virtual ICM 2022
International Congress of Mathematicians (ICM) | 6–14 July 2022

On Mathematical Modeling in Image Reconstruction and Beyond

Bin Dong
Beijing International Center for Mathematical Research
Center for Machine Learning Research
Peking University
北京大學 PEKING UNIVERSITY
北京国际数学研究中心 BEIJING INTERNATIONAL CENTER FOR MATHEMATICAL RESEARCH

Table of contents

01

Background

Image-based scientific discovery and computational imaging (sensing, reconstruction and analysis).

Background

- The development of natural science has been heavily relying on **visual examination**.
- To better analyze the phenomenon-of-interest, **images** are convenient tools.

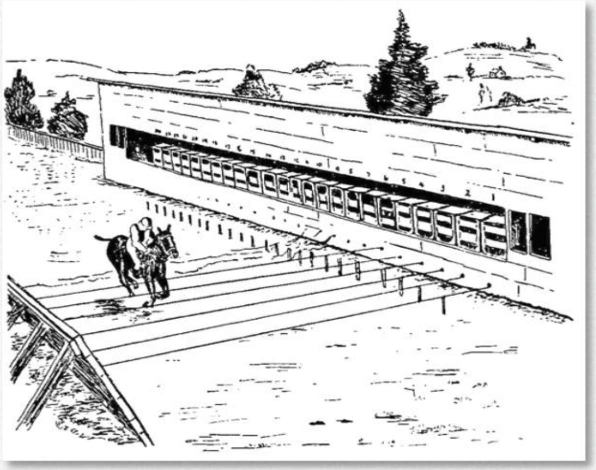
Background

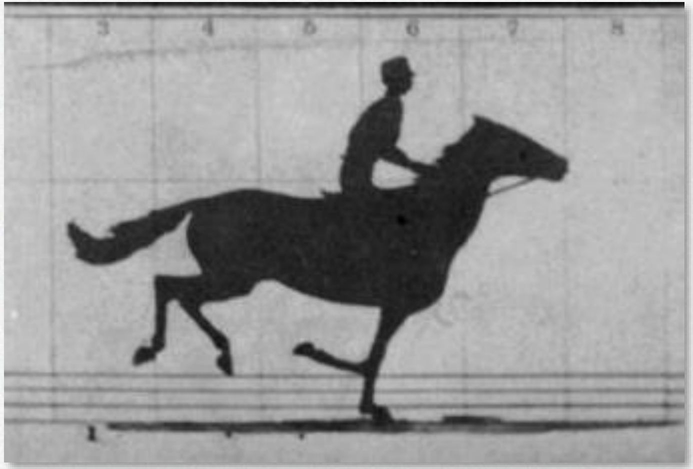
- The development of natural science has been heavily relying on **visual examination**.
- To better analyze the phenomenon-of-interest, **images** are convenient tools.
- A historical example: The Horse in Motion (Leland Stanford, 1874)
  - Hypothesis of “unsupported transit”: there were indeed moments in a horse’s stride in which all hooves were off the ground and the animal enjoyed “unsupported transit.”

EADWEARD MUYBRIDGE COLLECTION/Getty Images

Background

- The development of natural science has been heavily relying on **visual examination**.
- To better analyze the phenomenon-of-interest, **images** are convenient tools.
- A historical example: The Horse in Motion (Leland Stanford, 1874)
    - Hypothesis of "unsupported transit": there were indeed moments in a horse's stride in which all hooves were off the ground and the animal enjoyed "unsupported transit."





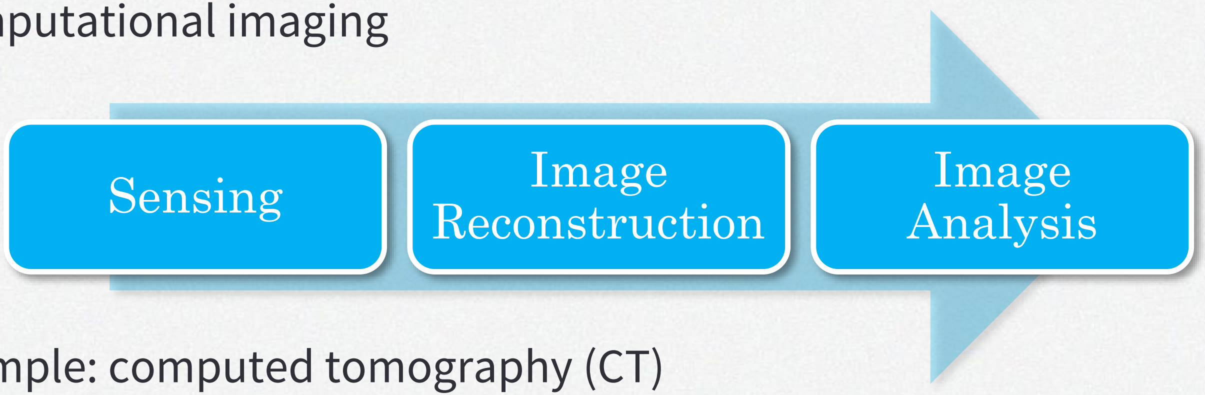
Eadweard Muybridge's imaging system, Palo Alto, 1878

Background

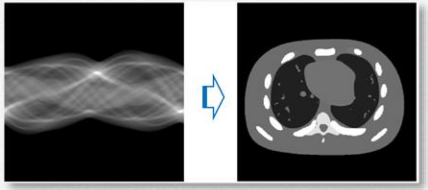
- Computational imaging

Background

- Computational imaging



- Example: computed tomography (CT)



02

Image Reconstruction

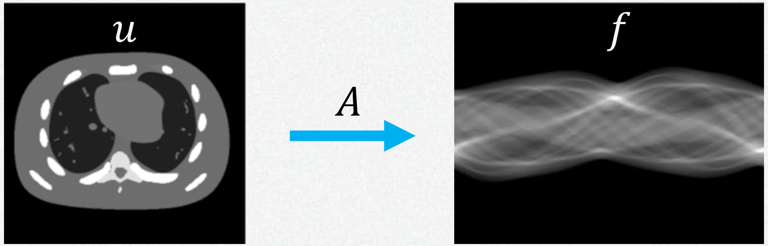
A review of PDE-based and wavelet frame-based approaches.

The Image Reconstruction Model

- The image reconstruction model as a linear inverse problem

$$f = Au + \eta$$

  - f is the observed image or measurement data;
  - A describes the image sensing process;
  - η is additive noise.



- The greatest challenge: **ill-posedness**
- An universal solution: **regularization/prior knowledge**

The PDE-Based Approach

- Designing PDEs that regularizes images

$$u_t = F\left(A, f, u, \nabla u, \nabla^2 u\right), \qquad u(0, x) = u_0(x).$$

  - Examples:
    - Shock-filters (Rudin-Osher 1990),
    - Perona-Malik equation (Perona-Malik 1990),
    - anisotropic diffusions (Weickert 1994) ,
    - fluid dynamics model (Bertalmio-Bertozzi-Sapiro 2001).

The PDE-Based Approach

- Designing PDEs that regularizes images

$$u_t = F\left(A, f, u, \nabla u, \nabla^2 u\right), \qquad u(0,x) = u_0(x).$$

  - Examples:
    - Shock-filters (Rudin-Osher 1990),
    - Perona-Malik equation (Perona-Malik 1990),
    - anisotropic diffusions (Weickert 1994) ,
    - fluid dynamics model (Bertalmio-Bertozzi-Sapiro 2001).
- Designing regularization in variational models

$$\inf_{u \in \mathcal{U}} L(A, f, u) + R(u)$$

  - Examples:
    - The total variation (TV) model (Rudin-Osher-Fatemi 1992): $R(u) = \|\nabla u\|_{L_1}$,
    - The Inf-convolution model (Chambolle-Lions 1997),
    - The total generalized variation (TGV) model (Bredies-Kunisch-Pock 2010).

The PDE-Based Approach

- Designing PDEs that regularizes images

$$u_t = F\left(A, f, u, \nabla u, \nabla^2 u\right), \qquad u(0,x) = u_0(x).$$

  - Examples:
    - Shock-filters (Rudin-Osher 1990),
    - Perona-Malik equation (Perona-Malik 1990),
    - anisotropic diffusions (Weickert 1994) ,
    - fluid dynamics model (Bertalmio-Bertozzi-Sapiro 2001).
- Designing regularization in variational models

$$\inf_{u \in \mathcal{U}} L(A, f, u) + R(u)$$

  - Examples:
    - The total variation (TV) model (Rudin-Osher-Fatemi 1992): $R(u) = \|\nabla u\|_{L_1}$,
    - The Inf-convolution model (Chambolle-Lions 1997),
    - The total generalized variation (TGV) model (Bredies-Kunisch-Pock 2010).
- Key to the PDE-based approach: image **geometry**, **edge** preservation.

The Wavelet Frame-Based Approach

- Denote $\boldsymbol{W}$: $\mathbb{R}^m \to \mathbb{R}^{m'}$ the wavelet frame transform.
- We have the following three typical wavelet frame-based models
  - The balanced model (Chan-Chan-Shen-Shen 2003, Cai-Chan-Shen 2008)

$$\min_{\boldsymbol{d}} L(\boldsymbol{A}, \boldsymbol{f}, \boldsymbol{u}) + \kappa \left\| \boldsymbol{I} - \boldsymbol{W}\boldsymbol{W}^\top \boldsymbol{d} \right\|_2^2 + \|\boldsymbol{\lambda} \cdot \boldsymbol{d}\|_1,$$

  - The analysis model (Starck-Elad-Donoho 2005, Cai-Osher-Shen 2009)

$$\min_{\boldsymbol{u}} L(\boldsymbol{A}, \boldsymbol{f}, \boldsymbol{u}) + \|\boldsymbol{\lambda} \cdot \boldsymbol{W}\boldsymbol{u}\|_1,$$

  - The synthesis model (Figueiredo-Nowak 2003, Daubechies-Teschke-Vese, 2007)

$$\min_{\boldsymbol{d}} L(\boldsymbol{A}, \boldsymbol{f}, \boldsymbol{W}^\top \boldsymbol{d}) + \|\boldsymbol{\lambda} \cdot \boldsymbol{d}\|_1.$$

The Wavelet Frame-Based Approach

- Denote $\boldsymbol{W}$: $\mathbb{R}^m \to \mathbb{R}^{m'}$ the wavelet frame transform.
- We have the following three typical wavelet frame-based models
  - The balanced model (Chan-Chan-Shen-Shen 2003, Cai-Chan-Shen 2008)

$$\min_{\boldsymbol{d}} L(\boldsymbol{A}, \boldsymbol{f}, \boldsymbol{u}) + \kappa \left\| \boldsymbol{I} - \boldsymbol{W}\boldsymbol{W}^\top \boldsymbol{d} \right\|_2^2 + \|\boldsymbol{\lambda} \cdot \boldsymbol{d}\|_1,$$

  - The analysis model (Starck-Elad-Donoho 2005, Cai-Osher-Shen 2009)

$$\min_{\boldsymbol{u}} L(\boldsymbol{A}, \boldsymbol{f}, \boldsymbol{u}) + \|\boldsymbol{\lambda} \cdot \boldsymbol{W}\boldsymbol{u}\|_1,$$

  - The synthesis model (Figueiredo-Nowak 2003, Daubechies-Teschke-Vese, 2007)

$$\min_{\boldsymbol{d}} L(\boldsymbol{A}, \boldsymbol{f}, \boldsymbol{W}^\top \boldsymbol{d}) + \|\boldsymbol{\lambda} \cdot \boldsymbol{d}\|_1.$$

- Optimization algorithms induce various wavelet shrinkage algorithms.

