Union of Learned Sparsifying Transforms Based Low-Dose 3D CT Image Reconstruction

Xuehang Zheng[1], Saiprasad Ravishankar[2], Yong Long[1], Jeff Fessler[2]

[1]University of Michigan - Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai, China
[2]Department of Electrical Engineering and Computer Science, University of Michigan, MI, USA

June 21, 2017

Outline

Dictionary Learning-Based LDCT Reconstruction

- Challenges in Low-Dose CT (LDCT):
  - significantly reduce patient radiation exposure
  - maintain high image quality
- Apply the Prior Information Learned from Big Datasets of Normal-Dose CT Images into LDCT Reconstruction.
  - Training Phase $\Longrightarrow$ Prior $\Longrightarrow$ Reconstruction Phase
- Dictionary Learning-Based Approaches[1]:
  - have shown promising results for LDCT
  - typically use an overcomplete dictionary
  - NP-Hard sparse coding

[1] [Xu et al., IEEE T-MI, 2012]

Dictionary Learning-Based LDCT Reconstruction

- Challenges in Low-Dose CT (LDCT):
  - significantly reduce patient radiation exposure
  - maintain high image quality
- Apply the Prior Information Learned from Big Datasets of Normal-Dose CT Images into LDCT Reconstruction.
  - Training Phase $\Longrightarrow$ Prior $\Longrightarrow$ Reconstruction Phase
- Dictionary Learning-Based Approaches[1]:
  - have shown promising results for LDCT
  - typically use an overcomplete dictionary
  - NP-Hard sparse coding

[1] [Xu et al., IEEE T-MI, 2012]

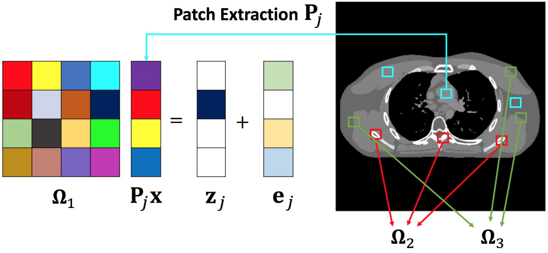
Dictionary Learning-Based LDCT Reconstruction

- Challenges in Low-Dose CT (LDCT):
  - significantly reduce patient radiation exposure
  - maintain high image quality
- Apply the Prior Information Learned from Big Datasets of Normal-Dose CT Images into LDCT Reconstruction.
  - Training Phase $\Longrightarrow$ Prior $\Longrightarrow$ Reconstruction Phase
- Dictionary Learning-Based Approaches[1]:
  - have shown promising results for LDCT
  - typically use an overcomplete dictionary
  - NP-Hard sparse coding

[1] [Xu et al., IEEE T-MI, 2012]

Union of Learned TRAnsforms (ULTRA)

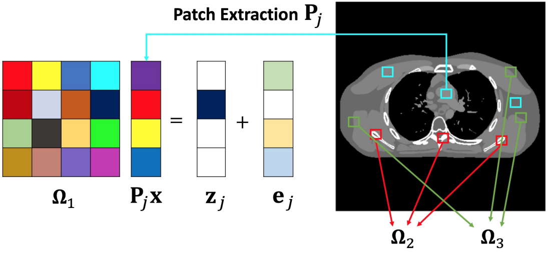
- **Sparsifying Transform Learning**[2]:
  a generalized analysis operator
- **Learning A Union of Transforms**[3]:
  one for each class of features (group of patches)
- Closed-form solutions for sparse coding (and clustering):
  Computational cost: $O(l^2 N)$ vs $O(l^3 N)$ for Dictionary



[2] [Ravishankar & Bresler, IEEE T-SP, 2015]
[3] [Wen et al., Int J Comput Vis, 2015]

Union of Learned TRAnsforms (ULTRA)

- **Sparsifying Transform Learning**[2]:
  a generalized analysis operator
- **Learning A Union of Transforms**[3]:
  one for each class of features (group of patches)
- Closed-form solutions for sparse coding (and clustering):
  Computational cost: $O(l^2N)$ vs $O(l^3N)$ for Dictionary

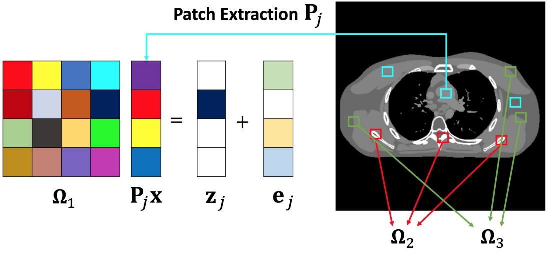


[2] [Ravishankar & Bresler, IEEE T-SP, 2015]

