Abstract—In nuclear imaging, limited resolution causes partial volume effects (PVEs) that affect image sharpness and quantitative accuracy. Partial volume correction (PVC) methods incorporating high-resolution anatomical information from CT or MRI have been demonstrated to be effective. However, such anatomical-guided methods typically require tedious image registration and segmentation steps. Accurately segmented organ templates are also hard to obtain, particularly in cardiac SPECT imaging, due to the lack of hybrid SPECT/CT scanners with high-end CT and associated motion artifacts. Slight mis-registration/mis-segmentation would result in severe degradation in image quality after PVC. In this work, we develop a deep-learning-based method for fast cardiac SPECT PVC without anatomical information and associated organ segmentation. The proposed network involves a densely-connected multi-dimensional dynamic mechanism, allowing the convolutional kernels to be adapted based on the input images, even after the network is fully trained. Intramyocardial blood volume (IMBV) is introduced as an additional clinical-relevant loss function for network optimization. The proposed network demonstrated promising performance on 28 canine studies acquired on a GE Discovery NM/CT 570c dedicated cardiac SPECT scanner with a 64-slice CT using Technetium-99m-labeled red blood cells. This work showed that the proposed network with densely-connected dynamic mechanism produced superior results compared with the same network without such mechanism. Results also showed that the proposed network without anatomical information could produce images with statistically comparable IMBV measurements to the images generated by anatomical-guided PVC methods, which could be helpful in clinical translation.

Index Terms—Cardiac SPECT, coronary microvascular disease, dynamic convolution, deep learning, intramyocardial blood volume, partial volume correction.

I. INTRODUCTION

SINGLE photon emission computed tomography (SPECT) is a nuclear imaging modality used to visualize and measure radiotracer activities within the body that is widely used for clinical purposes [1]. In SPECT imaging, the image quality is negatively affected by various factors, such as photon attenuation, photon scatterings, low photon sensitivities, and motion [2]. In addition to these, factors related to limited spatial resolution generally cause partial volume effects (PVEs) which contributions from different tissues may be combined into a single voxel. Because different tissues have different patterns and kinetics of tracer uptake, PVE will cause undesirable blurring in the reconstructed images. In addition, the size of reconstructed voxels of SPECT is usually larger than that of other higher-resolution imaging modalities (such as computed tomography (CT)), leading to additional PVEs. In SPECT imaging, PVEs are usually represented as spill-over of photon counts between different tissues [3]. For example, in cardiac imaging, photons emitted from the myocardium may spill-over to the blood pool and vice versa, resulting in under-estimation or over-estimation of tracer activities in the reconstructed images.
The spatial resolution of SPECT imaging systems is characterized by their point spread functions (PSFs), and the goal of partial volume corrections (PVCs) is to reverse the effects of the system PSF. Using only the SPECT data, this can be simply done by deconvolution in the image domain or by incorporating PSF into the system matrix for iterative reconstruction [4]. Nonetheless, both ways would result in undesirable artifacts due to loss of high-frequency information [3]. As discussed in [5], this issue can be effectively addressed by incorporating anatomical information for PVC. For cardiac SPECT imaging, our group has investigated a series of anatomical-guided PVC methods incorporating high-resolution contrast-enhanced CT angiography (CTA) images [6], [7]. Among these methods, we found that the iterative Yang (iY) [3], [8] method is preferred, as this approach is robust to image noise [9]. Despite its promising results, iY requires tedious image registration and segmentation steps between SPECT data and anatomical organ templates, which typically take about 10 hours per scan for the canine studies used in this work. iY also assumes perfect registration/segmentation between SPECT data and anatomical organ templates, which may not be practical in clinical settings due to motion artifacts, especially for cardiac imaging. Lastly, contrast-enhanced CT (CECT) is usually memory-consuming due to numerous non-rigid registration steps on high-resolution CECT image volumes. Also, its iterative Yang (iY-PVC) can be avoided. However, this method is time and memory-consuming due to numerous non-rigid registration steps on high-resolution CECT image volumes. Also, its performance is highly dependent on the atlas database. The atlas-based method needs to be individually optimized for different species, patient populations, and even different sizes of canine studies. The paper showed that the atlas-based method led to high variances for the PVC results [10]. Lastly, the atlas-based method only skips the manual segmentation step, but contrast CT is still needed.

Deep learning represents a new class of reconstruction algorithm [11] and may be an ideal tool to address the above-mentioned limitations of anatomical-guided or atlas-based PVC methods. In the past years, convolutional-based neural networks have been implemented in various medical imaging applications such as low-dose CT [12], few-view CT [13], [14], low-dose SPECT [15], few-view SPECT [16], [17], attenuation map generations [18] etc. However, little attention has been given to deep-learning-based PVC.

