edge driven model and the variational model in the continuum setting. Finally, the numerical simulation shows the results from our edge driven model are compared favorably against several existing image restoration models.

Problems:

- Didn't start with a big problem
- Too technical when describing the proposed method
- For a paper at the interface of theory and applications, didn't mention any potential impact or implications

摘要（ABSTRACT）

- Examples

  A relatively good one

Big picture

Non-technical description of the method

Highlights of the method

Potentials

Partial differential equations (PDEs) play a prominent role in many disciplines of science and engineering. PDEs are commonly derived based on empirical observations. However, with the rapid development of sensors, computational power, and data storage in the past decade, huge quantities of data can be easily collected and efficiently stored. Such vast quantity of data offers new opportunities for data-driven discovery of physical laws. Inspired by the latest development of neural network designs in deep learning, we propose a new feed-forward deep network, called PDE-Net, to fulfill two objectives at the same time: to accurately predict dynamics of complex systems and to uncover the underlying hidden PDE models. Comparing with existing approaches, our approach has the most flexibility by learning both differential operators and the nonlinear response function of the underlying PDE model. A special feature of the proposed PDE-Net is that all filters are properly constrained, which enables us to easily identify the governing PDE models while still maintaining the expressive and predictive power of the network. These constrains are carefully designed by fully exploiting the relation between the orders of differential operators and the orders of sum rules of filters (an important concept originated from wavelet theory). Numerical experiments show that the PDE-Net has the potential to uncover the hidden PDE of the observed dynamics, and predict the dynamical behavior for a relatively long time, even in a noisy environment.

论文内容

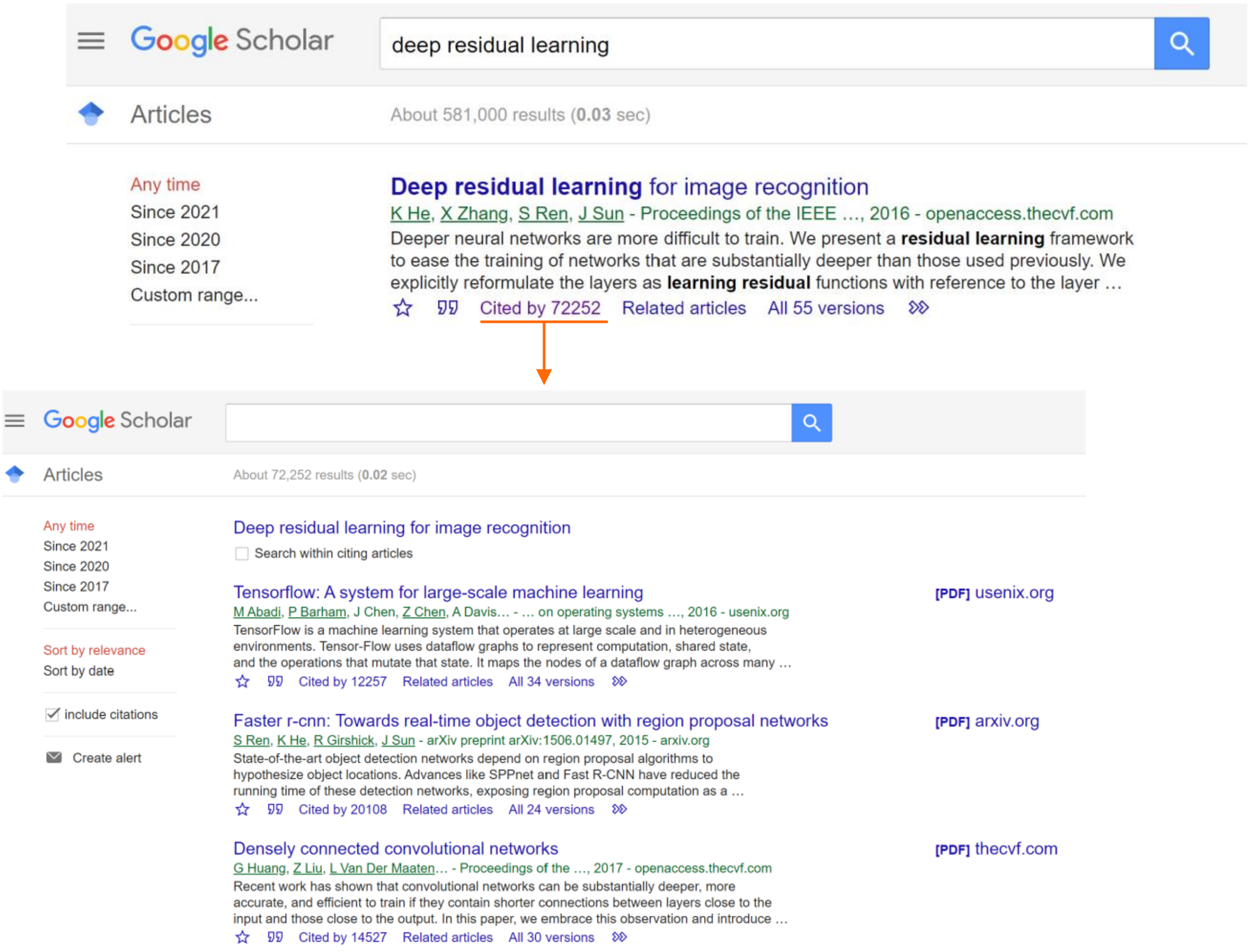
- 题目（Title）
- 摘要（Abstract）
- 绪论（Introduction）
- 背景/预备知识（Background/Preliminaries）
- 模型/方法/理论（Method/Theory）
- 数值结果和对结果的讨论（Results and Discussions）
- 结论（Conclusions）
- 致谢（Acknowledgements）