[3] [Wen et al., Int J Comput Vis, 2015]

Union of Learned TRAnsforms (ULTRA)

- **Sparsifying Transform Learning**[2]:
  a generalized analysis operator
- **Learning A Union of Transforms**[3]:
  one for each class of features (group of patches)
- Closed-form solutions for sparse coding (and clustering):
  Computational cost: $O(l^2 N)$ vs $O(l^3 N)$ for Dictionary



[2] [Ravishankar & Bresler, IEEE T-SP, 2015]

[3] [Wen et al., Int J Comput Vis, 2015]

Outline

Learning A Union of Transforms

$$\min_{\{\boldsymbol{\Omega}_k, \mathbf{Z}_i, C_k\}} \sum_{k=1}^{K} \sum_{i \in C_k} \Big\{ \overbrace{\|\boldsymbol{\Omega}_k \mathbf{X}_i - \mathbf{Z}_i\|_2^2}^{\text{Sparsification Error}} + \overbrace{\eta^2 \|\mathbf{Z}_i\|_0}^{\text{Sparsity Penalty}} \Big\} + \sum_{k=1}^{K} \lambda_k Q(\boldsymbol{\Omega}_k) \quad \text{(P0)}$$

- $\{\boldsymbol{\Omega}_k\}_{k=1}^{K}$: union of square transforms.
- $\mathbf{Z}_i$: sparse code of the training signal $\mathbf{X}_i$.
- $Q(\boldsymbol{\Omega}_k) \triangleq \|\boldsymbol{\Omega}_k\|_F^2 - \log|\det \boldsymbol{\Omega}_k|$: controls the properties of $\boldsymbol{\Omega}_k$[4].
- C_k: the set of indices of signals matched to the kth class.
- An efficient alternating algorithm is used for (P0).

[4] [Ravishankar & Bresler, IEEE T-SP, 2015]

Image Reconstruction

Penalized Weighted-Least Squares (PWLS):

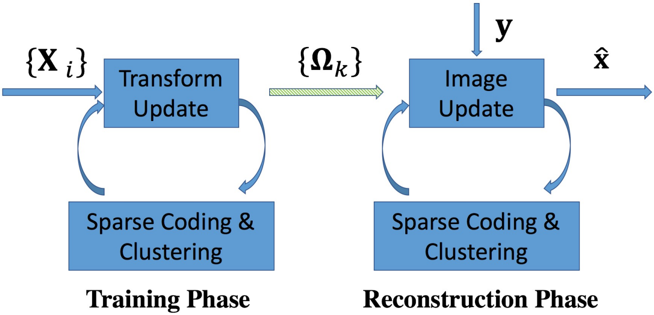
$$\min_{\mathbf{x} \succeq \mathbf{0}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \beta \mathrm{R}(\mathbf{x}) \tag{P1}$$

- $\mathbf{y}$: noisy sinogram (measurement)
- $\mathbf{A}$: system matrix
- $\mathbf{x}$: unknown image (volume)
- $\mathbf{W}$: diagonal weighting matrix
- $\mathrm{R}(\mathbf{x})$: regularizer
- β: regularization parameter

Image Reconstruction: PWLS-ULTRA

$$\min_{\mathbf{x}\succeq\mathbf{0}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \beta \mathrm{R}(\mathbf{x}) \tag{P1}$$

$$\mathrm{R}(\mathbf{x}) \triangleq \min_{\{\mathbf{z}_j, C_k\}} \sum_{k=1}^{K} \left\{ \sum_{j\in C_k} \|\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma^2 \|\mathbf{z}_j\|_0 \right\} \tag{1}$$



Outline

Image Update Step

$$\min_{\mathbf{x}\succeq\mathbf{0}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \beta \overbrace{\sum_{k=1}^{K}\sum_{j\in C_k}\|\mathbf{\Omega}_k\mathbf{P}_j\mathbf{x}-\mathbf{z}_j\|_2^2}^{\mathrm{R}_2(\mathbf{x})} \tag{2}$$

