DL-Recon: Combining 3D Deep Learning Image Synthesis and Model Uncertainty with Physics-Based Image Reconstruction

Xiaoxuan Zhang,[1] Pengwei Wu,[1] Wojciech B. Zbijewski,[1] Alejandro Sisniega,[1] Runze Han,[1] Craig K. Jones,[2] Prasad Vagdargi,[2] Ali Uneri,[1] Patrick A. Helm,[3] William S. Anderson,[4] Jeffrey H. Siewerdsen[1,2,4]

[1] Department of Biomedical Engineering, Johns Hopkins University, Baltimore MD
[2] Department of Computer Science, Johns Hopkins University, Baltimore MD
[3] Medtronic Inc., Littleton MA
[4] Department of Neurosurgery, Johns Hopkins Medical Institute, Baltimore MD

ABSTRACT

High-precision image-guided neurosurgery – especially in the presence of brain shift – would benefit from intraoperative image quality beyond the conventional contrast-resolution limits of cone-beam CT (CBCT) for visualization of the brain parenchyma, ventricles, and intracranial hemorrhage. Deep neural networks for 3D image reconstruction offer a promising basis for noise and artifact reduction, but generalizability can be challenged in scenarios involving features previously unseen in training data.

We propose a 3D deep learning reconstruction framework (termed "DL-Recon") that integrates learning-based image synthesis with physics-based reconstruction to leverage strengths of each. A 3D conditional GAN was developed to generate synthesized CT from CBCT images. Uncertainty in the synthesis image was estimated in a spatially varying, voxel-wise manner via Monte-Carlo dropout and was shown to correlate with abnormalities or pathology not present in training data. The DL-Recon approach improves the fidelity of the resulting image by combining the synthesized image ("DL-Synthesis") with physics-based reconstruction (filtered back-projection (FBP) or other approaches) in a manner weighted by uncertainty – i.e., drawing more from the physics-based method in regions where model uncertainty is high.

The performance of image synthesis, uncertainty estimation, and DL-Recon was investigated for the first time in real CBCT images of the brain. Variable input to the synthesis network was tested – including uncorrected FBP and precorrection with a simple (constant) scatter estimate – hypothesizing the latter to improve synthesis performance. The resulting uncertainty estimation was evaluated for the first time in real anatomical features not included in training (abnormalities and brain shift). The performance of DL-Recon was evaluated in terms of image uniformity, noise, and soft-tissue contrast-to-noise ratio in comparison to DL-Synthesis and FBP with a comprehensive artifact correction framework. DL-Recon was found to leverage the strengths of the learning-based and physics-based reconstruction approaches, providing a high degree of image uniformity similar to DL-Synthesis while accurately preserving soft-tissue contrast as in artifact-corrected FBP.

Keywords: Cone-beam CT, deep learning, artifact correction, image-guided intervention, image synthesis

1. INTRODUCTION

Neurosurgical approaches to cancer, trauma, or neuro-degenerative disease require a high degree of geometric precision to safely avoid vessels and eloquent brain and achieve effective treatment. The state of the art in intraoperative cone-beam CT (CBCT) is sufficient for visualization and registration of high-contrast objects (e.g., bone, surgical instruments), but it does not provide contrast resolution suitable to soft-tissue, brain parenchyma, or intracranial hemorrhage. Factors limiting CBCT image quality include image biases (e.g., scatter, beam hardening) and quantum and electronic noise.