All the proposed convolutional-based networks mentioned above were trying to learn static convolutional kernels throughout the networks. In the context of this paper, a convolutional kernel is defined as a set of filters in a convolutional layer. Recent studies showed that dynamic convolutional networks can significantly improve the performance for classification tasks [19]. Initially proposed by Chen et al. [20] and Yang et al. [21], in a dynamic convolutional layer, instead of applying the same convolutional kernel to all the input data after the network is fully trained, dynamic convolution learns a weighted combination of multiple convolutional kernels adapted based on the input features, with negligible extra computational cost. Recently, as pointed out by Li et al. [19], the previously proposed dynamic convolution limits the adaptive capability to only one dimension (i.e., the number of convolutional kernels), and other dimensions were omitted (i.e., kernel spatial size, input channel size, and output channel size). Li et al. proposed an omni-dimensional dynamic convolution, which generalizes the dynamic mechanism to all the dimensions in convolutional networks. However, the dynamic weights of all the previously proposed dynamic mechanisms are calculated based on feature maps of only the previous layer. In this work, inspired by the Dense-net structure [22], we proposed a densely-connected multi-dimensional dynamic (DC-Dy) mechanism, in which dynamic weights are obtained not only from the previous layer, but also all the preceding layers. This idea encourages the network to reuse feature maps from all the layers and improve adaptive capability of the dynamic mechanism. Compared with the omni-dimensional dynamic mechanism, the proposed DC-Dy mechanism does not introduce additional parameters. To the best of our knowledge, dynamic convolutional networks have not been investigated so far in the field of medical imaging.

When training a medical image reconstruction network, global image quality metrics, such as mean-absolute-error (MAE) and structural similarity index measurement (SSIM), are widely implemented as the loss functions to optimize network parameters. Even though using these metrics has achieved promising results in various deep-learning networks, it is a well-known issue that these metrics do not necessarily reflect the true image quality in clinical settings, especially for cardiac images acquired on dedicated cardiac scanners like the Discovery NM/CT 570c [23] or D-SPECT [24]. Since these scanners have a field-of-view (FOV) smaller than the reconstructed matrix size, there are some insignificant features outside the FOV. But these relatively unimportant features share the same weight in MAE or SSIM calculations. In this work, we proposed to incorporate a clinically-relevant image quantification metric as part of the loss function for network optimizations.

Coronary microvascular disease (CMVD) is a prevalent and critical global health problem, which is often unrecognized [25]. Non-invasive methods to evaluate myocardial micro-circulatory function to accurately diagnose CMVD remains complicated and time consuming. We previously proposed a methodological framework for intramyocardial blood volume (IMBV) quantification as a metric for microvascular function obtainable with SPECT imaging of 99mTc-labeled red blood cells (RBCs) [9]. 99mTc-RBCs is routinely used for cardiac blood-pool imaging in the assessment of regional and global left ventricular function. Previous results showed
that SPECT images with anatomical-guided PVC methods produced more accurate estimation of IMBV [9]. IMBV could served as a novel index to diagnose CMVD [26]. The noninvasive diagnosis of CMVD is a challenging clinical problem especially in the presence of coexisting epicardial coronary artery disease [27]. Here, we focused on investigating $^{99m}$Tc-RBCs canine studies as an example for cardiac SPECT PVC, while expect the proposed method can be applied to other imaging tracers/modalities in the future.

In this work, we proposed deep neural network with a densely-connected multi-dimensional dynamic (DC-Dy) mechanism for whole-volume cardiac SPECT PVC. Compared with the static counterpart, the proposed dynamic network demonstrated consistently better performance on canine studies acquired on the GE Discovery NM/CT 570c scanner for evaluation of IMBV. Additionally, based on the imaging tracer and medical application, IMBV was incorporated as a clinically-derived loss to optimize network parameters. Ablation experiments presented in this paper demonstrated the effectiveness of this clinically-derived loss function. Compared with the anatomical-guided iY method, the proposed deep-learning method demonstrated fast and comparable performance without CECT as prior knowledge. Lastly, we also tested the proposed network with both CECT and SPECT images as dual-channel input (but without segmented organ templates). The network trained with dual-channel input further enhanced the PVC quality without tedious image segmentation.

II. METHODOLOGY

A. Data Acquisition

A total of 28 resting canine studies scanned with $^{99m}$Tc-RBCs were included in this project to validate the effectiveness of the proposed deep-learning PVC method. Formulated in (1), IMBV is defined as the ratio between the mean activities of the entire myocardium ($M_{myo}$) and the mean activities of the left ventricular blood pool ($M_{lv-blp}$). IMBV served as the primary metric in this work to evaluate the image quality and quantitative accuracy before and after PVC. After PVC, IMBV values are expected to decrease as the $M_{myo}$ decreases and $M_{lv-blp}$ increases by reducing the spill-over effects. $M_{myo}$ and $M_{lv-blp}$ were calculated using the manually-segmented 3D organ templates obtained from the CECT image volumes.