绪论（Introduction）

- Conduct thorough literature review
- Source for references:
  - Google scholar
  - Arxiv
  - Scopus (Elsevier)
  - ResearchGate

绪论（Introduction）

- Surfing Google scholar



绪论（INTRODUCTION）

- Conduct thorough literature review
- Source of references:
  - Google scholar
  - Arxiv
  - Scopus (Elsevier)
  - ResearchGate
- How to estimate quality before reading it?
  - Examine the general format of the paper
  - Authors' institutes/company and citations
- Read introductions, settings, background sections carefully. (After 5 papers, you get a pretty good idea)

绪论（INTRODUCTION）

- Writing your introduction
  - It is the hardest part of the paper: from the Big Problems of the world, you select one, and solve part of it: like the trailer of a movie.
  - Zoom in from a big problem to a small solvable one, and you indicate how you solve it.
  - Know your audience.
- It is important to write a good introduction
  - Be accurate: make sure you truly understand the cited works.
  - Be professional: make fair comments on others' work.
- Start writing it when other parts are mostly done

绪论（INTRODUCTION）

- Example

1. **Introduction**

From the beginning of science, visual observations have been playing important roles. Advances in computer technology have made it possible to apply some of the most sophisticated developments in mathematics and the sciences to the design and implementation of fast algorithms running on a large number of processors to process image data. As a result, image processing and analysis techniques are now applied to virtually all natural sciences and technical disciplines ranging from computer sciences and electronic engineering to biology and medical sciences; and digital images have come into everyone's life. Image restoration, including image denoising, deblurring, inpainting, computed tomography, etc., is one of the most

The big problem: mathematics in imaging science

important areas in image processing and analysis. Its major purpose is to enhance the quality of a given image that is corrupted in various ways during the process of imaging, acquisition and communication, and enables us to see crucial but subtle objects reside in the image. Therefore, image restoration is an important step to take towards the accurate interpretations of the physical world and making the optimal decisions.

Mathematics has been playing an important role in image and signal processing from the very beginning; for example, Fourier analysis is one of the main tools in signal and image analysis, processing, and restoration. In fact, mathematics has been one of the driving forces of the modern development of image analysis, processing and restorations. At the same time, the interesting and challenging problems in imaging science also gave birth to new mathematical theories, techniques and methods. The variational methods (e.g., total variation based methods) and wavelets and wavelet frame based methods developed in the last few decades for image and signal processing are two successful recent examples among many. This paper is designed to establish connections between these two major image restoration approaches: variational methods and wavelet frame based methods. Such connections provide new interpretations and understanding of both approaches, and hence, lead to new applications for both approaches.

The solvable problem: connection between approach A & B

We start with an introduction of both the variational and wavelet frame based methods. The basic linear image restoration model used for variational methods is usually given as

$$f = Au + \eta, \tag{1.1}$$

where A is some linear bounded operator (not invertible in general) mapping one function space to another, e.g., the identity operator for image denoising, or a convolution operator for image deconvolution, and η denotes a perturbation caused by the additive noise in the observed image (or measurements), which is typically assumed to be a white Gaussian noise.

In order to recover the unknown image u from the equation (1.1), a typical variational approach takes the following form:

$$\inf_u \nu\|Du\|_* + \frac{1}{2}\|Au - f\|^2_{L_2(\Omega)}, \tag{1.2}$$

where D is a vector of some (weighted) differential operators, $\|\cdot\|_*$ is some properly chosen norm, Ω is some domain in $\mathbb{R}^2$, and ν is some scalar parameter. For example, when $D = \nabla$ and $\|\nabla u\|_* := \int_\Omega |\nabla u| \mathrm{d}x\mathrm{d}y$, then (1.2) is the well-known Rudin-Osher-Fatemi (ROF) model [1]:

Non-technical introduction of the mathematics for approach A

$$\inf_u \nu \int_\Omega |\nabla u| \mathrm{d}x\mathrm{d}y + \frac{1}{2}\|Au - f\|^2_{L_2(\Omega)}. \tag{1.3}$$

the ROF model works exceptionally well in terms of preserving edges while suppressing noise. After the ROF model was proposed, there were many variational methods developed in the literature. We shall provide a short review of variational methods in the next section.

and are numerically computable for all objective functionals considered in this paper. Therefore, we will not enforce additional conditions on $\boldsymbol{D}$ and $\boldsymbol{A}$ to ensure attainability of the infimum.