The Wavelet Frame-Based Approach

- Denote $\boldsymbol{W}$: $\mathbb{R}^m \to \mathbb{R}^{m'}$ the wavelet frame transform.
- We have the following three typical wavelet frame-based models
  - The balanced model (Chan-Chan-Shen-Shen 2003, Cai-Chan-Shen 2008)

$$\min_{\boldsymbol{d}} L(\boldsymbol{A}, \boldsymbol{f}, \boldsymbol{u}) + \kappa \left\| \boldsymbol{I} - \boldsymbol{W}\boldsymbol{W}^\top \boldsymbol{d} \right\|_2^2 + \|\boldsymbol{\lambda} \cdot \boldsymbol{d}\|_1,$$

  - The analysis model (Starck-Elad-Donoho 2005, Cai-Osher-Shen 2009)

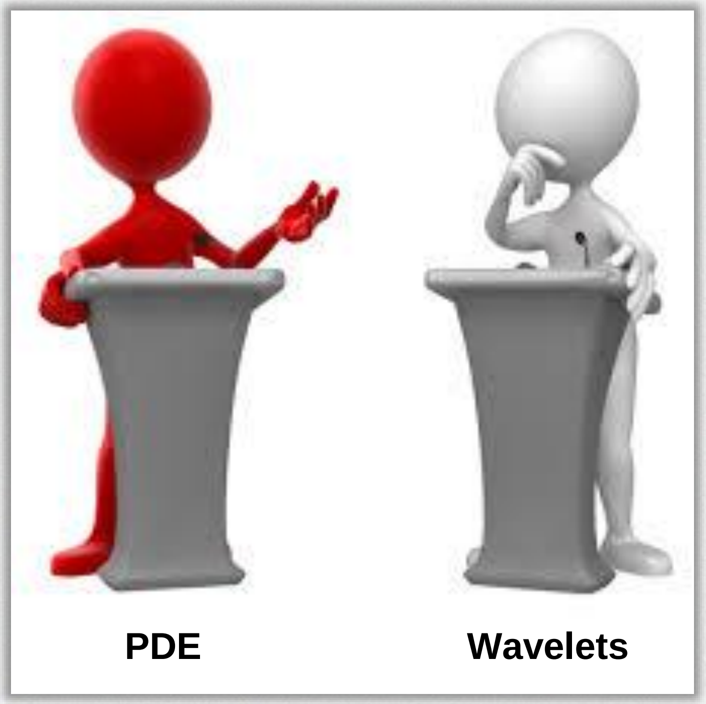
$$\min_{\boldsymbol{u}} L(\boldsymbol{A}, \boldsymbol{f}, \boldsymbol{u}) + \|\boldsymbol{\lambda} \cdot \boldsymbol{W}\boldsymbol{u}\|_1,$$

  - The synthesis model (Figueiredo-Nowak 2003, Daubechies-Teschke-Vese, 2007)

$$\min_{\boldsymbol{d}} L(\boldsymbol{A}, \boldsymbol{f}, \boldsymbol{W}^\top \boldsymbol{d}) + \|\boldsymbol{\lambda} \cdot \boldsymbol{d}\|_1.$$

- Optimization algorithms induce various wavelet shrinkage algorithms.
- Key to wavelet frame-based approach: **multiscale**, **sparse** approximation.

Relations between the Two Approaches?



Relations between the Two Approaches?

- Some early studies showed relations between
  - discrete 1D nonlinear diffusions and shift-invariant Haar shrinkage (Mrázek-Weickert-Steidl 2003)
  - discrete 1D nonlinear diffusions and wavelet frame shrinkage (Jiang 2012)

Relations between the Two Approaches?

- Some early studies showed relations between
  - discrete 1D nonlinear diffusions and shift-invariant Haar shrinkage (Mrázek-Weickert-Steidl 2003)
  - discrete 1D nonlinear diffusions and wavelet frame shrinkage (Jiang 2012)
- For most of the time, the two approaches compete with each other.
- What were still unclear:
  - Variational models (e.g., TV, TGV) v.s. wavelet-based optimization models?
  - More types of differential equations (e.g., fluid dynamics models)?
  - Generic wavelet frame shrinkage algorithms (e.g., Nesterov)?
  - New insights on both approaches?

03

Connections

between PDE-based and wavelet frame-based approach, and the implications of such connections.

Joint work with
Jian-Feng Cai (HKUST)
Qingtang Jiang (UMSL)
Stanley Osher (UCLA)
Zuowei Shen (NUS)

Wavelet Frame Transforms and Differential Operators

- A key observation is the link between vanishing moments of wavelet functions and the orders of differential operators.

Wavelet Frame Transforms and Differential Operators

- A key observation is the link between vanishing moments of wavelet functions and the orders of differential operators.
- Vanishing moments in the continuum and discrete setting.
  - We say ψ or $\boldsymbol{q}$ has vanishing moment of order $\boldsymbol{\alpha}$ if it annihilates polynomials up to degree $\boldsymbol{\alpha}$ as follows

$$\int_{\Omega} \boldsymbol{x}^{\boldsymbol{\beta}} \psi(\boldsymbol{x}) d\boldsymbol{x} = 0, \quad \forall \boldsymbol{\beta} \in \mathbb{Z}_+^2, \ \boldsymbol{\beta} < |\boldsymbol{\alpha}| \text{ or } |\boldsymbol{\beta}| = |\boldsymbol{\alpha}|, \boldsymbol{\beta} \neq \boldsymbol{\alpha}.$$

$$\sum_{\boldsymbol{k} \in \mathbb{Z}^2} \boldsymbol{k}^{\boldsymbol{\beta}} \boldsymbol{q}[\boldsymbol{k}] = 0, \quad \forall \boldsymbol{\beta} \in \mathbb{Z}_+^2, \ \boldsymbol{\beta} < |\boldsymbol{\alpha}| \text{ or } |\boldsymbol{\beta}| = |\boldsymbol{\alpha}|, \boldsymbol{\beta} \neq \boldsymbol{\alpha}.$$

Wavelet Frame Transforms and Differential Operators

- A key observation is the link between vanishing moments of wavelet functions and the orders of differential operators.
- A simple example of Haar wavelets:
  - Filters: $\boldsymbol{q}_{0,1} = \frac{1}{4}\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \boldsymbol{q}_{1,0} = \frac{1}{4}\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \boldsymbol{q}_{1,1} = \frac{1}{4}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$

Wavelet Frame Transforms and Differential Operators

- A key observation is the link between vanishing moments of wavelet functions and the orders of differential operators.
- A simple example of Haar wavelets:
  - Filters: $\boldsymbol{q}_{0,1} = \frac{1}{4}\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \boldsymbol{q}_{1,0} = \frac{1}{4}\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \boldsymbol{q}_{1,1} = \frac{1}{4}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$.
  - Haar framelet transform: $\boldsymbol{W}\boldsymbol{u} = \{\boldsymbol{q}_{i,j}[-\cdot] \circledast \boldsymbol{u}\}$. We also have

$$\boldsymbol{q}_{0,1}[-\cdot] \circledast \boldsymbol{u} \approx \frac{\delta}{2} u_x, \qquad \boldsymbol{q}_{1,0}[-\cdot] \circledast \boldsymbol{u} \approx \frac{\delta}{2} u_y, \qquad \boldsymbol{q}_{1,1}[-\cdot] \circledast \boldsymbol{u} \approx \frac{\delta^2}{4} u_{xy}.$$

Wavelet Frame Transforms and Differential Operators

- A key observation is the link between vanishing moments of wavelet functions and the orders of differential operators.
- A simple example of Haar wavelets:
  - Filters: $\boldsymbol{q}_{0,1} = \frac{1}{4}\begin{pmatrix}1 & -1\\ 1 & -1\end{pmatrix}, \boldsymbol{q}_{1,0} = \frac{1}{4}\begin{pmatrix}1 & 1\\ -1 & -1\end{pmatrix}, \boldsymbol{q}_{1,1} = \frac{1}{4}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}$.
  - Haar framelet transform: $\boldsymbol{W}\boldsymbol{u} = \{\boldsymbol{q}_{i,j}[-\cdot] \circledast \boldsymbol{u}\}$. We also have

$$\boldsymbol{q}_{0,1}[-\cdot] \circledast \boldsymbol{u} \approx \frac{\delta}{2} u_x, \qquad \boldsymbol{q}_{1,0}[-\cdot] \circledast \boldsymbol{u} \approx \frac{\delta}{2} u_y, \qquad \boldsymbol{q}_{1,1}[-\cdot] \circledast \boldsymbol{u} \approx \frac{\delta^2}{4} u_{xy}.$$

  - Thus, we have

$$|\nabla u| \approx \frac{2}{\delta}\boldsymbol{W}\boldsymbol{u} = \frac{1}{2}\Big(\Big[\big(D_x^+\boldsymbol{u}_{i,j}\big)^2 + \big(D_x^+\boldsymbol{u}_{i,j+1}\big)^2 + \big(D_y^+\boldsymbol{u}_{i,j}\big)^2 + \big(D_y^+\boldsymbol{u}_{i+1,j}\big)^2\Big]$$
$$+ \Big[\big(D_x^+\boldsymbol{u}_{i,j} + D_y^-\boldsymbol{u}_{i,j+1}\big)^2 + \Big(\big(D_x^+\boldsymbol{u}_{i,j} + D_y^+\boldsymbol{u}_{i+1,j}\big)^2\Big)\Big]\Big)^{1/2}$$

Wavelet Frame Transforms and Differential Operators

- A key observation is the link between vanishing moments of wavelet functions and the orders of differential operators.
- A simple example of Haar wavelets:
  - Filters: $\boldsymbol{q}_{0,1} = \frac{1}{4}\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \boldsymbol{q}_{1,0} = \frac{1}{4}\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \boldsymbol{q}_{1,1} = \frac{1}{4}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$.
  - Haar framelet transform: $\boldsymbol{W}\boldsymbol{u} = \{\boldsymbol{q}_{i,j}[-\cdot] \circledast \boldsymbol{u}\}$. We also have

$$\boldsymbol{q}_{0,1}[-\cdot] \circledast \boldsymbol{u} \approx \frac{\delta}{2} u_x, \qquad \boldsymbol{q}_{1,0}[-\cdot] \circledast \boldsymbol{u} \approx \frac{\delta}{2} u_y, \qquad \boldsymbol{q}_{1,1}[-\cdot] \circledast \boldsymbol{u} \approx \frac{\delta^2}{4} u_{xy}.$$

  - Thus, we have

$$|\nabla u| \approx \frac{2}{\delta}\boldsymbol{W}\boldsymbol{u} = \frac{1}{2}\Bigg(\left[\left(D_x^+ \boldsymbol{u}_{i,j}\right)^2 + \left(D_x^+ \boldsymbol{u}_{i,j+1}\right)^2 + \left(D_y^+ \boldsymbol{u}_{i,j}\right)^2 + \left(D_y^+ \boldsymbol{u}_{i+1,j}\right)^2\right]$$

$$+ \left[\left(D_x^+ \boldsymbol{u}_{i,j} + D_y^- \boldsymbol{u}_{i,j+1}\right)^2 + \left(\left(D_x^+ \boldsymbol{u}_{i,j} + D_y^+ \boldsymbol{u}_{i+1,j}\right)^2\right)\right]\Bigg)^{1/2}$$

Standard discretization

Specific to Haar, granting 45-degree rotation invariance

Wavelet Frame Transforms and Differential Operators

- A key observation is the link between vanishing moments of wavelet functions and the orders of differential operators.
- In the continuum setting (Cai-Dong-Osher-Shen 2012, Choi-Dong-Zhang 2020)

Proposition 1. Let a tensor product wavelet frame function $\psi_{\boldsymbol{\alpha}} \in L_2(\mathbb{R}^2)$ have vanishing moments of order $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| \leq s$, and let $\operatorname{supp}(\psi_{\boldsymbol{\alpha}}) = [a_1, a_2] \times [b_1, b_2]$. Then, there exists a unique $\varphi_{\boldsymbol{\alpha}} \in L_2(\mathbb{R}^2)$ such that $\varphi_{\boldsymbol{\alpha}}$ is differentiable up to order $\boldsymbol{\alpha}$ a.e.,

$$c_{\boldsymbol{\alpha}} = \int_{\mathbb{R}^2} \varphi_{\boldsymbol{\alpha}} \neq 0 \quad \text{and} \quad \psi_{\boldsymbol{\alpha}} = \partial^{\boldsymbol{\alpha}} \varphi_{\boldsymbol{\alpha}}.$$

Furthermore, for $n \in \mathbb{N}$ and $\boldsymbol{k} \in \mathbb{Z}^2$ with $\operatorname{supp}(\psi_{\boldsymbol{\alpha},n-1,\boldsymbol{k}}) \subseteq \overline{\Omega}$, we have

$$\left\langle u, \psi_{\boldsymbol{\alpha},n-1,\boldsymbol{k}} \right\rangle = (-1)^{|\boldsymbol{\alpha}|} 2^{|\boldsymbol{\alpha}|(1-n)} \left\langle \partial^{\boldsymbol{\alpha}} u, \varphi_{\boldsymbol{\alpha},n-1,\boldsymbol{k}} \right\rangle$$

for every u belonging to the Sobolev space $W_1^s(\Omega)$. Here, $\psi_{\boldsymbol{\alpha},n-1,\boldsymbol{k}} = 2^{n-2} \psi_{\boldsymbol{\alpha}}(2^{n-1} \cdot -2^{-1}\boldsymbol{k})$ and $\varphi_{\boldsymbol{\alpha},n-1,\boldsymbol{k}}$ is defined similarly.

