Review paper

Artificial intelligence in image reconstruction: The change is here

Ramandeep Singh\textsuperscript{a}, Weiwen Wu\textsuperscript{b}, Ge Wang\textsuperscript{b,1}, Mannudeep K. Kalra\textsuperscript{a,\textastemarrow,1}

\textsuperscript{a} Department of Radiology, Division of Thoracic Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

\textsuperscript{b} Biomedical Imaging Center, Rensselaer Polytechnic Institute, New York, USA

\textsuperscript{\textastemarrow} Corresponding author.

\textsuperscript{1} Equal contributing.

E-mail address: mkalra@mgh.harvard.edu (M.K. Kalra).

Keywords: Artificial Intelligence
Deep learning
Image reconstruction
Computed tomography

\textbf{A B S T R A C T}

Innovations in CT have been impressive among imaging and medical technologies in both the hardware and software domain. The range and speed of CT scanning improved from the introduction of multidetector-row CT scanners with wide-array detectors and faster gantry rotation speeds. To tackle concerns over rising radiation doses from its increasing use and to improve image quality, CT reconstruction techniques evolved from filtered back projection to commercial release of iterative reconstruction techniques, and recently, of deep learning (DL)-based image reconstruction. These newer reconstruction techniques enable improved or retained image quality versus filtered back projection at lower radiation doses. DL can aid in image reconstruction with training data without total reliance on the physical model of the imaging process, unique artifacts of PCD-CT due to charge sharing, K-escape, fluorescence x-ray emission, and pulse pileups can be handled in the data-driven fashion. With sufficiently reconstructed images, a well-designed network can be trained to upgrade image quality over a practical/clinical threshold or define new/killer applications. Besides, the much smaller detector pixel for PCD-CT can lead to huge computational costs with traditional model-based iterative reconstruction methods whereas deep networks can be much faster with training and validation. In this review, we present techniques, applications, uses, and limitations of deep learning-based image reconstruction methods in CT.

1. Introduction

Radiology has witnessed remarkable technological advancements beginning with two-dimensional projection radiography and progressing to three- and four-dimensional, high resolution, modern cross-sectional imaging capabilities. Amongst all imaging technologies, computed tomography (CT) has found an increasingly essential role for screening, diagnosis, and follow-up of diverse clinical conditions. CT has emerged as the imaging modality of choice in several pathologies, including trauma, infectious and oncologic conditions. Since its introduction as a medical imaging technology in 1972, CT software and hardware have been at the forefront of advancements at a rapid pace. As compared to CT scanners of today, past machines were slow since they took a long time for image acquisition and reconstruction, and yet generated poor image resolution \cite{1,2}. Historically, the first CT had acquisition and reconstruction times of several minutes, each with an image resolution of only $80 \times 80$ pixels. The state-of-the-art modern multidetector-row CT scanners can image the whole body in $< 5$ s and the entire chest-abdomen-pelvis in under two seconds for the most demanding patients and clinical settings at gantry rotation times down to 0.25 s and up to 16 cm scan coverage in one gantry rotation. Several scanners enable reconstruction of sub-millimeter section thickness with a matrix of $1024 \times 1024$ and on some scanners as high as $2048 \times 2048$ \cite{2-5}.

The technological revolution in CT can be judged from statistics on its use in modern medicine where some have dubbed CT as the "stethoscope of modern medicine." From about 3 million CT procedures in the year 1980, recent data suggest a staggering number of 88.7 million CT procedures in the United States alone in the year 2018 \cite{6}. These numbers illustrate a 30-fold increase in the use of CT scanners over the last four decades at an annual growth of 6.5% and accounting for about 245 exams per 1000 population in USA alone \cite{6}. For Europe, the number of CT scans per 1000 population range from 57 in Finland to 218 in Luxembourg, with about 13/27 countries reporting number above 100 per 1000 population \cite{7}.

In the late 1990 s, the increasing use of CT raised concerns over a substantial contribution of CT towards medical radiation dose from CT \cite{8}. The scientific community and the lay press focused on the potential...
of radiation-induced cancer from CT. Both justification of CT and optimization of associated radiation dose are important to address these concerns. CT does provide essential clinical information that helps in diagnosis, staging, and management decisions. Still, its overuse or inappropriate use is associated with unnecessary radiation dose and the risk of detecting incidental “innocent” findings, which can trigger further imaging and invasive interventions. For reduction and optimization of radiation dose with CT, it is imperative to ensure that the anticipated benefits of information obtained from CT outweigh the projected risks of radiation dose on an individual patient basis. When appropriate, the risk of radiation exposure can be mitigated with the use of alternative imaging modalities such as ultrasound and MRI, particularly in females of childbearing age group and children. Following justification of CT, dose optimization or reduction needs to be managed carefully, as for a CT performed with the same scan acquisition and reconstruction techniques, efforts to lower radiation dose can increase image noise and thus reduce image quality [9-11]. Images with little or no noise often involve a higher radiation dose without additional diagnostic information. Conversely, CT images acquired with too much noise or artifacts can decrease the diagnostic content. Within each body region, the amount of tolerable image noise depends on the clinical indication and the type of abnormality. For example, abdominal CT for renal colic can be performed at a lower radiation dose since high image noise does not impair the visibility of renal calculi. Abdominal CT for detection, staging, and characterization of low-contrast, focal liver or pancreatic lesions require images with lower noise and higher radiation dose compared to renal colic CT. Such variations in clinical indications and needs should be exploited to optimize or personalize radiation dose in CT for all body regions.

CT must be performed with the ALARA (as low as reasonably achievable) dose principle, which should involve setting the correct scan protocols based on the body region, body habitus, and most importantly, the clinical indication for performing the imaging study [12]. A multi-pronged approach is essential for radiation dose optimization to maintain acceptable image quality for clinical interpretation [9]. Broadly, this approach can be categorically summarized into patient and scanner dependent factors. It is important to establish body region and clinical indication-specific CT protocols. Within each protocol, appropriate use of customizable scan parameters with experience and expertise can help optimize radiation levels for the patient. For the scanner dependent factors, different CT vendors have improvised technologies to optimize radiation dose while maintaining or improving image quality by modifying scan and reconstruction parameters such as tube current (mA) and voltage (kV), detector configuration, gantry rotation time, number of gantry rotations, size of image voxel, slice thickness, beam filters, reconstruction kernels, and image reconstruction techniques [13-15].