We solve it using the **relaxed linearized augmented Lagrangian method with ordered-subsets** (relaxed OS-LALM)[5]:

$$\begin{cases} \mathbf{s}^{(k+1)} = \rho(\mathbf{D}_{\mathbf{A}}\mathbf{x}^{(k)} - \mathbf{h}^{(k)}) + (1-\rho)\mathbf{g}^{(k)} \\ \mathbf{x}^{(k+1)} = [\mathbf{x}^{(k)} - (\rho\mathbf{D}_{\mathbf{A}} + \mathbf{D}_{\mathrm{R}_2})^{-1}(\mathbf{s}^{(k+1)} + \nabla\mathrm{R}_2(\mathbf{x}^{(k)}))]_{\mathcal{C}} \\ \zeta^{(k+1)} \triangleq M\mathbf{A}_m^T\mathbf{W}_m(\mathbf{A}_m\mathbf{x}^{(k+1)} - \mathbf{y}_m) \\ \mathbf{g}^{(k+1)} = \frac{\rho}{\rho+1}(\alpha\zeta^{(k+1)} + (1-\alpha)\mathbf{g}^{(k)}) + \frac{1}{\rho+1}\mathbf{g}^{(k)} \\ \mathbf{h}^{(k+1)} = \alpha(\mathbf{D}_{\mathbf{A}}\mathbf{x}^{(k+1)} - \zeta^{(k+1)}) + (1-\alpha)\mathbf{h}^{(k)} \end{cases} \tag{3}$$

[5] [Nien & Fessler, IEEE T-MI, 2016]

Image Update Step

$$\min_{\mathbf{x}\succeq\mathbf{0}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \overbrace{\beta\sum_{k=1}^{K}\sum_{j\in C_k}\|\mathbf{\Omega}_k\mathbf{P}_j\mathbf{x}-\mathbf{z}_j\|_2^2}^{\mathrm{R}_2(\mathbf{x})} \tag{2}$$

We solve it using the **relaxed linearized augmented Lagrangian method with ordered-subsets** (relaxed OS-LALM)[5]:

$$\begin{cases} \mathbf{s}^{(k+1)} = \rho(\mathbf{D}_{\mathbf{A}}\mathbf{x}^{(k)} - \mathbf{h}^{(k)}) + (1-\rho)\mathbf{g}^{(k)} \\ \mathbf{x}^{(k+1)} = [\mathbf{x}^{(k)} - (\rho\mathbf{D}_{\mathbf{A}} + \mathbf{D}_{\mathrm{R}_2})^{-1}(\mathbf{s}^{(k+1)} + \nabla \mathrm{R}_2(\mathbf{x}^{(k)}))]_{\mathcal{C}} \\ \zeta^{(k+1)} \triangleq M\mathbf{A}_m^T\mathbf{W}_m(\mathbf{A}_m\mathbf{x}^{(k+1)} - \mathbf{y}_m) \\ \mathbf{g}^{(k+1)} = \frac{\rho}{\rho+1}(\alpha\zeta^{(k+1)} + (1-\alpha)\mathbf{g}^{(k)}) + \frac{1}{\rho+1}\mathbf{g}^{(k)} \\ \mathbf{h}^{(k+1)} = \alpha(\mathbf{D}_{\mathbf{A}}\mathbf{x}^{(k+1)} - \zeta^{(k+1)}) + (1-\alpha)\mathbf{h}^{(k)} \end{cases} \tag{3}$$

[5] [Nien & Fessler, IEEE T-MI, 2016]

Sparse Coding and Clustering Step

$$\min_{\{\mathbf{z}_j\},\{C_k\}} \sum_{k=1}^{K} \left\{ \sum_{j \in C_k} \|\boldsymbol{\Omega}_k \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma^2 \|\mathbf{z}_j\|_0 \right\} \tag{4}$$

- Hard-thresholding operator $H_\gamma(\cdot)$: sets entries with mag. $< \gamma$ to 0.
- For each patch, the optimal cluster assignment:

$$\hat{k}_j = \underset{1 \le k \le K}{\arg\min} \|\boldsymbol{\Omega}_k \mathbf{P}_j \mathbf{x} - H_\gamma(\boldsymbol{\Omega}_k \mathbf{P}_j \mathbf{x})\|_2^2 + \gamma^2 \|H_\gamma(\boldsymbol{\Omega}_k \mathbf{P}_j \mathbf{x})\|_0. \tag{5}$$

- The optimal sparse code: $\hat{\mathbf{z}}_j = H_\gamma(\boldsymbol{\Omega}_{\hat{k}_j} \mathbf{P}_j \mathbf{x})$.