Wavelet Frame Transforms and Differential Operators

- A key observation is the link between vanishing moments of wavelet functions and the orders of differential operators.
- In the continuum setting (Cai-Dong-Osher-Shen 2012, Choi-Dong-Zhang 2020)

Proposition 1. Let a tensor product wavelet frame function $\psi_\alpha \in L_2(\mathbb{R}^2)$ have vanishing moments of order $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| \leq s$, and let $\mathrm{supp}(\psi_\alpha) = [a_1, a_2] \times [b_1, b_2]$. Then, there exists a unique $\varphi_\alpha \in L_2(\mathbb{R}^2)$ such that φ_α is differentiable up to order $\boldsymbol{\alpha}$ a.e.,

$$c_\alpha = \int_{\mathbb{R}^2} \varphi_\alpha \neq 0 \quad \text{and} \quad \psi_\alpha = \partial^\alpha \varphi_\alpha .$$

Furthermore, for $n \in \mathbb{N}$ and $\boldsymbol{k} \in \mathbb{Z}^2$ with $\mathrm{supp}(\psi_{\alpha,n-1,\boldsymbol{k}}) \subseteq \overline{\Omega}$, we have

$$\left\langle u, \psi_{\alpha,n-1,\boldsymbol{k}} \right\rangle = (-1)^{|\alpha|} 2^{|\alpha|(1-n)} \left\langle \partial^\alpha u, \varphi_{\alpha,n-1,\boldsymbol{k}} \right\rangle$$

for every u belonging to the Sobolev space $W_1^s(\Omega)$. Here, $\psi_{\alpha,n-1,\boldsymbol{k}} = 2^{n-2}\psi_\alpha(2^{n-1} \cdot -2^{-1}\boldsymbol{k})$ and $\varphi_{\alpha,n-1,\boldsymbol{k}}$ is defined similarly.

Wavelet Frame Transforms and Differential Operators

- A key observation is the link between vanishing moments of wavelet functions and the orders of differential operators.
- In the discrete setting (Dong-Jiang-Shen 2017)

Proposition 2. Let $\boldsymbol{q}$ be a high-pass filter with vanishing moments of order $\boldsymbol{\alpha} \in \mathbb{Z}_+^2$. Then for a smooth function $F(\boldsymbol{x})$ on $\mathbb{R}^2$, we have

$$\frac{1}{\varepsilon^{|\alpha|}} \sum_{\boldsymbol{k} \in \mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}] F(\boldsymbol{x} + \varepsilon \boldsymbol{k}) = C_\alpha \frac{\partial^\alpha}{\partial \boldsymbol{x}^\alpha} F(\boldsymbol{x}) + O(\varepsilon),$$

where C_α is the constant defined by

$$C_\alpha = \frac{1}{\boldsymbol{\alpha}!} \sum_{\boldsymbol{k} \in \mathbb{Z}^2} \boldsymbol{k}^\alpha \boldsymbol{q}[\boldsymbol{k}] = \frac{i^{|\alpha|}}{\boldsymbol{\alpha}!} \frac{\partial^\alpha}{\partial \omega^\alpha} \widehat{\boldsymbol{q}}(\omega)\Big|_{\omega = \boldsymbol{0}}.$$

If, in addition, $\boldsymbol{q}$ has total vanishing moments of order $K \backslash \{|\alpha| + 1\}$ for some $K > |\alpha|$, then

$$\frac{1}{\varepsilon^{|\alpha|}} \sum_{\boldsymbol{k} \in \mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}] F(\boldsymbol{x} + \varepsilon \boldsymbol{k}) = C_\alpha \frac{\partial^\alpha}{\partial \boldsymbol{x}^\alpha} F(\boldsymbol{x}) + O(\varepsilon^{K - |\alpha|}).$$

Wavelet Frame Transforms and Differential Operators

- A key observation is the link between vanishing moments of wavelet functions and the orders of differential operators.
- In the discrete setting (Dong-Jiang-Shen 2017)

Proposition 2. Let $\boldsymbol{q}$ be a high-pass filter with vanishing moments of order $\boldsymbol{\alpha} \in \mathbb{Z}_+^2$. Then for a smooth function $F(\boldsymbol{x})$ on $\mathbb{R}^2$, we have

$$\frac{1}{\varepsilon^{|\alpha|}} \sum_{\boldsymbol{k} \in \mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}] F(\boldsymbol{x} + \varepsilon \boldsymbol{k}) = C_\alpha \frac{\partial^\alpha}{\partial \boldsymbol{x}^\alpha} F(\boldsymbol{x}) + O(\varepsilon),$$

where C_α is the constant defined by

$$C_\alpha = \frac{1}{\boldsymbol{\alpha}!} \sum_{\boldsymbol{k} \in \mathbb{Z}^2} \boldsymbol{k}^\alpha \boldsymbol{q}[\boldsymbol{k}] = \frac{i^{|\alpha|}}{\boldsymbol{\alpha}!} \frac{\partial^\alpha}{\partial \omega^\alpha} \widehat{\boldsymbol{q}}(\omega)\Big|_{\omega=\boldsymbol{0}}.$$

If, in addition, $\boldsymbol{q}$ has total vanishing moments of order $K \backslash \{|\boldsymbol{\alpha}| + 1\}$ for some $K > |\boldsymbol{\alpha}|$, then

$$\frac{1}{\varepsilon^{|\alpha|}} \sum_{\boldsymbol{k} \in \mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}] F(\boldsymbol{x} + \varepsilon \boldsymbol{k}) = C_\alpha \frac{\partial^\alpha}{\partial \boldsymbol{x}^\alpha} F(\boldsymbol{x}) + O(\varepsilon^{K-|\alpha|}).$$

Connections: Variational and Wavelet Models

- Consider the following two problems

$$\begin{cases} \inf_{u \in W_1^s(\Omega)} E_n(u) := \nu \|\boldsymbol{\lambda}_n \cdot \boldsymbol{W}\boldsymbol{T}_n u\|_1 + \frac{1}{2}\|\boldsymbol{A}_n \boldsymbol{T}_n u - \boldsymbol{T}_n f\|_2^2 & \text{Analysis model} \\ \inf_{u \in W_1^s(\Omega)} E(u) = \nu \|\boldsymbol{D}u\|_{L_1(\Omega)} + \frac{1}{2}\|Au - f\|_{L_2(\Omega)}^2 & \text{Variational model} \end{cases}$$

Connections: Variational and Wavelet Models

- Consider the following two problems

$$\begin{cases} \inf_{u \in W_1^s(\Omega)} E_n(u) := \nu \|\boldsymbol{\lambda}_n \cdot \boldsymbol{W}\boldsymbol{T}_n u\|_1 + \frac{1}{2}\|\boldsymbol{A}_n \boldsymbol{T}_n u - \boldsymbol{T}_n f\|_2^2 & \text{Analysis model} \\ \inf_{u \in W_1^s(\Omega)} E(u) = \nu \|\boldsymbol{D}u\|_{L_1(\Omega)} + \frac{1}{2}\|Au - f\|_{L_2(\Omega)}^2 & \text{Variational model} \end{cases}$$

- Then, we have (Cai-Dong-Osher-Shen 2012)

Theorem 1. Given a energy functional E as above, there exists a set of coefficients $\boldsymbol{\lambda}_n$, such that the functional E_n Γ-converges to E in $W_1^s(\Omega)$. Furthermore, let $u_n^\star$ be an ε-optimal solution to $\inf_u E_n(u)$, i.e. $E_n(u_n^\star) \le \inf_u E_n(u) + \epsilon$. We have that

$$\limsup_{n\to\infty} E_n(u_n^\star) \le \inf_u E(u) + \epsilon,$$

and any cluster point of $\{u_n^\star\}_n$ is an ε-optimal solution to $\inf_u E(u)$.

Connections: Variational and Wavelet Models

- Consider the following two problems

$$\begin{cases} \inf_{u\in W_1^s(\Omega)} E_n(u) := \nu\|\boldsymbol{\lambda}_n \cdot \boldsymbol{W}\boldsymbol{T}_n u\|_1 + \frac{1}{2}\|\boldsymbol{A}_n\boldsymbol{T}_n u - \boldsymbol{T}_n f\|_2^2 & \text{Analysis model} \\ \inf_{u\in W_1^s(\Omega)} E(u) = \nu\|\boldsymbol{D}u\|_{L_1(\Omega)} + \frac{1}{2}\|Au - f\|_{L_2(\Omega)}^2 & \text{Variational model} \end{cases}$$

- Then, we have (Cai-Dong-Osher-Shen 2012)

Theorem 1. Given a energy functional E as above, there exists a set of coefficients $\boldsymbol{\lambda}_n$, such that the functional E_n Γ-converges to E in $W_1^s(\Omega)$. Furthermore, let $u_n^\star$ be an ε-optimal solution to $\inf_u E_n(u)$, i.e. $E_n(u_n^\star) \le \inf_u E_n(u) + \epsilon$. We have that

$$\limsup_{n\to\infty} E_n(u_n^\star) \le \inf_u E(u) + \epsilon,$$

and any cluster point of $\{u_n^\star\}_n$ is an ε-optimal solution to $\inf_u E(u)$.

- Further extensions: Cai-Dong-Shen 2016, Dong-Shen-Xie 2017, Choi-Dong-Zhang 2020.

Connections: PDEs and Wavelet Shrinkage

- Consider the following two dynamics

$$\boldsymbol{u}^k = \widetilde{\boldsymbol{W}}^\top \boldsymbol{S}_{\boldsymbol{\lambda}^{k-1}}\left(\boldsymbol{W}\boldsymbol{u}^{k-1}\right), \qquad k = 1,2,\cdots.$$

Wavelet shrinkage

$$u_t = \sum_{\ell=1}^{L} \frac{\partial^{\alpha_\ell}}{\partial x^{\alpha_\ell}} \Phi_\ell(\boldsymbol{D}u, u), \quad \boldsymbol{D} = \left(\cdots \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial x^{\boldsymbol{\beta}_\ell}} \cdots\right), t \in (0, T].$$

Nonlinear evolution PDE

- It can be shown that (Dong-Jiang-Shen 2017) :

Theorem 2. Given a PDE as above, we can construct a pair of dual wavelet frames transforms $\boldsymbol{W}$ and $\widetilde{\boldsymbol{W}}$, and a shrinkage function $\boldsymbol{S}_\lambda$ such that the wavelet shrinkage algorithm is consistent with the PDE. For a nonlinear diffusion, the discrete solution generated by the wavelet frame shrinkage converges to that of the PDE.