Recent key innovation focus in CT has been towards image reconstruction, that advanced from the first algebraic reconstruction (AR) method of the 1970s to conventional and filtered back projection (FBP) techniques of 1980s to 2000s, and iterative reconstruction technique (I-RT) since late 2000s [16,17]. Although iterative reconstruction techniques allow users to reduce radiation dose compared to FBP while maintaining or improving image quality, there are concerns over the loss of conspicuity of low-contrast abnormalities and substantial changes in image appearance or texture with I-RT. In recent years, artificial intelligence (AI) has opened new possibilities in medicine and medical imaging [18]. Amongst the applications of AI in radiology, deep learning-based image reconstruction techniques have shown promise to lower radiation dose to unprecedented levels. In this review, we discuss techniques, applications, uses, and limitations of deep learning-based image reconstruction techniques in CT from a radiation dose optimization perspective.

2. CT radiation risks and optimization

Overall radiation dose due to medical exposure increased from 15% of total radiation exposure in the 1980 s to more than 50% in 2006 in USA and levelled off from 2006 to 2016. [19,20]. According to a prediction by Brenner et al. in 2007, about 1–2% of all cancers can be potentially caused in patients undergoing CT scans [21]. However, most such predictions are generated from the controversial dose–response curve based on linear no-threshold (LNT) relationship between low-radiation dose and carcinogenesis [22,23]. Detractors of LNT refer to effective biologic defenses against carcinogenesis from low radiation doses such as those from most diagnostic CT procedures. In fact, there is little evidence to support carcinogenesis in humans at radiation dose under 100 mSv [23]. Despite the controversies, per Biological Effects of Ionizing Radiation study (BEIR VII), radiation exposure and risk of cancer are directly proportional [24]. Two potentially vulnerable populations to radiation risks include younger patients who have higher cell turnover and longer life duration after exposure and females where certain cancers have a higher propensity to occur [25]. A recent study with greater than 160,000 children on the risk for leukemia and brain tumor risk following low-dose ionizing radiation from CT reported an increase in risk of brain tumor [26].

To optimize radiation dose and reduce its associated risks, it is important to understand how key patient-related factors affecting radiation dose are such as body habitus, scanned body part and clinical indication for CT scan influence the required diagnostic quality and radiation dose [8,27,28]. To maintain a constant image quality with identical image reconstruction technique, larger adults require scanning at higher energy output settings of the x-ray tube as compared to smaller adults and children. Body parts such as chest have higher contrast and cause less x-ray beam attenuation compared to the abdomen, and therefore chest CT should be scanned at a lower dose compared to abdomen CT. Clinical indications for imaging also determine the amount of CT radiation dose. High contrast conditions such as lung nodule follow-up and renal stone evaluation can be evaluated at a low radiation dose [29]. Even with the same clinical indication, large body habitus, or longer scan length of a clinically indicated body region can increase radiation dose. Another factor linked to clinical indications is imaging with multiple scan phases or for lesions that require better spatial resolution, both of which can compromise the extent to which radiation dose can be reduced [9,29-32].

Besides patient factors, users must also have an in-depth knowledge of how scanner technology and scan factors can be used to optimize radiation dose. Tube current and potential represent key scan factors when it comes to protocol and radiation dose optimization for CT. The radiation dose administered to the patient is directly proportional to tube current (milliampere) and tube current–time product (milliampere-seconds), modification of which remains the most common method for dose reduction [33,34]. These can either be automatically modified or by manual adjustment or selection of fixed values. Compared to manual selection of fixed tube current, the automatic techniques enable users to optimize radiation dose based on specified image quality parameters and regional attenuation of the patient in the body region being imaged. However, users must set the image quality parameters carefully since an inadvertent or intentional use of high image quality requirements can increase the radiation dose compared to the fixed tube current technique.

Compared to tube current, tube potential has a more profound effect on both radiation dose as well as CT attenuation numbers (HU) of iodinated contrast [35-38]. The automatic tube potential selection techniques differed on whether tube current dose to use for a given indication and the type of abnormality. For example, abdominal CT. Clinical indications for imaging also determine the amount of total radiation exposure in the 1980 s to more than 50% in 2006 in USA and levelled off from 2006 to 2016. [19,20]. According to a prediction by Brenner et al. in 2007, about 1–2% of all cancers can be potentially caused in patients undergoing CT scans [21]. However, most such predictions are generated from the controversial dose–response curve based on linear no-threshold (LNT) relationship between low-radiation dose and carcinogenesis [22,23]. Detractors of LNT refer to effective biologic defenses against carcinogenesis from low radiation doses such as those from most diagnostic CT procedures. In fact, there is little evidence to support carcinogenesis in humans at radiation dose under 100 mSv [23]. Despite the controversies, per Biological Effects of Ionizing Radiation study (BEIR VII), radiation exposure and risk of cancer are directly proportional [24]. Two potentially vulnerable populations to radiation risks include younger patients who have higher cell turnover and longer life duration after exposure and females where certain cancers have a higher propensity to occur [25]. A recent study with greater than 160,000 children on the risk for leukemia and brain tumor risk following low-dose ionizing radiation from CT reported an increase in risk of brain tumor [26].

To optimize radiation dose and reduce its associated risks, it is important to understand how key patient-related factors affecting radiation dose are such as body habitus, scanned body part and clinical indication for CT scan influence the required diagnostic quality and radiation dose [8,27,28]. To maintain a constant image quality with identical image reconstruction technique, larger adults require scanning at higher energy output settings of the x-ray tube as compared to smaller adults and children. Body parts such as chest have higher contrast and cause less x-ray beam attenuation compared to the abdomen, and therefore chest CT should be scanned at a lower dose compared to abdomen CT. Clinical indications for imaging also determine the amount of CT radiation dose. High contrast conditions such as lung nodule follow-up and renal stone evaluation can be evaluated at a low radiation dose [29]. Even with the same clinical indication, large body habitus, or longer scan length of a clinically indicated body region can increase radiation dose. Another factor linked to clinical indications is imaging with multiple scan phases or for lesions that require better spatial resolution, both of which can compromise the extent to which radiation dose can be reduced [9,29-32].