Sparse Coding and Clustering Step

$$\min_{\{\mathbf{z}_j\},\{C_k\}} \sum_{k=1}^{K} \left\{ \sum_{j \in C_k} \|\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma^2 \|\mathbf{z}_j\|_0 \right\} \quad (4)$$

- Hard-thresholding operator $H_\gamma(\cdot)$: sets entries with mag. $< \gamma$ to 0.
- For each patch, the optimal cluster assignment:

$$\hat{k}_j = \underset{1 \le k \le K}{\arg\min} \|\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - H_\gamma(\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x})\|_2^2 + \gamma^2 \|H_\gamma(\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x})\|_0. \quad (5)$$

- The optimal sparse code: $\hat{\mathbf{z}}_j = H_\gamma(\mathbf{\Omega}_{\hat{k}_j} \mathbf{P}_j \mathbf{x})$.

Sparse Coding and Clustering Step

$$\min_{\{\mathbf{z}_j\},\{C_k\}} \sum_{k=1}^{K} \left\{ \sum_{j \in C_k} \|\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma^2 \|\mathbf{z}_j\|_0 \right\} \tag{4}$$

- Hard-thresholding operator $H_\gamma(\cdot)$: sets entries with mag. $< \gamma$ to 0.
- For each patch, the optimal cluster assignment:

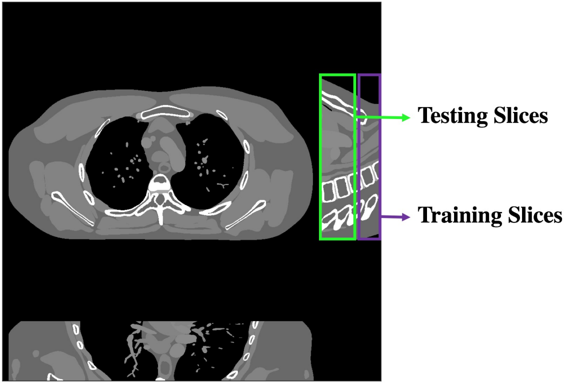
$$\hat{k}_j = \underset{1 \leq k \leq K}{\arg\min} \|\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - H_\gamma(\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x})\|_2^2 + \gamma^2 \|H_\gamma(\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x})\|_0. \tag{5}$$

- The optimal sparse code: $\hat{\mathbf{z}}_j = H_\gamma(\mathbf{\Omega}_{\hat{k}_j} \mathbf{P}_j \mathbf{x})$.

Outline

3D Axial Cone-beam CT with XCAT phantom[6]

- **Training:** $512 \times 512 \times 54$ XCAT image volume with patch size $8 \times 8 \times 8$ and patch stride $2 \times 2 \times 2$ ($\approx 1.5 \times 10^6$ patches).



[6] [Segars et al., MP, 2008]

3D Axial Cone-beam CT with XCAT phantom[7]

- **Testing**:
  - Sinogram size $888 \times 64 \times 984$;
  - Volume size $420 \times 420 \times 96$ (air cropped);
  - $\Delta_x = \Delta_y = 0.977$ and $\Delta_z = 0.625$ mm;
  - Patch size $8 \times 8 \times 8$ with stride $2 \times 2 \times 2$ ($\approx 2 \times 10^6$ patches).

- **Reconstruction Methods**:
  - **FDK** with a Hanning window.
  - **PWLS-EP** with "Lange3" **E**dge-**P**reserving regularizer.
  - **PWLS-ST** based on a learned single **S**quare **T**ransform ($K = 1$).
  - **PWLS-ULTRA** based on a **U**nion of **L**earned **TRA**nsforms.

[7] [Segars et al., MP, 2008]

RMSE & SSIM Comparison

Table: RMSE (HU) & SSIM of reconstructions for two incident photon intensities.

Intensity	FDK	EP	ST ($K = 1$)	ULTRA ($K = 15$)
1×10^4	67.8	33.7	31.9	**31.5**
	0.536	0.917	0.976	**0.979**
5×10^3	89.0	39.9	37.4	**37.2**
	0.463	0.894	0.967	**0.969**

- ULTRA scheme further improves the reconstruction than ST.