Connections: PDEs and Wavelet Shrinkage

- Consider the following two dynamics

$$\boldsymbol{u}^k = \widetilde{\boldsymbol{W}}^\top \boldsymbol{S}_{\boldsymbol{\lambda}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}), \quad k = 1,2,\cdots.$$ Wavelet shrinkage

$$u_t = \sum_{\ell=1}^{L} \frac{\partial^{\alpha_\ell}}{\partial x^{\alpha_\ell}} \Phi_\ell(\boldsymbol{D}u, u), \quad \boldsymbol{D} = \left(\cdots \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial x^{\boldsymbol{\beta}_\ell}} \cdots\right), t \in (0, T].$$ Nonlinear evolution PDE

- It can be shown that (Dong-Jiang-Shen 2017) :

Theorem 2. Given a PDE as above, we can construct a pair of dual wavelet frames transforms $\boldsymbol{W}$ and $\widetilde{\boldsymbol{W}}$, and a shrinkage function $\boldsymbol{S}_\lambda$ such that the wavelet shrinkage algorithm is consistent with the PDE. For a nonlinear diffusion, the discrete solution generated by the wavelet frame shrinkage converges to that of the PDE.

- This has led to new wavelet shrinkage and PDE models.
  - For example, the Nesterov accelerated wavelet frame shrinkage (Li-Fan-Ji-Shen 2014) :

$$\boldsymbol{u}^k = (I - \mu \boldsymbol{A}^\top \boldsymbol{A}) \boldsymbol{W}^\top \boldsymbol{S}_{\boldsymbol{\lambda}^{k-1}}\left[(1+\gamma^{k-1})\boldsymbol{W}\boldsymbol{u}^{k-1} - \gamma^{k-1}\boldsymbol{W}\boldsymbol{u}^{k-2}\right] + \mu \boldsymbol{A}^\top \boldsymbol{f}, \quad k = 1,2,\dots.$$

Connections: PDEs and Wavelet Shrinkage

- Consider the following two dynamics

$$\boldsymbol{u}^k = \widetilde{\boldsymbol{W}}^\top \boldsymbol{S}_{\boldsymbol{\lambda}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}), \qquad k = 1,2,\cdots.$$

Wavelet shrinkage

$$u_t = \sum_{\ell=1}^{L} \frac{\partial^{\alpha_\ell}}{\partial x^{\alpha_\ell}} \Phi_\ell(\boldsymbol{D}u, u), \quad \boldsymbol{D} = \left(\cdots \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial x^{\boldsymbol{\beta}_\ell}} \cdots\right), t \in (0, T].$$

Nonlinear evolution PDE

- It can be shown that (Dong-Jiang-Shen 2017) :

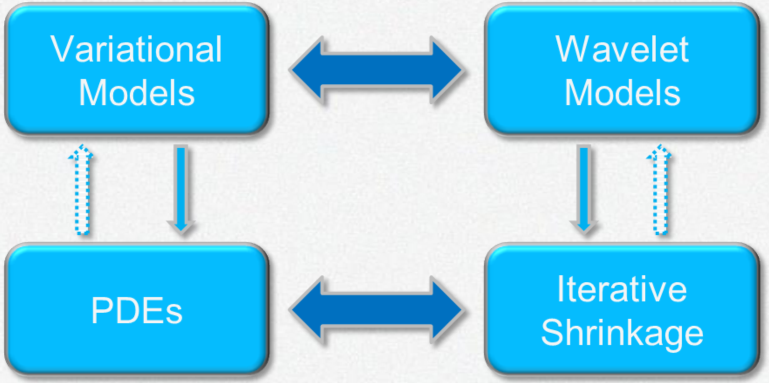
Theorem 2. Given a PDE as above, we can construct a pair of dual wavelet frames transforms $\boldsymbol{W}$ and $\widetilde{\boldsymbol{W}}$, and a shrinkage function $\boldsymbol{S}_\lambda$ such that the wavelet shrinkage algorithm is consistent with the PDE. For a nonlinear diffusion, the discrete solution generated by the wavelet frame shrinkage converges to that of the PDE.

- This has led to new wavelet shrinkage and PDE models.
  - For example, the Nesterov accelerated wavelet frame shrinkage (Li-Fan-Ji-Shen 2014) leads to the following PDE:

$$u_{tt} + Cu_t = \operatorname{div}(\boldsymbol{\Phi}(\boldsymbol{D}u, u)) - \kappa A^\top(Au - f).$$

  - Related works: Su-Boyd-Candes 2014, Wibisono-Wilson-Jordan 2016.

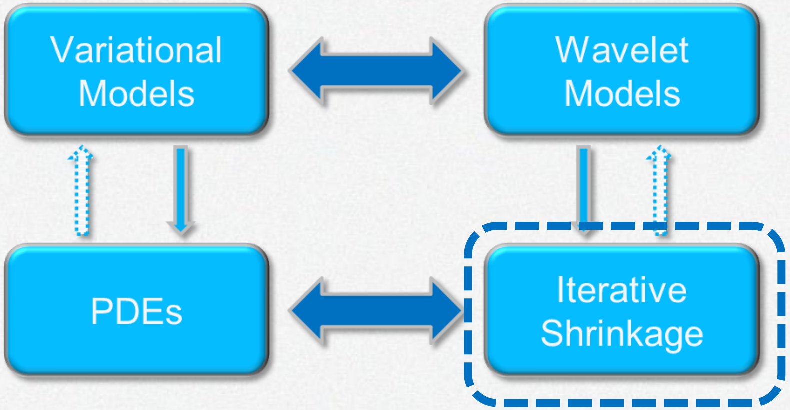
An Integrated Knowledge on Image Reconstruction Methods



What we know now:

- Wavelet methods have geometric meanings.
- PDE methods can be understood through the lens of sparsity.
- Giving birth to new models.

An Integrated Knowledge on Image Reconstruction Methods



What we know now:

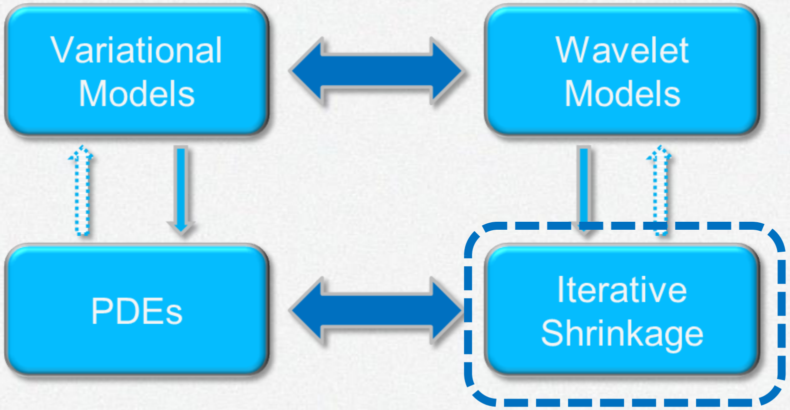
- Wavelet methods have geometric meanings.
- PDE methods can be understood through the lens of sparsity.
- Giving birth to new models.

- **Algorithmic level: solution mapping**

$$\mathcal{F}: X \to Y;\ \mathcal{F}(f) = \hat{u} \approx u$$

- Mappings $\mathcal{F}$ are often dynamic systems and mostly discrete approximations of differential equations.

An Integrated Knowledge on Image Reconstruction Methods



What we know now:

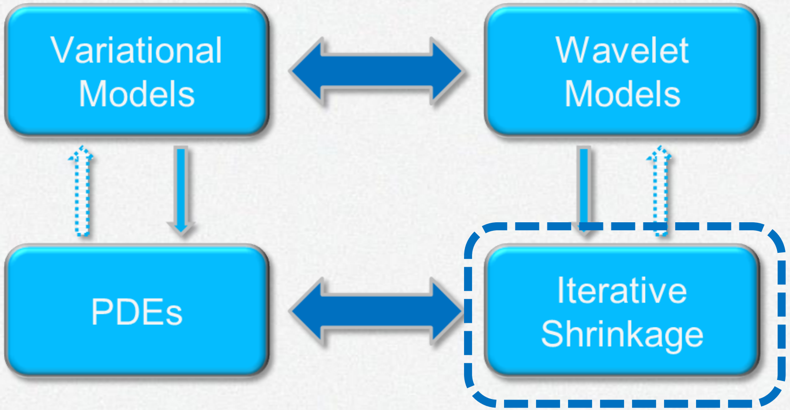
- Wavelet methods have geometric meanings.
- PDE methods can be understood through the lens of sparsity.
- Giving birth to new models.

- **Algorithmic level: solution mapping**

  $$\boldsymbol{\mathcal{F}}: X \to Y;\ \boldsymbol{\mathcal{F}}(f) = \hat{u} \approx u$$

- Mappings $\boldsymbol{\mathcal{F}}$ are often dynamic systems and mostly discrete approximations of differential equations.
- The mapping $\boldsymbol{\mathcal{F}}$ is "handcrafted"
  - Works generally well
  - Good stability and interpretability
- Still not good enough in practice:
  - Not everything can be "handcrafted"

An Integrated Knowledge on Image Reconstruction Methods



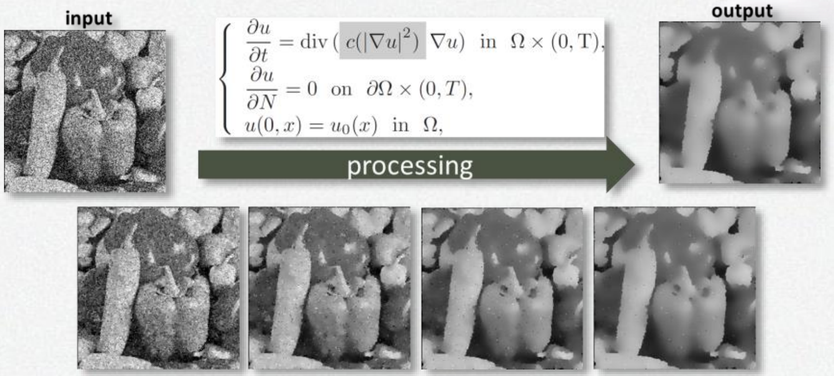
What we know now:

- Wavelet methods have geometric meanings.
- PDE methods can be understood through the lens of sparsity.
- Giving birth to new models.

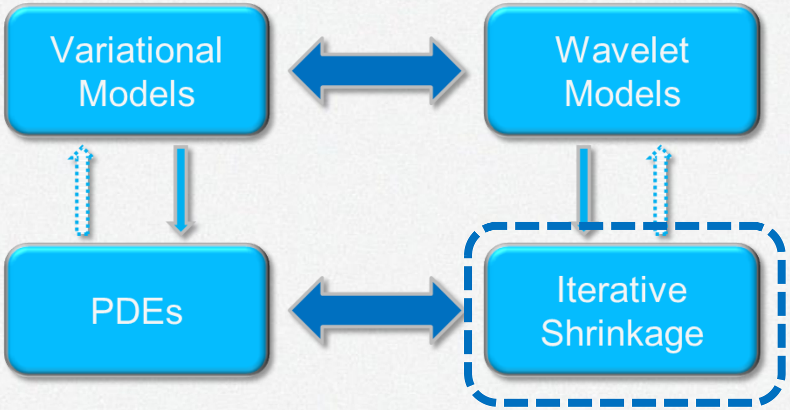
- **Algorithmic level:** **solution mapping**

$$\boldsymbol{\mathcal{F}}: X \to Y;\ \boldsymbol{\mathcal{F}}(f) = \hat{u} \approx u$$

- Mappings $\boldsymbol{\mathcal{F}}$ are often dynamic systems and mostly discrete approximations of differential equations.



