

# **MOTION AND RADIATION DOSE REDUCTION IN QUANTITATIVE CT PERFUSION IMAGING OF ACUTE STROKE**

**Mahdieh Dashtbani Moghari**

A thesis submitted in fulfilment of the requirements for the degree of  
Doctor of Philosophy

School of Biomedical Engineering  
Faculty of Engineering and Computer Science



The University of Sydney

2022



# Abstract

“Motion and radiation dose reduction in quantitative CT perfusion imaging of acute stroke”

by

Mahdiah Dashtbani Moghari

Doctor of Philosophy in Biomedical Engineering

The University of Sydney, 2022

Primary supervisor: Dr. Andre Z. Kyme

Co-supervisor: A/Prof. Noel Young

Co-supervisor: Prof. Roger Fulton

Computed tomography perfusion (CTP) imaging provides vital decision-support for physicians in the diagnosis and treatment planning for acute ischaemic stroke. Serial three-dimensional frames collected over 1-2 minutes during the transit of contrast agent enables visualisation of the integrity of the cerebral vasculature and underpins quantitative haemodynamic modelling to characterise stroke lesions.

Notwithstanding the value of CTP imaging for stroke management, there are two areas of fundamental limitation: the increased likelihood of motion-induced corruption of the serial (4D) data compared to conventional 3D neuroimaging CT scans that complete within seconds, and the noise-limiting radiation exposure to patients to ensure that robust haemodynamic modelling can be performed. The overarching aim of this thesis was to develop methods to address these key limitations in CTP imaging, thereby improving the accuracy of image-based stroke analysis and long-term outcomes for patients.

Our starting point was to characterise the prevalence, severity, temporal behaviour and dependencies of head movement during CTP imaging studies, and to quantify its clinical impact. Based on this understanding, a predictive model was established to identify patient-specific risk factors for motion. The model implicated stroke severity quantified by the National Institutes of Health Stroke Scale (NIHSS), patient age and time from stroke onset to imaging as the most important factors, all of

which can be used pre-emptively to mitigate motion risk in CTP imaging. The results also showed that the accuracy of image interpretation and treatment decision making can potentially be improved for at least a fifth of CTP studies by developing retrospective intra-frame motion correction methods to augment conventional inter-frame motion correction.

Although motion correction is well-recognised as an important pre-requisite to haemodynamic modelling in CTP image analysis, only inter-frame alignment is used and the impact of intra-frame corruption caused by continuous motion is ignored. We investigated the Intel RealSense D415 depth sensor, a compact, markerless and consumer-grade optical motion tracking device, for potential use in providing rapid and accurate pose estimates for continuous motion in CTP imaging. Suitability of the device was characterised with respect to thermal stability and jitter, static and dynamic six degree-of-freedom pose accuracy, and adaptability to the clinical setting. A conservative pose accuracy estimate for robotically controlled phantom motion was  $< 2$  mm and  $< 1^\circ$ , and for volunteer motion inside a clinical CT scanner was  $< 3$  mm and  $< 1^\circ$ . The device therefore shows promise for CTP motion correction but would likely need to be used in a multi-Intel D415 sensor configuration, or used to augment data-driven methods.

To simultaneously reduce the radiation dose and the likelihood of motion during a CTP acquisition, we attempted to reduce the scan duration by reducing the number of frames acquired. This was achieved using a novel application of a stochastic adversarial video prediction approach trained to predict late CTP image frames from early frames, thereby avoiding the truncation of the wash-out phase of contrast agent transit. Using this approach to predict the last 18 CTP frames resulted in bolus shape characteristics deviating by  $< 4 \pm 4\%$  compared to the ground-truth. Average volumetric error of the hypo-perfused region was overestimated by 28.36 mL (22%) and the corresponding spatial agreement was 83% (mean dice coefficient). The results showed that predicting the last 18 frames can preserve the majority of clinical content of the images while simultaneously reducing the scan duration and radiation dose by 65% and 54.5%, respectively.

The final strategy developed in this thesis was a radiation dose reduction method based on using a 3D generative adversarial network (GAN) to synthesise normal-dose CTP images from low-dose images. The method incorporated pre-processing aimed at

leveraging the full spatio-temporal (4D) information of CTP data within a 3D GAN architecture. The quality of GAN-denoised images was assessed via image quality metrics, expert quality rating, and the preservation of the lesion characteristics. The results showed that prioritising temporal information in adapting 4D CTP data to the 3D GAN model resulted in better restoration of tissue haemodynamic information. The average lesion volumetric error reduced significantly by 18 - 29% and dice coefficient improved significantly by 15 - 22% at 50% and 25% of normal radiation dose using the GAN model.

In summary, this thesis reports novel quantitative methods to improve our patient-specific understanding of the impact and dependencies of head motion during CTP imaging, the potential use of practical consumer-grade motion tracking devices for comprehensive motion-corrected CTP imaging, and two state-of-the-art deep learning-based approaches for radiation dose reduction in CTP imaging. The proposed methods lay the foundation for improved image-based stroke analysis and optimised CTP imaging workup and radiation dose, thereby providing more robust decision-support for physicians to improve patient outcomes.

# List of Publications & Presentations Arising from This Thesis

## Journal Publications

- 1) **Dashtbani Moghari, Mahdiah**, Luping Zhou, Biting Yu, Noel Young, Krystal Moore, Andrew Evans, Roger R. Fulton, and Andre Z. Kyme. "Efficient radiation dose reduction in whole-brain CT perfusion imaging using a 3D GAN: performance and clinical feasibility." *Physics in Medicine & Biology* 66, no. 7 (2021): 075008.
- 2) **Dashtbani Moghari, Mahdiah**, Noel Young, Roger R. Fulton, Andrew Evans, and Andre Z. Kyme. "Head Movement During Cerebral CT Perfusion Imaging of Acute Ischaemic Stroke: Characterisation and Correlation with Patient Baseline Features." *European Journal of Radiology* 144 (2021): 109979.
- 3) **Dashtbani Moghari, Mahdiah**, Philip Noonan, David L. Henry, Roger R. Fulton, Noel Young, Krystal Moore, and Andre Z. Kyme. "Characterisation of the Intel RealSense D415 Stereo Depth Camera for Motion-Corrected CT perfusion Imaging." – To be submitted.
- 4) **Dashtbani Moghari, Mahdiah**, Amirhossein Sanaat, Noel Young, Krystal Moore, Habib Zaidi, Andrew Evans, Roger R. Fulton, and Andre Z. Kyme. "Reducing scan duration and radiation dose in cerebral CT perfusion imaging of acute stroke using a recurrent neural network." – To be Submitted.

## Conference Proceedings

- 1) **Mahdiah Dashtbani Moghari**, et al, "Reducing scan duration in cerebral CT perfusion imaging of acute stroke using a recurrent neural network" *Proc 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*- Accepted
- 2) **Mahdiah Dashtbani Moghari**, et al, "Estimation of full-dose 4D CT perfusion images from low-dose images using conditional generative adversarial networks" *Proc 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1-3, DOI: [10.1109/NSS/MIC42101.2019.9059723](https://doi.org/10.1109/NSS/MIC42101.2019.9059723)
- 3) **Mahdiah Dashtbani Moghari**, et al, "Characterisation of the Intel RealSense D415 Stereo Depth Camera for Motion-Corrected CT Imaging" *Proc 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1-3, DOI: [10.1109/NSS/MIC42101.2019.9059935](https://doi.org/10.1109/NSS/MIC42101.2019.9059935)

## **Conference Presentation**

- 1) **Mahdieh Dashtbani Moghari**, et al, “Using AI to reduce scan duration in cerebral CT perfusion imaging of acute ischaemic stroke”. Royal Australian and New Zealand College of Radiologists (RANZCR) 71st Annual Scientific Meeting 2021. (Oral presentation).

# Table of Contents

Abstract .....	i
List of Publications & Presentations Arising from This Thesis .....	iv
Table of Contents .....	vi
List of Figures .....	viii
List of Tables.....	xiii
Statement of Original Authorship .....	xv
Acknowledgements .....	xvi
<b>Chapter 1: Introduction .....</b>	<b>1</b>
1.1 Computed Tomography Imaging in Stroke.....	1
1.2 Motivation and Aims of Thesis.....	5
1.3 Organisation of Thesis .....	6
<b>Chapter 2: Literature Review .....</b>	<b>7</b>
2.1 Stroke Overview .....	7
2.2 Medical Imaging in Stroke Diagnosis.....	8
2.3 Acute Ischaemic Stroke Treatments.....	12
2.4 Quantitative Analysis of CTP Images.....	13
2.5 Quantitative CTP Imaging: Challenges and Opportunities.....	20
2.6 Summary .....	25
<b>Chapter 3: Deep Learning: Concepts and Principles of Operation .....</b>	<b>26</b>
3.1 Deep Learning.....	26
3.2 Training a Deep Network.....	30
3.3 Deep Learning in Medical Image Analysis.....	33
3.4 U-Net.....	34
3.5 Generative Adversarial Networks .....	35
3.6 Recurrent Neural Networks.....	36
3.7 Summary .....	38
<b>Chapter 4: Head Movement During Cerebral CT Perfusion Imaging of Acute Ischaemic Stroke: Characterisation and Correlation with Patient Baseline Features.....</b>	<b>39</b>
4.1 Abstract.....	39
4.2 Introduction.....	40
4.3 Materials and Methods.....	41
4.4 Results.....	46
4.5 Discussion.....	49

4.6	Conclusion .....	52
<b>Chapter 5: Characterisation of the Intel RealSense D415 Stereo Depth Camera for Motion-Corrected CT perfusion Imaging..... 53</b>		
5.1	Abstract.....	53
5.2	Introduction .....	54
5.3	Material and methods .....	56
5.4	Results .....	64
5.5	Discussion.....	70
5.6	Conclusion .....	72
<b>Chapter 6: Reducing scan duration and radiation dose in cerebral CT perfusion imaging of acute stroke using a recurrent neural network..... 73</b>		
6.1	Abstract.....	73
6.2	Introduction .....	74
6.3	Materials and Methods .....	76
6.4	Results .....	84
6.5	Discussion.....	88
6.6	Conclusion .....	90
<b>Chapter 7: Efficient Radiation Dose Reduction in Whole-Brain CT Perfusion Imaging Using a 3D GAN: Performance and Clinical Feasibility .... 91</b>		
7.1	Abstract.....	91
7.2	Introduction .....	92
7.3	Material and methods .....	95
7.4	Results .....	104
7.5	Discussion.....	112
7.6	Conclusion .....	114
<b>Chapter 8: SUMMARY, CONCLUSIONS &amp; FURTHER WORK ..... 115</b>		
8.1	Summary and conclusions .....	115
8.2	Further Work .....	118
<b>References .....</b>		<b>121</b>

# List of Figures

Figure 2.1 A hypo-perfused region caused by a blood clot. The infarct core is irrecoverable, but the penumbra can potentially be saved by a timely reperfusion. The collateral flow provides microvascular perfusion to the hypo-perfused regions. MCA and ACA refer to the middle cerebral artery and anterior cerebral artery, respectively. Image courtesy to (Demuth, Dijkhuizen et al. 2017).....	8
Figure 2.2 A NCCT scan showing infarct core (red border) in the right hemisphere (left in the image) one week after the stroke. ....	10
Figure 2.3 The structure of 4D CTP imaging. Several brain volumes are collected repeatedly over time, in this case 33 volumes in total, to track the passage of contrast agent during wash-in and wash-out through the cerebral vasculature.....	10
Figure 2.4 A CTA scan showing absence of the middle cerebral artery (MCA) (red circle) on the right hemisphere (left in the image). Image courtesy to (Gao and Wang 2013) .....	11
Figure 2.5 Flowchart showing the steps of multimodal CT-based stroke diagnosis and treatment decision-making for an acute stroke patient.....	12
Figure 2.6 Overview of quantitative CTP imaging. Arterial input function and time-attenuation curves of tissue voxels are extracted from the CTP source images and deconvolved to obtain tissue response function. Haemodynamic parameters including CBF, CBV, and MTT are then computed from the tissue response function.....	16
Figure 2.7 An example of CTP haemodynamic maps and summary map. Haemodynamic maps include CBV, CBF, MTT, and TTP maps, indicating voxel-wise haemodynamic parameters. Summary map shows the extent of the infarct core and penumbra obtained by thresholding the haemodynamic parameters.....	16
Figure 2.8 Evidence of severe motion and motion-induced artifacts (indicated by white arrows) in (a) CTP images, and (b) the haemodynamic maps. ....	20
Figure 3.1 Information processing by a single neuron. $X_i$ and $W_i$ ( $i \in 1, 2, \dots, n$ ) are the inputs and their associated weights, respectively. $f(\cdot)$ performs linear or non-linear transformation. $b$ is referred to as the bias term, which is used to adjust the weighted sum of inputs.....	27
Figure 3.2 Basic architecture of a multi-layer ANN. Many neurons concatenate into successive layers, enabling higher-level features to be progressively extracted from the input data. Grey arrows show the connection between neurons of successive layers. ....	27
Figure 3.3 Basic architecture of a multi-layer feedforward CNN. BN, FC and F.maps represent batch normalization layer, fully-connected layer and feature maps, respectively.....	28

