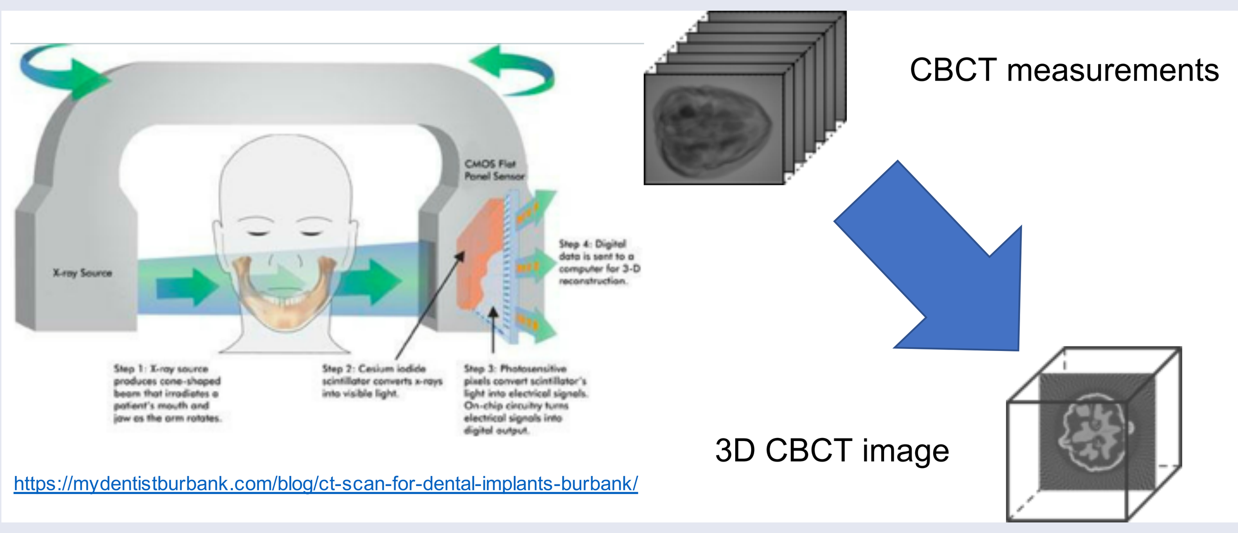
Stable Optimization for Large Vision Model Based Deep Image Prior in Cone-Beam CT Reconstruction

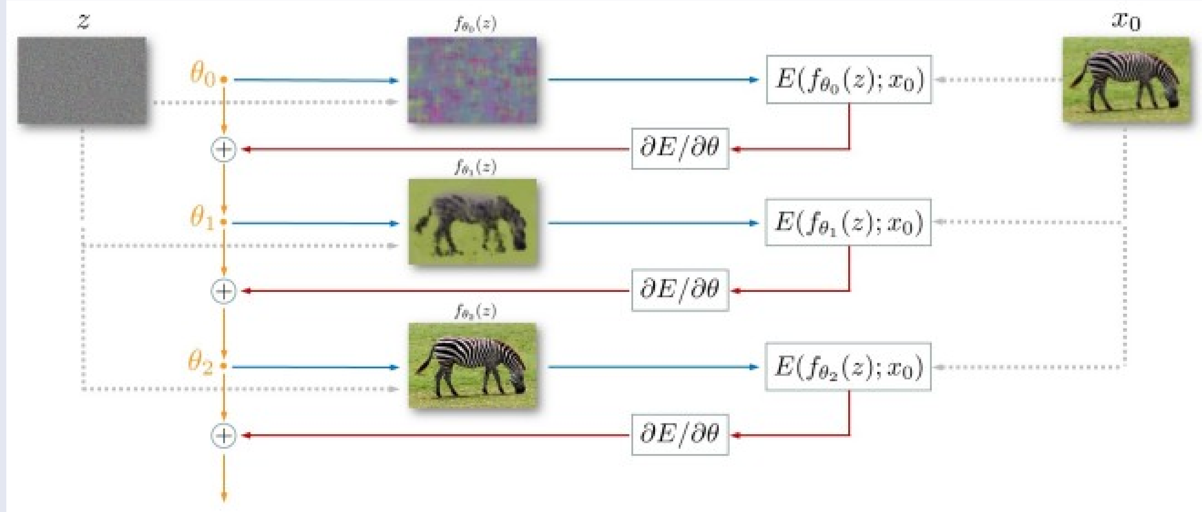
Minghui Wu[1], Yangdi Xu[1], Yingying Xu[1], Guangwei Wu[1], Qingqing Chen[2], Hongxiang Lin[1,*]

[1] Zhejiang Lab, China, [2] Sir Run Run Shaw Hospital, Zhejiang University College of Medicine, China

Background



- Cone-Beam Computed Tomography (CBCT) obtains 3D tomographic images at an equivalent radiation dose but with faster data acquisition process. (left)
- Deep image prior (DIP) can generate a high quality image for CBCT in a neural representation. (right, Ulyanov *et al.* IJCV 2020.)