An Integrated Knowledge on Image Reconstruction Methods



What we know now:

- Wavelet methods have geometric meanings.
- PDE methods can be understood through the lens of sparsity.
- Giving birth to new models.

- **Algorithmic level: solution mapping**
  $$\mathcal{F}: X \to Y;\ \mathcal{F}(f) = \hat{u} \approx u$$
- Mappings $\mathcal{F}$ are often dynamic systems and mostly discrete approximations of differential equations.
- The mapping $\mathcal{F}$ is "handcrafted"
  - Works generally well
  - Good stability and interpretability
- Still not good enough in practice:
  - Not everything can be "handcrafted"
- **Trending research direction:** Combination with deep learning

04

Deep Learning

Connections between CNNs and discrete differential equations (ODEs and PDEs).

Joint work with
Quanzheng Li (Harvard)
Zichao Long (Huawei)
Yiping Lu (Stanford)
Xianzhong Ma (Industry)
Aoxiao Zhong (Harvard)

Deep Learning

- Supervised Learning: given $(x_i, y_i) \sim \mathcal{P}$, find $\boldsymbol{\mathcal{F}}_{\widehat{\Theta}}: X \to Y$ through

$$\widehat{\Theta} = \operatorname{argmin}_{\Theta} \frac{1}{N} \sum_{i=1}^{N} L(\boldsymbol{\mathcal{F}}_{\Theta}(x_i), y_i)$$

  - with $\boldsymbol{\mathcal{F}}_{\Theta}$ a deep neural network, e.g., $\boldsymbol{\mathcal{F}}_{\Theta} = \boldsymbol{W}_3 \sigma(\boldsymbol{W}_2 \sigma(\boldsymbol{W}_1 x + \boldsymbol{b}_1) + \boldsymbol{b}_2)$

Deep Learning

- Supervised Learning: given $(x_i, y_i) \sim \mathcal{P}$, find $\boldsymbol{\mathcal{F}}_{\widehat{\Theta}}: X \to Y$ through

$$\widehat{\Theta} = \mathrm{argmin}_{\Theta} \frac{1}{N} \sum_{i=1}^{N} L(\boldsymbol{\mathcal{F}}_{\Theta}(x_i), y_i)$$

  - with $\boldsymbol{\mathcal{F}}_{\Theta}$ a deep neural network, e.g., $\boldsymbol{\mathcal{F}}_{\Theta} = \boldsymbol{W}_3 \sigma(\boldsymbol{W}_2 \sigma(\boldsymbol{W}_1 x + \boldsymbol{b}_1) + \boldsymbol{b}_2)$
- For image reconstruction, we can let $x = f, y = u$ and $\boldsymbol{\mathcal{F}}_{\widehat{\Theta}}(f) = \hat{u} \approx u$. This can be much better than handcrafted solution mapping.

Methods	BM3D	WNNM	EPLL	MLP	CSF	TNRD	DnCNN-S	DnCNN-B
$\sigma = 15$	31.07	31.37	31.21	-	31.24	31.42	**31.73**	31.61
$\sigma = 25$	28.57	28.83	28.68	28.96	28.74	28.92	**29.23**	29.16
$\sigma = 50$	25.62	25.87	25.67	26.03	-	25.97	**26.23**	**26.23**

Image denoising: average PSNR on BSD68 dataset (Zhang et al. 2017).

Deep Learning

- Supervised Learning: given $(x_i, y_i) \sim \mathcal{P}$, find $\boldsymbol{\mathcal{F}}_{\widehat{\Theta}}: X \to Y$ through

$$\widehat{\Theta} = \mathrm{argmin}_{\Theta} \frac{1}{N} \sum_{i=1}^{N} L(\boldsymbol{\mathcal{F}}_{\Theta}(x_i), y_i)$$

  - with $\boldsymbol{\mathcal{F}}_{\Theta}$ a deep neural network, e.g., $\boldsymbol{\mathcal{F}}_{\Theta} = \boldsymbol{W}_3 \sigma(\boldsymbol{W}_2 \sigma(\boldsymbol{W}_1 x + \boldsymbol{b}_1) + \boldsymbol{b}_2)$
- For image reconstruction, we can let $x = f, y = u$ and $\boldsymbol{\mathcal{F}}_{\widehat{\Theta}}(f) = \hat{u} \approx u$. This can be much better than handcrafted solution mapping.
- The impact of deep learning is much beyond image reconstruction.

Deep Learning

- Supervised Learning: given $(x_i, y_i) \sim \mathcal{P}$, find $\boldsymbol{\mathcal{F}}_{\widehat{\Theta}}: X \to Y$ through

$$\widehat{\Theta} = \mathrm{argmin}_{\Theta} \frac{1}{N} \sum_{i=1}^{N} L(\boldsymbol{\mathcal{F}}_{\Theta}(x_i), y_i)$$

  - with $\boldsymbol{\mathcal{F}}_{\Theta}$ a deep neural network, e.g., $\boldsymbol{\mathcal{F}}_{\Theta} = \boldsymbol{W}_3 \sigma(\boldsymbol{W}_2 \sigma(\boldsymbol{W}_1 x + \boldsymbol{b}_1) + \boldsymbol{b}_2)$
- For image reconstruction, we can let $x = f, y = u$ and $\boldsymbol{\mathcal{F}}_{\widehat{\Theta}}(f) = \hat{u} \approx u$. This can be much better than handcrafted solution mapping.
- The impact of deep learning is much beyond image reconstruction.
- How can we understand deep neural networks (DNNs) in comparison with the "handcrafted" mapping?

As discrete dynamic systems (ODEs and PDEs)

The ODE-Nets

- The success of residual networks (ResNet, He et al. 2015)

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{f}(\boldsymbol{x}^k, t_k), \qquad \boldsymbol{x}^0 = \boldsymbol{x}, \qquad k = 0,1, \dots$$

- ResNet can be interpreted as forward-Euler discretization of the dynamic system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$, and training ResNets can be viewed as an optimal control problem (E 2017).

The ODE-Nets

- The success of residual networks (ResNet, He et al. 2015)
$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{f}(\boldsymbol{x}^k, t_k), \qquad \boldsymbol{x}^0 = \boldsymbol{x}, \qquad k = 0,1,\ldots$$
- ResNet can be interpreted as forward-Euler discretization of the dynamic system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$, and training ResNets can be viewed as an optimal control problem (E 2017).
- More examples of deep networks that can be viewed as discrete form of ODEs or SDEs (Lu-Zhong-Li-Dong 2018):
    - e.g., PolyNet (Zhang-Li-Loy-Lin 2017) can be viewed as approximation to backward-Euler with truncated Neumann series.
    - e.g., ResNets with stochastic depth (Huang et al. 2016) strategy are discrete approximations of SDEs.
    - we can use discrete schemes of ODEs/SDEs to generate novel deep networks!

The ODE-Nets

- Using the linear two-step method (LM-ResNet):

$$\boldsymbol{x}^{k+1} = (1-\alpha_k)\boldsymbol{x}^k + \alpha_k \boldsymbol{x}^{k-1} + \boldsymbol{f}\left(\boldsymbol{x}^k, t_k\right), \qquad \boldsymbol{x}^0 = \boldsymbol{x}, \qquad k = 0,1,\ldots$$

- In comparison with ResNet via the modified equation analysis:

$$\begin{cases} \dot{\boldsymbol{x}}^k + \frac{\Delta t}{2}\ddot{\boldsymbol{x}}^k & = \boldsymbol{f}(\boldsymbol{x}^k, t_k). \quad \text{ResNet;} \\ (1+\alpha_k)\dot{\boldsymbol{x}}^k + (1-\alpha_k)\frac{\Delta t}{2}\ddot{\boldsymbol{x}}^k & = \boldsymbol{f}(\boldsymbol{x}^k, t_k), \quad \text{LM-ResNet.} \end{cases}$$

The ODE-Nets

- Using the linear two-step method (LM-ResNet):

$$\boldsymbol{x}^{k+1} = (1-\alpha_k)\boldsymbol{x}^k + \alpha_k \boldsymbol{x}^{k-1} + \boldsymbol{f}(\boldsymbol{x}^k, t_k), \qquad \boldsymbol{x}^0 = \boldsymbol{x}, \qquad k = 0,1,\ldots$$

- In comparison with ResNet via the modified equation analysis:

$$\begin{cases} \dot{\boldsymbol{x}}^k + \frac{\Delta t}{2}\ddot{\boldsymbol{x}}^k & = \boldsymbol{f}(\boldsymbol{x}^k, t_k). \quad \text{ResNet;} \\ (1+\alpha_k)\dot{\boldsymbol{x}}^k + (1-\alpha_k)\frac{\Delta t}{2}\ddot{\boldsymbol{x}}^k & = \boldsymbol{f}(\boldsymbol{x}^k, t_k), \quad \text{LM-ResNet.} \end{cases}$$

- Empirical results

Model	Layer	top-1	top-5
ResNet (He et al. (2015b))	50	24.7	7.8
ResNet (He et al. (2015b))	101	23.6	7.1
ResNet (He et al. (2015b))	152	23.0	6.7
LM-ResNet (Ours)	50, pre-act	23.8	7.0
LM-ResNet (Ours)	101, pre-act	**22.6**	**6.4**

ImageNet (1.28m training, 50k testing, 1k classes)

The ODE-Nets

- Using the linear two-step method (LM-ResNet):

$$\boldsymbol{x}^{k+1} = (1-\alpha_k)\boldsymbol{x}^k + \alpha_k \boldsymbol{x}^{k-1} + \boldsymbol{f}(\boldsymbol{x}^k, t_k), \qquad \boldsymbol{x}^0 = \boldsymbol{x}, \qquad k = 0,1,\dots$$

- In comparison with ResNet via the modified equation analysis:

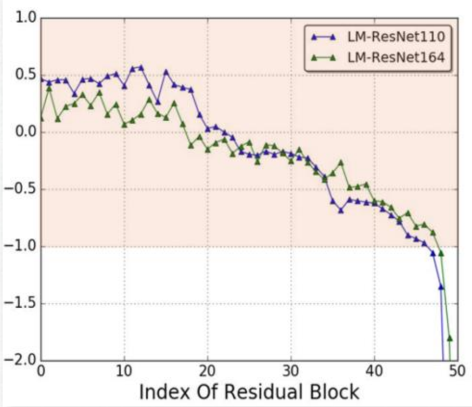
$$\begin{cases} \dot{\boldsymbol{x}}^k + \frac{\Delta t}{2}\ddot{\boldsymbol{x}}^k & = \boldsymbol{f}(\boldsymbol{x}^k, t_k). \quad \text{ResNet;} \\ (1+\alpha_k)\dot{\boldsymbol{x}}^k + (1-\alpha_k)\frac{\Delta t}{2}\ddot{\boldsymbol{x}}^k & = \boldsymbol{f}(\boldsymbol{x}^k, t_k), \quad \text{LM-ResNet.} \end{cases}$$

- Empirical results

Model	Layer	top-1	top-5
ResNet (He et al. (2015b))	50	24.7	7.8
ResNet (He et al. (2015b))	101	23.6	7.1
ResNet (He et al. (2015b))	152	23.0	6.7
LM-ResNet (Ours)	50, pre-act	23.8	7.0
LM-ResNet (Ours)	101, pre-act	**22.6**	**6.4**

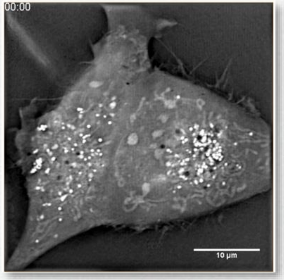
ImageNet (1.28m training, 50k testing, 1k classes)

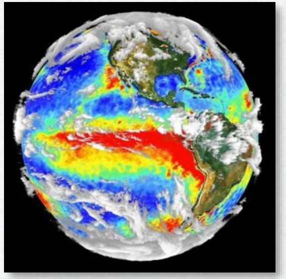
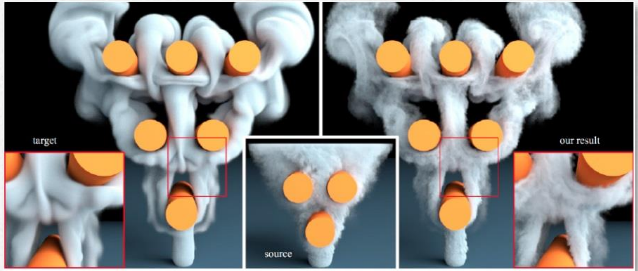
α_k



The PDE-Nets

- Motivation of PDE-Nets: given a sequence of observed dynamics
  - Find a PDE that best describes the observed data.
  - Enable fast simulations with the learned PDE.