Figure 3.4 An illustration of the convolution operation in a convolutional layer (a) and the max-pooling and average-pooling pooling operation (b). In (a), the source pixel is replaced with the weighted sum of itself and nearby pixels in the output. ....	28
Figure 3.5 Four commonly used activation functions in deep learning models. ....	30
Figure 3.6 Typical architecture of a U-Net. The architecture comprises the contraction path or “encoder” to capture image context and the symmetric expansion path or “decoder” that enables precise localisation of the region of interest in the image. Blue and white boxes represent feature maps and copied feature maps, respectively. The arrows denote different operations. Image courtesy to (Ronneberger, Fischer et al. 2015) .....	34
Figure 3.7 A generic GAN training framework. $G$ learns to map from random noise to a synthesised image resembling the real image. At the same time, $D$ tries to differentiate between the synthesised image and corresponding real image. ....	35
Figure 3.8 The pix2pix training framework. $G$ learns to map from the source image to a synthesised image resembling the real target image. $D$ receives two pairs of images: the synthesised pair, including the synthesised and source image; and the real pair, including the real target and source image, and tries to differentiate them. ....	36
Figure 3.9 Comparison of an ANN and RNN at different time steps $t_i$ ( $i \in 1, 2, \dots, T$ ). (a) An ANN takes in the input data $X_1, X_2$ and $X_3$ at once and passes them through the hidden layers to estimate the outputs $Y_1$ and $Y_2$ . (b) An RNN takes in the input data $X_i$ ( $i \in 1, 2, \dots, T$ ) at different time steps $t_i$ to estimate the output $Y_i$ . RNN units are connected via weight vectors. ....	37
Figure 3.10 One RNN unit. $X_t$ and $Y_t$ are the input and output at time step $t$ . $h_{t-1}$ and $h_t$ are the outputs of the previous and current hidden states, respectively. $W_{hh}$ , $W_{xh}$ and $W_{hy}$ are the weight vectors connecting hidden states to each other, the input and the output, respectively. ....	37
Figure 4.1 Evidence of motion and motion-induced artifacts over 3 consecutive CTP frames for a patient in Category 3 (top row) and Category 2 (bottom row). Each frame shows the axial (main panel), coronal (bottom-left panel) and sagittal (bottom-right panel) views of the middle slice of the brain. Frame-to-frame translations (b, e) and rotations (c, f) are shown to the right. Intra-frame motion artifacts are indicated by red arrows and inter-frame motion (relative rotation/shift of frames) is indicated by green arrows. ....	42
Figure 4.2 Example of residual motion artifacts (indicated by white arrows) in the haemodynamic maps after performing frame-to-frame motion correction. This example was from a study classified as Category 2 (“considerable” movement). ....	43
Figure 4.3 Rigid-body head movement and coordinate system definition. The X-ray source moves in a helical trajectory. $Z$ is along the bed direction. ....	44

Figure 4.4 Percentage of stroke patients in each motion category (0-4) ( $N = 80$ ). .....	46
Figure 4.5 Statistical comparisons between average motion parameters measured for the different motion categories. Comparisons were performed using an unpaired Student's t-test. Significance is indicated by: * ( $p < 0.05$ ), ** ( $p < 0.01$ ) and *** ( $p < 0.001$ ). Error bars represent 1 standard deviation.....	47
Figure 4.6 Box-whisker plots showing absolute translational (a-c) and rotational (d-f) head motion parameters with respect to the first frame, plotted as a function of CTP frame number for 78 patients. The bottom row shows the median absolute translational (g) and rotational (h) motion parameters against time.....	48
Figure 4.7 PCA-based correlation analysis using box-whisker plots. (a) First principal component (PC1) versus different motion categories; (b) absolute PC1 loading for different patient baseline features. ....	49
Figure 5.1 Intel RealSense D415 Stereo Depth Camera. ....	57
Figure 5.2 Robotic motion tracking, (a) Schematic of the experimental setup to control the Intel D415 and robotic phantom. The moving phantom frame coincided with the default UR3 tool frame ( $X'Y'Z'$ ). The robot base frame ( $XYZ$ ) is defined based on posterior to Anterior (PA), right to left (RL), and inferior to superior (IS). $xyz$ is the Intel D415 coordinate frame. (b) Rubber human face mask phantom mounted to the UR3 and viewed by the Intel D415. (c) Depth map generated by the Intel D415 using a region-of-interest confined to the phantom face. The coordinate frame shown in (b) is the robot base frame. Right/left in (a) is out/into the page. ....	59
Figure 5.3 Evaluating the Intel D415 performance in a clinical CT scanner. (a) Experimental setup showing head tracking of a volunteer using the mounted Intel D415 (white circle) and simultaneously using the IR-based OptiTrack system with 3 cameras; (b) IR-reflective markers affixed to the OptiTrack head cap (shown here for the human volunteer, but the same applied to the head phantom test); (c) depth map generated by the Intel D415 using a region-of-interest confined to the human volunteer face.....	63
Figure 5.4 A representative thermal stability of the Intel D415 pose measurements after start-up (green) and the corresponding internal temperature change (orange) over 70 min (approximately 130,000 frames). ....	64
Figure 5.5 Comparison of motion estimates for the Intel D415 (red) and UR3 (blue) for an arbitrary continuous motion trajectory traversed at a median speed of $0.022 \text{ ms}^{-1}$ (range $0 - 0.261 \text{ ms}^{-1}$ ).....	67
Figure 5.6 Comparison of motion estimates for the Intel D415 (red) and UR3 (blue) for a simulated head motion trajectory of a patient during a 60 s CTP scan. ....	67

Figure 5.7 Comparison of motion estimates for the Intel D415 (red) and OptiTrack (black) when tracking a head phantom inside a clinical CT scanner for 60 s. ....	69
Figure 5.8 Comparison of motion estimates for the Intel D415 (red) and OptiTrack (black) when tracking human head movement for 60 s inside a clinical CT scanner.....	69
Figure 6.1 Block diagram of the VAE-GANs model to predict late CTP frames ( $x_t$ ) from the sequence of previous (i.e. measured) frames ( $x_t$ ) shown in the training phase (a) and the testing phase (b), where $E$ , $G$ , $D$ , $p$ and $q$ refer to the encoder, generator, discriminator, and prior and posterior distributions, respectively, $L1$ is the $L1$ norm, and $DKL$ is the Kullback–Leibler divergence between $p$ and $q$ (see main text for details).....	79
Figure 6.2 A typical lesion summary map. The red border indicates the lesion estimate obtained using the proposed VAE-GANs maps and the green region indicates the lesion estimate obtained using the ground truth haemodynamic maps. The false positive ( $FP$ , healthy tissue misclassified as abnormal), false negative ( $FN$ , abnormally perfused tissue misclassified as healthy), true positive ( $TP$ , correctly classified abnormally perfused tissue) and true negative ( $TN$ , correctly classified normally perfused tissue) regions are shown in the image for this example. ....	83
Figure 6.3 Average PSNR, SSIM, and RMSE (mean $\pm$ SD) of the predicted CTP frames computed for 10 test studies. In each plot the green curve shows predicted CTP frames 26-33 estimated from frames 1-25, the purple curve represents predicted CTP frames 21-33 estimated from frames 1-20, and the black curve shows the metric for predicted CTP frames 16-33 estimated from frames 1-15. ....	84
Figure 6.4 Examples of image quality degradation from early predicted frames to late frames. Row (a) shows ground-truth CTP frames. Rows (b)-(d) show the corresponding predicted frames estimated from frames 1-15 (b), frames 1-20 (c), and frames 1-25 (d).....	85
Figure 6.5 The mean fitted venous output function (VOF) for the case of predicting frames 16-33 using frames 1-15, computed for 10 test cases: (a) ground-truth, (b) predicted, and (c) overlay of (a) and (b). The shaded regions indicate $\pm 1$ SD. The peak enhancement of all VOFs were aligned prior to averaging. ....	86
Figure 6.6 Ground-truth hemodynamic maps (top row) and the corresponding predicted maps using VAE-GANs (bottom row) for the case of predicting frames 16-33 using frames 1-15, shown for two patients (a) and (b). Columns from left to right show CBV, CBF, MTT and TTP maps. ....	87
Figure 6.7 Examples of the time-attenuation curves (TAC) of a representative voxel in a region including (a) the arterial input function (AIF), (b) normal grey matter, (c) penumbra, and (d) infarct. In each plot, the red and black curves represent the ground-truth TAC and the TAC estimated using the VAE-GANs, respectively. Left of the vertical dashed line shows the model input. ....	89

Figure 7.1 Overview of the 3D GAN training framework to estimate normal-dose CTP images from corresponding low-dose images. ....	97
Figure 7.2 Extraction of 3D input patches from 4D CTP data using the frame-based (a) and stacked (b) regimes. ....	99
Figure 7.3 An example lesion summary map. The red border indicates the lesion volume obtained using denoised haemodynamic maps and the green region using normal-dose haemodynamic maps (GT). Four regions including false positive ( <i>FP</i> , healthy tissue misclassified as abnormal), false negative ( <i>FN</i> , abnormally perfused tissue misclassified as healthy), true positive ( <i>TP</i> , correctly classified abnormally perfused tissue) and true negative ( <i>TN</i> , correctly classified normally perfused tissue) are shown in the image. ....	104
Figure 7.4 CTP images (top row) and haemodynamic maps (CBV, CBF, TTP) obtained using different denoising approaches (columns) at 25% dose. ...	106
Figure 7.5 Expert quality rating (mean $\pm$ SD) for haemodynamic maps obtained using different methods at 50% (a-c) and 25% (d-f) dose. Statistically significant scores with respect to the low-dose haemodynamic maps are indicated by *, **, and *** for GAN denoised images. ....	108
Figure 7.6 Estimated volume (mL) of infarct (a-c) and penumbra (d-f) from data obtained at 50% dose versus the volume from normal-dose (GT) images, shown for 30 acute stroke cases. Results are shown for low-dose CTP (a, d), <i>Gxyz</i> (b, e) and <i>Gxyt</i> (c, f) data. The dashed line shows 100% agreement with the GT. ....	109
Figure 7.7 Estimated volume (mL) of infarct (a-c) and penumbra (d-f) from data obtained at 25% dose versus the volume from normal-dose (GT) images, shown for 30 acute stroke cases. Results are shown for low-dose CTP (a, d), <i>Gxyz</i> (b, e) and <i>Gxyt</i> (c, f) data. The dashed line shows 100% agreement with the GT. ....	109
Figure 7.8 Estimated volume (mL) of hypo-perfused (infarct + penumbra) regions at 50% (a-c) and 25% (d-f) dose versus the volume in normal-dose (GT) images, shown for 30 acute stroke cases. Results are shown for low-dose CTP images (a, d), <i>Gxyz</i> (b, e) and <i>Gxyt</i> (c, f) data. The dashed line shows 100% agreement with the GT. ....	110