The image restoration model (1.1) views images as functions defined on a continuum, i.e. analog signals. What we observe in practice, however, are digital images which are discrete versions of their continuous counterparts. Given an analog data f, its digital/discrete version $\boldsymbol{f}$ can be obtained by various sampling schemes. However, digital images are never given in a function form. Hence, when model (1.1) is adopted, one needs to obtain an approximation of an observed function from the given digital image. Since digital images are discrete sequences and the restoration of a digital image is to restore a sequence from an observed sequence, it is more natural to view a digital image as a sequence $\boldsymbol{f}$ and establish a restoration model in a sequence space instead of a function space. This is what the wavelet frame based approach is for. In fact, as we will see, the wavelet frame based approach naturally fits the discrete setting of digital images.

The digital image restoration model is the discrete version of (1.1), which is to find the unknown true digital image $\boldsymbol{u}$ from an observed image (or measurements) $\boldsymbol{f}$ defined by

$$\boldsymbol{f} = \boldsymbol{A}\boldsymbol{u} + \boldsymbol{\eta}, \tag{1.4}$$

where $\boldsymbol{A}$ is a linear operator and $\boldsymbol{\eta}$ is a Gaussian noise. Here the linear operator $\boldsymbol{A}$ is the identity operator for image denoising, a convolution operator for image deconvolution, and a projection operator for image inpainting.

The problem (1.4) is a typical linear inverse problem. The frame based image restoration model that we shall focus on in this paper is the analysis based approach, which is to solve

Non-technical introduction of the mathematics for approach B

$$\inf_{\boldsymbol{u}} \nu\|\boldsymbol{\lambda} \cdot \boldsymbol{W}\boldsymbol{u}\|_* + \frac{1}{2}\|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{f}\|_2^2, \tag{1.5}$$

where $\boldsymbol{W}$ is the discrete wavelet frame transform using filters of some tight wavelet frame system, and $\|\cdot\|_*$ is some properly chosen norm that reflects the regularity or sparse properties of the underlying solutions, e.g., the weighted ℓ_1-norm is commonly used. We will provide details on wavelet frame transforms and the definition of $\|\cdot\|_*$ in the next section. In order to have a desired solution of (1.4) via solving (1.5), the underlying solution should have a good sparse approximation under the wavelet frame transform $\boldsymbol{W}$. One of the advantages of using tight wavelet frame systems is that they provide reasonably good sparse approximations to piecewise smooth functions, which form a large class of functions that images belong to.

When the wavelet frame based approach (1.5) is used, the digital image data is given in a sequence form and the minimization is also done in sequence space. Hence the solution is naturally a sequence. The underlying function from which the digital image is sampled does not appear explicitly. It appears implicitly when the regularity of the underlying solution is mentioned, which is stated in terms of the decay of the wavelet frame coefficients $\boldsymbol{W}u$. Furthermore, one can obtain an approximate solution in function form whenever it is needed. Indeed, since wavelet frame based approaches normally generate coefficients of a wavelet frame system (or coefficients of shifts of the corresponding refinable function that generates the wavelet frame system), it is easy to reconstruct a function from these coefficients.

In fact, as we will see, when wavelet frames are used, we can interpret the digital image $\boldsymbol{f}$ as $\boldsymbol{f}[\boldsymbol{k}] = \langle f, \phi(\cdot - \boldsymbol{k})\rangle$ for some function ϕ to link it to function

绪论（Introduction）

- Example

space. The operator $\boldsymbol{A}$ can be viewed as a discrete version of the continuous linear operator A of (1.1) through certain discretization. Furthermore, when special wavelet frames are chosen, the wavelet frame transform $\boldsymbol{\lambda}\cdot\boldsymbol{W}\boldsymbol{u}$ can be regarded as certain discretizations of $\boldsymbol{D}u$. This motivates us to explore whether the model (1.5) approximates (1.2) and whether the approximation becomes more accurate when image resolution increases. One of the major contributions of this paper is to establish a connection between the wavelet frame based image variational approach (1.5) and the differential operator based variational model (1.2), as well as their corresponding approximate minimizers. An approximate minimizer is the one on which the value of the objective functional is close to the infimum.