Existing methods for improving CBCT image quality include artifact corrections[1] and model-based iterative reconstruction (MBIR)[2] that leverages physical knowledge of the imaging chain and image formation process. Recent developments in deep learning approaches provide another means of mitigating artifacts and reducing noise, including image synthesis from CBCT to approximate diagnostic-quality CT.[3] Such approaches offer improvements in computational runtime compared to MBIR, but the performance of image synthesis is subject to uncertainties arising from features not present in training (e.g., pathology, anatomical variations, and unmodeled imaging conditions). The fidelity of the synthesized image hence cannot be guaranteed.[4]

Recognizing the potential pitfalls in generalizability of image synthesis to highly variable anatomical structures in image-guided surgery, we propose a deep learning reconstruction framework (referred to as "DL-Recon") that integrates image synthesis with physics-based reconstruction mediated by model uncertainty. Previous work[5] proposed a 2D U-Net for image synthesis and combined the result with FBP and MBIR reconstruction via model uncertainty in simulation studies. In this work, we developed a 3D generative adversarial network (GAN) for image synthesis and evaluated the performance of DL-Recon for the first time in real CBCT images, including anatomical abnormalities unseen in training data.

2. METHODS

A. Image synthesis and uncertainty estimation

A 3D conditional GAN was developed for CBCT-to-CT image synthesis. For training (Section II.C), a high-fidelity, physics-based forward projection framework (including an accurate beam model, absorption / scatter characteristics, and model of the imaging chain) was used to generate simulated CBCT images from corresponding CT images. Two alternative inputs to the synthesis network were investigated: (i) an uncorrected FBP (μ_{uncorr}^{CBCT}), and (ii) a precorrected FBP ($\mu_{precorr}^{CBCT}$) for which a simple (constant) scatter correction was applied, hypothesizing that the precorrection to improve synthesis performance.

As illustrated in Fig. 1, a 3D GAN was implemented with a U-Net with a residual block at each level of the encoding / decoding path as the generator, and a convolutional pixel-wise classifier[6] as the discriminator. The objective function combined GAN and L1 loss as follows:

$$\hat{G} = \arg \min_{G} \max_{D} \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \tag{1}$$

where

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}[\log D(\mu^{CBCT}, \mu^{CT})] + \mathbb{E}\left[\log\left(1 - D\left(\mu^{CBCT}, G(\mu^{CBCT})\right)\right)\right] \tag{2}$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}[\|\mu^{CT} - G(\mu^{CBCT})\|_1] \tag{3}$$

G and D denote the generator and discriminator, and μ^{CT} and μ^{CBCT} represent paired CT and CBCT images. The L1 loss helps avoid over smoothing, and the balance between the GAN and L1 loss is controlled by λ.
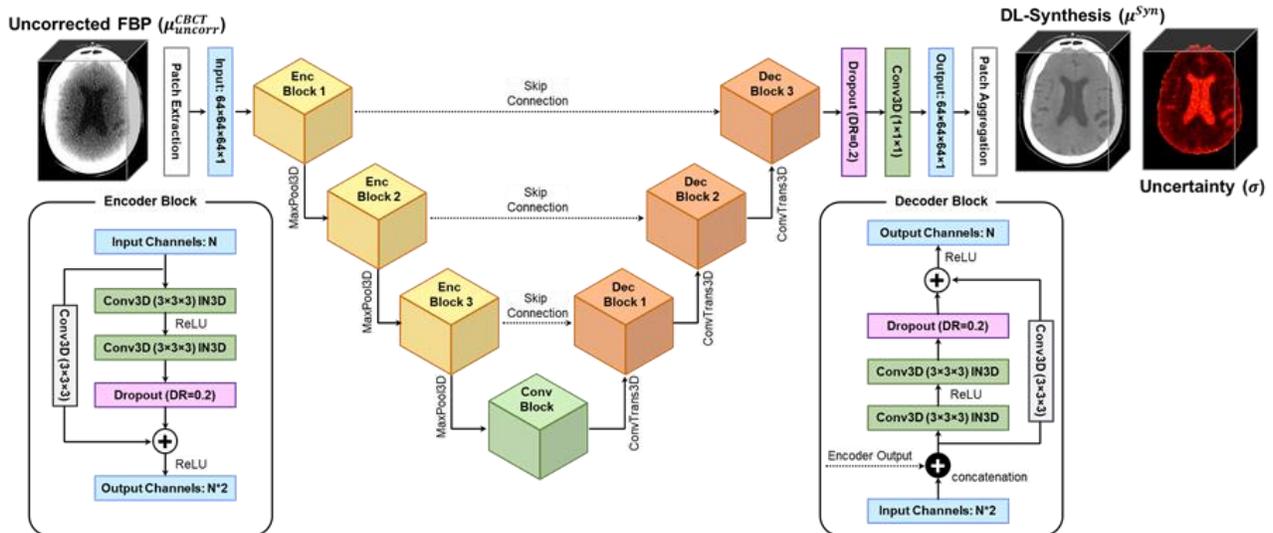
Figure 1. Network architecture of the generator for 3D image synthesis using a conditional GAN. Model uncertainty (σ) in the synthesis is computed via Monte-Carlo dropout layers.