Performance Across Slices

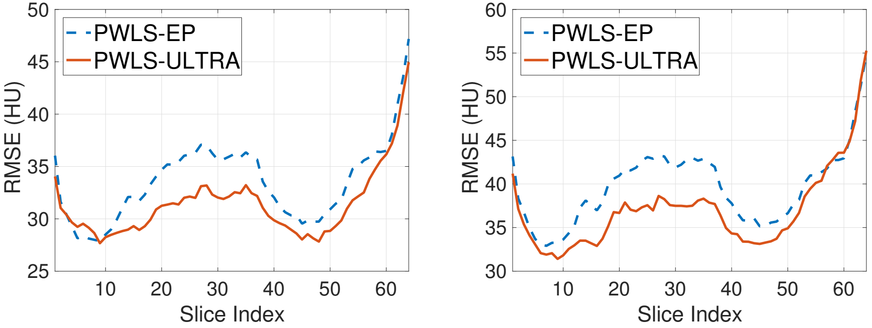


Figure: RMSE of axial slices for 1×10^4 (left) and 5×10^3 (right).

- ULTRA provides improvement for most of axial slices.

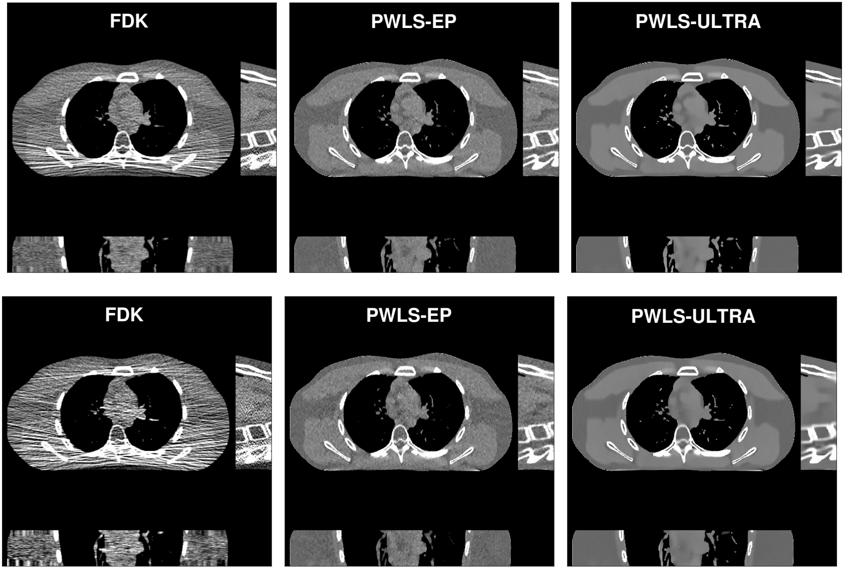


Figure: Photon intensity: 1×10^4 (top row) and 5×10^3 (bottom row).

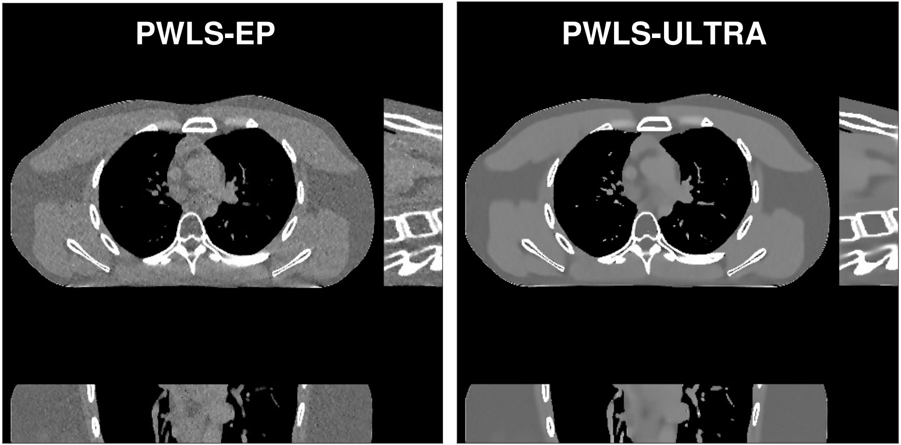


Figure: Photon intensity: 1×10^4.