# List of Tables

Table 4.1 Subjective head movement categorisation.....	43
Table 4.2 Patient baseline features and grouping (designated Gp1 to Gp4) used in this study.....	45
Table 4.3 Quantitative characterisation of head movement within the patient cohort for each motion category, using the method described in Section 4.3.2/Quantitative analysis. Values represent the mean motion amplitude with respect to the first frame. Errors represent $\pm 1$ standard deviation.....	47
Table 5.1 Specifications of the Intel RealSense D415.....	57
Table 5.2 1D static and dynamic pose accuracy of the Intel D415 compared to the UR3 robot. Measurements are reported as mean $\pm$ SD.....	65
Table 5.3 Dynamic pose accuracy of the Intel D415 compared to the UR3 robot. Measurements are reported as mean $\pm$ SD.....	66
Table 5.4 Dynamic pose accuracy of the Intel D415 compared to the OptiTrack for the phantom and volunteer motion inside a clinical CT scanner. Measurements are reported as mean $\pm$ SD.....	68
Table 6.1 Comparison of the fitted VOF characteristics (AUC, FWHM and $C_{max}$ ) in the ground-truth and predicted CTP images for the case of predicting frames 16-33 using frames 1-15, computed for 10 test cases. Values represent the mean $\pm 1$ standard deviation.....	85
Table 6.2 PSNR, RMSE and SSIM (mean $\pm$ SD) of haemodynamic maps computed for the case of predicting frames 16-33 from frames 1-15, averaged across 10 test studies. ....	87
Table 6.3 Lesion size characterisation (mean $\pm$ SD) computed for the case of predicting frames 16-33 from frames 1-15, averaged across the 10 test studies. ....	88
Table 7.1 PSNR and SSIM (mean $\pm$ SD calculated for 30 patients) for denoised CTP images at 50% and 25% of the reference dose. Bold font indicates the best performance. ....	105
Table 7.2 PSNR and SSIM (mean $\pm$ SD calculated for 30 patients) for haemodynamic maps obtained using different denoising methods applied to CTP images acquired at 50% and 25% of normal dose. Bold font indicates the best performance. ....	107
Table 7.3 Lesion size estimation error and dice coefficient (mean $\pm$ SD calculated for 30 cases) at 50% dose. Bold font indicates the best performance. Statistically significant differences ( $p < 0.05$ ) with respect to the low-dose data are indicated by *.....	111
Table 7.4 Lesion size estimation error and dice coefficient (mean $\pm$ SD calculated for 30 cases) at 25% dose. Bold font indicates the best	

performance. Statistically significant differences ( $p < 0.05$ ) with respect to the low-dose data are indicated by \*..... 111

# Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature: \_\_\_\_\_Mahdieh Dashtbani Moghari\_\_\_\_\_

Date: \_\_\_\_\_30/09/2021\_\_\_\_\_

# Acknowledgements

I was blessed to work with great and inspiring supervisors during my PhD journey. Firstly, I would like to express my sincere gratitude to my supervisor Dr. Andre Kyme for his continuous support and guidance throughout my PhD research. As my supervisor, he has taught me more than I could ever credit him for in this section. I am grateful not only for his immense knowledge and expertise in the field, but also for his approachability and patience. I could not have imagined having a better supervisor for my PhD study. My sincere thanks also go to my co-supervisor, A/Prof. Noel Young for his continuous support and valuable input throughout my PhD. His professional enthusiasm and invaluable advices on my PhD research have been a great inspiration and motivation for me. Despite his busy schedule at the University and Westmead hospital, he was always happy to quickly arrange a meeting and help me with my questions. Whenever I felt down or went through difficult times during my PhD, a trip to the hospital and meeting him made me motivated and encouraged to continue even stronger. Thanks for all you did for me! Special thanks to my co-supervisor Prof. Roger Fulton for all his support and help during my PhD. His passion for research has always been inspirational for me. I remember sitting at his office while he was working on a problem; I was mesmerised by his focus and passion for his work and wished one day I 'd become a researcher as good as him!

Others to thank are Dr. Philip Noonan and David Henry for their technical support and our scientific discussions, Dr. Andrew Evans for his valuable clinical advices on my research, Ms. Krystal Moore, for her help for data collection, and Ms. Pamela Dougan for all her help for ethic applications and administrative stuff.

I would like to thank the University of Sydney and Faculty of Engineering and Computer Science for funding my PhD studies through EITRS scholarship.

I am very thankful to my partner, Abbas, who is also my best friend and was like a mentor for me during my PhD. Thanks for all your support and encouragement during this journey. I am very grateful to my mother, Ozra, who has sacrificed selflessly to support my education. My grateful thanks go to my brother, Masoud, and my two sisters, Maryam and Marzieh, who always supported and encouraged me with my education and provided me with great emotional support whenever I went through difficult times. Thank you all!

Finally, I would like to dedicate this thesis to the memory of my father, Ali, who always encouraged me and believed in my ability to be successful in the academic area. You are gone but your belief in me has made this journey possible.

Mahdieh Dashtbani Moghari

Sydney, Sep 2021

# Chapter 1: Introduction

---

## 1.1 COMPUTED TOMOGRAPHY IMAGING IN STROKE

Death and taxes may be certainties in life, but being personally impacted by stroke is not far behind. As the second major cause of premature death globally, and the third leading cause of mortality in Australia, the likelihood that you or someone close to you will suffer from stroke during your lifetime is high (Wang, Naghavi et al. 2016, 2021).

Stroke is caused by a loss of blood supply to the major areas of the brain and can produce devastating effects on an individual, including cognitive impairment, loss of language and loss of motor control. Stroke is classified as either ischaemic, caused by a blockage in an artery, or haemorrhagic, caused by a burst in a vessel wall. Approximately 85% of stroke cases are ischaemic, and the focus of this thesis will be on this subtype.

When the brain is starved of blood, even for a matter of seconds, the cost can be extreme. Very quickly a region of tissue may become infarcted and die; this part of the brain is irrecoverable. The tissue surrounding the infarct typically has reduced blood flow and is at risk of also becoming infarcted unless adequate blood supply can be restored. Thus, rapid reperfusion of the hypo-perfused tissue-at-risk – known as the penumbra – is the chief aim of stroke treatment.

Deciding on the appropriate treatment to achieve successful reperfusion of the penumbra is non-trivial. Conventionally, this has been based chiefly on the time from stroke onset, called the time-window approach. If the time from stroke onset is less than 4.5 hours, thrombolysis using a “clot-busting” drug is prescribed to dissolve the occlusion, provided there is no contraindication. Beyond the accepted time-window, more invasive and higher risk endovascular treatments are prescribed (González 2013).

The time-window approach is not always successful. There have been cases of haemorrhage following thrombolysis despite being within the safety time-window, and other cases showing good clinical outcome beyond the accepted time-window using thrombolysis. Therefore, there is an important need for timely individual risk

assessment and the use of patient-specific criteria to make the best possible decision for patient treatment.

An ideal case for receiving treatment is a patient with a small infarct core and a large salvageable penumbra, termed a favourable mismatch pattern. Therefore, accurately identifying and quantifying favourable mismatch patterns from medical images will be a good indication of treatment outcome. This is where quantitative haemodynamic modelling based on medical imaging data becomes vital.

By far the most common imaging technique to diagnose stroke and inform treatment is computed tomography perfusion (CTP) imaging. Serial three-dimensional frames collected over 1-2 minutes during the transit of contrast agent enables visualisation of the integrity of the cerebral vasculature and underpins quantitative haemodynamic modelling to characterise stroke lesions. To ascertain the functional status of brain tissue and detect areas of abnormal perfusion, haemodynamic parameters including cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT) and time-to-peak (TTP) are derived from the CTP data. Based on the estimated perfusion maps of these parameters, the extent of the infarct core and penumbra – and hence the mismatch pattern – can be obtained. Therefore, CTP data and the haemodynamic maps derived from it serve as an important decision support for stroke physicians to recognise a favourable mismatch pattern and make time-critical decisions regarding patient eligibility for specific treatments.

Notwithstanding the value of CTP imaging for stroke management, there are two areas of fundamental limitation that can impact every scan acquired: the increased likelihood of motion-induced corruption of the serial (4D) data compared to conventional 3D neuroimaging CT scans that complete within seconds, and the noise-limiting radiation exposure to patients to ensure that robust haemodynamic modelling can be performed. The main aim of this thesis was to develop methods to address these key limitations in CTP imaging, thereby improving the accuracy of image-based stroke analysis and long-term outcomes for patients.

Patient head motion is a problem in CT since a fundamental assumption of the image reconstruction is that the object remains stationary. In practice, this assumption is easily violated due to patient movement, resulting in artifacts such as blurring, streaking, ghosting, loss of spatial resolution and contrast, and increased noise in the reconstructed image (Yazdi and Beaulieu 2008, Boas and Fleischmann 2012).

Irrespective of whether motion manifests as image artifacts, the data corruption caused by motion will impact haemodynamic modelling and lesion characterisation, potentially compromising the treatment strategy (Liu, Cuevas et al. 2008). Thus, finding ways to mitigate and/or correct for motion is vital.

Even the rapid speed of CTP imaging does not make it immune to motion artifacts (Moghari, Noonan et al. , Fahmi, Riordan et al. 2014). Indeed CTP imaging, which takes 1-2 min to complete, is far more susceptible to motion than standard CT neuroimaging that completes within seconds. The serial (4D) nature of CTP imaging also makes it susceptible to both inter-frame and intra-frame motion (Moghari, Noonan et al.). Inter-frame motion refers to patient head movement between the reconstructed frames of the CTP acquisition. Intra-frame motion refers to motion occurring during the acquisition of a single frame.

Several methods can reduce the amount of motion during CTP imaging. The most common is passive head restraint and improving patient comfort. Sedation can prevent motion but is not recommended in acute stroke imaging since it may cause haemodynamic effects which bias clinical assessment of neurological status (Oddo, Crippa et al. 2016). Another option is to use faster CT scanners with faster gantry rotation or more detector rows and X-ray sources to reduce the probability of motion artifacts.

In practice, it is difficult to fully eliminate the impact of head motion without some form of motion correction. Motion correction is well-recognised as an important pre-requisite to haemodynamic modelling in CTP image analysis. By far the most common correction strategy is frame-to-frame alignment, in which each frame is registered to a reference image such as the initial frame or another image (e.g. non-contrast CT) collected as part of the multi-modal imaging workup. However, this regime implies that motion is fast and discrete and occurs chiefly between frames, which is unlikely to be true. Indeed, motion is much more likely to be continuous and to cause corruption throughout multiple individual frames. Interestingly, almost nothing is known about the impact of intra-frame motion in CTP image analysis.

Intra-frame motion correction necessitates rapid and accurate motion estimates. Motion estimates can be obtained using data-driven approaches or external tracking devices. Compared to data-driven techniques, external tracking can deliver continuous and more accurate motion estimates because they are independent of the motion-

corrupted data. The vast majority of external tracking devices rely on optical tracking of markers attached to the patient head. However, challenges related to the hardware and calibrations, marker attachment and cooperation of the subject have led to poor translation of these methods (Kyme and Fulton 2021). Low-cost, compact, and markerless tracking devices are an attractive option for integrating into medical imaging applications provided they meet the accuracy requirements.

Not only are the current motion correction strategies for CTP imaging very limited, there are scant data on the prevalence, severity and dependencies of motion that might inform the development of more advanced correction strategies. This is a key gap which this thesis aims to address.

The second key limitation of CTP imaging addressed in this thesis is the cumulative radiation exposure to the patient during serial imaging of the brain volume. The CTP acquisition delivers a radiation dose of 5-6 mSv to the patient (Manniesing, Oei et al. 2015), not excessive on its own, but large when considered alongside the dose from non-contrast CT (NCCT), CT angiography (CTA) and other potential follow-up CT scans post treatment. Thus, reducing the radiation dose is highly desirable, especially for younger adult and paediatric patients for whom the radiation dose carries a greater risk to long-term health due to the harmful stochastic effects of ionising radiation (Wolterink, Leiner et al. 2017).

Reducing the radiation dose in CTP imaging is not trivial. For instance, it is not as simple as reducing the overall scan duration since this results in truncating the washout phase of the contrast agent transit which are necessary for accurate haemodynamic modelling. Nor is it as simple as collect fewer frames by increasing the frame-to-frame time interval since inadequate temporal sampling can cause erroneous perfusion estimates. An obvious alternative is to reduce tube current throughout a conventional protocol in order to lessen the radiation exposure per frame – but this necessitates a robust means of counteracting the noise in what are already count-limited data.