As described in Gal and Ghahramani,[7] dropout applied during network training is equivalent to a Bayesian approximation of the Gaussian process, and uncertainty in the model output can be estimated by computing the voxel-wise variance of multiple forward passes. Following such an approach, we added dropout layers (dropout rate = 0.2) prior to

the skip connection in each encoder and decoder block and to the final output. Both training and inference were performed with dropout. The predictive mean computed from a collection of 8 network outputs yields the synthesized image (DL-Synthesis, μ^{Syn}), and the predictive variance (σ^2) serves as a proxy for model uncertainty.

B. The DL-Recon framework

The proposed method (termed DL-Recon) integrates 3D image synthesis with physics-based reconstruction via uncertainty associated with the synthesis model. The method involves three steps: (i) generation of a 3D synthetic CT image (μ^{Syn}) from a CBCT volume with estimation of model uncertainty (σ) as described above; (ii) physics-based 3D image reconstruction of projection data, including artifact corrections – for example, the pipeline described in [1] – to yield an artifact-corrected CBCT image (denoted μ_{corr}^{CBCT}); and (iii) voxel-wise combination of μ^{Syn} and μ_{corr}^{CBCT} weighted by the estimated uncertainty to yield the DL-Recon image (denoted $\mu^{DL-Recon}$). The resulting image is:

$$\mu^{DL-Recon} = [1 - \beta(\sigma)]\mu^{Syn} + \beta(\sigma)\mu_{corr}^{CBCT} \tag{4}$$

where uncertainty is contained within a spatially varying map (β, with values in the range [0, 1]) related by a sigmoid function:

$$\beta(\sigma) = \frac{1}{1 + e^{-(c_1 \cdot \sigma + c_2)}} \tag{5}$$

where c_1 and c_2 specify the range and level, respectively, of the sigmoid, and β controls the contribution of μ^{Syn} and μ_{corr}^{CBCT} in a voxel-wise manner. When predictive uncertainty is high, the β map draws more from the physics-based reconstruction.

The underlying premise in this approach is that the synthesis image (μ^{Syn}) carries particular benefits (e.g., uniformity and noise reduction) but may be subject to systematic error – for example, in structures unseen in the training data. The uncertainty map [$\sigma(x, y, z)$, alternatively $\beta(x, y, z)$] were shown previously in simulation studies [5] to correlate with deviations from ground truth. The "uncertainty map" therefore offers insight on where the synthesis image may be subject to error and where it is advantageous to draw more from the physics-based 3D image reconstruction (μ_{corr}^{CBCT}).

Note that the physics-based method incorporated in DL-Recon could be FBP or any particular form of MBIR, recognizing that the latter may invite disadvantages of computational load associated with conventional iterative optimization. Alternatively, the synthesis image could be incorporated as a prior within a penalized optimization, as in [5]. In any of these scenarios, the voxel-wise weighting of synthesis and physics-based image reconstructions is intended to leverage the strengths of each, mediated by the model uncertainty. In the work reported below, DL-Recon incorporates (artifact-corrected) FBP reconstruction as a practical implementation that may be compatible with the rapid runtime requirements of image-guided surgery, focusing here on intracranial neurosurgery.

