Real-Time Deep-Learned Reconstruction for a Scanning Intraoperative Probe

Joshua Moo, Paul K. Marsden, Kunal Vyas and Andrew J. Reader

Abstract—Accurate delineation of the boundary between cancerous and healthy tissue during cancer resection surgeries is important to ensure complete removal of cancerous cells while preserving healthy tissue. Labelling cancer cells with radiotracers, and then using a probe during surgery to detect the radiotracer distribution, is a potential solution for accurate tumour localisation and hence better surgical outcomes. This work explores the feasibility of using deep learning to reconstruct a radiotracer distribution from data acquired by an intraoperative probe. The probe’s sensor array outputs (SAOs), obtained by scanning the probe over a region of interest, are supplied to the deep network, which then outputs a reconstructed radiotracer distribution for the region of interest. This initial work demonstrates that the deep network used here, a convolutional encoder decoder (CED), can successfully reconstruct simulated 2D radiotracer distributions from synthesised input data. However, the network was unable to generalise reliably when tested with count levels not present in the training set. Therefore the network must be trained with desired count levels or else should include estimation of epistemic uncertainty to avoid misleading outcomes. We also show that test-time augmentation can improve reconstructed image quality, and hence can also be used to reduce the amount of training data required.

Index Terms— Intraoperative imaging probes, cancer surgeries, CMOS, Convolutional Encoder Decoder

I. INTRODUCTION

One of the main challenges faced in cancer resection surgery is the difficulty in delineating the boundary between cancerous and healthy tissue. Overestimation of the extent of tumour margins leads to excessive removal of healthy tissue. On the other hand, if tumour excision is incomplete, subsequent resection operations and extensive postoperative adjuvant radiotherapy are required which in turn will reduce patient survival rates [1]. Current tumour margin evaluation involves the pathological analysis of biopsies taken by the surgeon during surgery, and it can take weeks to know whether re-excision is required. Therefore, there is a need for real-time detection of cancerous tissue during surgery, which can be achieved through intraoperative technology. This has led to the utilisation of radiation-detecting intraoperative probes to guide the resection of tumours labelled with radiotracers.

Currently, most probes for detecting gamma and beta radiation during surgery use the count rate as a measure of the average activity within a region of interest in order to guide excision. However, the susceptibility of these probes to highly penetrating gamma rays originating from regions irrelevant to the surgery remains one of the main limiting factors for real-time tumour localisation.

On the other hand, probes which are able to provide an image of the distribution of the radioactivity within the region, allow better visualisation of the tumour margins. Optical image-guided cancer surgery, which involves the detection of photons emitted by fluorescence tumour-specific agents, was presented as a solution for real-time tumour margin delineation [2]. However, the need for the development of these tumour-specific agents has limited clinical adoption for reasons mentioned in [3, 11].

Gamma-radiation imaging probes have also been proposed [4], which allow the detection of tumours that are located deep under the tissue surface. However, these gamma-imaging probes are susceptible to the highly-penetrating gamma rays originating from distant organs due to the non-specific uptake of the radiotracer. This degrades the tumour-to-background ratio which leads to ambiguity of tumour location. These gamma imaging probes also suffer from greatly reduced sensitivity due to the required shielding and collimation.

More recent research has primarily focused on the development of beta-imaging probes, which provide improved...
localisation due to the shorter range of beta particles in tissue and air. However, many beta-imaging probes are designed to reduce the impact of background gamma rays, and in turn this can compromise beta-imaging sensitivity. Several beta-imaging probes utilise scintillators coupled to photodetectors via optic fibres, but these probes can suffer from low signal strength due to the low transmission efficiency of the optic fibres [5, 6, 7]. To address this limitation, several silicon-based probes were proposed in [8, 9]. Such probes use only a single layer detector, which is beneficial in addressing the physical constraints of the probe for a surgical environment, while providing a superior performance in terms of sensitivity.