Problem

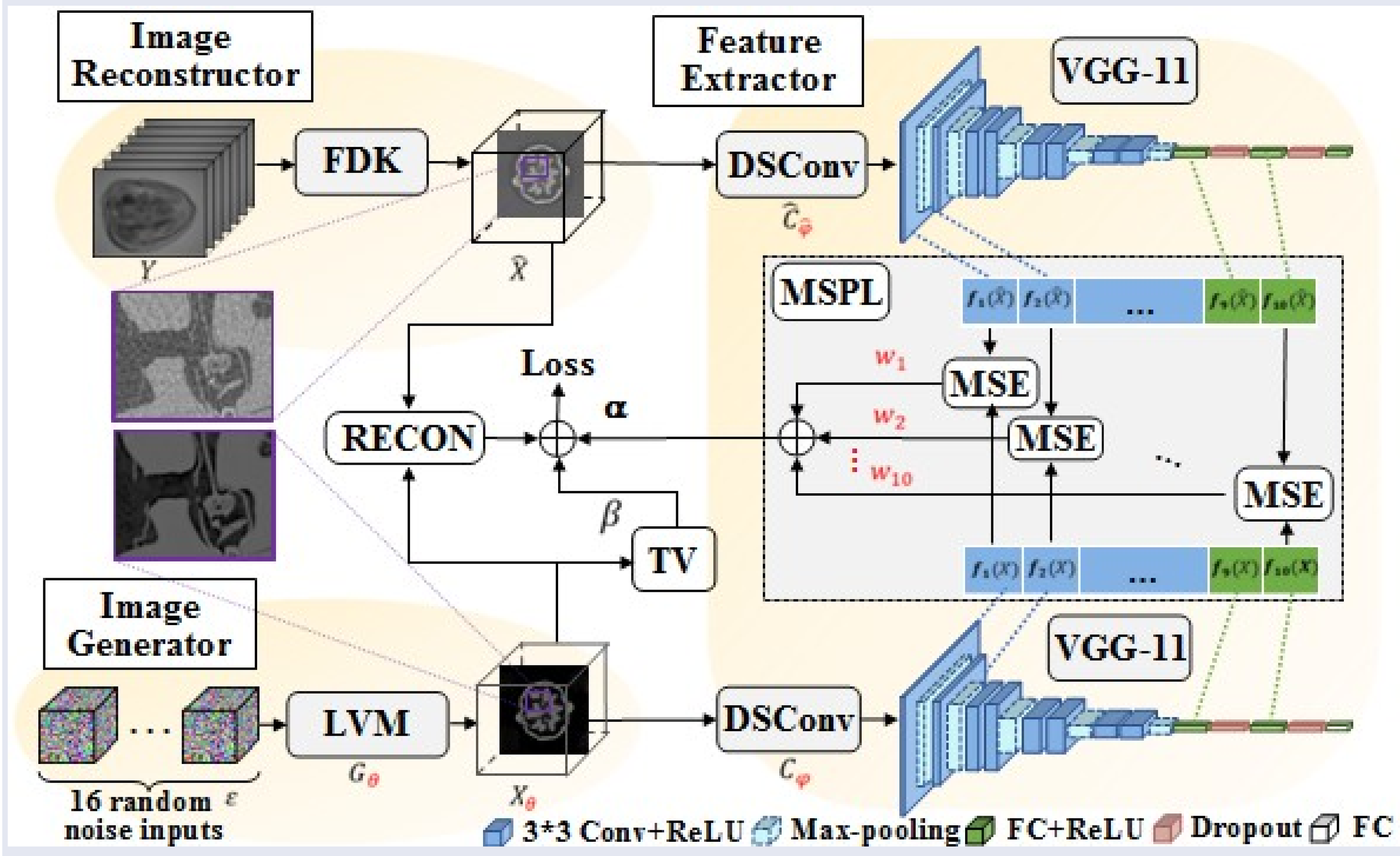
We propose an unsupervised forward-model-free large vision model (LVM)-based DIP for CBCT reconstruction without the need of large number of training data. But it was an open challenging:

- DIP requires a well-defined forward model.
- The classical DIP was expected to increase its model capacity and to apply to LVM.
- LVM having the transformer module is usually hard to converge.