C. Training data generation

To obtain a large training dataset of matched CT and CBCT images, CBCT projection data were simulated from 35 real, helical CT volumes of 35 healthy subjects using a high-fidelity forward projector [5]. CBCT system geometry and image acquisition were simulated to match data (~745 views over 360°) acquired from the O-arm (the O-arm[TM] "O2" imaging system, Medtronic) using nominal head scan protocols (100–120 kV and 75–240 mAs). Volumes were reconstructed with isotropic 0.7 mm voxels via FBP without artifact correction. Signal normalization linearly transformed the CBCT intensity histogram within the brain parenchyma to [-1, 1]. Volumetric patches (64×64×64 voxels) were stochastically sampled from the brain volume and fed to the network, and a total of 875 patches were used for training. The Adam optimizer (learning rate = 5×10^{-5}, $\beta_1 = 0.5$, $\beta_2 = 0.999$, L1 regularization λ=100, and batch size = 2) was used and early stopping at 800 epochs was applied.
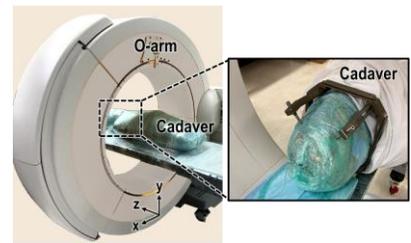
Figure 2. Experimental setup for cadaver studies using the O-arm.

D. Experimental studies

D.1. Image synthesis of simulated and real brain CBCT images

The proposed image synthesis method was validated on both simulated and real CBCT data. Simulated CBCT projections of 5 test CT volumes were generated and reconstructed in the same manner as the training set. Intensity

differences between synthesized images and ground truth were measured within the brain region for each volume. Experiments were conducted using the O-arm™ system illustrated in Fig. 2. Real projection data for 3 cadaveric heads (denoted below as cadaver #1-3) were collected at 120 kV and 150 mAs. Volumetric images were reconstructed on a grid of 320×320×280 voxels with isotropic 0.7 mm voxels. The runtime of DL-Synthesis was ~1 min per prediction (NVIDIA TITAN Xp). DL-Synthesis images were evaluated with uncorrected CBCT as input (denoted μ_{uncorr}^{Syn}) and with a basic (constant-scatter) precorrection (denoted $\mu_{precorr}^{Syn}$). Method performance was quantified in terms of image non-uniformity (NU), the difference in mean voxel value between region of interests (ROIs) in the parenchyma near the dural surface / sphenoid bone and about the lateral ventricles.

D.2 Uncertainty estimation in real anatomical abnormalities

Previous work [5] has shown correlation between synthesis error and uncertainty for simulated lesions (not exist in the training cohort) of difference location, size, and contrast. In this work, the accuracy of uncertainty estimation was evaluated in cadaver images, including specimens exhibiting true abnormalities that were not present in the training data. Specifically, abnormalities included a large intraparenchymal calcification, a loss of cerebrospinal fluid, and brain shift in which the brain cortex collapsed from the interior surface of the cranium.

D.3 Cadaver studies on an intraoperative CBCT system

Imaging performance was evaluated in terms of visual image quality as well as image uniformity, noise, and soft-tissue contrast-to-noise ratio (CNR) in cadavers imaged on the O-arm™ system (Fig. 2). FBP reconstructions were evaluated with and without artifact correction. DL-Recon was evaluated in comparison to FBP and DL-Synthesis, and uncertainty maps were displayed to understand how physics-based and deep learning-based approaches contributed to the final result.

3. RESULTS

A. Performance of image synthesis

Fig. 3 shows results of image synthesis on simulated data (high-fidelity CBCT projections generated from CT). DL-Synthesis demonstrated good overall correspondence with the ground truth CT, yielding high image uniformity and reduced noise compared to the uncorrected FBP image. In 5 test volume images, DL-Synthesis exhibited a difference in overall mean intensity (in the brain) of less than 1 HU (compared to > 12 HU for FBP) to the ground truth, with residual differences owing mainly to image noise. The estimated uncertainty highlights regions with anatomical variations such as the lateral ventricles and sulci in the cerebral cortex, which is susceptible to error (e.g., contrast loss) in the synthesis mage.