$$\text{IMBV} = \frac{M_{myo}}{M_{lv-blp}} \quad (1)$$

Prior to imaging, about 3 ml arterial blood was withdrawn from each animal from a femoral artery catheter into a heparinized vacutainer. RBCs were then labeled in vitro with sodium pertechnetate. About 30 minutes after RBC labeling, $^{99m}$Tc-RBCs (18.8 ± 4.3 mCi) were intravenously injected into the imaging subjects. SPECT scans were then performed 15 minutes after the injections. The SPECT scan time was about 10 minutes. A low-dose non-contrast CT was performed afterward for attenuation correction. Then, CECT scans were performed with retrospective electrocardiogram (ECG) gating during end-expiration. The animals were mechanically ventilated and were under general anesthesia (1-2% isoflurane and 55-60% nitrous oxide) during the procedure [28]. The use of animal data in this study was approved by the Institutional Animal Care & Use Committee (IACUC) of Yale University. All the scans were performed on a GE Discovery NM/CT 570c dedicated cardiac SPECT/CT system. The SPECT scanner consists of 19 solid-state cadmium zinc telluride (CZT) detector modules to generate projections $P \in \mathbb{R}^{32 \times 32 \times 19}$ with pixel size 2.46 × 2.46 mm$^2$. The system has pinhole collimators focusing on the heart with a FOV of about 19 cm in diameter. Images ($I \in \mathbb{R}^{50 \times 70 \times 50}$) were reconstructed using the maximum likelihood expected-maximization (MLEM) algorithm [29] with 80 iterations and voxel size 4 × 4 × 4 mm$^3$. No post-filtering was applied.

B. Iterative Yang PVC

The acquired SPECT list-mode data were rebinned into 8 cardiac gates during end-expiration to ensure alignments between CECT and SPECT. End-diastolic and end-systolic gates were selected for iY-PVC and network training/validation/testing. SPECT images were first resized to the dimension of CECT images. CECT images were then manually registered to SPECT images. 5 organ templates, including myocardium ($ROI_{myo}$), blood pool ($ROI_{blp}$), liver ($ROI_{liver}$), lung ($ROI_{lung}$), and background ($ROI_{bkg}$) were then generated by manual segmentations from contrast CT for iY-PVC. The generated templates are binary and cover the entire 3D volumes. iY-PVC was achieved by calculating voxel-wise correction factors through the process summarized in algorithm 1.

Algorithm 1 iY-PVC Method

Require: $I$  \quad \triangleright Initial reconstructed image using MLEM

Require: $S$  \quad \triangleright The system matrix

Require: $ROI_0$  \quad \triangleright Segmented organ templates $O \in \{myo, blp, lung, liver, bkg\}$

$k \leftarrow 0$

$\ell_k \leftarrow I$

while $k < 10$ do

\hspace{1em} $O_{mean} \leftarrow \text{mean}(\ell_k(ROI_0))$ \quad \triangleright Mean values in each organ

\hspace{1em} $T \leftarrow \sum O_{mean} \cdot ROI_0$ \quad \triangleright Noise-free template

\hspace{1em} $T_{proj} \leftarrow S \cdot T$ \quad \triangleright Forward projection

\hspace{1em} $T_{recon} \leftarrow \text{MLEM}(T_{proj})$ \quad \triangleright MLEM reconstruction

\hspace{1em} $F \leftarrow T/T_{recon}$ \quad \triangleright Correction factors

\hspace{1em} $\ell_{k+1} \leftarrow I \cdot F$ \quad \triangleright Corrected image

\hspace{1em} $k \leftarrow k + 1$

end while

In iY-PVC, a noise-free organ template is generated first using the mean values within the segmented templates. The voxel-wise correction factors are equal to the ratio between the noise-free template and the MLEM-reconstructed template after forward-projection using the system matrix. This method typically requires a few iterations to converge. In this work, 10 iterations were used for iY-PVC. The manual
registration took a few minutes per scan and the manual segmentation for five organs took around 10 hours per scan. The time-consuming process of generating accurate organ templates demonstrates the need for segmentation-free PVC method.

C. Densely-Connected Multi-Dimensional Dynamic Convolution

Mathematically, a static 3D convolutional operations [30] can be formulated as:

\[ y = W \odot x + B \]  \hspace{1cm} (2)

where \( x \in \mathbb{R}^{d \times w \times h \times C_{in}} \) and \( y \in \mathbb{R}^{d \times w \times h \times C_{out}} \) represent input feature maps and output feature maps of the convolutional layer, respectively. \( d \), \( w \), and \( h \) denote the spatial dimensions of input/output feature maps, which may be different depending on the chosen parameters of the convolutional layer. \( C_{in} \) and \( C_{out} \) represent input and output channel dimensions. \( W \in \mathbb{R}^{k \times k \times k \times C_{in} \times C_{out}} \) represents the weights in convolutional kernel. \( k \) is the kernel spatial dimension. \( B \in \mathbb{R}^{C_{out}} \) represents the bias term. \( \odot \) denotes the convolutional operations.

In the proposed DC-Dy convolutional layer, the kernel weight \( W \) becomes adaptive based on the feature maps of all preceding layers on three kernel dimensions (i.e., spatial, input channel, and output channel dimensions). Specifically, three attention values, \( a_{spa}, a_{in}, \) and \( a_{out} \), are obtained via an attention mechanism. \( a_{spa}, a_{in}, \) and \( a_{out} \) are different for different input image volumes. Mathematically, (2) becomes:

\[ y = \left( \frac{1}{3} W \odot (a_{spa} + a_{in} + a_{out}) \right) \odot x + B \]  \hspace{1cm} (3)

The kernel weight \( W \) is progressively weighted using \( a_{spa}, a_{in}, \) and \( a_{out} \), and make it different for different input image volumes along with all spatial locations, input channels, and output channels. With this design, dynamic convolutional layer can capture more informative features than its static counterpart.