Besides patient factors, users must also have an in-depth knowledge of how scanner technology and scan factors can be used to optimize radiation dose. Tube current and potential represent key scan factors when it comes to protocol and radiation dose optimization for CT. The radiation dose administered to the patient is directly proportional to tube current (milliampere) and tube current–time product (milliampere-seconds), modification of which remains the most common method for dose reduction [33,34]. These can either be automatically modified or by manual adjustment or selection of fixed values. Compared to manual selection of fixed tube current, the automatic techniques enable users to optimize radiation dose based on specified image quality parameters and regional attenuation of the patient in the body region being imaged. However, users must set the image quality parameters carefully since an inadvertent or intentional use of high image quality requirements can increase the radiation dose compared to the fixed tube current technique.

Compared to tube current, tube potential has a more profound effect on both radiation dose as well as CT attenuation numbers (HU) of iodinated contrast [35-38]. The automatic tube potential selection techniques differed on whether tube current dose to use for a given indication and the type of abnormality. For example, abdominal CT. Clinical indications for imaging also determine the amount of total radiation exposure in the 1980 s to more than 50% in 2006 in USA and levelled off from 2006 to 2016. [19,20]. According to a prediction by Brenner et al. in 2007, about 1–2% of all cancers can be potentially caused in patients undergoing CT scans [21]. However, most such predictions are generated from the controversial dose–response curve based on linear no-threshold (LNT) relationship between low-radiation dose and carcinogenesis [22,23]. Detractors of LNT refer to effective biologic defenses against carcinogenesis from low radiation doses such as those from most diagnostic CT procedures. In fact, there is little evidence to support carcinogenesis in humans at radiation dose under 100 mSv [23]. Despite the controversies, per Biological Effects of Ionizing Radiation study (BEIR VII), radiation exposure and risk of cancer are directly proportional [24]. Two potentially vulnerable populations to radiation risks include younger patients who have higher cell turnover and longer life duration after exposure and females where certain cancers have a higher propensity to occur [25]. A recent study with greater than 160,000 children on the risk for leukemia and brain tumor risk following low-dose ionizing radiation from CT reported an increase in risk of brain tumor [26].

This unique advantage can be used to lower radiation dose for
contrast-enhanced CT studies such as CT angiography and perfusion exams while improving image quality. However, in larger patients, inter-vendor heterogeneity and limited signal-to-noise ratio on reduced dose CT examinations can at times, result in inadequate visualization of structures [8,9,11].

Other scan factors such as gantry rotation time, pitch, and detector configuration should be selected based on the speed and resolution requirements rather than for radiation dose reduction. For children and patients who are unable to hold still during image acquisition, faster scanning can help reduce motion artifacts and the need for repeat scanning. For such patients, faster gantry rotation speed, higher pitch and/or wider beam collimation can expedite scan duration and help obtain images with less motion artifacts. In large patients, particularly those with morbid obesity, scan duration may need to be prolonged with longer gantry rotation time, and/or slower pitch to deliver adequate radiation dose and obtain diagnostic quality images. Likewise, the choice of beam collimation and detector configuration depends on the scan mode and clinical requirements of section thickness. On older CT scanners with variable thickness of detector rows, wider beam collimation may not provide thin sections, which may be critical for certain scan protocols such as CT angiography, high-resolution chest CT, and temporal bone CT. The choice of optimal detector configuration also depends on the body region and scan length.

3. Image reconstruction techniques

FBP techniques were and on many current scanners remain the preferred method of image reconstruction since discovery and introduction of CT scanning into clinical practice. FBP provides sufficient image quality without a substantial demand on computational power which was at a premium in the last century. However, with low radiation dose CT, image noise and artifacts on FBP images can increase substantially. To address this limitation of FBP at low radiation doses, advances in computational power in the early 2000s led to introduction of I-RT into clinical practice. Since their introduction, I-RT is the mainstay for image reconstruction for all CT protocols and examinations at our and several other sites. Yet, some sites use their concerns over changes in the image appearance or texture with I-RT as a reason to continue with the use of FBP. To address concerns over altered image appearance with I-RT and enable further dose reduction, CT vendors recently introduced DL-based image reconstruction techniques. This section provides an overview of FBP, I-RT, and DL-based methods of image reconstruction in CT.

3.1. Filtered back projection

The standard method of image reconstruction in CT used back-projection of multiple image data at the same angle as it was acquired. However, this method resulted in the blurring of generated images, and thus FBP technique was developed to alter sinogram data using high-pass filters before back projection [40]. By creating negative image pixels at the image margin, FBP-based image reconstruction ignores flat areas to select sharp edges. This enhances image sharpness and reduces blurry margins. It represented a significant advancement over the first AR technique formulated in 1970 and found increasing use in clinical imaging [17]. Despite simple mathematics used, FBP works on assumptions and does not account for the cone-beam geometry of the calculated data considering the individual detector cell to be infinitely small. Even today, FBP remains the dominant method of image reconstruction on most early single- or few-detector-row CT scanners to larger detector arrays due to its speed and acceptable image quality. However, with the emergence of scanners capable of generating thinner sections at a finer matrix and the urgency to reduce associated radiation dose, FBP based image reconstruction techniques associated with higher image noise and artifacts are no longer ideal [41,42].

3.2. Iterative reconstruction

To address the limitations of FBP-based image reconstruction, both sinogram-based and image-based iterative reconstruction methods have been developed. The iterative reconstruction methods require higher computational power, and therefore, are typically implemented on programmable graphical processing units (GPUs) [1,42-45]. Iterative reconstruction techniques can be classified as hybrid (HI-RT) and model-based I-RTs (MI-RT) [46]. In HI-RT, iterative filtering of the projection data and image data reduces artifacts and image noise. The first HI-RT including iterative reconstruction in space (IRIS, Siemens Healthineers) and adaptive statistical iterative reconstruction (ASIR, GE Healthcare) developed in 2008, used a single back-projection step on raw data, instead of multiple steps in FBP to create less noisy images [47,48]. Subsequently, other HI-RT developed included sinogram-affirmed iterative reconstruction (SAFIRE, Siemens Healthineers), ASIR-V (GE, Healthcare), iDose® (Philips Healthcare), KARL 3D iterative denoising (United Imaging, Shangai, China), ClearView (Neusoft Medical Systems, Shenyang, China), and adaptive iterative dose reduction 3D (AIDR3D, Canon Healthcare) [47-49].