Fig. 4 illustrates the performance of image synthesis on real data, in which the input to the synthesis network was either uncorrected or precorrected image data. DL-Synthesis acting on uncorrected FBP input exhibits performance degradation in regions affected by severe artifacts, yielding a higher degree of non-uniformity near the sphenoid bone (yellow arrow). A simple (constant) scatter correction was shown to partially account for biases that were not modeled by the forward projector (e.g., variation in bone density) and improve the overall image uniformity (2–4 HU). As a result, precorrected FBP yielded more accurate synthesis, reducing image NU by ~50% compared to synthesis acting on uncorrected FBP. However, DL-Synthesis exhibited a loss in contrast in structures such as the lateral ventricles (cadaver #1, magenta arrows), demonstrating potential pitfalls in the generalizability of image synthesis to real and highly variable image data.
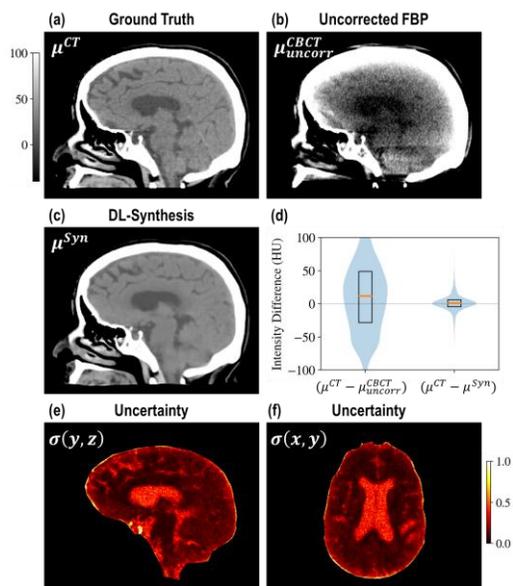
Figure 3. Synthesis performance in simulated CBCT data. (a) Sagittal slice of a test CT image volume. (b) Corresponding CBCT reconstruction (network input). (c) Resulting synthesized image. (d) Violin plot quantifying the respective difference in voxel values of uncorrected FBP and DL-Synthesis to the ground truth measured for 5 test data. (e) Sagittal and (f) axial slice of the estimated uncertainty.
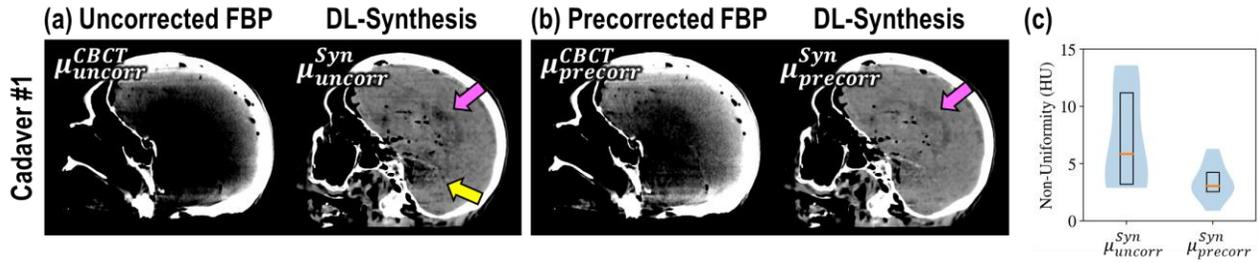
Figure 4. Synthesis performance for (a) uncorrected and (b) precorrected FBP of real CBCT images. (c) Boxplot quantifying image non-uniformity in synthesized images of 3 cadavers.