The attention mechanism used to obtain \( a_{spa}, a_{in}, \) and \( a_{out} \) is adapted from the squeeze-and-excitation attention proposed by Hu et al. [31]. The input features \( x \) are first squeezed along spatial dimension using a global average pooling (GAP) operation \( (\text{GAP}(x) \in \mathbb{R}^{1 \times 1 \times 1 \times C_{in}}) \), followed by a fully connected (FC) layer and a rectified linear unit (ReLU) as the activation function. Another FC layer is used to project the output from the previous FC layer to the size of the attention values, followed by a sigmoid activation function to squeeze the attention values between 0 and 1. The 2 FC layers contain 2n and \( n \) neurons, where \( n = k^3 \) for \( a_{spa}, n = C_{in} \) for \( a_{in}, \) and \( n = C_{out} \) for \( a_{out} \). A graphical illustration of the proposed multi-dimensional dynamic convolutional layer with kernel size \( 3 \times 3 \times 3 \) (i.e., \( k = 3 \) ) is presented in Fig. 1. The attention mechanism used for the convolutional kernels is also presented in Fig. 1.

With proposed DC-Dy mechanism, three attention values are obtained based on input features of all preceding layers. Specifically, with the DC mechanism, the three attention weights (i.e., \( a_{spa}, a_{in}, \) and \( a_{out} \)) are obtained using feature maps from the previous layer \( (l-1)^{th} \) and all the preceding layers \( (1^{st} \) to \( (l-2)^{th} \) layers). Note that the DC mechanism only affects the three attention weights for convolutional kernels but not the input to the convolutional layers. Formulated in (4), to avoid introducing additional parameters and memory burden, features for attention weights calculations are combined using additions:

\[ x_{l}^{in} = \frac{1}{2} (x_{l-1} + x_{prev}) \]  \hspace{1cm} (4)

where \( x_{l}^{in} \) and \( x_{l} \) denote input features to the \( l^{th} \) layer for attention weights calculations, and output features from
the \( l^{th} \), respectively. Note that by using (4), features from the \((l-1)^{th}\) layer has the largest weight when calculating the three attention values for the \(l^{th}\) layer, and weights of all preceding layers gradually diminish. Intuitively, features that are the closest to the \(l^{th}\) layer should contribute most to the attention calculations. For the \(1^{st}\) layer, the input image volume is used to calculate three attention values. Feature maps are resized for dimension matching before (4).

**D. Network Structure**

The proposed network adapts a U-net-like structure [32], and it is presented in Fig. 2. The proposed network takes a batch of 3D non-PVC image volumes as input and tries to perform PVC without anatomical information from CECT. The dimensions for both input and output are \( N_b \times 50 \times 70 \times 70 \), where \( N_b \) represents the input batch size. As depicted in Fig. 2, the network contains 4 down-sampling blocks and 4 up-sampling blocks. Each of the down-sampling/up-sampling block has either a dynamic convolutional (Dy-Conv) or dynamic de-convolutional (Dy-DeConv) layer, followed by a dense-net block [22]. The size of dynamic convolutional kernel used for down-sampling or up-sampling is \( 1 \times 3 \times 3 \) without zero-padding, so that the dimension along z-axis remains the same throughout the network. The dynamic convolutional kernel size used in dense-net block is \( 5 \times 3 \times 3 \) with zero-padding to allow the network to capture contextual information for adjacent slices. ReLU activation functions were used after all the convolutional layers. Conveying paths are used to connect earlier layers and later layers. All convolutional layers contain 32 filters.

**E. Optimization and Training**

The proposed network was trained in a supervised manner with the MLEM-reconstructed images post-processed using iY as the reference. The objective function used to optimize all the trained networks in this study includes MAE and SSIM [33]. MAE loss can be formulated as:

\[
\ell_{\text{MAE}}(Y, X) = \frac{1}{N_b W H D} \sum_{i=1}^{N_b} ||Y_i - X_i||_1,
\]

where \( N_b \) is input batch size, \( W = 70 \), \( H = 70 \), and \( D = 50 \) are the height, width, and depth of the input/output image volumes, respectively. \( X \) and \( Y \) represent output image volumes and the corresponding training label, respectively.

SSIM measures the structural similarity between two images, and it equals to 1 when two images are identical. In this work, the convolutional window used to calculate SSIM is set as \( 11 \times 11 \). The SSIM formula is expressed as:

\[
\text{SSIM}(Y, X) = \frac{2\mu_Y \mu_X + C_1}{\mu^2_Y + \mu^2_X + C_1} \frac{2\sigma_{YX} + C_2}{\sigma^2_Y + \sigma^2_X + C_2}
\]

where \( C_1 = (0.01 \cdot R)^2 \) and \( C_2 = (0.03 \cdot R)^2 \) are constants to stabilize the ratios when the denominator is too small. \( R \) stands for the dynamic range of pixel values. \( \mu_Y \), \( \mu_X \), \( \sigma^2_Y \), \( \sigma^2_X \), and \( \sigma_{YX} \) are the means of \( Y \) and \( X \), deviations of \( Y \) and \( X \), and the correlation between \( Y \) and \( X \) respectively.