This present work concerns how to reconstruct superficial radiotracer distributions from use of a probe originally designed towards the detection of internal conversion (IC) electrons from $^{99m}$Tc [10] scanned over a region of interest during surgery. It has already been well demonstrated that image reconstruction is feasible for many different imaging modalities through use of a deep neural network feed-forward architecture. Such networks map from the measurement sensor domain to the image domain data [11, 25]. The direct reconstruction of PET images from sinogram data using a convolutional encoder-decoder (CED) architecture has also been demonstrated in [12].

Based on a CED, this study explores the feasibility of reconstructing a superficial radiotracer distribution within a region of interest from sensor-domain probe data, referred to here as sensor array outputs (SAOs). In this work these SAOs were acquired with the intraoperative probe utilising CMOS monolithic active pixel sensors. The use of a deep-learned CED architecture i) provides potential to utilise the entirety of the information contained within the measured SAO data, ii) obviates signal/data rejection methods used in current processing, and iii) opens the way for real-time reconstruction during surgery. Finally, it is worth noting that the proposed methodology also provides scope for understanding the capabilities of supervised deep learning in image reconstruction due to the relative ease of obtaining ground truth data to pair with SAO measurements in the training sets.

II. METHODS AND MATERIALS

A. Detector

The prototype intraoperative probe under consideration in this work (Fig. 1) was developed by Lightpoint Medical [10]. The probe utilises a CMOS sensor with 480 × 640 pixels, pixel size 6 µm × 6 µm, and a sensitive area of 3.84 mm × 2.99 mm. The thickness of the sensitive layer, which is the epitaxial layer of the CMOS sensor is 4 µm (similar model used in [9]). The probe casing dimensions are 6.3 cm × 3.4 cm × 3.4 cm and is composed of aluminium to shield against visible light. The sensitive area is also covered with a layer of 10 µm thick aluminium and a layer of 12 µm thick Mylar to shield against visible light and provide protection to the sensor. The current probe casing was used to accommodate for the prototype electronics, and will be further miniaturised in future models. The detection efficiency for the CMOS sensor is estimated to be 0.0675% for the 140 keV gammas using the Beer-Lambert law. The probe is connected to a PC via USB for SAO acquisition, transmission, processing and analysis. The acquired raw SAOs contain event clusters which correspond to the charge distribution over several detector pixels caused by the radioactive detections.

B. Data collection

To account for the dark current or electronic noise of each pixel, 50 blank frames were initially acquired to obtain an average dark image which was then subtracted from each image acquired in the presence of a radioactive source. A threshold was applied to the dark current-corrected image which results in a binary image containing the clusters. Connected components labelling (8-connectivity) was then applied to the binary image which is required for the regionprops function in Matlab (ver 2020b) in order to isolate the clusters from the original dark current corrected image according to the size of their bounding box. The isolated clusters were then saved to form a dictionary. SAOs with $^{14}$C were used to generate a beta cluster dictionary, whereas SAOs from $^{99m}$Tc with a 2 mm PMMA blocker to shield against internal conversion (IC) electrons were used to generate a pure gamma cluster dictionary. In this study, β clusters from $^{14}$C were used as a surrogate for the IC electrons from $^{99m}$Tc. This is due to the difficulty in obtaining a pure IC electron-only dictionary as well as their similar energies, allowing the synthesised SAOs (which will be described in section II.C), to have ratios of beta to gamma clusters reflecting the branching ratios of the IC electron and gamma emissions from $^{99m}$Tc.

C. Image synthesis

Annular $^{99m}$Tc radiotracer distributions with different radii and positions within a 9.6 mm × 9.6 mm region (1600 × 1600 image) were simulated (Fig. 2) where the hot regions are assigned a value of 0.9 and cold regions with a value of 0.1 (Fig. 2a). The rejection sampling method [13] was then applied on the simulated distribution to sample coordinates for cluster centres. Initially, the clusters were added to a zero image using the sampled coordinates. The remaining zeros in the resulting image were then replaced with values sampled from a fitted normal distribution of background noise intensities where the mean and variance were determined from SAOs acquired experimentally without the presence of radioactive sources. SAOs of size 480 × 640 were then acquired in a rectilinear
trajectory with a step size of 160 pixels in both the vertical and horizontal directions across the simulated radiotracer distribution. The SAOs were then added to an empty 1600 × 1600 matrix according to their acquisition position to form a superimposed SAO (SSAO) containing the event clusters corresponding to the simulated distribution. To account for the overlapped regions within the SSAO due to a step size that is smaller than the size of an SAO, a sensitivity map was also generated, and the SSAO was divided by this. A threshold was then applied to the sensitivity-corrected SSAO (Fig. 2b) which was used as the input to the CED network.