B. Uncertainty estimation in cadaver studies

Fig. 5 demonstrates the performance of uncertainty estimation on real data with unseen features (calcium deposit in cadaver #2 and brain shift in cadaver #3). For both cases, the uncertainty map highlights the location of the unseen structure as well as at the lateral ventricles, suggesting a lack of reliability in the synthesis result and the need for input from physics-based reconstruction.
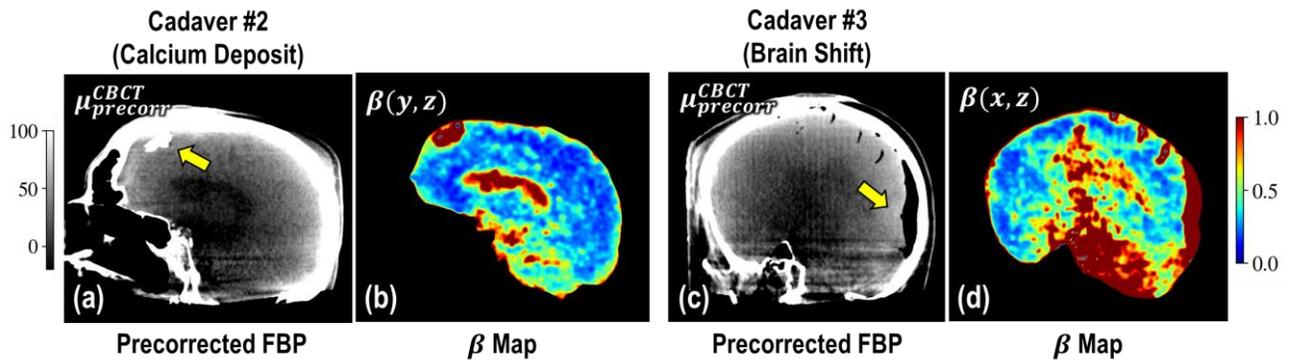
Figure 5. Uncertainty estimation in cadaver CBCT head images. Precorrected FBP and the estimated uncertainty (β map) within the brain parenchyma for (a-b) cadaver #1 with calcium deposit and (c-d) cadaver #3 with brain shift.

C. Performance of DL-Recon

Fig. 6 shows reconstructed images from conventional methods (FBP and DL-Synthesis) and the proposed DL-Recon framework. As shown in Fig. 6(b), the comprehensive artifact correction pipeline reduced NU by 59%, but led to 38% increase in image noise. DL-Synthesis yielded the lowest NU value and noise but suffered from loss in soft-tissue contrast. In comparison, DL-Recon was able to reduce both NU and noise while preserving image contrast of the ventricles, providing ~15% increase in soft-tissue CNR compared to fully corrected FBP.