Since the network output and the training label are both 3D image volumes, we took advantage of that by including the SSIM values along 3 planes (transverse, coronal, and sagittal) into the objective function for network optimization. SSIM loss used to optimize the network parameters is expressed as:

\[
\ell_{\text{SSIM}} = 1 - \text{SSIM}(Y, X)
\]

To emphasize the edges for different organs, the sobel operator (SO) [34] is used to obtain edge images from the network output and training label. The SO uses two separable filters to produce two gradient vector maps at each 2D spatial location. The MAEs between the gradient vector maps from network output and the training label are included as part of the objective function. Similar to the SSIM calculations,
SO is applied for 3 planes, and the corresponding MAEs are included in the objective function. MAE loss calculated from the gradient vector maps can be expressed as:

$$\ell_{SO}(Y, X) = MAE(SO(X) - SO(Y))$$  \hspace{1cm} (7)

Since the GE Discovery NM Alcyone scanner used in this work has a FOV smaller than the reconstructed matrix, features inside and outside the FOV share the same weights in the SSIM and MAE calculations. This is not ideal because features inside the FOV should be more clinically relevant. As the anatomical information is available from CECT, and IMBV is a more relevant image quality metric before and after PVC, we took advantage of that by including the IMBV values into the objective function. IMBV loss function can be formulated as:

$$\ell_{IMBV}(Y, X) = |IMBV(X) - IMBV(Y)|$$  \hspace{1cm} (8)

IMBV values were calculated using the manually-segmented organ templates. Note that IMBV calculations were required only during the training stage. At the testing stage, no segmented template was needed.

SSIM values and MAE for gradient vector maps along 3 planes have the same weight in the composite loss function, which can be formulated as:

$$\min \theta_{Net} L = \ell_{MAE}(Y, X) + \lambda_a \ell_{SSIM}(Y, X)$$

$$+ \lambda_b \ell_{SO}(Y, X) + \lambda_c \ell_{IMBV}(Y, X)$$  \hspace{1cm} (9)

where $\lambda_a = 0.8$, $\lambda_b = 0.1$, $\lambda_c = 0.1$ are hyper-parameters used to balance different loss functions. $\theta_{Net}$ represents all the trainable parameters in the network.

The Adam method [35] was used to optimize all the trainable parameters in the network with 2 exponential decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Xavier method [36] was used to initialize all the convolutional kernel weights. All bias terms were initialized to 0. Among all the 28 canine studies, 15 were used for network training, 3 were used for validation, and the remaining 10 were used for network testing. Since end-diastolic and end-systolic gates are available for all the canine studies, there are a total of 30, 6, and 20 image volumes for network training, validation, and testing, respectively. Due to the limited amount of training data, 30 image volumes were augmented by rotating the image volume $30^\circ$ in the interval $0^\circ, 360^\circ$ along all 3 planes, resulting in a total of $30 \times 11 \times 3 + 30 = 1,020$ image volumes for network training. Batch size $N_b = 6$ was used.

F. Evaluations

In this work, image quality was quantitatively evaluated using SSIM, root-mean-squared-error (RMSE), peak signal-to-noise ratio (PSNR), and IMBV. One-way analysis of variance (ANOVA) and Tukey multiple comparison test [37] were used to evaluate the statistical significance in this study. In this work, p-value $p < 0.05$ indicates statistical significance. Pairings were assumed in statistical testings. After PVC, the spill-over effects should be corrected, leading to a higher $M_{blp}$ and lower $M_{myo}$. Hence, lower IMBV values typically represent better image quality in this work. The IMBV values obtained from iY-PVC results served as the gold standard for comparison.

Six networks were trained in this work to evaluate the effectiveness of all the network components and the proposed IMBV-derived loss function $\ell_{IMBV}$. Six networks include: (1) a U-net (network depicted in Fig. 2 without the dynamic mechanism); (2) a Dynamic U-net (U-net with the proposed multi-dimensional dynamic mechanism but without $\ell_{IMBV}$, denoted as Dy-U); (3) a Dynamic U-net with CECT as the second-channel input but without $\ell_{IMBV}$ (denoted as Dy-U-CT); (4) a Dynamic U-net with $\ell_{IMBV}$ as an additional loss function (denoted as Dy-U-BV); (5) a Dynamic U-net with CECT and $\ell_{IMBV}$ as an additional loss function (denoted as Dy-U-CT-BV); (6) a densely-connected dynamic U-net with $\ell_{IMBV}$ as an additional loss function (denoted as DC-Dy-U-BV). Note that dynamic networks (2)-(5) did not implement the DC mechanism and the three attention values were obtained using features only from the previous layer. For networks with CECT as the second-channel input, CECT images were resized to SPECT dimension before concatenation.

It is expected that networks with dynamic mechanism would produce better results than the network without it, and networks with $\ell_{IMBV}$ would produce images with better IMBV quantification than the networks without $\ell_{IMBV}$. Images post-processed with iY served as the reference in this work.