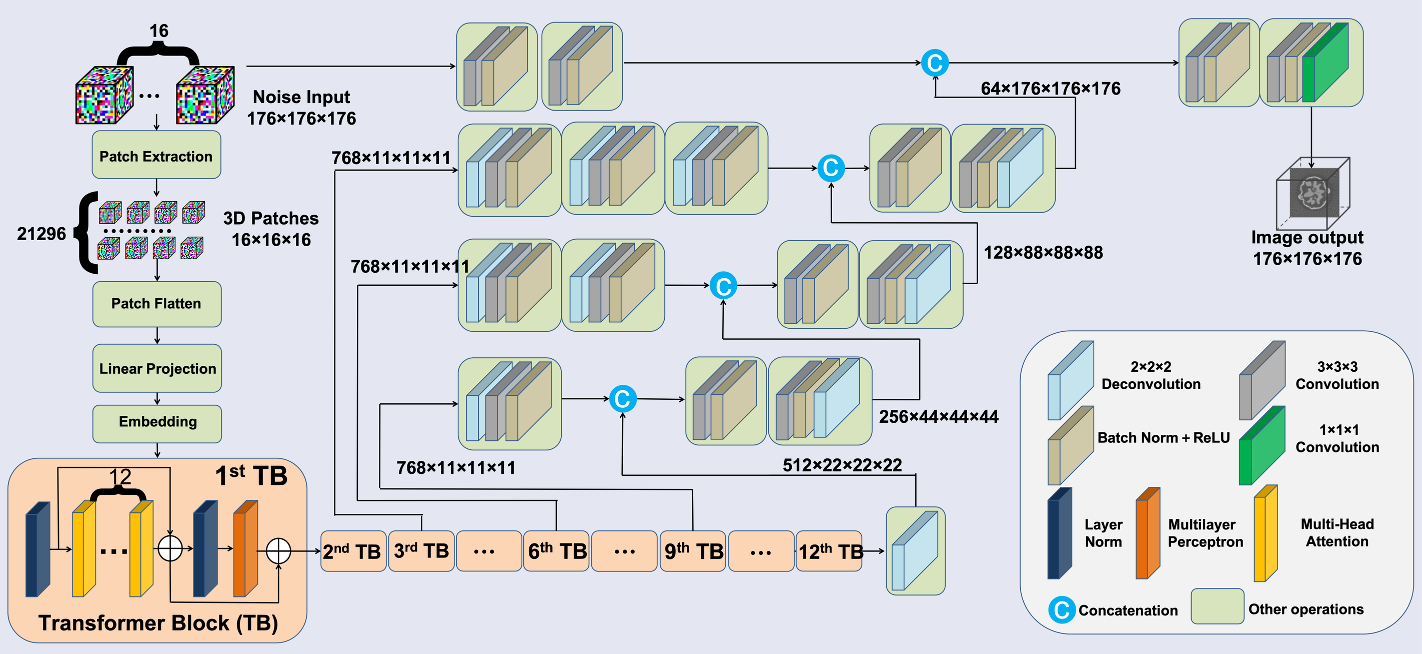
Main Contribution: Stable Optimization

- We derive the first DIP method with an LVM backbone for 3D CBCT.
- We devise the multi-scale perceptual loss (MSPL), measures the similarity of perceptual features between the reference and output images at multiple resolutions without the need for any forward model.
- The reweighting mechanism which stabilizes the iteration trajectory of MSPL.