Veo (GE Healthcare) was the first MI-RT where a priori model for acquiring images, was used in combination with statistical noise and system geometry models while removing the image noise in a regularization step [50]. MI-RT incorporates three steps: 1) backward projection of raw data in cross-sectional image space, 2) forward projection of image space data calculate the artificial raw data, and 3) comparison of synthesized and true raw data to update the cross-sectional image [51]. The process is repeated several times to minimize the difference between true and synthesized raw data [52]. Other MI-RT algorithms include iterative model reconstruction (IMR, Philips Healthcare), and advanced modeled iterative Reconstruction (ADMIRE, Siemens Healthineers). MI-RT require higher computational power and more reconstruction time than HI-RT [53-55]. In 2016, a Forward-projected model-based Iterative Reconstruction Solution (FIRST, Canon Healthcare) was developed, which uses a three-dimensional (3D) image reconstruction approach [56]. FIRST uses a forward and statistical model in the projection data to improve spatial resolution and reduce the image noise. FIRST optimizes image quality jointly in sinogram and image spaces. Also, FIRST adapts image quality based on organ types such as lung, heart, abdomen, and bones. Several single- and multi-vendor studies have reported advantages of iterative reconstruction techniques over FBP with improved image quality at 23–76% lower radiation dose burden [51]. Conversely, some studies have reported that CT images generated from iterative reconstruction methods are too smooth (less noisy) and, therefore, not superior to FBP [52,57,58].

Compared to HI-RT, MI-RT have greater noise reduction and artifact reduction potential, and therefore, can enable a greater reduction in radiation dose associated with CT scanning. Several studies have reported that MI-RT can reduce radiation dose further than HI-RT and FBP while maintaining image quality for several routine and non-routine clinical indications and body regions in both adult and pediatric patients [57-59]. Although most studies have found better image quality at lower radiation dose with iterative reconstruction techniques compared to FBP, some investigators have described a loss of image resolution and low-contrast resolution at low radiation dose [60,61]. Due to their greater de-noising capabilities, MI-RT is also associated with a greater change in image texture or appearance compared to both FBP and HI-RT.

Apart from variations in their nomenclature and strength settings, there are substantial variations in the I-RT methods between all CT manufacturers in terms of domain of data processing. Most commercial I-RT use a HI-RT approach with blending of I-RT and FBP whereas few apply I-RT on the raw data or sinogram domain. These variations result in differences in image appearance and pose a challenge to both the physicists and radiologists from perspective of objective image quality.
Traditional methods of assessing and estimating image quality with image noise, contrast-to-noise and signal-to-noise ratios do not provide a sufficient way of assessing image quality with I-RT. Image noise can be virtually close to zero with some IM-RT and at high strength of HI-RT but such images are often regarded as too plastic or pixelated for clinical use.

### 3.3. Deep Learning-based reconstruction

Concerns over use of traditional methods of image quality evaluation with I-RT also extend to the non-linear image output from DL-based image reconstruction techniques. Physicists must employ new methods for assessing clinically relevant quality based on objective quantitative methods [62]. This may involve using more complex phantoms and metrics than traditional image quality phantoms.

Altered image texture and visual appearance with several iterative reconstruction techniques have been variably reported as blotchy, artifical, waxy, or plastic appearance which worsens with increasing strength of iterative reconstruction [58,63]. Loss of low-contrast resolution on reduced-dose iterative reconstruction can affect the evaluation of lesions in solid abdominal organs. With better computational capabilities in both hardware and software domain, data availability for training, validation, and testing, artificial intelligence-based reconstruction techniques have emerged as viable options for CT image reconstruction [64].

ML systems work on pattern recognition and require expertise and experience in designing the first component of the system, termed as a feature extractor [65]. The image pixels can be changed into a suitable data representation or feature vector using the feature extractor. Following feature extraction, a classifier, another key component, can detect and classify input patterns. In deep learning (DL), the mainstream of machine learning, deep neural networks use multiple layers of artificial neurons capable of performing mathematical calculations for an assigned task. Different mathematical calculations are separately applied for quantitative imaging and radiomics, along with unsupervised and supervised machine learning to add diagnostic information [66–68]. DL can handle many data, learn, and utilize different model structures beyond human capability. When presented with large labeled and unlabeled data, DL can extract important features, which are data representations and input for a specific task. Unlike traditional ML, the potential to learn data representations obviates the need to use handcrafted features for DL. Therefore, DL applications are increasing in radiology for object detection, image segmentation, and classification [69,70]. While most applications of DL focus on image analysis, which is invaluable to radiology, DL has found a new turf in CT image reconstruction [71,72].

DL algorithms can be subclassified into stacked autoencoders, deep Boltzmann machines, deep neural networks (DNNs), convolutional neural networks (CNNs), and others [73]. Pre-trained DL-based algorithms use data representations to generate image encoding for classification. DL-based image reconstruction techniques can generate high image quality at low noise levels and perform better image reconstruction even at thinner slices, which have been an inherent limitation of FBP and some iterative reconstruction techniques [74].

Encouraged by promising results of DL-based image reconstruction algorithms, some have advocated their use to replace the current image reconstruction techniques [75,76]. There are reports of multiple approaches for devising DL-based image reconstruction algorithms, which include image-based reconstruction using CNNs trained on low dose noisy CT images to generate images with the quality of routine standard noise acquisition and using prior complex functions to derive low-noise images with maintained resolution [65,77].

There are at least two commercially available and United States Food and Drug Administration (FDA)-approved DL-based image reconstruction algorithms - Advanced Intelligent Clear-IQ engine (AiCE, Canon Medical Systems) and TrueFidelity (GE Healthcare). The AiCE is based on a deep CNN trained on high-quality CT images acquired at high radiation dose and reconstructed with an MI-RT technique [78]. During training with projection data, DCNN learns the statistical method of differentiating signals and noise and then generates a high signal image as the output. These learned features are incorporated in the DCNN for future use on test data. Since this DCNN is used in the regularization step of iterative reconstruction, it maintains a noise texture similar to FBP and generates images with better spatial resolution than HI-RT. The average reconstruction time for AiCE is 30–40 frames per second for routine CT but can vary based on the applied scan factors such as the number of sections, section thickness, reconstruction kernel, pitch, and image noise.