Ideally, a denoising strategy for CTP data should be applicable to the reconstructed data and not requiring access to the raw data, have short computation time, respect and leverage the spatio-temporal correlations in four-dimensional (4D) data and preserve the clinical content underpinning haemodynamic modelling and lesion characterisation. Conventional denoising techniques – sinogram domain

filtration (Wang, Lu et al. 2005), statistical iterative reconstruction (SIR) (Li, Yu et al. 2014) and traditional image post-processing (Pisana, Henzler et al. 2017) – do not satisfy these requirements. Deep learning-based methods have shown promising performance in denoising low-dose CT (Wolterink, Leiner et al. 2017, Shan, Zhang et al. 2018). Compared to conventional approaches, deep learning methods have shorter computation time (during the testing phase) and in most cases, have shown better restoration of the texture and resolution of CT images (Zhao, Martin et al. 2020). However, the problem of denoising 4D data, such as CTP data, is more challenging compared to two- dimensional (2D) or three-dimensional (3D) data since both the spatial and temporal information must be considered within the deep network.

Following the success of deep learning models to predict the pattern within the data and denoise low-dose images, the power of these models can be harnessed to achieve radiation dose reduction in CTP imaging. In light of the capability of these models in other applications, the suitability of deep learning-based models to predict the trend in the CTP data – e.g. predicting a portion of CTP frames from truncated acquisitions – and generating new images from a learned distribution – e.g. low-noise images from high-noise images – can be investigated.

## **1.2 MOTIVATION AND AIMS OF THESIS**

The overarching objective of this thesis is to develop novel methods to overcome the limitations of motion and radiation dose in quantitative CTP imaging. The four specific aims to achieve this goal are:

- i. To investigate the prevalence, severity and dependencies of head motion during CTP imaging;
- ii. To investigate the suitability of compact, markerless, consumer-grade motion tracking devices for use in motion-corrected CTP imaging;
- iii. To develop and validate an accurate and reliable deep learning-based image processing approach to reduce CTP imaging scan time, thereby simultaneously reducing both the radiation dose and the likelihood of motion;

- iv. To investigate the feasibility of deep learning-based noise reduction in CTP imaging to reduce radiation dose without compromising the accuracy of haemodynamic modelling.

### **1.3 ORGANISATION OF THESIS**

The remainder of the thesis is organised as follows:

Chapter 2 introduces essential background on acute ischaemic stroke and quantitative CTP imaging. The focus of this chapter is on the current challenges and limitations in stroke analysis and identifying areas for improvement. Current approaches to address the limitations are critically surveyed.

Chapter 3 introduces the key principles and concepts of the deep learning approaches used in later chapters. We also discuss applications of deep learning in medical image analysis and describe the architectures of the deep learning techniques used in this thesis.

Chapter 4 addresses Aim 1 by characterising head motion prevalence, severity, temporal behaviour and dependencies in CTP imaging. This forms the basis of a predictive model for the identification of patient-based risk factors for motion.

Chapter 5 addresses Aim 2 by characterising the Intel RealSense D415 depth sensor as a candidate motion tracking device for CTP imaging. Performance of the Intel D415 is evaluated with respect to thermal stability and jitter, static and dynamic six degree-of-freedom (DoF) pose accuracy, and its adaptability to the clinical setting.

Chapter 6 addresses Aim 3 by presenting a novel application of a stochastic adversarial video prediction (SAVP) approach to predict late CTP image frames from early frames. This simultaneously reduces the likelihood of motion and radiation dose by reducing scan duration.

Chapter 7 addresses Aim 4 by developing and validating a 3D generative adversarial network (GAN) for synthesising normal-dose CTP images from low-dose images. To achieve this goal, practical and effective ways to exploit the spatio-temporal relationships in 4D CTP data within the 3D GAN architecture are investigated.

Chapter 8 summarises the main outcomes and conclusions of the thesis and suggests avenues for further work.

# Chapter 2: Literature Review

---

Stroke is the leading cause of long-term disability and the second major cause of death in the world (Ritchie and Roser 2018). It results from a loss of blood supply to some regions of the brain, which can potentially impair cognition, motor control and language. Stroke is broadly classified as ischaemic or haemorrhagic (Benjamin, Virani et al. 2018). Ischaemic stroke is by far the most common (87%) and is caused by a blockage in an artery to the brain, leading to reduced blood flow (hypo-perfusion) in the brain tissue. Haemorrhagic stroke occurs less frequently (13%) and results from a rupture in the blood vessel wall, causing bleeding inside the brain (Australian Stroke Foundation 2017). This thesis focuses on acute ischaemic stroke and its diagnosis using CTP imaging.

In this chapter, we briefly review stroke imaging techniques, stroke treatments and quantitative CTP imaging and its challenges and opportunities.

## 2.1 STROKE OVERVIEW

Figure 2.1 shows an overview of the hypo-perfused brain regions caused by a blood clot. A hypo-perfused region (Saita, Chen et al. 2004) is typically composed of an unsalvageable dead area in the middle, called the infarct core, surrounded by an extremely low perfused tissue that is potentially salvageable, called the penumbra or tissue-at-risk. Structurally, the penumbra is similar to healthy tissue, but its electrical and protein synthetic functions have been suppressed (Mies, Ishimaru et al. 1991). These functions are potentially restorable with timely reperfusion. Otherwise, penumbral tissue evolves into infarction. The amount of collateral flow—the alternate circulation around a blocked artery through nearby minor vessels—indicates the rate at which the penumbra can convert to the infarct core (Bang, Saver et al. 2008).

Penumbral tissue is the target of acute stroke treatments. Age, vascular risk factors such as hypertension, the degree of collateral flow, and, most importantly, the time from stroke onset can influence the penumbra size (Jung, Gilgen et al. 2013). For every minute a large-vessel acute ischaemic stroke is left untreated, 1.9 million neurons, 14 billion synapses and 12 kilometres (7.5 miles) of myelinated fibres are destroyed (Saver 2006). For each hour in which treatment is delayed, the neuron loss

in the brain is roughly equivalent to 3.6 years of normal aging. The phrase ‘time is brain’, first introduced in 1993 (Gomez 1993), highlights the significance of time in stroke management since the neurological system can be damaged quickly and permanently for acute ischaemic stroke patients.

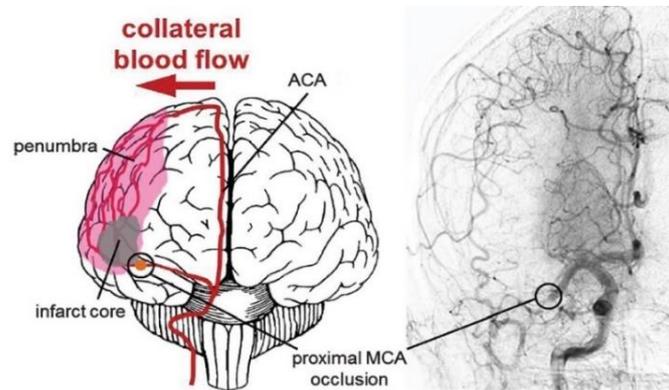


Figure 2.1 A hypo-perfused region caused by a blood clot. The infarct core is irrecoverable, but the penumbra can potentially be saved by a timely reperfusion. The collateral flow provides microvascular perfusion to the hypo-perfused regions. MCA and ACA refer to the middle cerebral artery and anterior cerebral artery, respectively. Image courtesy to (Demuth, Dijkhuizen et al. 2017)

## 2.2 MEDICAL IMAGING IN STROKE DIAGNOSIS

The goal of acute neuroimaging is to detect haemorrhage, occluded vessels and the extent and ratio of dead and salvageable tissue (González 2013). Although the ratio of the infarct and penumbra plays an important role in decision-making in stroke treatment, reliable quantification of this ratio from medical images is challenging.

Multi-tracer positron emission tomography (PET) imaging is the gold-standard imaging modality for ischaemic tissue quantification (Heiss, Sobesky et al. 2004). In PET, the areas of reduced blood flow, increased oxygen extraction fraction and a cerebral metabolic rate of oxygen (CMRO<sub>2</sub>) above the threshold of cell viability are considered tissue-at-risk (Østergaard, Gjedde et al. 2001). However, since PET imaging needs a nearby cyclotron for radiotracer production, requires invasive arterial sampling and is less common compared to other scanning machines in hospitals, it is not feasible in the emergency setting of acute stroke; instead, PET is mainly used for research purposes (Wintermark, Sesay et al. 2005). Magnetic resonance (MR) perfusion-weighted imaging (PWI) and diffusion-weighted imaging (DWI) mismatch and CTP are more commonly used due to their availability and being less technically demanding (Heiss, Sobesky et al. 2004). However, reliably distinguishing the infarct

core, true tissue-at-risk (penumbra) and false tissue-at-risk (benign oligemia) in the acute phase of stroke is challenging using Magnetic resonance (MRI) or CTP imaging (González 2013).

### **2.2.1 MRI**

In terms of accuracy (after PET), MRI is the best available method for early infarct core detection (González 2013). Cytotoxic oedema, caused by the passage of extracellular water into cells, restricts diffusional water movement, reducing the apparent diffusion coefficient (ADC) (Reid, Graham et al. 2012). Areas of cytotoxic oedema correspond to the infarct core and become apparent as hyperintensity in DWI as a high-contrast lesion. In MR PWI, a contrast agent bolus is injected into the patient, and the impaired perfusion regions in both the ischaemic core and surrounding tissue can be identified by tracking the contrast agent bolus through the brain. The mismatch between DWI-defined core and PWI-defined hypo-perfused area is used as an estimation of penumbra region.

However, MRI is not usually implemented for acute ischaemic stroke patients for several reasons. A full clinical multimodal MRI regime in an acute stroke setting takes 20–25 minutes (Hoffmann, Zhu et al. 2012), which may reduce the chance of salvaging the penumbra. Moreover, there is no consensus on the intensity threshold for best describing the penumbra in MRI images (Reid, Graham et al. 2012). MRI is also costly and usually unavailable in minor local hospitals. Therefore, although MRI is radiation-free and more accurate for lesion detection than CT, multimodal CT is usually performed as a baseline imaging modality for acute stroke.

### **2.2.2 Multimodal CT**

A multimodal CT imaging regime, including non-contrast CT (NCCT), CTP, and CT angiography (CTA), is the most common imaging procedure in acute stroke.

**NCCT:** NCCT is used to detect haemorrhage and is considered a reliable method for distinguishing haemorrhagic from ischaemic cases in the acute phase (González 2013). However, the early infarct core is not detectable by NCCT. It can only be used to detect the infarct core in the late stage of stroke when a hypodensity is clearly visible (Figure 2.2).

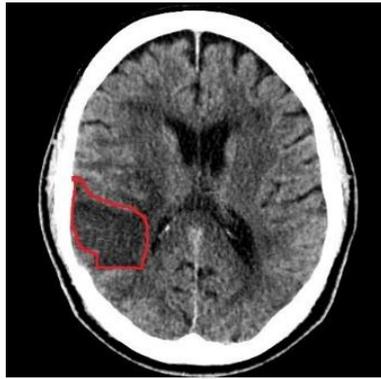


Figure 2.2 A NCCT scan showing infarct core (red border) in the right hemisphere (left in the image) one week after the stroke.

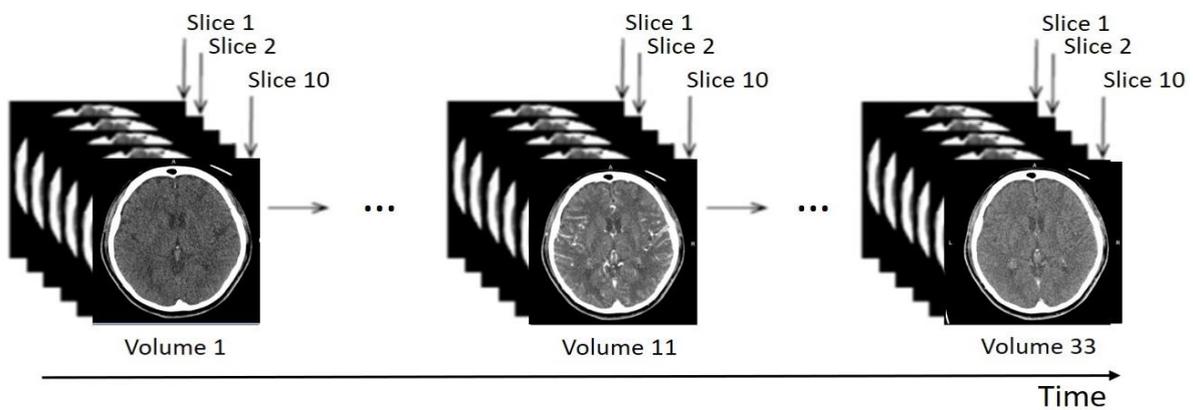


Figure 2.3 The structure of 4D CTP imaging. Several brain volumes are collected repeatedly over time, in this case 33 volumes in total, to track the passage of contrast agent during wash-in and wash-out through the cerebral vasculature.