The intensity profile of a curve across the brain [yellow dashed curve shown in Fig. 6(b)] was plotted in Fig. 7 for fully corrected FBP, DL-Synthesis, and DL-Recon. Fully corrected FBP exhibited residual nonuniformity, especially just inside the cranium due to residual beam-hardening effects, as indicated by the nonuniform intensity profile between the ventricle and cranium. DL-Synthesis improved uniformity in these regions but reduced the contrast in the ventricle, similar to the effects shown above in relation to model uncertainty. By comparison, DL-Recon maintained the benefits of image uniformity from DL-Synthesis while achieving contrast in the ventricles similar to the fully corrected FBP.
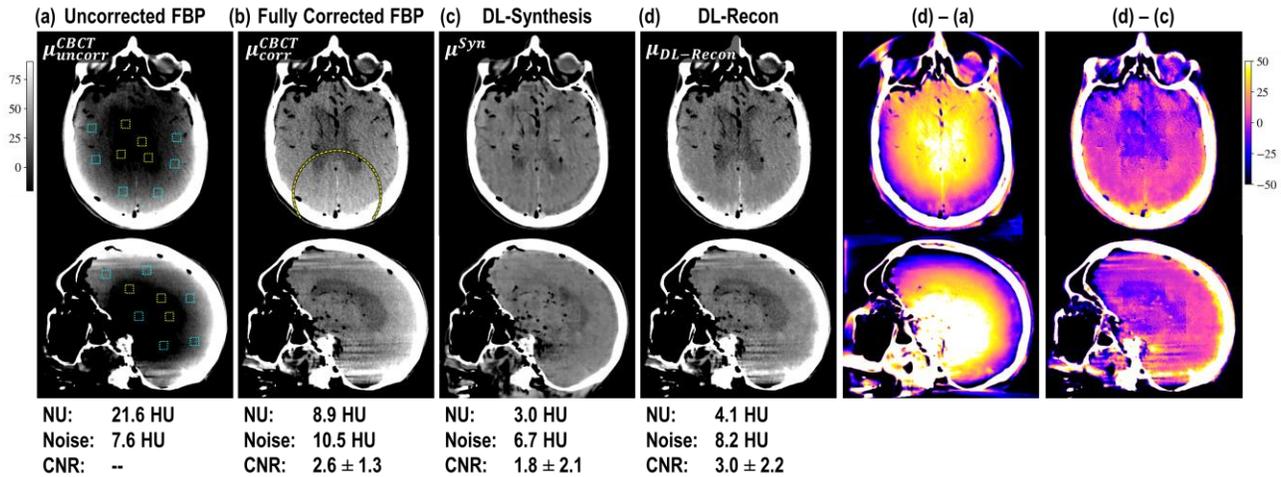
Figure 6. Example axial and sagittal slices of FBP, DL synthesis, and DL-Recon in cadaver CBCT data. Measurements of image non-uniformity (NU), noise, and contrast-to-noise ratio (CNR) of the lateral ventricles are listed below each image. Difference images show the contributions of the physics-based [(d)-(a)] and image synthesis [(d)-(c)] methods to the DL-Recon image (approximate Hounsfield Units).
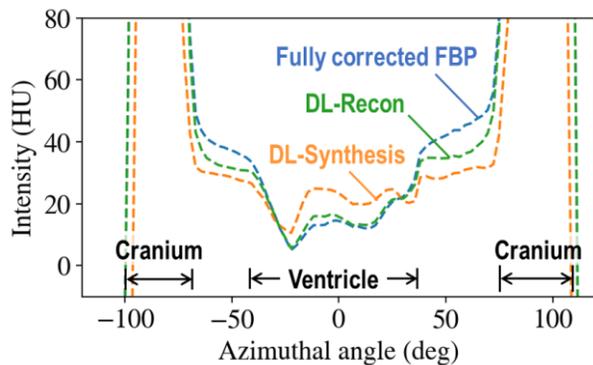
Figure 7. Intensity profiles of FBP, DL-Synthesis, and DL-Recon. The DL-Recon image leverages the improved uniformity of DL-Synthesis (region just inside the cranium) and the improved (accurate) contrast of fully corrected FBP (in the ventricles).

REFERENCES

1. A. Sisniega et al., "High-fidelity artifact correction for cone-beam CT imaging of the brain," Phys. Med. Biol. **60**(4), 1415–1439 (2015) [doi:10.1088/0031-9155/60/4/1415].
2. I. A. Elbakri and J. A. Fessler, "Statistical Image Reconstruction for Polyenergetic X-Ray Computed Tomography," IEEE Trans. Med. Imaging **21**(2), 89–99 (2002) [doi:10.1109/42.993128].
3. X. Liang et al., "Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy," Phys. Med. Biol. **64**(12), IOP Publishing (2019) [doi:10.1088/1361-6560/ab22f9].
4. Q. Yang et al., "Low Dose CT Image Denoising Using a Generative Adversarial Network with Wasserstein Distance and Perceptual Loss," IEEE Trans Med Imaging **37**(6), 1348–1357 (2018) [doi:10.1109/TMI.2018.2827462.Low].
5. P. Wu et al., "Using Uncertainty in Deep Learning Reconstruction for Cone-Beam CT of the Brain."
6. P. Isola et al., "Image-to-image translation with conditional adversarial networks," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 **2017-January**, 5967–5976 (2017) [doi:10.1109/CVPR.2017.632].
7. Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," 33rd Int. Conf. Mach. Learn. ICML 2016 **3**, 1651–1660 (2016).