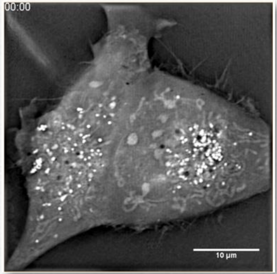



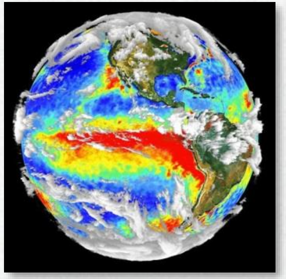
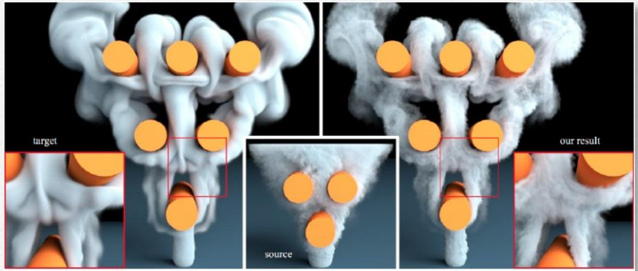
S. Sato et al. 2018

- Key to PDE-Nets:
  - Exploiting the structural similarity between deep convolutional neural networks (CNNs) and discrete schemes of PDEs.

The PDE-Nets

- Motivation of PDE-Nets: given a sequence of observed dynamics
  - Find a PDE that best describes the observed data.
  - Enable fast simulations with the learned PDE.



S. Sato et al. 2018

- Key to PDE-Nets:
  - Exploiting the structural similarity between deep convolutional neural networks (CNNs) and discrete schemes of PDEs.
- Related works on system identification:
  - Schmidt-Lipson 2009, Brunton-Proctor-Kutz 2016 (SINDy), Raissi-Perdikaris-Karniadakis 2019 (PINNs).

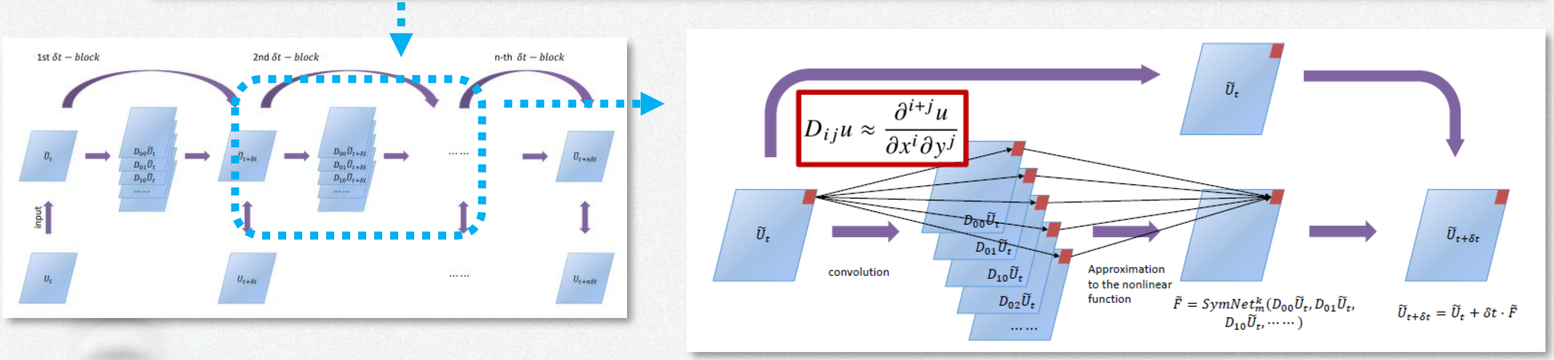
The PDE-Nets

- We first assume
  - $\frac{\partial u}{\partial t} = F\left(u, \nabla u, \nabla^2 u, \ldots\right),\ u \in \mathbb{R}^m$;
  - Maximum order of the PDE is known;
  - F is a multivariate function with simple operations (e.g., a polynomial).

The PDE-Nets

- We first assume
  - $\frac{\partial u}{\partial t} = F(u, \nabla u, \nabla^2 u, \dots),\ u \in \mathbb{R}^m$;
  - Maximum order of the PDE is known;
  - F is a multivariate function with simple operations (e.g., a polynomial).
- The PDE-Nets (Long-Lu-Ma-Dong 2018, Long-Lu-Dong 2019):

$$\boldsymbol{u}^{k+1} = \boldsymbol{u}^k + \Delta t \cdot SymNet_m^n(D_{00}\boldsymbol{u}^k, D_{01}\boldsymbol{u}^k, D_{10}\boldsymbol{u}^k, \cdots), \quad k = 0, 1, \dots, K-1.$$



The PDE-Nets

- Enforcing $D_{ij}u \approx \frac{\partial^{i+j}u}{\partial x^i \partial y^j}$ by applying **Proposition 2**.
  - Define the moment matrix

$$M(\boldsymbol{q}) = (m_{i,j})_{N\times N}, \quad m_{i,j} = \frac{1}{i!j!} \sum_{k_1,k_2=-\frac{N-1}{2}}^{\frac{N-1}{2}} k_1^i k_2^j \boldsymbol{q}[k_1,k_2], i,j = 0,1,\ldots,N-1.$$

Proposition 2. Let $\boldsymbol{q}$ be a high-pass filter with vanishing moments of order $\boldsymbol{\alpha} \in \mathbb{Z}_+^2$. Then for a smooth function $F(\boldsymbol{x})$ on $\mathbb{R}^2$, we have

$$\frac{1}{\varepsilon^{|\alpha|}} \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}]F(\boldsymbol{x}+\varepsilon\boldsymbol{k}) = C_\alpha \frac{\partial^\alpha}{\partial \boldsymbol{x}^\alpha} F(\boldsymbol{x}) + O(\varepsilon),$$

where C_α is the constant defined by

$$C_\alpha = \frac{1}{\alpha!} \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{k}^\alpha \boldsymbol{q}[\boldsymbol{k}] = \frac{i^{|\alpha|}}{\alpha!} \frac{\partial^\alpha}{\partial \omega^\alpha} \widehat{\boldsymbol{q}}(\omega)\Big|_{\omega=\boldsymbol{0}}.$$

If, in addition, $\boldsymbol{q}$ has total vanishing moments of order $K\backslash\{|\alpha|+1\}$ for some $K > |\alpha|$, then

$$\frac{1}{\varepsilon^{|\alpha|}} \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}]F(\boldsymbol{x}+\varepsilon\boldsymbol{k}) = C_\alpha \frac{\partial^\alpha}{\partial \boldsymbol{x}^\alpha} F(\boldsymbol{x}) + O(\varepsilon^{K-|\alpha|}).$$

The PDE-Nets

- Enforcing $D_{ij}u \approx \frac{\partial^{i+j}u}{\partial x^i \partial y^j}$ by applying **Proposition 2**.
  - Define the moment matrix

$$M(\boldsymbol{q}) = (m_{i,j})_{N\times N}, \quad m_{i,j} = \frac{1}{i!j!}\sum_{k_1,k_2=-\frac{N-1}{2}}^{\frac{N-1}{2}} k_1^i k_2^j \boldsymbol{q}[k_1,k_2], i,j = 0,1,\ldots,N-1.$$

  - We can approximate any differential operator at any prescribed order by constraining $M(\boldsymbol{q})$!
  - For example: approximation of $\frac{\partial u}{\partial x}$ by $\boldsymbol{q} \circledast \boldsymbol{u}$ with a 3×3 kernel $\boldsymbol{q}$

$$\begin{pmatrix} 0 & 0 & \star \\ 1 & \star & \star \\ \star & \star & \star \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & \star \\ 0 & \star & \star \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The PDE-Nets

- Enforcing $D_{ij}u \approx \frac{\partial^{i+j}u}{\partial x^i \partial y^j}$ by applying **Proposition 2**.
  - Define the moment matrix

$$M(\boldsymbol{q}) = (m_{i,j})_{N\times N}, \quad m_{i,j} = \frac{1}{i!j!} \sum_{k_1,k_2=-\frac{N-1}{2}}^{\frac{N-1}{2}} k_1^i k_2^j \boldsymbol{q}[k_1,k_2], i,j = 0,1,\ldots,N-1.$$

  - We can approximate any differential operator at any prescribed order by constraining $M(\boldsymbol{q})$!
  - For example: approximation of $\frac{\partial u}{\partial x}$ by $\boldsymbol{q} \circledast \boldsymbol{u}$ with a 3×3 kernel $\boldsymbol{q}$

$$\begin{pmatrix} 0 & 0 & \star \\ 1 & \star & \star \\ \star & \star & \star \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & \star \\ 0 & \star & \star \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

  - Similar idea was also adopted by (Bar-Sinai-Hoyer-Hickey-Brenner 2019, Chambolle-Pock 2021, Alt et al. 2021).