An Example of Clustering Result

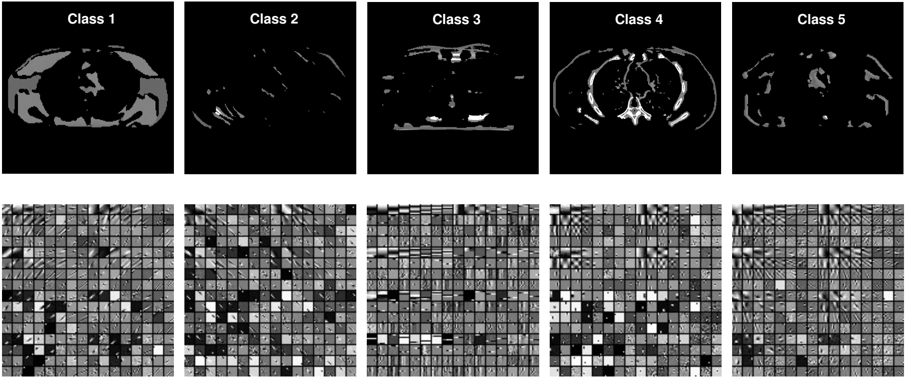
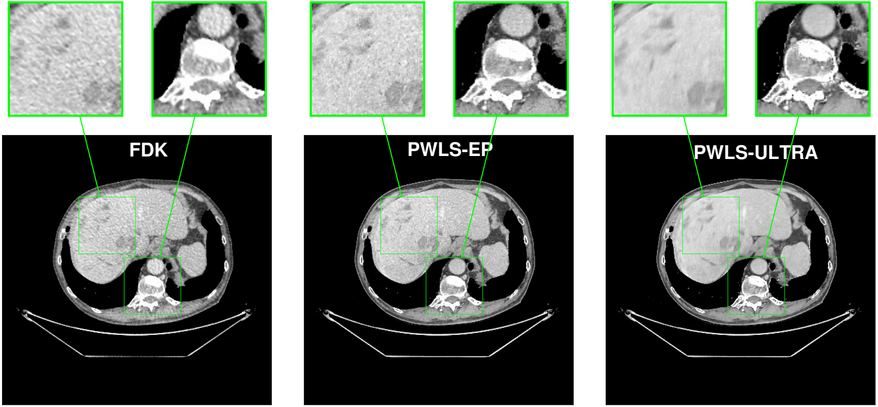


Figure: Pixel clustering results (top) for the central axial slice of PWLS-ULTRA ($K = 5$) for 1×10^4, and a slice of the corresponding 3D transforms (bottom).

3D Reconstructions of a Helical Chest Scan

- Sinogram size $888 \times 64 \times 3611$;
- Pitch 1.0 (about 3.7 rotations with rotation time 0.4s);
- Volume size $420 \times 420 \times 222$, $\Delta_x = \Delta_y = 1.167$ and $\Delta_z = 0.625$ mm;
- Patch size $8 \times 8 \times 8$ with stride $3 \times 3 \times 3$ ($\approx 1.5 \times 10^6$ patches);

3D Reconstructions of a Helical Chest Scan



- Use the transforms learned from XCAT phantom! ($K = 5$)
- Might not need closely matched dataset for training.

[7] FDK Reconstruction is provided by GE Healthcare.

Outline

- **Conclusion**
  - We proposed **PWLS-ST** and **PWLS-ULTRA** for 3D LDCT imaging, which combine PWLS reconstruction with **regularization based on learned sparsifying transforms**.
  - Both PWLS-ST and PWLS-ULTRA significantly improve the reconstruction quality compared to PWLS-EP.
  - The ULTRA scheme with a richer union of transforms model provides better reconstruction of various features such as bones, specific soft tissues, and edges, compared to a single transform model.
- **Future Work**
  - Convergence guarantees and automating the parameter selection.
  - New transform learning-based LDCT reconstruction methods, such as involving rotationally invariant transforms, or online transform learning[8], etc.

[8] [Ravishankar et al., IEEE J-STSP, 2015]

- **Conclusion**
  - We proposed **PWLS-ST** and **PWLS-ULTRA** for 3D LDCT imaging, which combine PWLS reconstruction with **regularization based on learned sparsifying transforms**.
  - Both PWLS-ST and PWLS-ULTRA significantly improve the reconstruction quality compared to PWLS-EP.
  - The ULTRA scheme with a richer union of transforms model provides better reconstruction of various features such as bones, specific soft tissues, and edges, compared to a single transform model.
- **Future Work**
  - Convergence guarantees and automating the parameter selection.
  - New transform learning-based LDCT reconstruction methods, such as involving rotationally invariant transforms, or online transform learning[8], etc.

[8] [Ravishankar et al., IEEE J-STSP, 2015]

Thanks for your attention!

While the total runtime for the 200 iterations (using a machine with two 2.80 GHz 10-core Intel Xeon E5-2680 processors) was 110 minutes for PWLS-DL, it was only 56 minutes for PWLS-ST and 60 minutes for PWLS-ULTRA($K = 15$).