In order to briefly summarize our results, we need to introduce some notation. The details can be found in later sections. First, define the objective function

$$F_n(\boldsymbol{u}) := \nu\|\boldsymbol{\lambda}\cdot\boldsymbol{W}\boldsymbol{u}\|_* + \frac{1}{2}\|\boldsymbol{A}_n\boldsymbol{u}-\boldsymbol{f}\|_2^2,$$

where $\boldsymbol{u}$ and $\boldsymbol{f}$ are $N\times N$ with $N = 2^n+1$. The operator $\boldsymbol{A}_n$ should be understood as a proper discretization of some operator A defined on a function space. We will first show that the approximate minimizers of F_n can be put into correspondence with those of the following objective functional:

Non-technical forecast of the main findings

$$E_n(u) := \nu\|\boldsymbol{\lambda}\cdot\boldsymbol{W}_n\boldsymbol{T}_n u\|_* + \frac{1}{2}\|\boldsymbol{A}_n\boldsymbol{T}_n u - \boldsymbol{T}_n f\|_2^2,$$

where $\boldsymbol{T}_n$ is some discretization operator associated to the given wavelet frame system. In particular, if $\boldsymbol{u}_n^\star$ is an (approximate) minimizer of F_n, then we can explicitly construct a $u_n^\star$ such that $u_n^\star$ is an (approximate) minimizer of E_n. The converse argument is also true. Note that E_n is a functional defined on some function space, as well as the objective functional

$$E(u) := \nu\|\boldsymbol{D}u\|_* + \frac{1}{2}\|Au-f\|_{L_2(\Omega)}^2.$$

For clarity of the presentation here, we postpone the detailed definition of the function space of u and the domains of the operators $\boldsymbol{D}$, $\boldsymbol{W}$ and $\boldsymbol{T}_n$ to later sections.

While E_n is essentially a reformulation of F_n, it can also be regarded as an approximation of E. For this, we first show that E_n pointwise converges to E. Then, we prove that E_n Γ-converges to E (see Definition 3.1 for the definition of Γ-convergence; also see, e.g., [4] for an introduction of Γ-convergence). In fact, we shall prove a convergence result that is stronger than the Γ-convergence. The analysis is based on approximation of functions through multiresolution structures generated by B-splines and their associated spline wavelet frames constructed from [2]. In numerical computation for both the variational and wavelet frame based approaches, the computed solutions are often those whose values of the corresponding objective functional E or E_n are ϵ close to the infimum. We refer to such solutions as ϵ-optimal solutions to E or E_n (see (3.7) for the definition). Denote by $u_n^\star$ an ϵ-optimal solution to E_n, which can be constructed by the ϵ-optimal solution to F_n, denoted as $\boldsymbol{u}_n^\star$. As a result of Γ-convergence, we are able to connect the functional E_n with E, hence F_n with E, in the following sense:

$$\limsup_{n\to\infty} F_n(\boldsymbol{u}_n^\star) = \limsup_{n\to\infty} E_n(u_n^\star) \le \inf_u E(u) + \epsilon.$$

The above inequality shows that $\{u_n^\star\}$ is a sequence that enables E_n to approximate the optimal value of E from below. Furthermore, any cluster point of $\{u_n^\star\}$ is an ϵ-optimal solution to E.

The discussions of the similarity and difference between the total variation method and wavelet method for image denoising have already been given through the discussions of the space of bounded variation functions (BV space) and the Besov space (e.g., $B_1^{1,1}$ space) in [5], which provides a fundamental understanding of the two approaches for image denoising. The results given here not only consider more general image restoration problems, but also reveal close connections between the solutions of wavelet frame based image restoration, which is used in numerical computing, and those of the total variation method. Furthermore, since the number of levels used in wavelet decomposition is fixed independent of the resolution (though the resolution can be higher when the give data is denser) and since the parameters for the low pass filter are always set to be zero in numerical implementation of (1.5), the weighted ℓ_1-norm of wavelet coefficients here is not equivalent to the Besov norm as discussed in [5]. All these indicate that the study here takes a different approach and leads to a different set of results which will extend our understanding for both the wavelet frame based and variational approaches. Our conclusion goes beyond the theoretic justifications of the linkage of the two approaches. Since the total variation approach has a strong geometric interpretation, this connection gives geometric interpretations to the wavelet frame based approach (1.5) as well as its minimizers, as it can be understood as the discrete form of (1.2). This also leads to even wider applications of the wavelet frame based approach, e.g., image segmentation [6] and 3D surface reconstruction from unorganized point sets [7]. On the other hand, for any given variational model (1.2), (1.5) provides various and flexible discretizations, as well as fast numerical algorithms. Using a wavelet frame based approach, we can approximate various differential operators for model (1.2) by choosing a proper wavelet frame transform and parameters $\boldsymbol{\lambda}$. Furthermore, by putting (1.2) into a wavelet frame setting, one can use a multiresolution structure to adaptively choose proper differential operators in different regions of a given image according to the orders of the singularities of underlying solutions. It should be pointed out here that if one wants to use a more general differential operator in (1.2), the ability of applying different differential operators according to where various singularities are located is the key to make such a generalization successful. The wavelet frame based approach has a built-in adaptive mechanism via the multiresolution analysis that provides a natural tool for this very purpose. Finally, we can use more general wavelet frame based approaches, e.g., the wavelet frame based balanced approach, or using two-system models to solve various generalizations of (1.2), as will be shown in a later section.