III. RESULTS

In this section, various network components are progressively added to demonstrate their effectiveness for cardiac SPECT PVC. Numbers are presented as MEAN ± STD.

A. Results on Dynamic U-Net

In this sub-section, the first 3 networks (U-net, Dy-U, and Dy-U-CT) are compared. Both U-net and Dy-U used only the SPECT image data for PVC without CECT and segmented organ templates. Dy-U-CT concatenated CECT and SPECT data together as input without segmented organ templates. Since CECT served as the prior knowledge in the iY-PVC method, we expected Dy-U-CT should capture more anatomical information from CECT and perform better than the networks without CECT.

One sample canine study was selected and presented in Fig. 3. All three deep learning methods produced images with reduced PVEs. Especially in the coronal slice, the contour of the left-ventricular blood-pool is fused with the myocardium region and is not clearly visible due to spill-over effects. The deep learning methods significantly improved reconstruction results, and the contours of each organ became more aligned with the CECT image. In addition to visual observations in the figure, the lower IMBV values and the profile plots (Fig. 4) demonstrated that networks with dynamic mechanisms produced even better results than the U-net. As indicated by the blue arrows in the transverse slice in Fig. 3, U-net did not perform well at the septal wall between the right and left ventricular blood-pools. Using the same network...
Fig. 3. Transverse and coronal slices of a canine study reconstructed using different methods. Profile plots were generated along the dashed blue lines to present the partial volume effects and are presented in Fig. 4. Arrows with the same color point to the same regions in the images. Blue arrows point to regions that are over-estimated in the non-PVC images. Green arrows point to regions that Dy-U-CT performed better (closer to iY) than Dy-U. Corresponding IMBV values are also included in the figure.

Fig. 4. Profile plots along the dashed blue lines in Fig. 3. The upper and lower plots correspond to the images on the first and second rows of Fig. 3, respectively. Arrows with the same color point to the same regions in the images presented in Fig. 3.

B. Results on Dynamic U-Net With IMBV Loss Function

The Dy-U and Dy-U-CT were re-trained with the \( \ell_{\text{IMBV}} \) (denoted as Dy-U-BV, and Dy-U-CT-BV, respectively) as an additional loss function to demonstrate the effectiveness of the proposed IMBV-derived loss function \( \ell_{\text{IMBV}} \). Ideally, using \( \ell_{\text{IMBV}} \) would force the network to concentrate more on the cardiac region and produce images with IMBV measurements closer to those of iY-PVC images than images generated by the networks without \( \ell_{\text{IMBV}} \). With the additional \( \ell_{\text{IMBV}} \), the average IMBV values of the 20 testing canine studies were 0.174 ± 0.036 and 0.166 ± 0.033 for Dy-U-BV and Dy-U-CT-BV, respectively. Compared with iY-PVC, the p-values were \( p = 0.009 \) and \( p = 0.986 \) for Dy-U-BV and Dy-U-CT-BV, respectively. Another canine study was selected and presented in Fig. 5. For the SPECT-only networks, the network with IMBV (Dy-U-BV) generated images with better reconstructed anatomical structure of the heart (blue arrows in Fig. 5), and less PVEs (green arrows in Fig. 5), as compared with Dy-U. For the networks using CECT as a prior knowledge (Dy-U-CT and Dy-U-CT-BV), despite the improved quantitative accuracy, they produced visually similar reconstructions.

C. Results on Densely-Connected Dynamic U-Net

The proposed Dy-U-CT-BV network already produces images with no statistically significant different IMBV quantification using the multi-dimensional dynamic mechanism and \( \ell_{\text{IMBV}} \) but without the DC mechanism. However, most available GE Alcyone scanners installed globally do not have...
Fig. 5. Coronal slices of a canine study reconstructed using different methods. Blue arrows point to anatomical features of the heart. Corresponding IMBV values are also included in the figure.

Fig. 6. Coronal slices of a canine study reconstructed using different methods. Profile plot was generated along the dashed blue lines to present the partial volume effects and is presented in Fig. 7. Blue arrows point to regions with noticeable spill-over effects. Corresponding IMBV values are also included in the figure.

Fig. 7. Profile plot along the dashed blue lines in Fig. 6. Blue arrows point to the same regions in the images presented in Fig. 6.

D. Additional Quantitative Evaluations

In order to compare results across all the 20 testing canine studies, 7 Bland-Altman plots and 7 linear fitting plots were generated using the IMBV values obtained from all the testing canine studies and are presented in Fig. 8-9, using the values obtained from iY-PVC results as the reference.

The effectiveness of different network components can be clearly observed in Fig. 8-9. A box plot summarizing the IMBV values of all the networks is also presented in Fig. 10. After applying the $\ell_{IMBV}$ and using CECT as the second-channel input, no statistical difference was observed for IMBV quantification between Dy-U-CT-BV results and the iY-PVC results ($p = 0.986$). Without CECT as network input, the p-value between Dy-U-BV and iY-PVC results was $p = 0.0089$. The difference between Dy-U-CT and iY-PVC was also not significant ($p = 0.153$). With only the SPECT data, network with the proposed DC-Dy mechanism and $\ell_{IMBV}$ also results similar IMBV quantification to iY ($0.169 \pm 0.034$, $p = 0.946$).