**CTP:** After NCCT, if a haemorrhagic stroke has been ruled out, CTP imaging is performed. In CTP imaging, a bolus of iodinated contrast material is injected into the patient intravenously. Since the contrast agent has high absorbance, the passage of the contrast agent during wash-in and wash-out can be tracked through the cerebral vasculature over time by dynamic sequential scanning of the brain (usually within 60 seconds) (González 2013). Figure 2.3 shows the structure of 4D CTP imaging. Perfusion maps can be produced from CTP images using tracer kinetic analysis (Section 2.4.1). Evaluating the perfusion maps can distinguish healthy, salvageable and irreversibly damaged tissues.

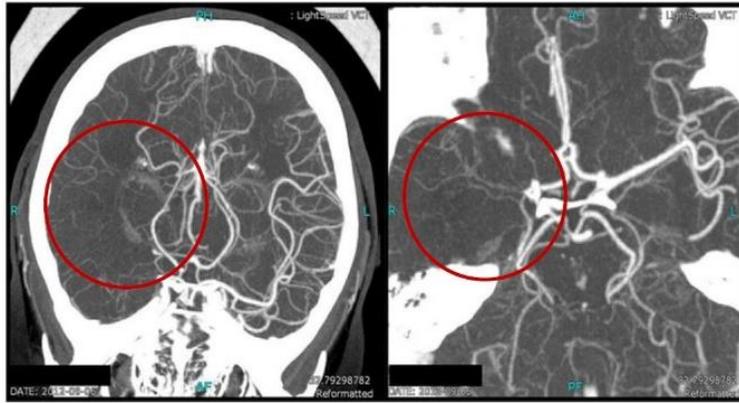


Figure 2.4 A CTA scan showing absence of the middle cerebral artery (MCA) (red circle) on the right hemisphere (left in the image). Image courtesy to (Gao and Wang 2013)

**CTA:** Analogous to CTP, CTA starts by injecting a bolus of iodinated contrast agent, but instead of acquiring a series of images, a single scan is obtained. CTA is performed after CTP to highlight the location and the extent of occlusion (thrombus) (Verro, Tanenbaum et al. 2002). Moreover, it is feasible to determine the degree of collateral flow at the occlusion site using CTA, which helps physicians predict a patient's response to the treatment. Figure 2.4 shows a typical CTA scan.

#### *Advantages and challenges of multimodal CT in acute stroke*

Multimodal CT is advantageous as a diagnostic tool in the emergency setting of acute stroke because of its relatively short acquisition time (Jia, Scalzo et al. 2016), thereby reducing the time to diagnosis. Multimodal CT, including NCCT, CTP and CTA, can be acquired within approximately 10 minutes. The lower cost of CT imaging compared to PET and MRI and the widespread availability of CT scanners in hospitals are further advantages for using multimodal CT in an acute stroke context. Further, the linear relationship between contrast agent concentration and attenuation in CT facilitates quantitative analysis of images.

However, disadvantages with CTP exist. These include the delivery of a relatively high radiation dose to the patient during serial imaging of the brain, low signal-to-noise ratio of the images compared to PET and MRI, and high variability in data-processing methods resulting in different quantitative values for perfusion parameters (González 2013).

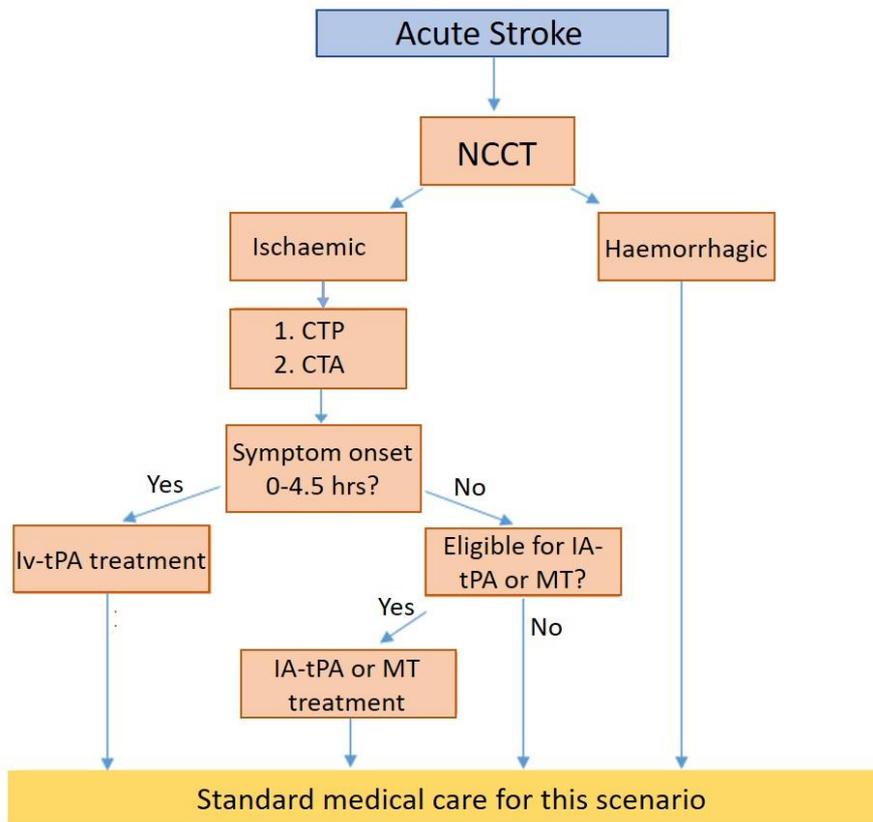


Figure 2.5 Flowchart showing the steps of multimodal CT-based stroke diagnosis and treatment decision-making for an acute stroke patient.

### 2.3 ACUTE ISCHAEMIC STROKE TREATMENTS

The aim of treating acute ischaemic stroke is to stop the conversion of the penumbra to infarction by timely restoration of blood flow to the salvageable tissue (Bhaskar, Stanwell et al. 2018). This way, the final infarct size is limited, and the functional outcome is improved.

Figure 2.5 shows multimodal CT steps and possible treatments that acute stroke patients may receive. Typically, after performing a neurological assessment on an acute stroke patient, NCCT is performed. If no haemorrhage is detected by NCCT, then CTP and CTA are performed. If the patient has a favourable perfusion pattern of ischaemic stroke, the time from stroke onset is less than 4.5 hours (accepted time window) and there is no large artery occlusion, then thrombolysis or intravenous tissue-type plasminogen activator (IV-tPA) is prescribed, provided there is no contraindication for it (González 2013). tPA is a ‘clot-busting’ drug that can dissolve the occlusion. For a large artery occlusion or being beyond the accepted time-window (up to 6–8 hours), endovascular treatment including intra-arterial (IA) thrombolysis or

mechanical thrombectomy (MT) is prescribed (González 2013). In IA thrombolysis, the clot is mechanically manipulated to deliver higher concentrations of the drug to the clot. MT is an invasive surgical procedure involving a microcatheter and other thrombectomy devices for trapping and removing the blood clot.

Although endovascular therapy is beneficial, this is only prescribed in 5–10% of cases (González 2013). Only a small number of specialised stroke centres offer such treatment, and there are risks involved, such as injuring adjacent blood vessels while manipulating a clot. The thrombolysis (using IV-tPA) is easy to perform and can be administered in most medical centres. Therefore, thrombolysis is much more preferable and effective if there is no large artery occlusion and a high probability of a good tissue outcome. However, the narrow window of 4.5 hours, which is established to reduce the risk of haemorrhagic transformation, means that around 85% of acute stroke patients are ineligible for receiving thrombolysis (González 2013). Therefore, if the thrombolysis time window can be extended, many patients could benefit from timely reperfusion, and, as a result, the extent of brain injury could be minimised.

The safety time-window for thrombolysis was developed based on average patient responses to therapy, but the effect of treatment out of this window is not fully understood (Scalzo, Hao et al. 2012). There are reported cases of HT for patients within the safety window and also reports of good clinical outcomes for patients beyond the window. This highlights a need for timely individual risk assessment and the use of patient-specific criteria to make the best possible decision for acute ischaemic stroke patient treatment.

It seems that quantitatively defining favourable mismatch patterns (patients with small infarct core and large penumbra) (Tsai and Albers 2017), finding perfusion markers capable of identifying salvageable tissue and improving methods for accurate quantitative measurements of infarct and penumbra size at the baseline could help move beyond the simple time window approach, enabling patient-specific risk assessment.

## **2.4 QUANTITATIVE ANALYSIS OF CTP IMAGES**

### **2.4.1 CTP imaging analysis overview**

The diagnosis of the ischaemic region in the brain is very challenging, and only trained neuroradiologists can reliably interpret the clinical images (Saad, Abu-Bakar

et al. 2014). Therefore, methods to quantitatively calculate the extent of the injury using computer-aided diagnosis (CAD) are very helpful, potentially differentiating damaged and healthy tissue in cases that clinicians may overlook. Moreover, it can improve the accuracy of lesion detection by providing an alternative opinion for clinical validation (Saad, Bakar et al. 2015).

CTP images are obtained by scanning the brain every one to three seconds for approximately one minute after contrast agent injection. By tracking the passage of the contrast agent, a time attenuation curve (TAC) representing the concentration of the contrast agent over time can be obtained for every voxel in the brain. The shape of the TAC characterises the perfusion properties of brain tissue (Lin, Bivard et al. 2013). Hence, by processing TACs through different mathematical models, various perfusion parameters can be derived quantitatively. Such haemodynamic parameters include (Cuenod and Balvay 2013):

- cerebral blood flow (CBF): defined as the volume of blood passing through 100 g of brain tissue per unit of time (mL/100 g/min)
- cerebral blood volume (CBV): defined as the volume of blood in 100 g of brain tissue (mL/100 g)
- mean transit time (MTT): defined as the mean time (s) from wash-in to wash-out of contrast agent in the region of interest,
- time to peak (TTP): defined as the time (s) taken to reach maximal contrast enhancement in the tissue, which reflects arrival time of the contrast bolus to the tissue.

Mathematical methods for calculating perfusion parameters are divided into deconvolution and non-deconvolution approaches (Konstas, Goldmakher et al. 2009). Deconvolution methods are more accurate but also mathematically more complicated and time-consuming. Non-deconvolution methods are based on simplified assumptions and are easier to implement; however, they are less reliable than deconvolution methods.

According to (Meier and Zierler 1954), the relationship between TACs of tissue voxels ( $C_{tissue}(t)$ ) and TACs of the related feeding artery (arterial input function ( $AIF(t)$ )) can be expressed by:

$$\begin{aligned}
C_{tissue}(t) &= \frac{\rho}{CH} \cdot CBF \cdot (AIF(t) * R(t)) \\
&= \frac{\rho}{CH} \cdot CBF \cdot \int_{\tau}^t AIF(\tau) R(t - \tau) d\tau
\end{aligned} \tag{2.1}$$

Here,  $\rho$  is the attenuation of the brain tissue,  $CH$  is a correction factor for haematocrit levels of capillaries, and ‘\*’ denotes the convolution operator.  $R(t)$  is a residue function varying between zero and one and represents the fraction of contrast agent remaining in a voxel at a specific time point.