Another commercial DL algorithm (True Fidelity (GE Healthcare) based on a deep neural network inputs high-noise projection data through its deep neural network and compares the output image with its low-noise version across multiple attributes such as image noise, low contrast resolution, low contrast detectability, and noise texture as measured in a MITA IQ LCD phantom with a body ring (Phantom Laboratory, Salem, NY) [79]. The comparison helps improve and strengthen the output image datasets. However, such phantom-based deductions on noise texture and detection of low contrast lesions need validation with comprehensive patient studies. Fig. 1 shows the artifact and noise reduction using a DL-based reconstruction of low-dose chest and abdomen CT as compared to normal-dose FBP images. The chest (low dose: CTIdvol 1.1 mgY; normal dose: 9 mgY) and abdomen (low dose: CTIdvol 1.2 mgY; normal dose: 10 mgY) were performed with 120 kV, 0.9:1 pitch, gantry rotation time of 0.5 s, and automatic exposure control technique. Both patients were scanned twice at normal- and low-dose following institutional review board approval and written informed consent.

Both AiCE (mild, standard and strong) and True Fidelity (low, medium and high) techniques offer three strength presets to control the strength of noise reduction in the reconstructed images. Reconstruction time can fluctuate based on the selected strength of these DL-based algorithms.

Tian et al. found better image quality, high SNR and contrast-to-noise ratio (CNR) on contrast-enhanced arterial phase CT images for pelvis using PixelShine (AlgoMedica PS), a DL-based algorithm that uses high amounts of both high and low dose CT training data [80]. The vendor-agnostic, PixelShine uses ML-based approach to reduce noise on FBP images without inducing changes in image texture noted with I-RT [81]. Pan et al reported that this technique had lower central frequency shift in the noise power spectrum for the same amount of noise reduction or most reduction in for same central frequency shift when compared with some HI-RT [81].

In addition to the commercially available DL-based reconstruction techniques, there are many other exciting approaches to CT image reconstruction. Some techniques focus on improving image quality from under-sampled measurements. Generally, DL reconstruction methods can be grouped into the three catalogues: image-to-image, data-to-image, and hybrid methods. Regarding the image-to-image mapping network, the trained neural network transforms suboptimal images to an acceptable image quality without accessing raw or sinogram data; for examples, with use of image-based convolutional neural network (CNN) [82], deep convolutional neural network (FBPConvNet) [83], wavelet transform based U-net [84,85], residual encoder-decoder CNN (RED-CNN) [86], and denseNet deconvolution network (DD-Net) [87]. These methods obtain good results with fast speed. However, such image-to-image networks usually predict output from a corrupted input with image-prior which may result in the data inconsistency between the original measurement and that synthesized based on the network output.

The data-to-image reconstruction methods can directly transform the measurement into images such as AUTOMAP [88]. Since AUTOMAP utilizes fully connected layers, it is difficult to implement this in practice with high-dimensional measurements. For inverting the Radon data to an underlying image, the iDronMap method was proposed for CT [89].
However, these deep networks are vulnerable to perturbation which gives unstable results [90]. Inspired by advantages of iterative algorithms such as compressed sensing methods, it is feasible to unfold a specific iteration scheme into a network taking imaging geometry into account, such as the learned experts’ assessment-based reconstruction network (LEARN) [91], manifold and graph integrative convolutional network [92], alternating direction method of multipliers (ADMM) framework [93], and so on. Besides, the trained networks can be used iteratively, such as the learned primal–dual hybrid gradient (PDHG) algorithm [94], momentum-Net [95], and deep BCD-net [96]. A new iteration DL by combining analytic, compressed sensing, iteration refinement and deep learning (ACID) was recently proposed [97], to systemically overcome the instabilities reported by Antun et al [90].

Fig. 1. Low dose FBP-DL chest (C) and abdomen (G) CT images show better denoising and artifact reduction compared to low-dose FBP (A, E) and low-dose C-IR (commercial/clinical iterative reconstruction) images (B, F). Chest (low dose: CTDIvol 1.1 mGy; normal dose: 9 mGy) and abdomen (low dose: CTDIvol 1.2 mGy; normal dose: 10 mGy) were performed with 120 kV, 0.9:1 pitch, gantry rotation time of 0.5 s, and automatic exposure control technique.

Fig. 2. Methodology of denoising of post contrast low dose images of CT abdomen reconstructed with filtered back projection (FBP; A) to generate better image quality with deep learning-based image reconstruction (DLR; C). Transverse CT images at standard radiation dose (B) show acceptable image quality. The standard/normal dose image was acquired on a 64-slice MDCT with 100 kV and 213 mA with a CTDIvol of 5.2 mGy. The low dose CT image was obtained from projecting noise in the sonogram domain of the normal dose CT to simulate data at 25% of normal dose.
3.3.1. Evidence for DL-based image reconstruction

Image denoising is often the first step in CT image processing and needs to be performed with little signal aberration. Several DL methods for noise reduction are available for diagnostic imaging [98]. Chen et al. have described high signal-to-noise ratio (SNR) using DCNN and virtual low dose routine CT images for training (n = 200) and separate testing (n = 100) data [82]. Du et al. reported an improved DCNN capability using several stacked competitive blocks to formulate wide competitive network and used the publicly available Cancer Imaging Archive as training data [99]. Both Chen et al. and Du et al. used randomized images of CT head, chest and abdomen derived from National Cancer Imaging Archive as training data [99]. Although both studies used Poisson noise added to the training raw data, the highest SNR and lowest mean square error were reported with the stacked competitive network. Another approach by Kang et al. used the wavelet transformation of an input image to remove image noise in the wavelet domain through a DCNN [100]. The image output is obtained with wavelet recompositing. Although image noise using this method was reduced at even 25% radiation dose for abdomen CT with maintained organ boundaries, radiologists reported qualitative image degradation with denoised image texture. Fan et al. proposed a new method by changing the artificial neurons in the inner product with quadratic neurons and showed better efficiency [98]. Fig. 2 illustrates network-based denoising of low dose CT images.