Existing work and scientific impact of the findings

The rest of this paper is organized as follows. In Section 2, we will first review some of the classical as well as some most recent variational approaches. Then we will review some basic notions of wavelet frames, fast framelet transforms, and wavelet frame based image restoration approaches. Finally, we will motivate the readers by showing the link between the wavelet frame based approach with Haar framelet and the ROF model with $A = I$. Theoretical justifications of the link between (1.2) and (1.5) will be given in Section 3 with some technical details given later in Section 4. In particular, in Section 3, we will prove that E_n pointwise converges to E, which indicates that E_n can be used to approximate E. In order to

Organization of the rest of the paper

论文内容

- 题目（Title）
- 摘要（Abstract）
- 绪论（Introduction）
- 背景/预备知识（Background/Preliminaries）
- 模型/方法/理论（Method/Theory）
- 数值结果和对结果的讨论（Results and Discussions）
- 结论（Conclusions）
- 致谢（Acknowledgements）

背景/预备知识（Background/Preliminaries）

- Include ALL information needed to understand the paper.
- Include ALL information related to the problem you selected.
- Include NO information that is not needed or relevant: don't show off how much you have read
- Tips: you may start writing it AFTER you finish writing the section on Method/Theory

背景/预备知识（Background/Preliminaries）

- Example

2. Models and Motivations

In this section, we first recall some variational methods, as well as their history and recent developments. Then we review some basic notation of wavelet frames and wavelet frame based image restoration approaches. At the end, we shall give an example to illustrate that these two approaches are closely related.

2.1. Total Variation Based Models and Their Generalizations.

2.1.1. *Total Variation Based Models.* The trend of variational methods and partial differential equation (PDE) based image processing started with the refined Rudin-Osher-Fatemi (ROF) model [1]:

$$\inf_u \nu \int_\Omega |\nabla u| \mathrm{d}x\mathrm{d}y + \frac{1}{2}\|Au - f\|^2_{L_2(\Omega)}. \tag{2.1}$$

The ROF model aims at finding a desirable solution of (1.1). The choice of regularization $\int_\Omega |\nabla u|\mathrm{d}x\mathrm{d}y$ is the total variation (TV) of u, which is much more effective than the classical choice $\int_\Omega |\nabla u|^2 \mathrm{d}x\mathrm{d}y$ in terms of preserving sharp edges, which are key features for images. Many of the current PDE based methods for image denoising and decomposition utilize TV regularization for its beneficial edge-preserving property (see, e.g., [5, 8, 9]). The ROF model is especially effective on restoring images that are piecewise constant, e.g., binary images.

Approach A

To solve the ROF model (2.1), one can solve the corresponding Euler-Lagrange equation:

$$-\nu\nabla \cdot \left(\frac{\nabla u}{|\nabla u|}\right) + A^\top(Au - f) = 0. \tag{2.2}$$

To solve the Euler-Lagrange equation (2.2), it was proposed in [1] to use a time evolution PDE which is essentially the steepest descent equation for the objective functional (2.1). This time evolution PDE was later modified by [10] to improve the Courant-Friedrichs-Lewy (CFL) condition when solving the PDE using explicit finite difference schemes. To completely remove the CFL restriction, a fixed point iteration scheme which solves the Euler-Lagrange equation (2.2) directly was proposed in [11].

All the numerical methods in [1, 10, 11] aimed at solving the Euler-Lagrange equation (2.2), which becomes singular when $|\nabla u| = 0$. To avoid this, [1, 10, 11] used a regularized parabolic term for (2.2) instead, i.e.

$$-\nu\nabla \cdot \left(\frac{\nabla u}{|\nabla u|_\epsilon}\right) + A^\top(Au - f) = 0, \quad \text{where } |\nabla u|_\epsilon := \sqrt{|\nabla u|^2 + \epsilon^2}.$$

Another approach reducing the staircase effect without multi-layer assumptions of the images is adding a higher order regularization term to the original ROF model by [29]. Thanks to the higher order term, the solution prefers "ramps" over "staircases" (see [29] for details).

More recently, a more general variational approach was proposed by [30], where the authors introduced total generalized variation (TGV) as the regularization term. Here we will recall a modified version of the general TGV model (we shall refer to it as the TGV model in this paper):

$$\inf_{u,v} \nu_1\|\nabla u - v\|_{L_1(\Omega)} + \nu_2\|\nabla \cdot v\|_{L_1(\Omega)} + \frac{1}{2}\|Au - f\|^2_{L_2(\Omega)}, \tag{2.5}$$

where the variable $v = (v_1, v_2)$ varies in the space of all continuously differential 2-tensors and

$$\|\nabla \cdot v\|_{L_1(\Omega)} := \int_\Omega \sqrt{(\partial_x v_1)^2 + (\partial_y v_1)^2 + (\partial_x v_2)^2 + (\partial_y v_2)^2}\mathrm{d}x\mathrm{d}y.$$

The TGV model generalizes the inf-convolution model in the sense that it coincides with the inf-convolution model when v only varies in the range of ∇. The TGV model improves the results for image denoising over the ROF model and inf-convolution model. The interested reader should consult [30] for details.