To evaluate the ability of the network to reconstruct quantitative and more complex radiotracer distributions, the Shepp-Logan (SL) and Brainweb (BW) phantoms [14, 15, 16, 17] were used to generate SSAOs using the same method as described above, with the background voxels outside of the phantom having 10% of the maximum activity present within the phantom. More general and unbiased radiotracer distributions were also used to evaluate the reconstructions without any objects or real-world structures in the training set. These random distributions were generated from images containing random values and smoothed with a Gaussian kernel.

The number of clusters added to SAOs at each step was based upon the Poisson distributed nature of the emission process, the total activity within the field-of-view of the probe, detector efficiency and dwell time. In our simulation of these radiotracer distributions, we have assumed an injected activity of 700 MBq, with the intraoperative imaging occurring 3 hours after injection, which corresponds to 495 MBq being present within the patient during surgery. With a patient weight of 70 kg and a voxel activity concentration of 250 kBq/ml, the resulting standardised uptake value (SUV) was calculated to be at 35.4, which is in agreement with the mean SUV in [18]. The following is a step-by-step description on how the number of clusters added to the SAOs was determined:

1) Calculate the total activity within the field of view (FOV) using a voxel activity concentration of 250 kBq/ml. (The thickness of the phantoms is 1 cm in the simulations).
2) Determine the expected flux using the total activity within the FOV, at a given detector distance and detector area (the distance was kept at 0.3 cm)
3) Multiply the expected detector flux by the dwell time and the detector efficiency (8%) which was determined through experimental data. This gives the expected number of clusters to be added, C.
4) Sample the final number of clusters to be added through poissrnd(C), where C is used as the parameter (mean and variance) of the Poisson distribution.

Transformations including rotation, scaling and shear were applied to the simulated distributions before synthesizing the SAOs to generate examples which exhibit more variability within the dataset for training the CED network. Overall, 4 CED networks were trained, one with only BW phantoms, one with only SL phantoms, one with only noise distributions and a final one with all 3 types of objects. In all cases, there were 60 samples for each type of object with 80% used for training and 20% used for validation. The networks were then cross tested with all object types and the normalised root mean squared error (NRMSE) was found.

Given a network that is trained at a particular count level, it is implied that higher count levels in the test set would yield higher quality reconstructions. However, the ability of the trained network to generalise to different count levels is not guaranteed as shown in [19] for the case of MRI imaging, where several deep learned reconstruction networks yielded a counterintuitive result, showing a decrease in quality of the reconstructions as the sampling rate increased in the test set. Therefore, to investigate the generalisability of the reconstruction framework to different count levels, 3 CED networks were trained, one with SL and BW SSAOs with dwell times of 2s, 10s and 20s, which corresponds to low, mid and high count levels respectively. A final CED network was also trained with all 3 count levels and the target distributions for the networks were also not normalised to assess generalisability. The networks were then tested with SSAOs at different count levels and the NRMSE were found.