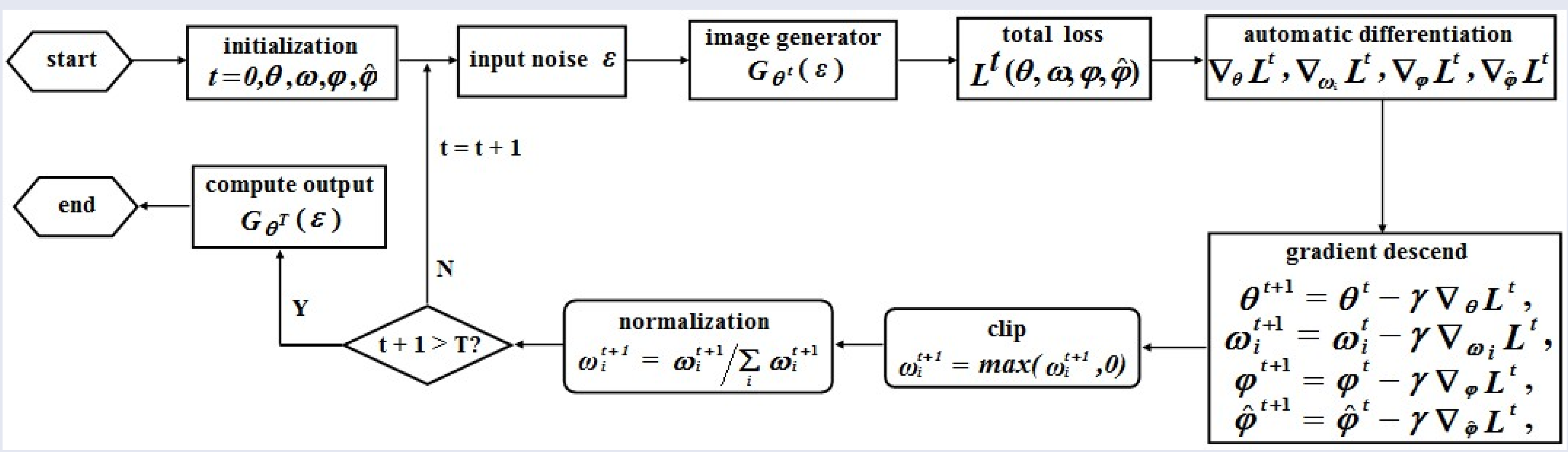
Flowchart of Ours



LVM Architecture: A Modified 3D UNETR



Reweighting Mechanism by One-Shot Optimization

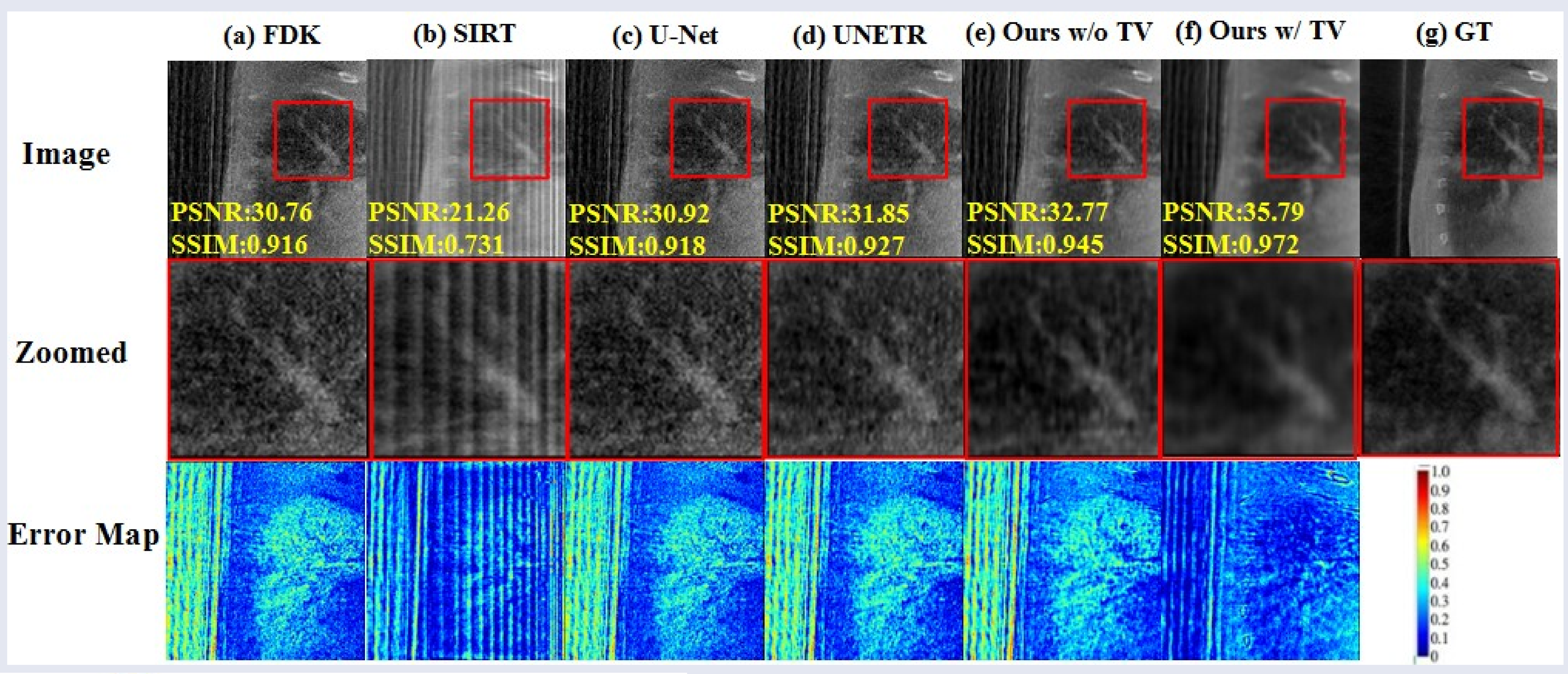


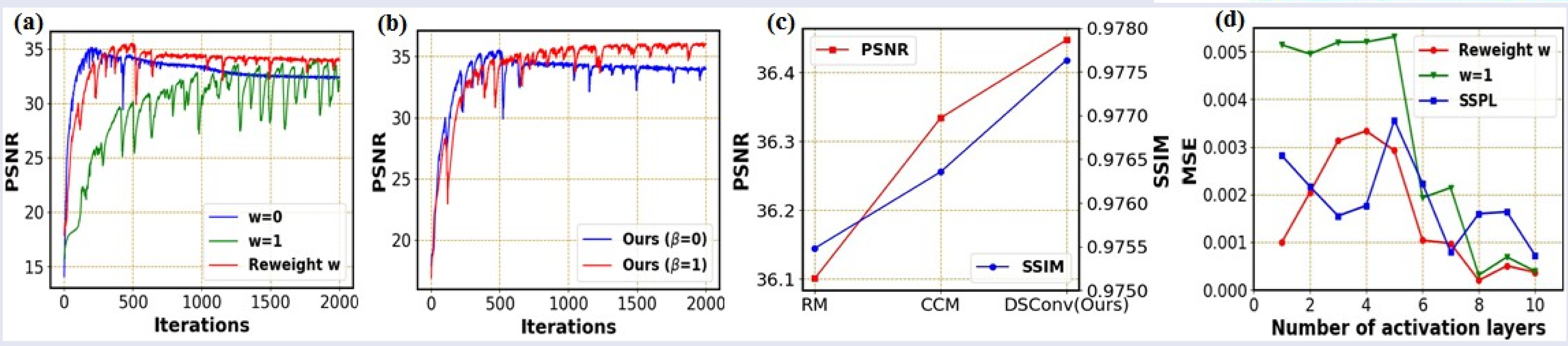
The clip/gradient clip and the normalization on w avoid the negative loss phenomenon.

Experiment and Result

- Models for comparison: two model-based FDK and SIRT, the original DIP using 3D U-Net and UNETR, and ours (+MSPL) with and without TV penalty.
- The SPARE and the Walnut dataset.

Dataset	SPARE		Walnut	
Metric	PSNR (dB)	SSIM	PSNR (dB)	SSIM
FDK	31.15 ± 0.31	0.929 ± 0.010	40.74 ± 1.08	0.977 ± 0.008
SIRT	22.12 ± 0.69	0.753 ± 0.025	30.75 ± 1.97	0.873 ± 0.031
3D U-Net	31.34 ± 0.32	0.931 ± 0.010	41.41 ± 1.19	0.979 ± 0.008
UNETR	32.85 ± 0.93	0.947 ± 0.018	40.91 ± 0.91	$\mathbf{0.982} \pm 0.005$
Ours	$\mathbf{36.47} \pm 0.58^*$	$\mathbf{0.977} \pm 0.005^*$	$\mathbf{43.02} \pm 1.30^*$	$\mathbf{0.982} \pm 0.007$





- (a) 3 different weighting for MSPL;
- (b) W/ and w/o TV while using Reweight;
- (c) 3 downsampling operations: Resampling method (RM), center-clipping method (CCM), and DSConv;
- (d) Evaluation of representative features corresponding to activation layers.

Conclusion

- Ours is a novel forward-model-free LVM-based DIP framework with MSPL for sparse-view 3D CBCT reconstruction using the reweighting strategy.
- Quantitative/qualitative evaluations demonstrate ours effectively enhances the reconstructed image quality and ensures the convergence to the full-view GT image.

Corresponding Author: harryhxlin@gmail.com