Similar to total variation based methods, we will establish the connections between the inf-convolution model (2.4) and the TGV model (2.5) with some wavelet frame based approach. Both the inf-convolution model (2.4) and the TGV model (2.5) can be approximated by some tight wavelet frame based approach, while it is more natural and efficient to have a frame based approach for these generalizations as well. We shall postpone the detailed discussions to Section 6.

2.2. Wavelet Frame Based Models. This section is devoted to recalling the wavelet frame based image restoration approaches. We start with the concept of a tight frame and a tight wavelet frame, and then introduce the analysis based image restoration approach and other more general wavelet frame based approaches.

2.2.1. *MRA-Based Tight Frames.* In this subsection, we briefly introduce the concept of tight frames and tight wavelet frames. The interested readers should consult [2, 31, 32] for theories of frames and wavelet frames, [33] for a short survey on the theory and applications of frames, and [34] for a more detailed survey.

Approach B

A countable set $X \subset L_2(\mathbb{R}^d)$, with $d \in \mathbb{Z}^+$, is called a tight frame of $L_2(\mathbb{R}^d)$ if

$$f = \sum_{g\in X}\langle f, g\rangle g \quad \forall f \in L_2(\mathbb{R}^d), \tag{2.6}$$

where $\langle\cdot,\cdot\rangle$ is the inner product of $L_2(\mathbb{R}^d)$.

For given $\Psi := \{\psi_1, \ldots, \psi_q\} \subset L_2(\mathbb{R}^d)$, the corresponding quasi-affine system $X^J(\Psi)$ generated by Ψ is defined by the collection of the dilations and the shifts of Ψ as

$$X^J(\Psi) = \{\psi_{\ell,n,\boldsymbol{k}} : 1 \le \ell \le q; n \in \mathbb{Z}, \boldsymbol{k} \in \mathbb{Z}^d\}, \tag{2.7}$$

where $\psi_{\ell,n,\boldsymbol{k}}$ is defined by

$$\psi_{\ell,n,\boldsymbol{k}} := \begin{cases} 2^{\frac{nd}{2}}\psi_\ell(2^n \cdot - \boldsymbol{k}), & n \ge J; \\ 2^{(n-\frac{J}{2})d}\psi_\ell(2^n \cdot - 2^{n-J}\boldsymbol{k}), & n < J. \end{cases} \tag{2.8}$$

It should be noted that the wavelet frame based approaches solve image restoration problems in digital domain directly, and they give a wide variety of models, (e.g., the synthesis based, analysis based and balanced approaches, and approaches with multiple frame systems), as well as the associated efficient algorithms which utilize the multilevel nature of wavelet frame transforms to achieve better sparse approximation of the underlying solutions. The interested readers should consult, for example, the survey articles [33, 34] for further details.

2.3. Motivation by an Example. As stated, we are aiming at establishing the connections between the wavelet frame based approach and differential operator based variational approaches. The connection is done by using the analysis based approach (2.18) to approximate variational models when a proper $\boldsymbol{\lambda}$ is chosen. In this section, we use the ROF denoising model, i.e.,

$$\min_u E(u) := \nu\int_\Omega |\nabla u|\mathrm{d}x\mathrm{d}y + \frac{1}{2}\|u - f\|^2_{L_2(\Omega)} \tag{2.20}$$

Motivation under a simpler setting

as an example. Note that "min" is used here because the minimal value of E defined above is attainable since E is coercive [3].

A complete analysis of more general cases is given in Section 3. The emphasis here is to use a simple example to motivate us and ready ourselves for more complicated ones. Some of the technical details may be left out and a complete analysis will be given in Sections 3 and 4.

For simplicity, we start with a $u \in L_2(\Omega)$ that is sufficiently smooth. Let $\boldsymbol{u}_|$ denote the restriction of u on the discrete mesh with meshsize $h = 1/(N-1)$, i.e.,

$$(\boldsymbol{u}_|)[i,j] = u(x_i, y_j), \quad \text{with } (x_i, y_j) = (ih, jh) \text{ and } 0 \le i, j \le N-1.$$

For the data fidelity term, it is straightforward to use $\frac{1}{2}\|\boldsymbol{u}_| - \boldsymbol{f}_|\|_2^2$. Here the ℓ_2-norm, as well as other norms involved, are scaled by taking the meshsize h into account. In particular, we define $\|\boldsymbol{x}\|_p^p := h^2\sum_{i,j=0}^{N-1}|\boldsymbol{x}[i,j]|^p$. It can be verified that

$$\frac{1}{2}\|\boldsymbol{u}_| - \boldsymbol{f}_|\|_2^2 \to \frac{1}{2}\|u - f\|^2_{L_2(\Omega)}, \qquad \text{as } h \to 0. \tag{2.21}$$