A grid search on the training set size was also conducted with mid count SSAOs to evaluate whether the number of images used in this study is sufficient to optimise the reconstruction task. As a further optimisation to improve the reconstruction performance, we also assessed the impact of applying test-time augmentation (TTA) as this technique has shown to be successful for image classification and segmentation tasks in preventing overconfident incorrect predictions [20, 26]. In this
D. Convolutional encoder-decoder network

The CED network architecture is adapted from [12] as shown in Fig. 3. The encoder contains sequential blocks of convolutional layers with kernel size of $3 \times 3$ with stride 1 and stride 2 for spatial sampling reduction, followed by a rectified linear unit (ReLU) activation and a factor of 2 increase in number of features at the output of each block. Due to the large size of the input image, the mini batch size was set to 1 to reduce computational costs which also removed the need for batch normalisation layers (which were used in the DeepPET architecture). The encoded features of the input SSAO are then fed into the decoder consisting of upsampling layers which increased the spatial size of the feature maps by a factor of 2 at each stage with bilinear interpolation, followed by convolutional layers with kernel size of $3 \times 3$ with stride 1 and ReLU activation. The upsampling layers double the spatial size of the feature maps and the number of features is also halved at each block. The loss function used was the mean squared error (MSE),

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2$$

which was calculated between the reconstructed distribution, $x_i$ and the ground truth simulated target distribution, $y_i$ across $N$ number of pixels. The proposed reconstruction framework was also compared to a benchmark conventional processing method where the SSAO was smoothed with a Gaussian kernel with an optimised full-width half-maximum value determined by solving a nonlinear least-squares problem,

$$\min_{\sigma} ||g(x, \sigma) - y||^2$$

where $g(\cdot)$ corresponds to the Gaussian operator with a $25 \times 25$ kernel size, $\sigma$ corresponding to the standard deviation of the filter, $x$ corresponding to the SSAO and $y$ corresponding to the target radiotracer distribution.

III. Results

A. Reconstruction

The training and validation MSE losses with the CED networks trained with the annular distributions and the SL and BW phantoms are shown in Fig. 4. The network trained on annular distributions is able to reconstruct the simulated distributions from the synthesised SSAOS and also suppress signals from the cold regions (Fig. 5). The network trained on the phantoms is also able to quantitatively match the general structures within the target distribution. However, the network is unable to reconstruct the smaller ellipses within the SL phantom. Fig. 6 also shows the comparison of a horizontal profile between the benchmark method and the CED network. A similar comparison is also shown in Fig. 7 with a BW phantom.

B. Count level

The reconstructions from the models trained on a particular count level when tested with SSAOs with different count levels are shown in Fig. 8. Models that were tested with SSAOs of higher count level compared to the count level that they were trained on provided saturated reconstructions whereas the models were not able to fully reconstruct the phantom when provided with SSAOS of lower count level. The mean and standard deviation NRMSE values across 10 test images are also shown in Fig. 9.

C. Cross Object / Domain testing

The reconstructions from the models trained on a particular object when tested with SSAOs from all domains are shown in Fig. 10. All models were able to reconstruct the general
structures of the SL and BW phantoms but not the noise distributions as seen with model SL and model BW when tested with an SSAO generated from the radiotracer distribution corresponding to noise. The model trained with all 3 types of objects also yielded reconstructions with NRMSE values comparable, and in some cases better, than the reconstructions with models tested with their respective trained objects. The mean and standard deviation NRMSE values across 10 test images are also shown in Fig. 11.

IV. DISCUSSION AND CONCLUSION

This study has shown that the CED network is able to reconstruct the radiotracer distributions from synthesised SAOs containing event clusters that were obtained experimentally from beta and gamma emitting sources. The proposed reconstruction framework avoids any of the signal rejection methods which are conventionally implemented in several beta intraoperative imaging probe studies [21, 22]. The proposed method is able to provide a qualitative as well as potentially a quantitative evaluation of the radiotracer distribution, providing a more detailed measure of activity within a region of interest. The CED network was also able to reconstruct the radiotracer distributions with background activity included in the simulated SSAOs. This demonstrates the network can learn to differentiate between emissions originating from the target radiotracer distribution and the background. Further evaluation can be done with SSAOs of different levels of background activities. This opens the way for visualising tumour margins during surgery without compromising the signal sensitivity through collimation or introducing any new tumour-specific agents for beta/electron-targeting radioguided surgeries.