In addition to IMBV calculations, other image quality measurements, including SSIM, PSNR, and RMSE are presented in Table I. Using the same loss function and training strategy, Dy-U benefited from the proposed dynamic convolution and outperformed the U-net with superior quantitative measurements ($p = 0.003$, $p = 0.018$, $p = 0.039$, and $p < 0.0001$ for SSIM, RMSE, PSNR, and IMBV, respectively).

Dy-U-CT, the network using CECT as additional input, produced results with better quantitative results than Dy-U did ($p < 0.0001$ for SSIM, RMSE, PSNR, and IMBV).

With $\ell_{IMBV}$ as an additional loss function, both Dy-U-BV and Dy-U-CT-BV produced images with more accurate IMBV measurements compared with the networks without $\ell_{IMBV}$ ($p < 0.0001$ for both networks when compared with their counterparts).

As presented in Table I, compared with the networks without $\ell_{IMBV}$, networks with $\ell_{IMBV}$ produced images with superior IMBV measurements but similar SSIM, PSNR, and RMSE values. We believe such a small difference in the measured values will not affect overall image quality, especially in clinical settings. The DC mechanism also resulted in similar SSIM, PSNR, and RMSE values compared with its counterpart.
TABLE I

| Metric | Method  | Mean ± Std | IMBV | Dy-U | Dy-U-BV | Dy-CT | Dy-CT-BV | Y-PVC |
|--------|---------|------------|------|------|---------|-------|----------|-------|
|        | Non-PVC | 0.205 ± 0.042 | 0.195 ± 0.038 | 0.194 ± 0.037 | 0.174 ± 0.056 | 0.167 ± 0.034 | 0.158 ± 0.035 | 0.167 ± 0.033 (Reference) |
|        | Dy-U    | 0.205 ± 0.042 | 0.195 ± 0.038 | 0.194 ± 0.037 | 0.174 ± 0.056 | 0.167 ± 0.034 | 0.158 ± 0.035 | 0.167 ± 0.033 (Reference) |
|        | Dy-U-BV | 0.205 ± 0.042 | 0.195 ± 0.038 | 0.194 ± 0.037 | 0.174 ± 0.056 | 0.167 ± 0.034 | 0.158 ± 0.035 | 0.167 ± 0.033 (Reference) |
|        | Dy-CT   | 0.205 ± 0.042 | 0.195 ± 0.038 | 0.194 ± 0.037 | 0.174 ± 0.056 | 0.167 ± 0.034 | 0.158 ± 0.035 | 0.167 ± 0.033 (Reference) |
|        | Dy-CT-BV| 0.205 ± 0.042 | 0.195 ± 0.038 | 0.194 ± 0.037 | 0.174 ± 0.056 | 0.167 ± 0.034 | 0.158 ± 0.035 | 0.167 ± 0.033 (Reference) |
|        | Y-PVC   | 0.205 ± 0.042 | 0.195 ± 0.038 | 0.194 ± 0.037 | 0.174 ± 0.056 | 0.167 ± 0.034 | 0.158 ± 0.035 | 0.167 ± 0.033 (Reference) |

TABLE II

| Metric | Method  | Mean ± Std | IMBV | Dy-U | Dy-U-BV | Dy-CT | Dy-CT-BV | Y-PVC |
|--------|---------|------------|------|------|---------|-------|----------|-------|
|        | U-net-BV| 0.180 ± 0.034 | 0.177 ± 0.033 | 0.177 ± 0.033 | 0.177 ± 0.033 | 0.177 ± 0.033 | 0.177 ± 0.033 | 0.177 ± 0.033 |
|        | U-net-CT| 0.180 ± 0.034 | 0.177 ± 0.033 | 0.177 ± 0.033 | 0.177 ± 0.033 | 0.177 ± 0.033 | 0.177 ± 0.033 | 0.177 ± 0.033 |

Ischemic heart disease (IHD) is considered a leading cause of mortality globally, accounting for more than 9 million deaths in 2016 [38]. IHD can be divided into obstructive coronary artery disease (CAD) and/or CMVD. CAD affects larger epicardial arteries (> 500 μm), while CMVD affects smaller ones (<500 μm). CMVD exists in a large proportion of patients with IHD and/or other cardiovascular diseases, and it may or may not co-exist with CAD [25]. IMBV could serve as a novel index for micro-vascular function to diagnose CMVD independent of CAD. Our previous work [9] showed that SPECT imaging using 99mTc-RBCs could be implemented as a non-invasive imaging technique for IMBV quantification. However, physical limitations of SPECT imaging result in PVEs and compromise the quantitative accuracy. We previously demonstrated that SPECT images reconstructed with iY significantly improved IMBV quantification [9].

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Fig. 9. Linear fitting plots for comparing non-PVC and deep learning methods. Results obtained from iY-PVC were used as the reference. IMBV values were used as the data points in the plots. Coefficient of determination ($R^2$) and Pearson correlation coefficient (Corr. Coef.) are included in the linear fitting plots.

Fig. 10. A box plot summarizing IMBV values obtained from all the methods.