Phantom studies suggested that DL-based image reconstruction is superior to other iterative reconstruction techniques for image quality and lesion detection on low dose CT due to improved detectability of low contrast lesions not easily seen on low dose MI-RT images [101,102]. Conversely, when CT is performed at high radiation dose, high contrast lesions are better delineated on MI-RT images, because MI-RT images at high dose show a sharp margin for high contrast lesions in comparison to blurring by DL-based image reconstruction in generated images [101,102]. Racine et al compared FBP, I-RT (ASiR-V), and DL-based reconstruction (DLIR, GE) with a customized 25-cm diameter phantom containing three 10-cm inserts simulating –100 to 1250 HU at 100–120 kV [103]. They estimated noise power spectra, noise magnitude ratio, target transfer function (contrast-dependent spatial resolution) and detectability indexes (to assess joint effect of noise and contrast-dependent spatial resolution) for simulated abdominal abnormalities such as appendicitis, diverticulitis, and calcium-based renal calculi. The latter was estimated with a nonprewhitening with eye filter (NPWE) model observer method (107). DLIR outperformed ASiR-V for the simulated clinical findings at different dose levels (CTDIvol of 1, 3 and 7 mGy).

Watanabe et al compared FBP, HI-RT (ASIR, GE) and a DL-based image-based noise reduction technique (iNoir) using a pediatric body size cylindrical water phantom with two inserts simulating soft tissue attenuation and 12 mg iodine/mL [104]. They estimated noise texture change, task-based system performance function, and peak frequency difference, and concluded that iNoir had better image quality without unacceptable changes in noise texture compared to the I-RT [104]. When MI-RT, HI-RT, and DL-based image reconstruction techniques were compared in the phantom studies, it was found that small structures are best seen on MI-RT at high-dose and with DL-based image reconstruction at low doses. Thus, both MI-RT and DL reconstruction techniques were cited as better than HI-RT [76,101,102].

In patient studies, DL-based image reconstruction consistently demonstrated lower qualitative and quantitative image noise than HI-RT in low dose CT angiography with no deterioration of spatial resolution [77]. Specifically, when DL-based image reconstruction was performed based on the MI-RT, it was found that the low-frequency noise on abdominal MI-RT was suppressed by DL-based image reconstruction, suggesting its application in body regions where both MI-RT and HI-RT struggle for optimum image quality at low-dose [77].

Ultra-high-resolution CT images with MI-RT have better spatial resolution due to smaller detector elements and fine x-ray tube focal spot. Oostveen et al found the modulation transfer function of a novel ultra-high-resolution CT to be twice as high as a multidetector CT [105]. Thus, the high resolution and ultra-high-resolution scanning modes can increase the spatial resolution by 200% with only 23% increase in radiation dose for same detectability. However, due to smaller detector elements in ultra-high-resolution CT, the number of photons striking these detectors drop, which can lead to increased image noise with MI-RT. This limitation for MI-RT at low dose CT is inconsequential for chest CT since noise does not affect the visibility of high-contrast lung findings, but it is of concern in abdomen CT [65,106,107]. The DL-based image reconstruction can circumvent this disadvantage of MI-RT and generate higher image quality for CT abdomen as compared with other reconstruction methods [77]. In this manner, DL-based image reconstruction helps reduce image noise in ultra-high-resolution CT images. Despite this beneficial impact on reconstructed CT images, the vessel conspicuity on MI-RT was found to be better [77]. The reason for this effect was the non-optimization of the algorithm for blood vessels. Although statistically significant, the shift in attenuation values (HU) between HI-RT, MI-RT, and DL-based image reconstruction techniques was <3% [91].

Another study by Tatsugami et al. used CNN for DL-based image reconstruction in 30 adult patients who underwent clinically-indicated CT coronary angiography [108]. In this study, DL-based image reconstruction had lower image noise as well as improved CNR and image quality. A comparison was done between the HI-RT and DL-based image reconstruction. Singh et al. assessed the effect of DL-based image reconstruction in 59 adult patients who underwent chest (n = 22) and abdomen CT (n = 37) on a 320-slice multidetector-row CT scanner (Aquilion ONE, Canon Medical Systems) and compared the image quality and detection of the clinically significant lesions at submillisievert radiation dose levels [78]. In this study, AIDR3D, FBP, FIRST, FBP, and AiCE methods at submillisievert and AIDR3D at standard-dose levels were compared. All four radiologists detected clinically significant abdominal lesions on standard-dose AIDR3D as well as low-dose AiCE images. Radiologists missed 20–80% of lesions in the chest and abdomen on low-dose iterative reconstruction and FBP images. Low-dose AiCE images were acceptable for diagnostic interpretation in 97% of low-dose abdominal CT and 95–100% of low-dose CT dose CT studies. Figs. 3 and 4 illustrate the comparative image quality of low dose CT images for the chest and abdomen with different image reconstruction techniques. Relative to the images presented in Figs. 3, 4, there was no significant change in HU values (<5 HU difference) for low-dose chest and abdomen CT images reconstructed with HI-RT (AIDR3D), MI-RT (FIRST), and DL-IR (AICE) there was no significant change in the HU value (<5 HU difference between the three techniques) [78].

Motion artifacts can compromise image quality, and apart from techniques such as pre-scan breath-hold instructions and non-breath hold scan, deep neural networks can enable 4D-CT artifact reduction and improve geometrical accuracy of structures such as diaphragm [109,110,111].

3.3.2. Emerging technologies and DL-based reconstruction

One of the emerging methods is sparse-sampling CT where the number of image projections are reduced while radiation dose is high for each projection. Using this technique, the signal-to-noise ratio (SNR) is improved and the radiation dose reduced by more than 50%. Due to sparse projection angles, these datasets required fully iterative reconstruction techniques for image reconstruction [112–117]. Despite several studies in this field citing potential clinical benefits of sparse-sampling CT, a large-scale application into routine practice is dependent on hardware availability. Recent studies have reported that DL-based image reconstruction can be beneficial for improving the image quality of sparse-sampling CT [91]. The authors described the use of LEARN (Learned Expert Assessment-based Reconstruction Network) for sparse-data CT for generating images with fewer artifacts, preserved features, and faster computational speed compared to other reconstruction methods. By using only 50 layers in the LEARN network, the
investigators were able to reduce computational complexity by several folds.