The PDE-Nets

- Empirical results: learning Burgers' equation

$$\boldsymbol{u}_t + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = \varepsilon\, \nabla^2 \boldsymbol{u}, \qquad \boldsymbol{u} = (u, v), \varepsilon = 0.05.$$

The PDE-Nets

- Empirical results: learning Burgers' equation

$$\boldsymbol{u}_t + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = \varepsilon \nabla^2 \boldsymbol{u}, \qquad \boldsymbol{u} = (u, v), \varepsilon = 0.05.$$

Ground truth	$u_t = -uu_x - vu_y + 0.05(u_{xx} + u_{yy})$ $v_t = -uv_x - vv_y + 0.05(v_{xx} + v_{yy})$
Frozen PDE-Net	$u_t = -0.906uu_x - 0.901vu_y + 0.033u_{xx} + 0.037u_{yy}$ $v_t = -0.907vv_y - 0.902uv_x + 0.039v_{xx} + 0.032v_{yy}$
PDE-Net	$u_t = -0.979uu_x - 0.973u_yv + 0.052u_{xx} + 0.051u_{yy}$ $v_t = -0.973uv_x - 0.977vv_y + 0.053v_{xx} + 0.051v_{yy}$

Model recovery

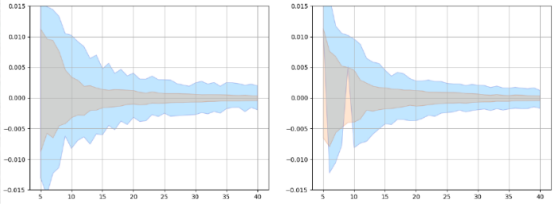
The PDE-Nets

- Empirical results: learning Burgers' equation

$$\boldsymbol{u}_t + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = \varepsilon\,\nabla^2\boldsymbol{u}, \qquad \boldsymbol{u} = (u, v), \varepsilon = 0.05.$$

Ground truth	$u_t = -uu_x - vu_y + 0.05(u_{xx} + u_{yy})$ $v_t = -uv_x - vv_y + 0.05(v_{xx} + v_{yy})$
Frozen PDE-Net	$u_t = -0.906uu_x - 0.901vu_y + 0.033u_{xx} + 0.037u_{yy}$ $v_t = -0.907vv_y - 0.902uv_x + 0.039v_{xx} + 0.032v_{yy}$
PDE-Net	$u_t = -0.979uu_x - 0.973u_yv + 0.052u_{xx} + 0.051u_{yy}$ $v_t = -0.973uv_x - 0.977vv_y + 0.053v_{xx} + 0.051v_{yy}$

Model recovery



Remainer weights of u, v

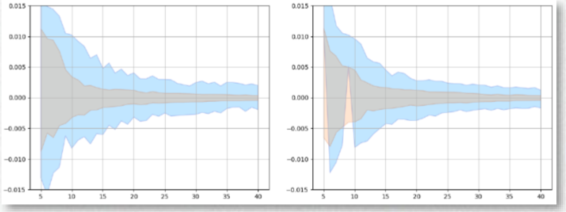
The PDE-Nets

- Empirical results: learning Burgers' equation

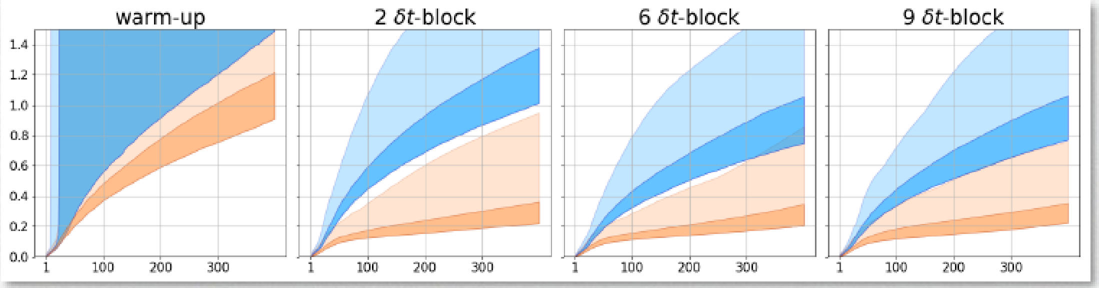
$$\boldsymbol{u}_t + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = \varepsilon\, \nabla^2 \boldsymbol{u}, \qquad \boldsymbol{u} = (u, v), \varepsilon = 0.05.$$

Ground truth	$u_t = -uu_x - vu_y + 0.05(u_{xx} + u_{yy})$ $v_t = -uv_x - vv_y + 0.05(v_{xx} + v_{yy})$
Frozen PDE-Net	$u_t = -0.906uu_x - 0.901vu_y + 0.033u_{xx} + 0.037u_{yy}$ $v_t = -0.907vv_y - 0.902uv_x + 0.039v_{xx} + 0.032v_{yy}$
PDE-Net	$u_t = -0.979uu_x - 0.973u_y v + 0.052u_{xx} + 0.051u_{yy}$ $v_t = -0.973uv_x - 0.977vv_y + 0.053v_{xx} + 0.051v_{yy}$

Model recovery



Remainer weights of u, v



Frozen PDE-Net

PDE-Net

Simulations

The PDE-Nets

- Incorporation of PDEs (physics) into the architecture of the network is beneficial!

The PDE-Nets

- Incorporation of PDEs (physics) into the architecture of the network is beneficial!
  - Machine learning accelerated simulations: e.g., learning an operator

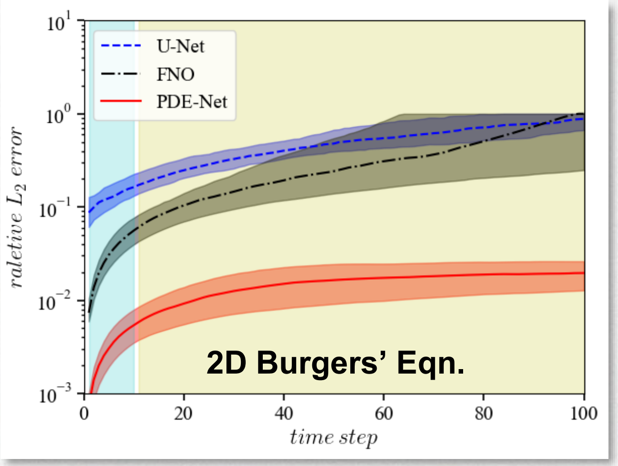
$$\mathcal{F}_{\Theta}: u(t,\cdot) \to u(t + \Delta\mathrm{t},\cdot)$$

The PDE-Nets

- Incorporation of PDEs (physics) into the architecture of the network is beneficial!
    - Machine learning accelerated simulations: e.g., learning an operator
    $$\mathcal{F}_{\Theta}: u(t,\cdot) \to u(t+\Delta t,\cdot)$$
    - A comparison between PDE-Nets, U-Net (Ronneberger-Fischer-Brox 2015) and Fourier Neural Operator (FNO, Li et al. 2020)

The PDE-Nets

- Incorporation of PDEs (physics) into the architecture of the network is beneficial!
  - Machine learning accelerated simulations: e.g., learning an operator
$$\mathcal{F}_{\Theta}: u(t,\cdot) \to u(t+\Delta t,\cdot)$$
  - A comparison between PDE-Nets, U-Net (Ronneberger-Fischer-Brox 2015) and Fourier Neural Operator (FNO, Li et al. 2020)



 # Parameters

- PDE-Net: $\sim 10^2$
- FNO: $\sim 10^5$
- U-Net: $\sim 10^7$

Modeling + Learning: A General Strategy

- A typical workflow:
  - For a given problem of interest, start with your favorite algorithm which is most likely a discrete dynamic system, e.g.,
    - solution mapping $\mathcal{F}$ for image reconstruction.
  - Identify the component(s) that is hard to handcraft, e.g.,
    - dependence of hyperparameters with input image.
  - Approximate the it with a properly designed deep neural network, making the solution mapping $\mathcal{F}_\Theta$ learnable.
  - Select a loss function and training algorithm, e.g.,

$$\begin{cases} \min_{\boldsymbol{\Theta}} \ \mathbb{E}_{(\boldsymbol{u},\boldsymbol{f})\sim\mathcal{P}} \ \ell(\mathcal{F}_{\boldsymbol{\Theta}}(\boldsymbol{f}),\boldsymbol{u}) + r(\mathcal{F}_{\boldsymbol{\Theta}}), & \text{supervised;} \\ \min_{\boldsymbol{\Theta}} \ \mathbb{E}_{\boldsymbol{f}\sim\mathcal{P}} \ \ell(\boldsymbol{A}\mathcal{F}_{\boldsymbol{\Theta}}(\boldsymbol{f}),\boldsymbol{f}) + r(\mathcal{F}_{\boldsymbol{\Theta}}), & \text{unsupervised.} \end{cases}$$

- Same workflow can be applied to problems in other areas as well, e.g., scientific computing, AI for science.

05

Computational Imaging

Integrating sensing, reconstruction and analysis.

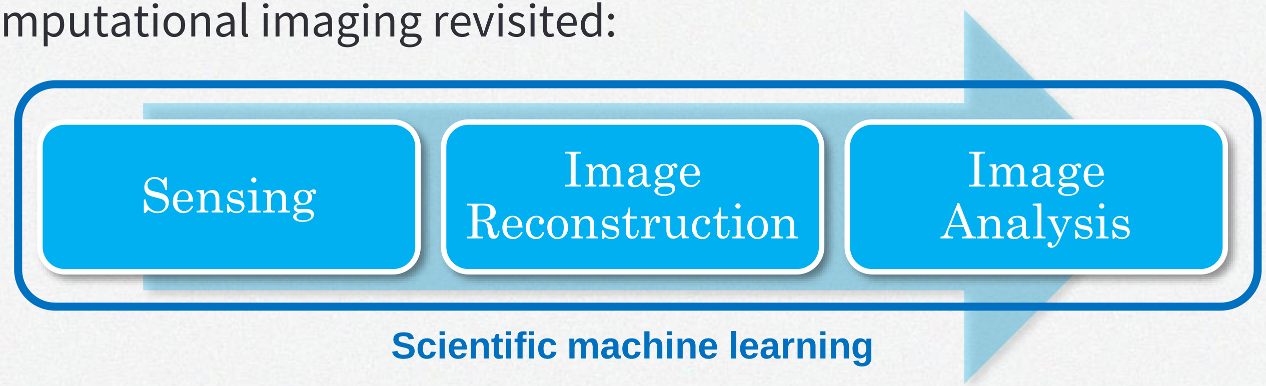
Joint work with:

Harvard University:
Georges El Fakhri
Kyungsang Kim
Quanzheng Li
Dufan Wu

Ziju Shen (PKU)
Yufei Wang (CMU)
Xu Yang (UCSB)

Integration of Computational Imaging

- Computational imaging revisited:

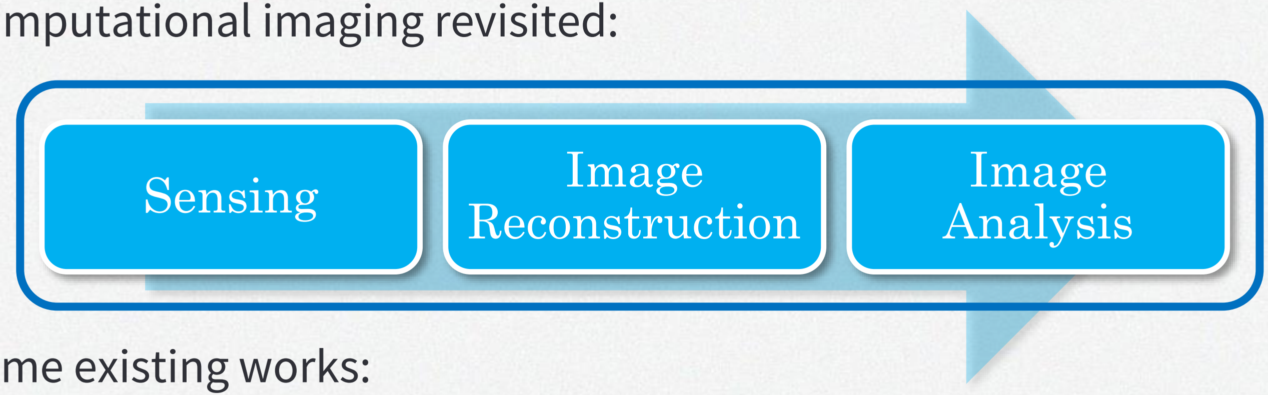


$$\min_{\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2,\boldsymbol{\Theta}_3} \mathbb{E}_{(\boldsymbol{u},\boldsymbol{z})\sim\mathcal{P}}\, \ell\left(\mathcal{G}_{\boldsymbol{\Theta}_3} \circ \mathcal{F}_{\boldsymbol{\Theta}_2} \circ \mathcal{M}_{\boldsymbol{\Theta}_1}(\boldsymbol{u})\right), \boldsymbol{z}).$$