论文内容

- 题目（Title）
- 摘要（Abstract）
- 绪论（Introduction）
- 背景/预备知识（Background/Preliminaries）
- 模型/方法/理论（Method/Theory）
- 数值结果和对结果的讨论（Results and Discussions）
- 结论（Conclusions）
- 致谢（Acknowledgements）

模型/方法/理论（METHOD/THEORY）

- This is the easiest part of the paper. Can be written first.
- Your idea or method may seem simple and don't have much to write. However,
  - Theory paper:
    - Corollaries: theorems may be abstract, but corollaries are meant to be more intuitive
    - Discussions on the assumptions and settings: reasonable?
    - Variants of your main theorem: changed assumptions or settings
  - Methodology paper:
    - Motivations: may use illustrative examples
    - Possible variants: ablations or alternative opt. algorithms

模型/方法/理论（METHOD/THEORY）

- Example: theory paper

Theorem 3.2. *Suppose that Assumptions (A1)–(A3) are satisfied. Then, for every sequence* $u_n \to u$ *in* $W_1^s(\Omega)$*, we have* $\lim_{n\to+\infty} E_n(u_n) = E(u)$*. Consequently,* E_n Γ*-converges to* E *in* $W_1^s(\Omega)$*.*

Corollary 3.1. *Suppose that Assumptions (A1)–(A3) are satisfied. Let* $u_n^\star$ *be an* ϵ*-optimal solution of* E_n *for a given* $\epsilon > 0$ *and for all* n*.*

(1) *We have*

$$\limsup_{n\to\infty} E_n(u_n^\star) \leq \inf_u E(u) + \epsilon. \tag{3.16}$$

In particular, when $u_n^\star$ *is a minimizer of* E_n*, we have*

$$\limsup_{n\to\infty} E_n(u_n^\star) \leq \inf_u E(u).$$

(2) *If, in addition, the set* $\{u_n^\star\}$ *has a cluster point* $u^\star$*, then* $u^\star$ *is an* ϵ*-optimal solution to* E*. In particular, when* $u_n^\star$ *is a minimizer of* E_n *and* $u^\star$ *a cluster point of the set* $\{u_n^\star\}$*, then*

$$E(u^\star) = \limsup_{n\to\infty} (E_n(u_n^\star)) = \inf_u E(u)$$

and $u^\star$ *is a minimizer of* E*.*

模型/方法/理论（METHOD/THEORY）

- Example: theory paper

Theorem 3.2. *Suppose that Assumptions (A1)–(A3) are satisfied. Then, for every sequence* $u_n \to u$ *in* $W_1^s(\Omega)$, *we have* $\lim_{n\to+\infty} E_n(u_n) = E(u)$. *Consequently,* E_n Γ*-converges to* E *in* $W_1^s(\Omega)$.

4.2. **Discretization of Linear Operator** A**.** We assume that A is a bounded linear operator that maps $L_2(\Omega)$ into $L_2(\Omega)$. In this subsection, we demonstrate how to discretize A to get a matrix $\boldsymbol{A}_n$. As stated in Section 3.2.2, we require that A and $\boldsymbol{A}_n$ satisfy Assumption (A1); i.e., the discretization $\boldsymbol{A}_n$ should be consistent with A asymptotically. Obviously, not all discretizations of A would satisfy the assumption (A1). In this subsection, we demonstrate how to properly discretize A such that Assumption (A1) is indeed satisfied. We shall focus on image denoising, image inpainting, and image deblurring problems.

4.3. **Discretization of Differential Operator** $\boldsymbol{D}$**.** We imposed Assumptions (A2) and (A3) in Section 3.2.2 for the discretization of the differential operators. In this section, we verify that all B-spline tight frame systems constructed by [2] do satisfy these assumptions.

模型/方法/理论（METHOD/THEORY）

- Example: methodology paper

3.2. **Our Model.** Here we introduce our modified illusory surface model based on equation (3),

$$\frac{\partial \phi}{\partial t} = |\nabla \phi| \Big(A(\psi) \nabla d \cdot \frac{\nabla \phi}{|\nabla \phi|} + d \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} - \alpha H(\psi) + \beta \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} \Big), \qquad (10)$$

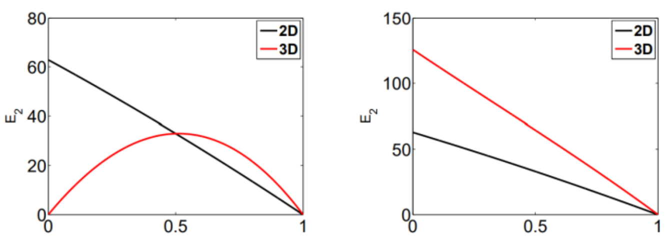
$$A(\psi) = 1 + \mu \kappa^{+}(\psi), \qquad (11)$$

where μ is a constant parameter and $\kappa^{+}(\psi)$ is the positive part of the Gaussian curvature of ψ.