Another technology that has seen increasing use in radiology and clinical practice is dual-energy CT (DECT), where different degrees of beam attenuation of low and high kV energy by various materials based on their K-edge values enable material differentiation [118,119]. DECT images can provide useful diagnostic information with better image contrast and, in some cases, fewer artifacts compared to single-energy CT images. Recent studies have reported that DECT can be performed at similar or lower radiation doses as compared to single-energy CT in both children and adults while maintaining image quality [120,121].

Most DECT applications use iterative reconstruction techniques to denoise and reduce artifacts. In some patients, particularly those with a large body habitus and those scanned with low- dose, the iterative reconstruction methods can have high noise and artifacts. A recent study reported that physics-informed and anatomic information-trained CNN-based DL could help reconstruct non-contrast single-energy CT from DECT scans with higher fidelity than the processed DECT image datasets [122]. At the time of writing this manuscript, AiCE has been FDA approved for commercial use. In future, it will be interesting to assess if DL-based algorithms can permit dose reduction for DECT and help enhance the quality of DECT imaging in patients with large body habitus undergoing DECT. Currently, per the standard of care, department policy of our hospital, we do not use DECT when performing chest or abdomen CT in patients with body weight greater than 120 kg.

Lung cancer accounts for 1.6 million deaths worldwide, out of which...
40% occur in developed countries such as USA, Canada, Europe, Australia, New Zealand and Japan [123]. Kim et al compared image quality and noise on a low dose chest CT between ASiR-V and DL-based imaged reconstruction [124]. While maintaining better image quality, DL-based image reconstruction images were found to be less noisy compared to iterative reconstruction with ASiR-V. Other applications of CT screening include, low dose coronary CT angiography, where Liu et al found that DL-based algorithm can decrease image noise and thus improve the image quality [125]. Such DL-based image reconstruction can help further reduce radiation dose and improve image quality for low-dose CT based lung cancer screening.

DL-based image reconstruction can also reduce metal artifacts by incorporation of computer-aided detection of the implant(s) into image reconstruction [126,127]. This can not only help in single energy CT, but also in dual-energy CT scans, as compared to existent approach to reduce metal artifacts in DECT with an increase in keV, which reduces image contrast [128,129]. In future, DECT dictionaries, a current area of research, can be integrated into CT scanners [130-132].

The subjects of spectral CT and dose reduction are incomplete without discussing the latest clinical imaging technology using the photon-counting detector CT (PCD-CT), where individual x-ray photons can be counted in an energy discriminative fashion, without the complication of the electrical noise in the current-integrating detector [133-141]. PCD holds a revolutionary potential since it adds a spectral dimension to data collected with x-rays. It is well known that the linear attenuation coefficient of a pixel/voxel is a function of x-ray energy. With conventional CT, one approximately measures the area under the curve of linear attenuation versus x-ray energy. With photon-counting CT, one measures the curve itself. Hence, PCD can reveal and quantify multiple materials with x-ray data within several narrow energy bins. PCD-CT offers more information than dual-energy CT. More information in raw data means better opportunities for radiation dose reduction, and PCD-CT is more advantageous than DECT in soft tissue characterization [142], material discrimination [143], x-ray beam hardening reduction, and even molecular imaging with functionalized nanoparticles [144]. Due to the hardware concerns and physical limits, the number of photons each PCD unit can count is usually limited, which will be further distributed into several energy bins. This explains that the resultant projections have a low signal to noise ratio, and the reconstructed image quality and material decomposition results could be compromised if a traditional reconstruction algorithm is directly applied. Besides, it is also difficult to maintain consistent responses from different PCD elements, leading to errors between measurements and theoretic values. This results in spectral distortions, including various effects such as charge sharing, K-escape, fluorescence x-ray emission, and pulse pileups. All of them can make artifacts in projections, and limit the utilities of spectral CT in practice [145].

To achieve satisfactory results from noisy measurements, many advanced MI-RTs and material decomposition methods were proposed. First, the use of a high-quality image in the MI-IR model is feasible, including total image constrained diffusion tensor [146] and spectral prior image constrained compressed sensing [147]. However, obtaining such high-quality prior images in practice is not always possible in practice. Considering sparsity and low-rank properties in the spatial-spectral domain, multiple dedicated algorithms were developed using tensor-based nuclear norm [148], prior rank, intensity and sparsity model [149,150], total nuclear variation [151], patch-based low-rank [152], structure tensor TV [153], and tensor dictionary learning [154,155]. To further improve the reconstructed image quality, the non-local image similarity was explored in spectral CT reconstruction, including nonlocal low-rank and sparse matrix decomposition [156], spatial-spectral non-local means [157], nonlocal spectral similarity [158], spatial-spectral cube matching frame (SSCMF) [159], non-local low-rank cube-based tensor factorization (NLCTF) [160], and so on. All these methods were designed for two-dimensional cases. However, it remains challenging in practice due to its high computational costs, especially for cone-beam geometry. Besides, these methods usually involve several tunable parameters that need to be chosen for specific clinical applications. Finally, image structures and details are often still blurred. Up to now, performing such advanced reconstructions is still pre-mature because of the computational complexity and needs of validation and optimization.

Regarding the material decomposition algorithms, they can be divided into two categories: direct and indirect. Specifically, a direct material decomposition method can directly obtain material components from projections [161,162] with a known x-ray spectrum. However, due to x-ray beam hardening and scattering during the x-ray transmission, it is difficult to have an accurate x-ray transmission spectral model in practice. Some regularization priors have been considered, and the direct material decomposition results are often sensitive to noise. Indirect material decomposition methods can be further classified into projection-based and image-based methods [163]. For the projection-based methods, the projections are first decomposed into sinograms of basis materials. Then, an image reconstruction algorithm (for example, FBP and MI-RTs [164]) is employed. For such a material decomposition method, the errors from the specific basis material sinograms can be magnified in the material maps, resulting in a compromised decomposition accuracy.