$C_{tissue}(t)$  and  $AIF(t)$  are extracted from the CTP source images. Then, by deconvolution of equation 2.1, we can obtain the tissue impulse response function ( $IRF(t)$ ), defined as the multiplication of CBF and  $R(t)$ . Perfusion parameters can then be calculated from  $IRF(t)$  according to:

$$CBF = Max(IRF(t)) \tag{2.2}$$

$$CBV = \int_0^{\infty} IRF(t) dt \tag{2.3}$$

$$MTT = CBV / CBF \tag{2.4}$$

Equation 2.4 is known as the central volume principle. Figure 2.6 shows an overview of different steps in quantitative CTP imaging.

After calculating CBF, CBV and MTT for each voxel of the brain, haemodynamic maps (summary maps) showing the perfusion status of the brain can be generated (Figure 2.7). The other common perfusion parameters, including TTP or  $T_{max}$ , represent the time of peak enhancement for brain tissue voxels, reflecting the delay in contrast agent arrival to the tissue (Kurz, Ringstad et al. 2016). Relative CBF (rCBF) and relative CBV (rCBV) maps can be generated by normalising CBF or CBV for each voxel in the ipsilateral hemisphere to the mean CBF or CBV in the contralateral brain hemisphere. This normalisation provides a useful comparison between the values of haemodynamic parameters in the abnormal and normal brain hemispheres.

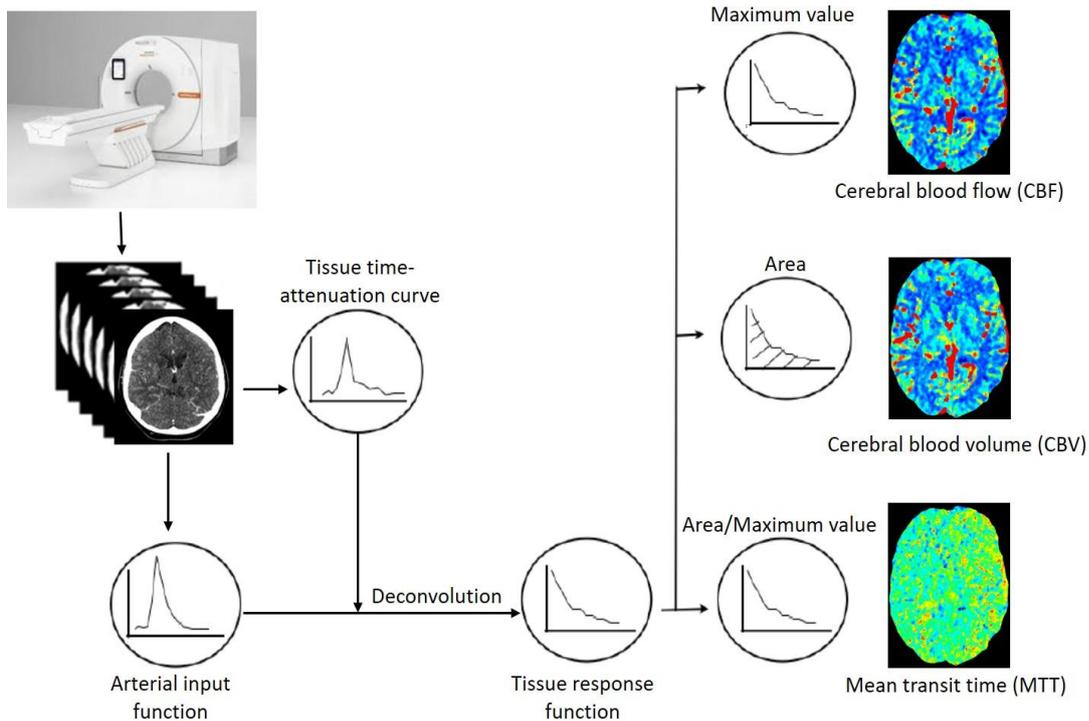


Figure 2.6 Overview of quantitative CTP imaging. Arterial input function and time-attenuation curves of tissue voxels are extracted from the CTP source images and deconvolved to obtain tissue response function. Haemodynamic parameters including CBF, CBV, and MTT are then computed from the tissue response function.

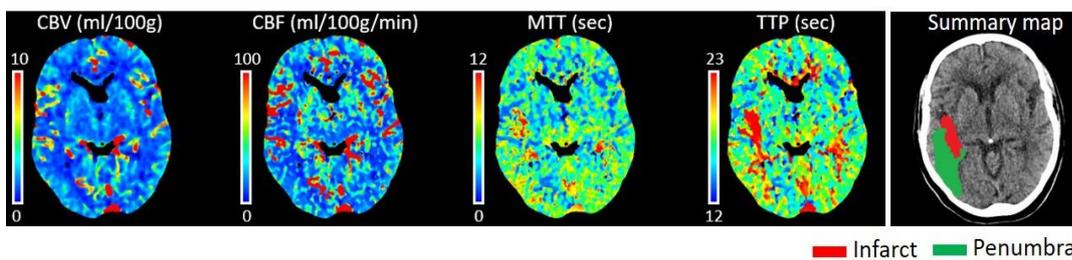


Figure 2.7 An example of CTP haemodynamic maps and summary map. Haemodynamic maps include CBV, CBF, MTT, and TTP maps, indicating voxel-wise haemodynamic parameters. Summary map shows the extent of the infarct core and penumbra obtained by thresholding the haemodynamic parameters.

Infarcted and penumbral areas are typically characterised quantitatively based on thresholding the perfusion maps (Section 2.4.4). For example, the infarct core is defined as the region with severely reduced CBF and CBV and increased TTP,  $T_{max}$  or MTT. The penumbra is defined as the region with moderately decreased CBF, approximately normal CBV and increased TTP,  $T_{max}$  or MTT (Paciaroni, Caso et al. 2009).

## **2.4.2 Arterial input function selection**

The selection of the arterial input function (AIF) influences the value of perfusion parameters and, thus, the estimated size of infarct and penumbral volumes. AIF can be selected globally or locally (Zhu 2013). In global AIF selection, a single artery is used for the whole brain. In the local AIF technique, different AIFs are selected based on a specific metric associated with the volumes surrounding the voxels of interest. Both global and local AIFs are selected from voxels showing high absorption due to a high contrast agent ratio. The peak TAC value at the arrival time is used to calculate a score for each voxel, which is then used to find arteries in the different volumes.

Using a global AIF, calculating perfusion parameters becomes faster and simpler. However, this technique is prone to error due to the delay in the arrival of the contrast agent between the chosen artery and specific voxels. By using a local artery, the error in arrival time can be minimised, and haemodynamic parameters can be estimated more accurately; however, the calculation process is more time-consuming and complex. Another challenge in AIF selection is that manual and automated selection of AIF can cause variation in perfusion parameters (Heit and Wintermark 2016). Manual selection is subject to inter-operator variability; however, automated placement may result in the unsuitable selection of vascular structure or cause volume averaging errors.

In this thesis, Vitrea CTP software (Vital Images, Minneapolis, Minnesota, USA) was used to compute haemodynamic maps. In Vitrea, a global AIF is automatically chosen from a major artery validated by a radiographer or radiologist, and then perfusion maps are calculated via the approach described in Section 2.4.1.

## **2.4.3 Deconvolution methods**

As shown in several studies (Bivard, Levi et al. 2013, Kudo, Christensen et al. 2013), deconvolution methods such as singular value decomposition (SVD), delay-corrected or delay-insensitive SVD (called SVD plus or SVD+), and block-circulant deconvolution result in different perfusion maps and up to 100 mL discrepancy in the infarct or penumbra size estimation. Each method has its pros and cons, and there is a lack of agreement on the most accurate technique for calculating haemodynamic parameters. This is mainly because characterising the accuracy and limitations of the

different mathematical methods is limited by the lack of a well-defined gold standard (Riordan, Prokop et al. 2011). The major problem is that a simultaneous acquisition is needed to correctly compare the perfusion measurements by CTP with other more accurate modalities such as PET or MRI, which is not currently possible.

Some studies have tried alternative solutions to estimate haemodynamic parameters and lesion volumes. In (Ho, Scalzo et al. 2016), a deep learning approach based on pattern recognition from MR images was suggested as an alternative to deconvolution for estimating perfusion parameters. In (Zhu, Gonzalez et al. 2013) and (Wang, Gao et al. 2009), the similarity score between AIF and TACs of brain voxels was used to measure tissue status. They used the correlation coefficient test to compare the patterns of TACs rather than intensity values of haemodynamic parameters. These alternative techniques could potentially improve perfusion image analysis, but their superiority over deconvolution methods is yet to be demonstrated.

In this thesis, the Vitrea software was used to compute perfusion maps using an SVD+ deconvolution algorithm (Yang, Bal et al. 2020). SVD+ is considered more accurate than SVD (Yang, Bal et al. 2020) because it shifts the artery curve to always begin before the arrival of the contrast agent in the tissue curve (Angel 2010).

#### **2.4.4 Infarct and penumbra quantification**

##### *Optimal parameter choice*

The physiology behind impaired perfusion areas and the characteristics of the brain tissue should be taken into account to determine infarct and penumbra regions (Balaji, Babu et al. 2015). In response to a perfusion deficit in the brain tissue, vessels around the perfusion deficit dilate (i.e., autoregulatory vasodilation or autoregulation) to compensate for the decreased perfusion in the tissue. Vasodilation causes a rise in CBV and, as a result, a rise in MTT and TTP due to the central volume principle (equation 2.4). Vasodilation eventually reaches a maximum, and CBF continues to decrease, reducing CBV and causing the tissue to infarct. Since CBV is subject to the effects of autoregulation, MTT and TTP can be better parameters for detecting hypoperfusion. CBF is widely used for infarct and penumbra differentiation (Yu, Han et al. 2016, Peretz, Orion et al. 2017, Chen, Bivard et al. 2019), and the accuracy of differentiation increases when the CBF threshold is adjusted based on the metabolic rate of different brain tissue types. In this thesis, regions with reduced CBF inside the

hypoperfused volume (reduced TTP or MTT volume) are considered infarcted tissue, in line with (Yu, Han et al. 2016).

Although CBV, CBF, MTT and TTP are by far the most common parameters used for infarct and penumbra characterisation, there are alternatives. In (Horsch, Dankbaar et al. 2016) and (Scalzo, Alger et al. 2013), the change in capillary permeability due to blood–brain barrier damage during hypoperfusion was suggested as a potential feature for tissue status assessment. However, permeability is not used in practice due to the long scanning time required for permeability measurement (Hom, Dankbaar et al. 2009) (around four times longer scanning time compared to a normal scan) and its unknown sensitivity for infarct and penumbra differentiation (Fernández-López, Faustino et al. 2012). In (Engedal, Hjort et al. 2017), a change in microscopic aspects of blood flow, such as relative transit time heterogeneity (RTH), is considered a clue to perfusion deficit in the brain. However, RTH computation was based on simplified assumptions, limiting its practicality.

#### *Optimal threshold choice*

The optimal threshold for defining the infarct and penumbral regions is an ongoing topic of research (Wintermark, Flanders et al. 2006, Yu, Han et al. 2016, Chen, Bivard et al. 2019). In (Bivard, Spratt et al. 2011), DWI lesion volume at 24 hours was used to define the extent of penumbra (in patients without reperfusion at 24 hours) and infarct core (in patients with complete reperfusion at 24 hours). DWI at 24 hours was co-registered to baseline (acute phase) haemodynamic maps to define the optimum threshold of haemodynamic parameters for the penumbra and infarct core.

Recently, by applying both voxel and volume-based analysis to the whole brain, traditional thresholds were partially optimised (Yu, Han et al. 2016). In this thesis, I applied the threshold established in (Yu, Han et al. 2016) to the perfusion maps generated by Vitrea to determine the infarct and penumbral regions. The penumbra was defined as the region with a delay of  $\geq 3$  s relative to the healthy hemisphere, and the infarct core as the subregion of the penumbra for which rCBF was  $\leq 30\%$ . The delay was obtained by calculating the difference between the TTP value of each voxel of the ipsilateral hemisphere and the mean TTP of the contralateral hemisphere.