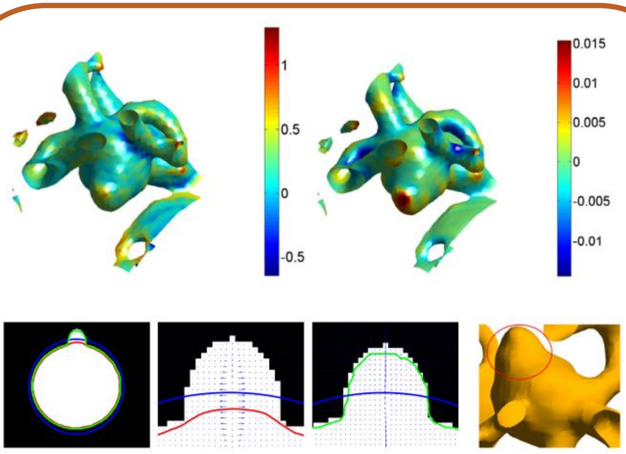
One key novelty is the use of the Gaussian curvature. Thus I used simplified example and numerical demonstrations.

$$E_2(r) = \int_S (1 + \mu \kappa_m) d(s) ds = \begin{cases} 2\pi(1 + \frac{\mu}{r})(1 - r) r, & \text{2D} \\ 4\pi(1 + \frac{\mu}{r})(1 - r) r^2, & \text{3D}; \end{cases}$$

$$E_2(r) = \int_S (1 + \mu \kappa_g) d(s) ds = \begin{cases} 2\pi(1 + \frac{\mu}{r})(1 - r) r, & \text{2D} \\ 4\pi(1 + \frac{\mu}{r^2})(1 - r) r^2, & \text{3D}. \end{cases}$$



Simplified example



Numerical demonstration

论文内容

- 题目（Title）
- 摘要（Abstract）
- 绪论（Introduction）
- 背景/预备知识（Background/Preliminaries）
- 模型/方法/理论（Method/Theory）
- 数值结果和对结果的讨论（Results and Discussions）
- 结论（Conclusions）
- 致谢（Acknowledgements）

结果与讨论（Results and Discussions）

- Only present your own, new results. Anything published should be in the "Background" section.
- Only present relevant results: in support of your theory or methodology.
- Give representative examples: e.g. one for each type of PDE, instead of multiple for one type
- Present important results (not the boring stuffs)
- Be honest: do not use the best results as the typical examples
- When discuss your results, do not just repeat the results. Give your own interpretations.

Best way to improve on these is by reading others' papers (good ones).

论文内容

- 题目（Title）
- 摘要（Abstract）
- 绪论（Introduction）
- 背景/预备知识（Background/Preliminaries）
- 模型/方法/理论（Method/Theory）
- 数值结果和对结果的讨论（Results and Discussions）
- 结论（Conclusions）
- 致谢（Acknowledgements）

结论（CONCLUSIONS）

- Brief summary of the main findings of the paper
- Make the conclusion as short as possible
- Draw SMALL conclusions (avoid overclaiming)
- Highlight future works

5. Conclusion and Discussion

In this paper, we designed a deep feed-forward network, called the PDE-Net, to discover the hidden PDE model from the observed dynamics and to predict the dynamical behavior. The PDE-Net consists of two major components which are jointly trained: to approximate differential operations by convolutions with properly constrained filters, and to approximate the nonlinear response by deep neural networks or other machine learning methods. The PDE-Net is suitable for learning PDEs as general as in (1). As an example, we considered a linear variable-coefficient convection-diffusion equation. The results show that the PDE-Net can uncover the hidden equation of the observed dynamics, and predict the dynamical behavior for a relatively long time, even in a noisy environment. PyTorch codes of the PDE-Net are available at `https://github.com/ZichaoLong/PDE-Net`. As part of the future work, we will try the proposed framework on real data sets. One of the important directions is to uncover hidden variables which cannot be measured by sensors directly, such as in data assimilation. Another interesting direction which is worth exploring is to learn stable and consistent numerical schemes for a given PDE model based on the architecture of the PDE-Net.

论文内容

- 题目（Title）
- 摘要（Abstract）
- 绪论（Introduction）
- 背景/预备知识（Background/Preliminaries）
- 模型/方法/理论（Method/Theory）
- 数值结果和对结果的讨论（Results and Discussions）
- 结论（Conclusions）
- 致谢（Acknowledgements）

致谢（Acknowledgements）

- Who to thank?
  - Those who helped you, but not a co-author
  - The reviewers
  - The funding agencies
- Example:
  - We would like to thank X1 and X2 for the helpful discussions during the preparation of the paper, and X3 for providing us with her source codes for comparison. We would also like to thank the anonymous reviewers for their valuable suggestions. Author1 is supported in part by NSF xxxx and NIH xxxx. Author 2 is supported in part by NSF yyyy and NIH yyyy.