Analysis | Reconstruction | Sensing

Integration of Computational Imaging

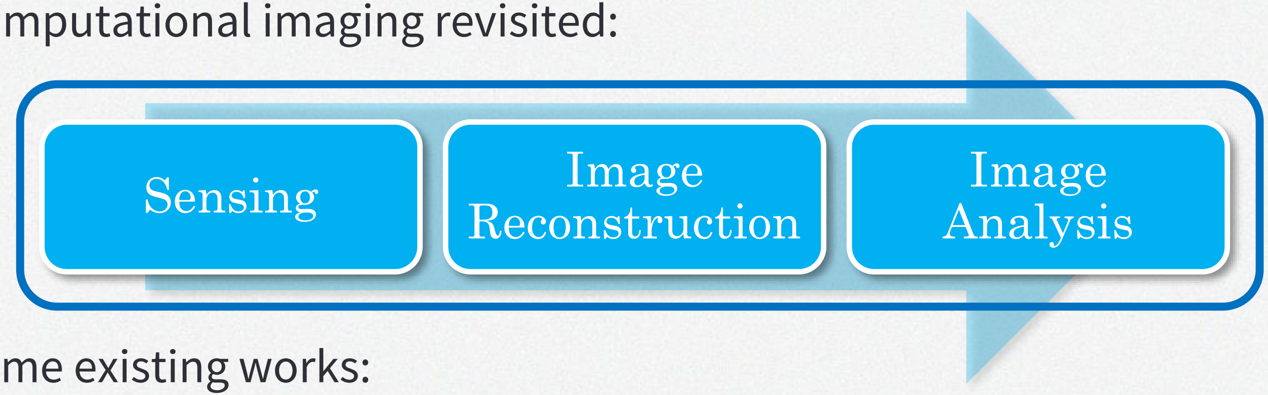
- Computational imaging revisited:



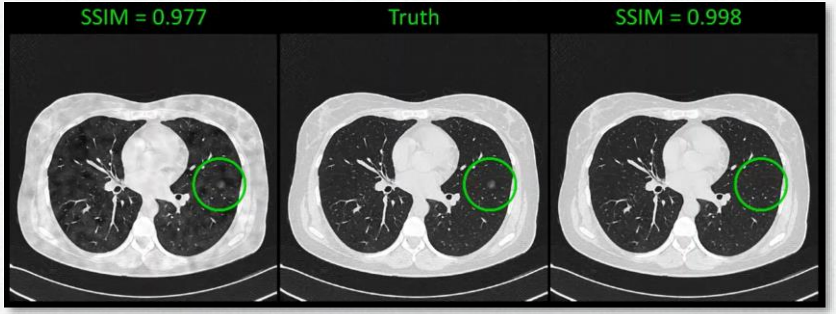
- Some existing works:
  - Reconstruction + Analysis: Liu et al. 2018, **Wu et al. 2018**, Huang et al. 2019

Integration of Computational Imaging

- Computational imaging revisited:



- Some existing works:
  - Reconstruction + Analysis: Liu et al. 2018, **Wu et al. 2018**, Huang et al. 2019

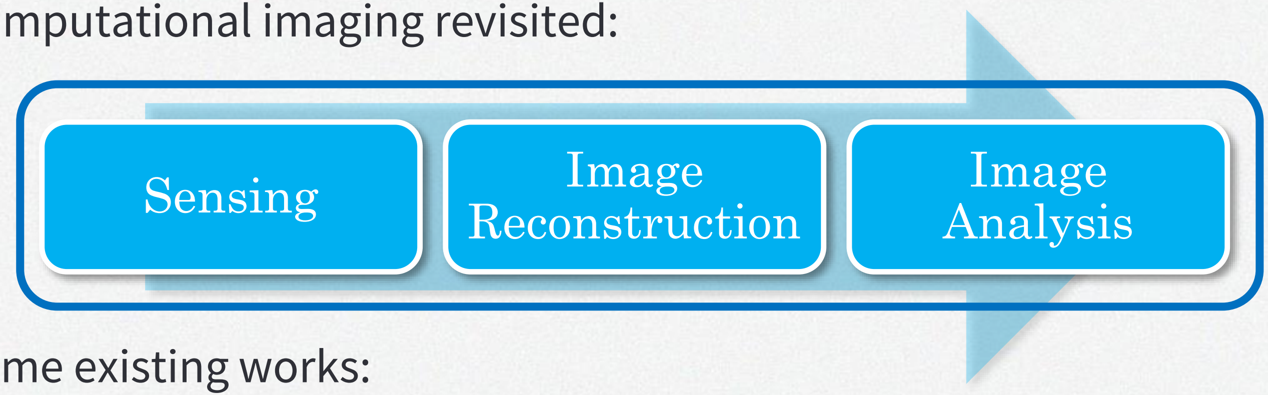


J. Webster Stayman, JHU

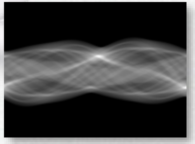
Slides from “Deep Recon Workshop, 2021”

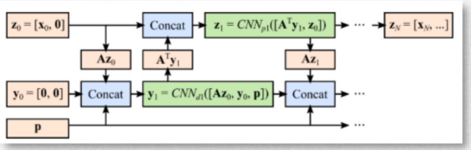
Integration of Computational Imaging

- Computational imaging revisited:

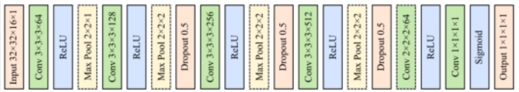


- Some existing works:
    - Reconstruction + Analysis: Liu et al. 2018, **Wu et al. 2018**, Huang et al. 2019

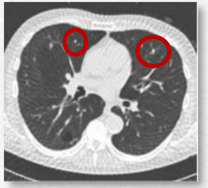




A network for reconstruction

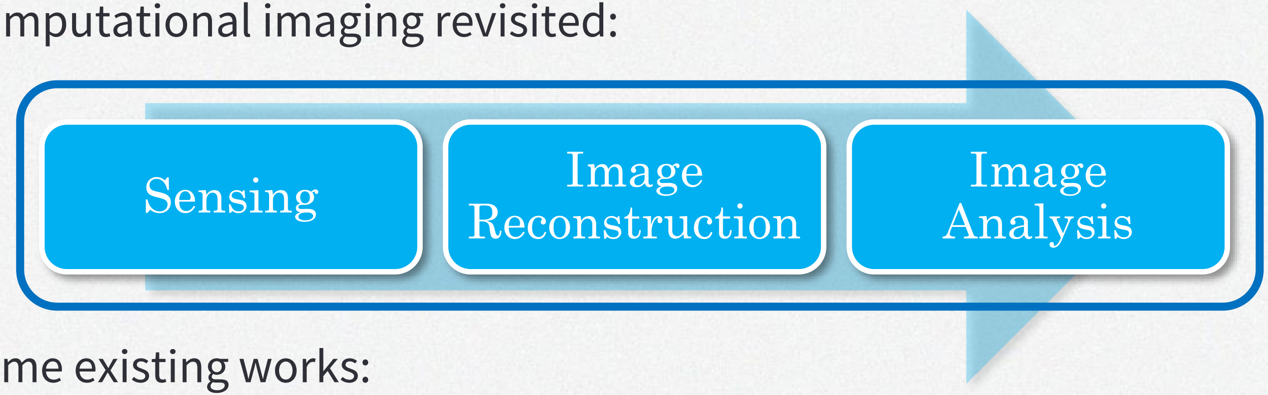


A network for nodule detection



Integration of Computational Imaging

- Computational imaging revisited:

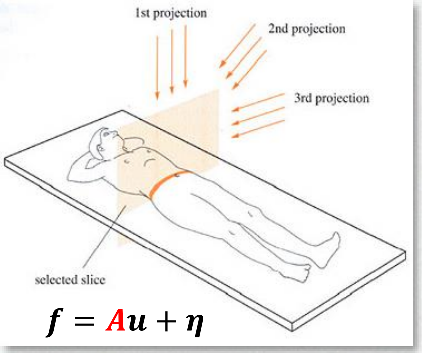


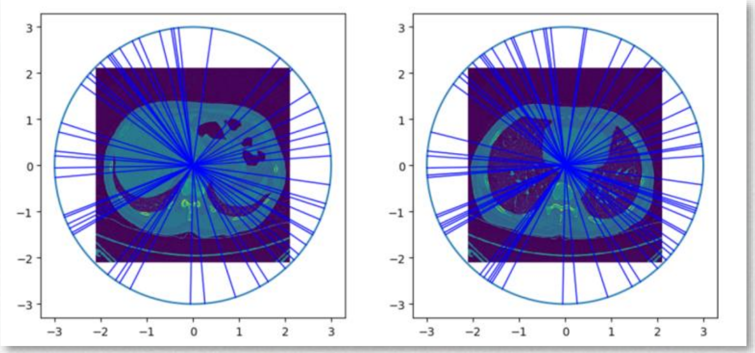
- Some existing works:
  - Reconstruction + Analysis: Liu et al. 2018, **Wu et al. 2018**, Huang et al. 2019
  - Sensing + Reconstruction: Jin-Unser-Yi 2019, Pineda et al. 2020, Ede 2021, Yin et al. 2021, **Shen et al. 2022**.

Integration of Computational Imaging

- Computational imaging revisited:

- Personalized CT scanning (Shen et al. 2022)



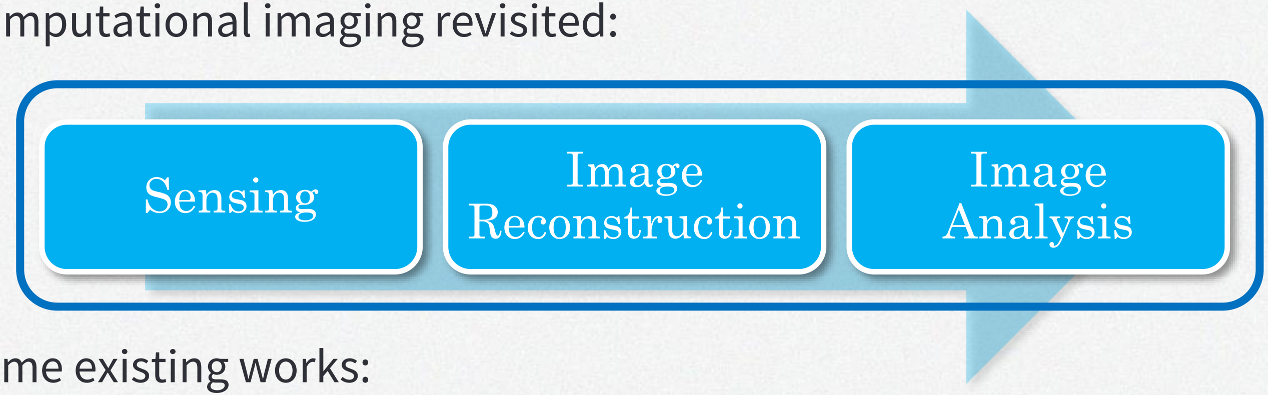


Goal: to optimize
1. projection angles;
2. dose allocation,

so that image quality is maximized for a given total dose.

Integration of Computational Imaging

- Computational imaging revisited:



- Some existing works:
  - Reconstruction + Analysis: Liu et al. 2018, **Wu et al. 2018**, Huang et al. 2019
  - Sensing + Reconstruction: Jin-Unser-Yi 2019, Pineda et al. 2020, Ede 2021, Yin et al. 2021, **Shen et al. 2022**.
  - Sensing + Reconstruction + Analysis: Wetzstein et al. 2020.

06

Conclusions

Concluding Remarks

- What I have covered in this talk
    - Importance of images and the role of computational imaging.
    - Two prevailing mathematical approaches for image reconstruction, and their connections.
    - Understanding of deep learning, and how we can work with it together with the other tools we have.

Concluding Remarks

- What I have covered in this talk
  - Importance of images and the role of computational imaging.
  - Two prevailing mathematical approaches for image reconstruction, and their connections.
  - Understanding of deep learning, and how we can work with it together with the other tools we have.
- Looking into the (near) future
  - Combination handcraft and data-driven modeling.
  - More advancements in the integrated computational imaging.

Thanks!

Questions?

You can reach me at
dongbin@math.pku.edu.cn

https://bicmr.pku.edu.cn/~dongbin/