However, iY is not practical to be implemented in reality due to tedious image registration and segmentation steps. Additional radiation dose introduced from CT scans with or without contrast administration is also not ideal in human studies. In this work, we proposed a deep-learning technique to perform fast and consistent PVC without requiring segmented organ templates from CECT. The proposed network is featured with a U-net-like structure, the DC-Dy convolution, and an IMBV-derived loss function. Conventional convolutional-based networks aim to learn static convolutional kernels throughout the network. After training, the same set of convolutional kernels is applied to all the testing images, which is not ideal due to the large variability between imaging subjects. In this work, the proposed DC-Dy convolution allows the network to have adaptive convolutional kernels based on the input data to capture more informative contextual features for better performance and compensate for the variances between imaging subjects. Specifically, the DC-Dy mechanism in the network aims to assign different weights to the convolutional kernels along different dimensions for each input image volume. By doing so, the convolutional kernels in the proposed network can be adjusted even after the network is fully trained. Compared with the recently proposed omnidi-dimensional dynamic convolution [19], the proposed DC-Dy mechanism allows the network to learn the attention values based on features not only from the previous layer, but also all the preceding layers, without introducing additional parameters. This newly proposed mechanism further strengthens the adaptive capability of the network and improves IMBV quantification. To the best of our knowledge, this is the first work to investigate dynamic network in the medical imaging field.

The proposed network with both SPECT and CECT (without segmented organ templates) was validated with 28 canine studies, and produced images with comparable IMBV measurements to the iY-PVC images ($p = 0.986$). Since most installed GE Alcyone scanners are SPECT-only systems, the use of co-registered CECT images is not as easily accomplished. We also demonstrated that using only the SPECT data, the network with DC-Dy mechanism and $f_{\text{IMBV}}$ (DC-Dy-U-BV) produced images with comparable IMBV measurements to the iY-PVC images ($p = 0.946$).

Note that iY-PVC method assumes perfect alignments between SPECT images and segmented organ masks. In this work, to ensure perfect alignments, CECT images were...
The proposed loss function $\ell_{IMBV}$ and the DC mechanism produced images with statistically better IMBV quantification but did not lead to noticeable improvements in SSIM, RMSE, and PSNR measurements. We suspect that it may be attributed to the fact that IMBV calculations only focus on the heart, while other metrics are based on the entire image volumes. Also, SSIM, RMSE, and PSNR were calculated using the iY-PVC images as the gold-standard. However, ifY is not a perfect method for PVC as it relies on numerous assumptions [3]. Thus, we believe IMBV is a more relevant metric in this work to evaluate the image quality before and after PVC.

The GE Discovery NM/CT570c and 530c scanners are equipped with 19 CZT detectors mounted on an L-shape arc that covers a nearly 180° range for stationary imaging. Therefore, these scanners essentially represent limited-view imaging systems with truncated projections. Our previous work [16] proposed a multi-angle reconstruction approach to acquire multi-angle projections for improved image quality with detector gantry rotations. But obtaining multi-angle projections is not always feasible and complicates and extends the acquisition time. Combined corrections for both limited-view and partial volume artifacts will be incorporated in future studies to further improve the resolution and accuracy of reconstructions on scanners of this type.

V. CONCLUSION

In conclusion, we propose a deep-learning method for fast and robust PVC for cardiac SPECT imaging. The proposed network overcomes the limitations of the anatomical-guided PVC method. The network is featured with a densely-connected multi-dimension dynamic convolutions that allow the network to have adaptive convolutional kernels for each input image volume, even after the network is fully trained. The proposed network produced images with statistically comparable IMBV measurements to the gold-standard iY-PVC method, which demonstrates a strong potential for clinical implementations. However, the results presented in this paper are limited to canine studies. Further analysis is needed to evaluate the network performance in human studies for the diagnosis of CMVD. In our future studies, we also plan to validate the proposed method on other perfusion imaging tracers.

REFERENCES

[1] A. Bockisch, L. S. Freudenberg, D. Schmidt, and T. Kuwert, “Hybrid imaging by SPECT/CT and PET/CT: Proven outcomes in cancer imaging,” Seminars Nucl. Med., vol. 39, no. 4, pp. 276–289, Jul. 2009.
[2] P. Ritt, H. Vija, J. Hornegger, and T. Kuwert, “Absolute quantification in SPECT,” Eur. J. Nucl. Med. Mol. Imag., vol. 38, no. S1, pp. 69–77, May 2011.
[3] K. Erlandsson, I. Buvat, P. H. Pretorius, B. A. Thomas, and B. F. Hutton, “A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology,” Phys. Med. Biol., vol. 57, no. 21, pp. R119–R159, Nov. 2012.
[4] A. J. Reader, P. J. Juhyan, H. Williams, D. L. Hastings, and J. Zweit, “EM algorithm system modeling by image-space techniques for PET reconstruction,” IEEE Trans. Nucl. Sci., vol. 50, no. 5, pp. 1392–1397, Oct. 2003.
[5] D. Strul and B. Bendriem, “Robustness of anatomically guided Pixel-by-Pixel algorithms for partial volume effect correction in positron emission tomography,” J. Cerebral Blood Flow Metabolism, vol. 19, no. 5, pp. 547–559, May 1999.