## 2.5 QUANTITATIVE CTP IMAGING: CHALLENGES AND OPPORTUNITIES

### 2.5.1 Motion in CT perfusion

A fundamental assumption of CT image reconstruction is that the object being imaged remains stationary during the image acquisition. In practice, this assumption is easily violated due to patient movement, leading to inconsistent projection data. Inconsistencies between acquired projections result in reconstructed image artifacts (Figure 2.8a) such as blurring, streaking, ghosting, loss of spatial resolution and contrast and increased noise (Yazdi and Beaulieu 2008, Boas and Fleischmann 2012). Such artifacts affect haemodynamic modelling and lesion characterisation (Figure 2.8b), potentially compromising the diagnosis, image interpretation or treatment decisions (Liu, Cuevas et al. 2008). CTP imaging is susceptible to both inter-frame and intra-frame motions (Moghari, Noonan et al. 2019). Inter-frame motion refers to the patient's head movement between the reconstructed frames of the CTP acquisition. Intra-frame motion refers to the motion occurring during the acquisition of a single frame.

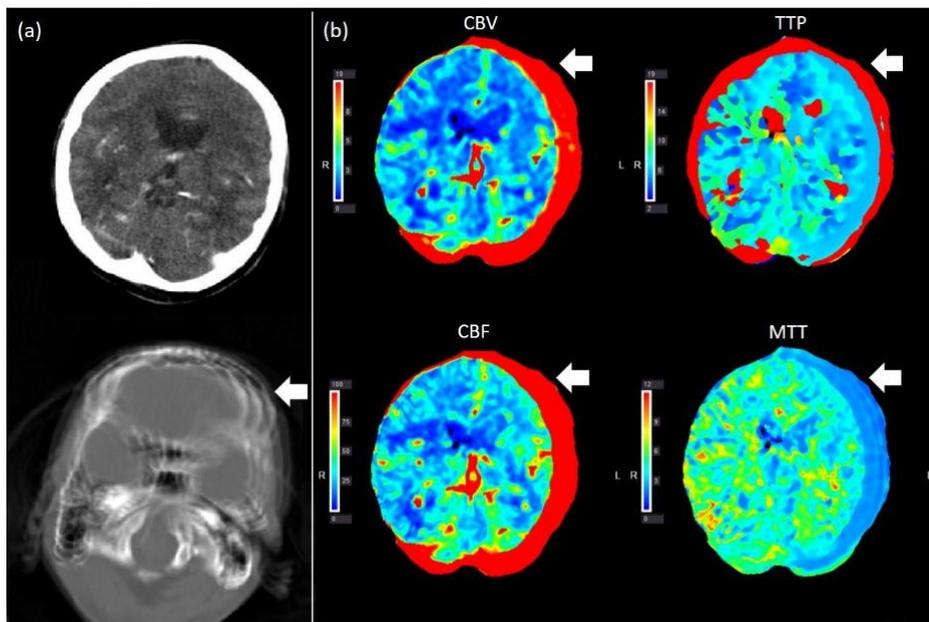


Figure 2.8 Evidence of severe motion and motion-induced artifacts (indicated by white arrows) in (a) CTP images, and (b) the haemodynamic maps.

There are two types of object motion: rigid and non-rigid. Rigid motion describes a situation where distances and angles between all points within an object are maintained; thus, there is no change in the object's size or shape. Non-rigid movement results in a change in the shape or size of the object. Human head movement, which

occurs in CT perfusion imaging, is well approximated as a rigid body motion. Thus, motion correction methods for CT perfusion have exclusively focused on rigid body approaches.

The effects of motion can be reduced using passive head restraint. The development of CT scanners with more detector rows and X-ray sources, which can cover the whole brain at once due to their large axial field of view (FoV), and scanners with faster gantry rotation, can also reduce the likelihood of motion. However, neither passive head restraint nor faster scans can eliminate motion artifacts (Boas and Fleischmann 2012, Kim, Nuyts et al. 2015).

More effective methods should be developed to control or reduce the likelihood of motion during CTP imaging. Achieving this aim will require understanding the motion prevalence, pattern, effects and risk factors (Moghari, Young et al. 2021). Chapter 4 of this thesis addresses this need by developing quantitative methods for motion characterisation and identifying its risk factors. The results of Chapter 4 guided Chapter 6 of this thesis, where a deep learning-based method was developed to optimise the CTP imaging protocol and reduce the probability of motion in the terminal phase of the scan.

### ***Motion correction***

The most common motion correction method in CTP imaging is data-driven, frame-to-frame motion correction. CTP software packages such as RAPID (Menlo Park, California, USA) or Vitrea include inter-frame motion correction algorithms that estimate and correct head movement by registering all CTP timeframes to a reference such as the first timeframe, mean image (Klein, Staring et al. 2010, FILgroup 2014) or NCCT data (Fahmi, Marquering et al. 2014). However, the inter-frame motion correction approaches can only partially compensate for motion, and the underlying continuous (intra-frame) motion in CTP data is left uncorrected (Moghari, Noonan et al. 2019).

The effective, retrospective, intra-frame head-motion-correction methods for rigid body motion in helical CT (Kim, Nuyts et al. 2015, Kim, Sun et al. 2016, Sun, Kim et al. 2016, Sun, Clackdoyle et al. 2017) could have potential value in CTP imaging as well. These methods restore the projection consistency by assigning a modified virtual orbit to the source and detector and applying a 3D reconstruction

algorithm with the assumption that the source and detector follow this modified orbit, not the conventional one that results in motion artifacts.

An accurate estimate of head motion during the image acquisition is needed to construct the modified orbit. Rigid motion can be estimated via data-driven or motion-tracking methods. Data-driven methods estimate the motion directly from the acquired data, while motion-tracking methods use an external tracking device to record the object motion.

There is currently no practical approach for intra-frame motion correction in the time-critical situation of acute stroke CTP imaging. In (Sun, Kim et al. 2016), a data-driven method was developed to simultaneously approximate the head motion and the motion-corrected NCCT image from the raw CT projection data. The algorithm finds the mismatch between the computed forward projection and the measured projection and performs the required pose adjustment via an iterative optimisation algorithm for each view angle. However, the proposed approach was computationally expensive and too slow to be useful in the clinical situation. The performance of data-driven methods can also be affected by the type and severity of motion and other degrading factors such as noise and attenuation (Mukherjee, Pretorius et al. 2010).

### ***External motion tracking***

Motion correction can be more accurate when it is based on motion estimates independent of the raw data. Various external motion-tracking devices provide continuous and real-time motion monitoring, which can be fed into motion correction algorithms.

Stereovision systems have been the most popular external tracking method for estimating head and body motion in CT and PET (Kyme and Fulton 2021). They rely on detecting sparse or dense object features in two or more camera views. Several stereovision systems are commercially available. Some are marker-based, relying on tracking markers (i.e., reflective discs or spheres) attached to the patient's head (Kim, Nuyts et al. 2015, Kim, Sun et al. 2016). Examples of marker-based systems include infrared-based binocular systems and trinocular Optotrak systems (Northern Digital, Ontario, Canada). However, the detachment or non-rigid movement of the markers relative to the head can render the motion data unreliable and hinder the workflow.

The challenges related to the marker-based tracking approaches have motivated a push to markerless systems. Stereovision systems are also suitable for generating depth maps of the object surface, which can be registered across consecutive frames for markerless motion estimation. Examples of markerless systems that have been applied in medical imaging include depth cameras such as the Intel SR522 and D53X family and the Kinect system (Microsoft Corp., USA). Markerless tracking via an accurate, cheap and compact device that can be easily integrated into the clinical CT scanner is desirable for head motion tracking during CTP imaging. Chapter 4 of this thesis characterises a device for this purpose.

### **2.5.2 Radiation dose and noise**

The primary concern with CTP imaging is the radiation dose of the procedure. During CTP imaging, multiple radiation exposures are used to track the contrast agent through the vasculature of the brain. Therefore, the CTP imaging protocol is associated with a radiation dose of around 5–6 mSv (Manniesing, Oei et al. 2015), not excessive on its own, but significant when considered along with the accumulated radiation dose from the multimodal CT regime and other potential follow-up CT scans. Reducing radiation is highly desirable to improve the safety of ionising radiation (Wolterink, Leiner et al. 2017), especially for younger adults and paediatric patients who are more vulnerable to the harmful stochastic effects of ionising radiation (Wolterink, Leiner et al. 2017). However, in CT scanning, there is a trade-off between the amount of X-ray radiation delivered to the patients and the noise in the acquired images. A lower radiation dose corresponds to higher noise in the reconstructed frames.

Several techniques have been suggested to reduce radiation exposure in CTP exams. Related works can be categorised into (i) changing in scanning procedure and (ii) denoising the low-dose images. Examples of changing the image acquisition procedure include reducing the overall scan duration (or truncating the scan) or collecting fewer frames by increasing the frame-to-frame time interval (Wiesmann, Berg et al. 2008, Mo Kim, Haider et al. 2011). The advantage of these methods is that the quality of each image frame is maintained. The disadvantage is that inadequate temporal sampling could cause erroneous perfusion estimations. Applying deep learning-based methods to predict the downsampled frames can improve the estimation of perfusion parameters (Zhu, Tong et al. 2020). In Chapter 6 of this thesis, a deep learning method is developed to predict the late frames from a sequence of early

frames in a truncated scan. Denoising low-dose images has been the more widely reported approach. Related works can be categorised into sinogram domain filtration, statistical iterative reconstruction (SIR) and image post-processing approaches.

Sinogram filtering methods smooth the raw data before image reconstruction, for example, by using a bilateral or statistical nonlinear filter (Wang, Lu et al. 2005). This may result in blurred edges or resolution loss in the reconstructed image. SIR approaches can significantly improve the low-dose CT image quality by modelling the noise behaviour and statistics of the raw data during the reconstruction process and, optionally, incorporating prior information using methods such as nonlocal means (Li, Yu et al. 2014), dictionary learning (Xu, Yu et al. 2012) and total variation (Sidky and Pan 2008). However, SIR approaches are computationally expensive, causing long delays between image acquisition and reconstruction that are not consistent with the time-critical context of stroke. Moreover, both sinogram filtering and SIR methods require access to raw sinogram data from commercial scanners, which is rarely feasible.

Image post-processing approaches that denoise the reconstructed low-dose CT images do not need access to the raw sinogram data. Traditional image-based approaches used for denoising CTP images include filter-based techniques such as k-means-clustering-guided bilateral filtering (Pisana, Henzler et al. 2017) and time-intensity profile similarity (TIPS) filters (Mendrik, Vonken et al. 2010, Mendrik, Vonken et al. 2011). However, filter-based approaches are usually time-consuming, may induce artifacts or cause over-smoothing or edge blurring.

Recently, deep learning-based strategies have appeared promising compared to conventional solutions for addressing various challenging tasks in medical imaging, including denoising (Moghari, Zhou et al. 2019, Moghari, Zhou et al. 2021). Several DL approaches have shown promising results in image restoration and low-dose 2D/3D CT denoising (Wolterink, Leiner et al. 2017, Shan, Zhang et al. 2018). Examples of progress in this area include the use of a generative adversarial network (GAN) to estimate standard-dose CT images from low-dose images (Yang, Yan et al. 2018, Choi, Vania et al. 2019, Yin, Xia et al. 2021), a 2D smoothed dense-convolution network (SDCNet) for standard-dose CTP image synthesis (Liu and Fang 2018) and a convolutional neural network (CNN) model to directly approximate normal-dose CTP haemodynamic maps from low-dose maps (Kadimesetty, Gutta et al. 2018). Although

these methods are promising, there is not enough evidence that they are preferable to traditional approaches. In Chapter 7 of this thesis, we apply and adapt a state-of-the-art deep learning approach to denoising low-dose CTP images.

## **2.6 SUMMARY**

In this chapter, we have reviewed AIS and the use of quantitative CTP imaging for diagnosis and treatment decisions. We have also identified ongoing challenges with the accuracy of haemodynamic analysis in CTP imaging: motion, radiation dose and noise. Each of these areas is addressed in the following chapters. However, first we present a general background on the deep learning approaches and methods used in this thesis.

# Chapter 3: Deep Learning: Concepts and Principles of Operation

---

This chapter introduces the basic concepts of a deep neural network and the main architectural components of these models applied to images. Some key aspects of implementation and the basic architecture of some of the deep learning models used in this thesis are explored. This background lays the groundwork for Chapters 6 and 7, which describe the implementation of different deep neural network-based models for CTP data processing.

## 3.1 DEEP LEARNING

Deep learning is a branch of machine learning that allows the computer to learn complex concepts and patterns directly from data. In deep learning, the features required for a prediction or classification task are automatically learned from the data by optimising many parameters within the model.