One of the motivations for using this supervised deep learning approach is that the imaging forward model has not even been implemented, and conventional statistical image reconstruction methods would need at least some model before even being considered. In contrast, a CED can be trained to learn an accurate forward model from example data, thereby
implicitly including unique hardware effects such as detector imperfections which can be hard to model.

The CED network was not able to generalise to SSAOs of count levels that were not present in the training dataset. The saturated reconstructions (which were obtained when testing the CED with higher count level inputs) demonstrates the ability of the CED to learn the association between the count density and activity. Therefore, the network must be retrained with training data containing the intended count levels or, the target radiotracer distribution must be normalised according to the count levels.

The reconstructions from cross-domain testing also highlight the importance of object variability within the training dataset. As seen in column 1 of Fig. 10, the network trained only on SL phantoms reconstructed a BW phantom containing features only observed in SL phantoms, demonstrating overfitting.

Fig. 10. Comparison of reconstructions from models when tested with objects not present in the training data. The models were trained on specific object types and were applied to SSAOs of other object types; The first column under “Model SL” corresponds to the reconstructions from the model trained only on Shepp-Logan phantoms.

Fig. 11. Mean and standard deviation NRMSE values across 10 test images for each model with all 3 object types. The model that was trained on all object types “Model_all” was able to match the performance of the models that were trained on a single object type.

Fig. 12. Reconstruction performance as a function of training set size with and without test time augmentation. The mean and standard deviation values are calculated across 3 training run repeats.
some cases, training with multiple objects improves the reconstructions as seen in row 1 of Fig. 10 for the SL phantoms, which shows that using diversified training set manifolds has the potential to outperform individualised object training as explored in [11].

The NRMSE begins to plateau at a training set size of 60 (Fig. 12) which shows that the number of images used in this study is sufficient for the reconstruction task. TTA was also able to consistently provide better quality reconstructions due to the merging of multiple reconstructions after dis- augmentation and match the NRMSE values when higher amounts of training data were used. For example, the NRMSE value with TTA at a training set size of 40 images is comparable to and even lower than the NRMSE value without TTA at a training set size of 60 images. Therefore, this can serve as a method to reduce the training data required to train the network.

Although a rectilinear scanning trajectory was used when simulating the acquisition of the SAOs, this method will also work with any scanning trajectory provided that the position of the probe can be tracked during surgery, as shown to be possible in [23]. Further development of this method would involve the acquisition of SAOs on more realistic radiotracer distributions through Monte Carlo simulations (GATE) [24]. Experimental assessment of this method can be done using a dried $^{99m}$Tc solution, where the delineation of the ground truth distribution can be optically estimated. This provides a means of assessing reconstruction performance and can help provide paired datasets for supervised learning from purely real data to train the CED. This presents an excellent opportunity for exploration of deep learning capabilities in reconstruction, where normally ground truth distributions for non-emitters (e.g. emission tomography) it is impossible to pair data with ground truth distributions for non-trivial imaging scenarios.

Assuming that the dwell times for collecting the SAOs can be kept low at 2 s and that the probe is able to collect images at a rate of 10 frames per second, real-time capabilities of the reconstruction framework is feasible given that the trained network can map a single SSAO into a reconstruction in approximately 0.27 s. Therefore, the factors contributing to longer overall reconstructions would be related to the size of the region of interest and the count level of the SSAO. In our simulations, the overall area of the region of interest is approximately 1 cm $\times$ 1 cm, which requires 112 s to generate a low count SSAO, 560 s for a mid count SSAO and 1120 s for a high count SSAO. However in practice, a live reconstruction can be done as the surgeon is scanning to build the overall SSAO, where the reconstructed image is constantly updated as more SAOs are acquired over the region of interest (Fig. 13). The reconstructions were also performed on an 8 GB GPU, therefore the frame rates for the live reconstructions will improve with greater computational power.

Overall, the proposed reconstruction framework together with the CMOS intraoperative probe serves as a potential solution for the visualisation of tumour margins with a single layer detector. This method is also compatible with existing PET and SPECT radiotracers which presents as an advantage over current existing counting and imaging probes.

ACKNOWLEDGMENT

For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

REFERENCES