DL-based image reconstruction and material decomposition can be improved by learning from raw data to help improve imaging performance. Fig. 5 demonstrates the overall procedure of deep PCD-CT imaging. First, we obtain information from PCD projection data by using a DL-based method on measurements directly. In this case, features from projections should be limited so that it can satisfy the requirement of low-level tasks. As for complicated tasks, the network may be very complicated; any instability of the network may lead to inaccurate predictions. It is widely known that DL-based image reconstruction can be used for the reconstruction of PCD-CT images from measurements. For that purpose, the physical effects can be corrected through pre-processing, including x-ray scatter removal [165] and ring artifacts suppression [166]. Different from traditional CT imaging, the correlation between different energy bins can be incorporated into the network to improve the performance, and the attention characteristics in the image-spectral domain may be best learned by the network. Once we have image reconstruction results, the DL-based methods depending on different tasks can be designed to satisfy the specific requirements. Fig. 6 demonstrates a fully convolution convolutional neural network for PCD-
CT image denoising. Regarding material decomposition, there are two DL based strategies. One is to obtain material decomposition maps from projections directly using a DL based method, as shown in Fig. 7. The advantages of this method include removal of the x-ray beam hardening and other artifacts explicitly in a data-driven manner. The other DL based method for material decomposition component can be based on image reconstruction results and train a material decomposition network that takes image reconstruction results as input [167], as shown in Fig. 8.

Since DL-based methods can be potent in pushing the applications of PCD-CT, there are two key issues that should be addressed well. First, how to generate training data with adequate ground truth (label)? This challenge could be partially addressed by modeling PCD-CT imaging with full consideration of key physical and technical factors [168]. Second, how to design advanced deep learning networks to obtain the best performance? Classic PCD-CT methods can be utilized in this context. For example, we can incorporate the correlation between different energy bins and spatial properties of image context into a network.

Other latest CT scanner developments include dark-field CT and phase-contrast CT scanning [107,169–186]. These techniques use electromagnetic wave interactions of x-rays by applying interferometric gratings and sensing small-angle scattering and phase shift. While phase-contrast CT images show higher soft-tissue contrast, dark-field CT shows better morphological information on nano-scale features critically important in distinguishing benign and malignant tumors. In either of these two techniques, iterative reconstruction techniques have been found to be potentially useful for diagnostic image quality. However, the limitation of a non-stop rotating gantry has been a practical concern of instrumentation and scanning, although many MI-RTs still can generate acceptable image quality under well-controlled conditions. Future applications with DL-based image reconstruction might help improve image quality in these scanners.

3.3.3. Limitations of DL-based reconstruction

DL-based image reconstruction methods have certain limitations. These methods of image reconstruction need ground truth and a large training dataset. Such training dataset must be homogenized to account for the inter-regional differences such as age, race, body weight, and gender and derive high quality diagnostic images. Although this step requires substantial modifications, it can be used to perform demographic specific cross-validation of DL-based algorithm according to the regional patient population. The appetite of DL-based applications for large amounts of patient data for training, testing and validation of reliable models DL applications can be an expensive and time-consuming [187]. Samei et al recommended use of virtual clinical trial data as an efficient alternative to clinical trials in imaging where virtual imaging in virtual or digital subjects enable virtual interpretation [188]. Although such virtual datasets can expedite and lessen the financial burden with true patient imaging data, from the point of view of image reconstruction, such virtual data must undergo robust validation to ensure coverage of multifold variations in patient size, geometry, centering, positioning, physiology (affecting contrast enhancement), as well as in size, shape, attenuation, margins and heterogeneity of focal and diffuse abnormalities detected in true clinical data.

Antun et al described instability of DL-based image reconstruction which can either cause severe artifacts with indetectable image perturbation or miss lesions from subtle structural variations in images.
Another limitation pertains to reduction in algorithm performance with an increase in samples [189]. Weiwan et al have described an Analytic, Compressive, Iterative Deep (ACID) network to address these instabilities by combining deep reconstruction, compression sensing, analytic mapping and iterative refinement [97].

DL methods can reduce image artifacts from imperfect data, but this is limited by a certain level of the signal-to-noise ratio. This implies that in order to drastically lower the artifacts, the resultant smoothened appearance of DL-based image reconstruction may not be optimum for qualitative evaluation of subtle structures, such as tiny lung nodules or small vessels. Additionally, any training data needs to be clearly defined to avoid data misrepresentation. For optimal image quality, training, validation, and testing data should be separated and effectively utilized. Thus, enough training data should be available and well-annotated for supervised and unsupervised learning, for best reproducibility of deep reconstruction algorithms on testing data. Also, when using a DL-based method to improve reconstruction quality, high quality raw data reduce the difficulty of the reconstruction task, facilitating images quality improvement [190 191]. However, it is still difficult to eliminate artifacts unless the fundamental cause of the artifacts is directly addressed and the strong prior in the specific domain is utilized.

Primary limitation of some commercial DL methods is the slight loss of sharpness of small structures or blurring of fine details when compared with CT images reconstructed with FBP and iterative reconstruction techniques. Although this did not compromise lesion detection in some clinical studies published to date [76–78], it can potentially affect visualization or reader confidence when assessing subtle lung and bone abnormalities. Another limitation of DL-based methods is that currently, they cannot be applied with DECT scan mode. Presently, these methods are only available from a few commercial CT vendors on their high-end or latest CT scanners. It is unclear when these methods can be made available on other scanners. Large, randomized studies on the extent of dose reduction, best scan factors, advantages, and limitations of DL-based methods compared to well-established iterative reconstruction techniques are necessary. As with iterative reconstruction techniques, CT vendors will likely create their different DL-based methods with distinct image appearance and textures, which could further increase the complexity in CT image reconstruction. With a rapidly expanding number of DL-based image reconstruction techniques, there is a need for enhanced clinical and vendor partnerships to reduce variations and standardize terminology. It is not inconceivable that a separate AI algorithm might be needed to select the correct scan factors and DL-based reconstruction settings!

4. Conclusion

In conclusion, deep learning-based image reconstruction methods are increasing in popularity due to their advantages compared to old and current methods of image reconstruction based on filtered back-projection and iterative reconstruction techniques. The advantages of deep learning include greater image denoising, image texture improvement, and artifact reduction compared to other techniques of image reconstruction. Users must remember the limitations of deep learning methods when applying them to their clinical practice. The potential of enabling the next generation of CT hardware improvements such as photon-counting detector CT with the aid of deep learning methods is exciting but requires further vetting. As we celebrate and toast 125 years of x-rays in medical imaging, applications of artificial intelligence and deep learning in image reconstruction and beyond forecast an equally interesting and exciting future of radiology, which must outlast another 125 years of cutting-edge medical developments in the future.

References