Most modern deep learning models are based on artificial neural networks (ANN). The components of these networks are described in the following section.

### 3.1.1 Artificial neural networks

A single building block of deep learning models is an artificial neuron that receives inputs and generates an output. Artificial neurons generate output by performing a linear or nonlinear transformation of the weighted sum of their inputs (Figure 3.1). The weighting of each input is based on its relative importance to the output. Deep artificial neural networks (ANN) typically comprise many neurons concatenated into successive layers, enabling progressively higher-level features to be extracted from the input data (Figure 3.2).

The process of training a deep ANN can be inefficient due to a large number of trainable parameters. This limitation motivated the development of convolutional neural networks (CNN) (Goodfellow, Bengio et al. 2016), which have less trainable parameters and superior performance in various image estimation tasks.

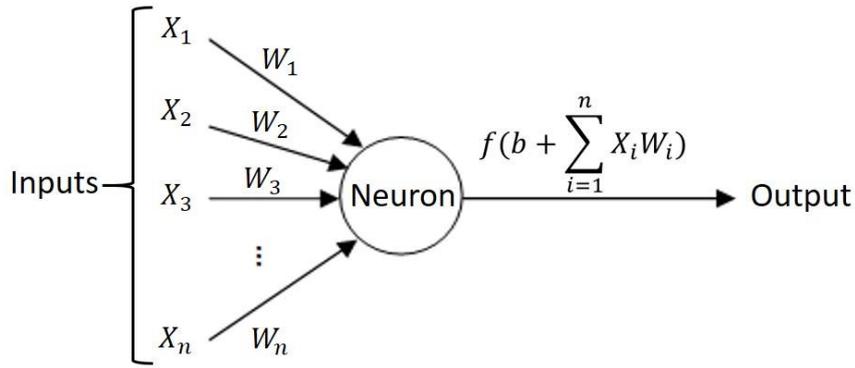


Figure 3.1 Information processing by a single neuron.  $X_i$  and  $W_i$  ( $i \in \{1, 2, \dots, n\}$ ) are the inputs and their associated weights, respectively.  $f(\cdot)$  performs linear or non-linear transformation.  $b$  is referred to as the bias term, which is used to adjust the weighted sum of inputs.

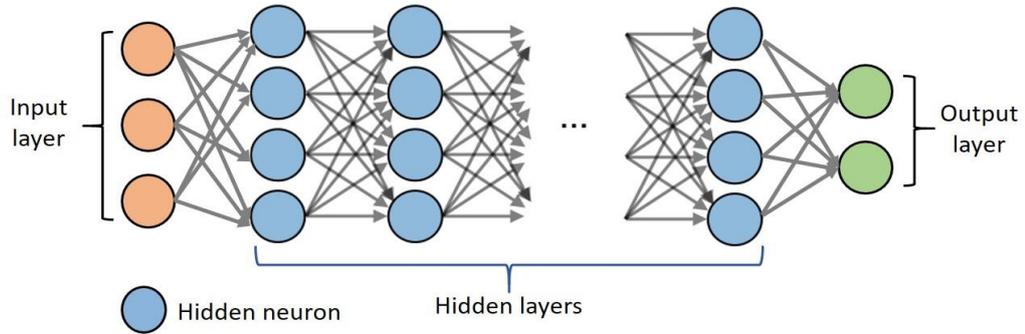


Figure 3.2 Basic architecture of a multi-layer ANN. Many neurons concatenate into successive layers, enabling higher-level features to be progressively extracted from the input data. Grey arrows show the connection between neurons of successive layers.

### 3.1.2 Convolutional neural networks

A CNN (Goodfellow, Bengio et al. 2016) is a variant of the ANN designed to process data with a grid structure, most commonly two- or three-dimensional image data. Unlike a regular ANN, the layers of a CNN have neurons arranged in two dimensions (width and height) or three dimensions (width, height and depth). The CNN layers between the input and output consist of a set of filter kernels, matrices of adjustable weight parameters, which are learned during model training. Compared to ANNs, the unique architecture of CNNs allows for easier training, better generalisability and higher computational efficiency.

A CNN consists of an input layer, hidden layers and an output layer. Like a standard ANN, as shown in Figure 3.2, a CNN maps the model input (images) to the output, representing some prediction from the data. The features are learned through the hidden layers located between the input and output layers. Hidden layers are

typically composed of five standard types of layers, including convolutional, batch normalisation, activation, pooling and fully connected layers, which are described in the following sections. The basic architecture of a multi-layer feedforward CNN is shown in Figure 3.3.

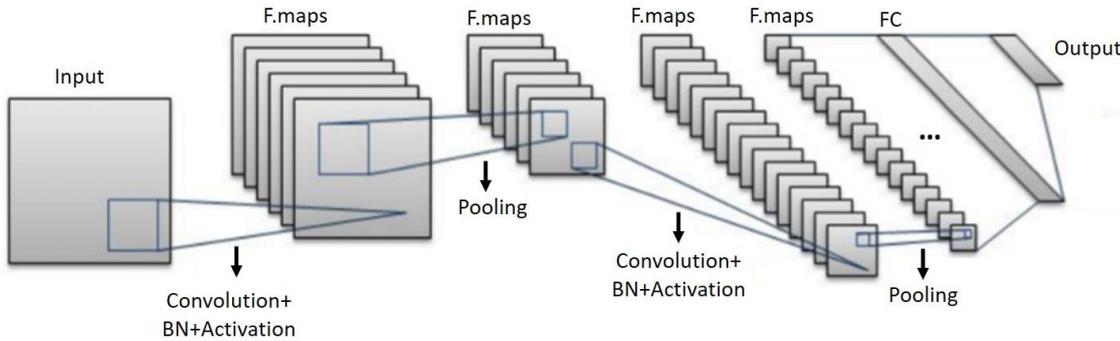


Figure 3.3 Basic architecture of a multi-layer feedforward CNN. BN, FC and F.maps represent batch normalization layer, fully-connected layer and feature maps, respectively.

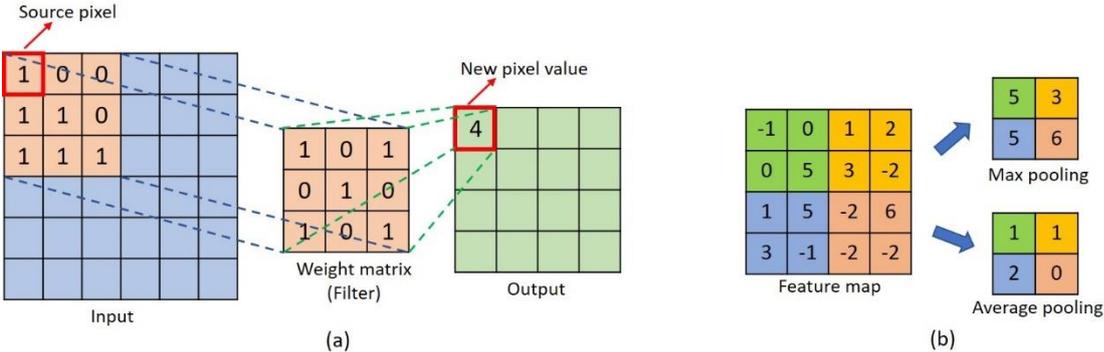


Figure 3.4 An illustration of the convolution operation in a convolutional layer (a) and the max-pooling and average-pooling pooling operation (b). In (a), the source pixel is replaced with the weighted sum of itself and nearby pixels in the output.

**Convolutional layer**

Convolutional layers contain a set of filter kernels with learnable parameters. Each filter is convolved with the input matrix—either an image or the output from a preceding layer—to extract specific features from the input and compute a feature map. Practically, this involves sliding the filter kernel across the input and computing the dot product between the filter and input at each spatial position (Figure 3.4a). The weights in the filter kernel are shared for all local positions, and a fixed number of filters are learned, reducing the number of model parameters for efficiency of learning

and generalisation. The feature maps of all filters are stacked along the depth dimension to build the output of the convolutional layer.

The size of the output feature map ( $I_{out}$ ) of a convolutional layer along each dimension can be calculated from (Goodfellow, Bengio et al. 2016):

$$I_{out} = \frac{I_{in} - F + 2P}{S} + 1 \quad (3.1)$$

where  $I_{in}$ ,  $F$ ,  $P$  and  $S$  are the sizes of the input matrix at each convolutional layer, filter, padding and stride, respectively. Padding is the process of symmetrically adding zeros around the border of the input matrix to control the spatial size of the output. Stride is defined as the pixel shift between applications of a filter kernel over the input matrix during convolution. Filter size, stride value and zero-padding value are three important hyperparameters to be optimised in a convolutional layer (Krizhevsky, Sutskever et al. 2012).

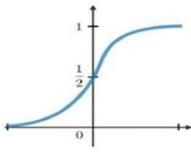
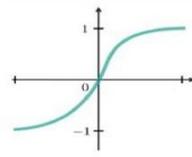
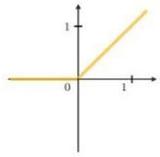
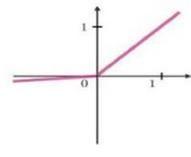
In deep learning models, stacking convolutional layers in succession allows earlier layers near the model input to learn low-level features such as lines and layers deeper in the model to learn higher-level features such as shapes and specific objects.

#### ***Batch normalisation layer***

A batch normalisation layer usually follows each convolutional layer to normalise the output, allow each layer to learn more independently and speed up training.

#### ***Activation layers***

Following the batch normalisation layer, activation functions that are nonlinear operations are applied to the data and feature maps are generated. Activation functions make the data statistically balanced to prevent significantly different ranges of values. Widely used activation functions are sigmoid (logistic), hyperbolic tangent (tanh), the rectified linear unit (ReLU) and Leaky ReLU (Figure 3.5).

Sigmoid (Logistic)	tanh	ReLU	Leaky ReLU
$f(z) = \frac{1}{1 + e^{-z}}$	$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$f(z) = \begin{cases} z & z > 0 \\ 0 & z \leq 0 \end{cases}$	$f(z) = \begin{cases} z & z > 0 \\ \alpha z & z \leq 0 \end{cases}$
			

\*  $z$  is the output of convolutional layer

Figure 3.5 Four commonly used activation functions in deep learning models.

### ***Pooling layer***

Pooling operations are typically applied to feature maps to reduce their dimensionality by down-sampling their spatial size to remove redundant spatial information. Pooling helps to reduce the risk of overfitting and improve the computational efficiency of the model by allowing the use of more filters in deeper convolutional layers without increasing computation per layer. The two common pooling operations are mean-pooling and max-pooling, outputting the average and maximum value for each local patch of the feature map, respectively. An example of mean-pooling and max-pooling is shown in Figure 3.4b.

### ***Fully connected layer***

Fully connected layers are typically used before an output layer. The output of the final pooling or convolutional layer is flattened into a one-dimensional vector and fed into a fully connected layer, where every neuron connects to all neurons in the next layer through a weight vector. This layer combines all the features learned by the network to identify the larger patterns. The features are then mapped into the final output layer, where activation functions such as sigmoid (for binary classification) or SoftMax (for multi-class classification) are applied to compute the probability of each class. In a typical CNN, only two to three FC layers are normally used due to many trainable weights.

## **3.2 TRAINING A DEEP NETWORK**

Training a deep learning model refers to learning the set of model parameters that optimally map the input to the output. ‘Optimal’ here is defined as minimising an

objective function—also called a loss—characterising the difference between the model output and ground-truth data. The model takes in the input, and the weights are iteratively refined over time via a series of forward and backward propagation. The network encodes the input data by passing it through hidden layers during forward propagation, and the loss between the model output and the ground-truth data is measured. During backward propagation, which refers to moving from the output to the input layer, the model parameters are updated via an optimisation algorithm (described in 3.2.1) to reduce the loss. An epoch occurs when the entire training set has passed backward and forward through the model once. The training procedure involves multiple epochs to learn an optimised set of weights that minimises the loss.

Choosing a suitable loss function depends on several factors, including the dataset, the framework of the predictive modelling problem such as classification or regression, and the optimisation method. Commonly used loss functions are the  $L1$ -norm (mean absolute error),  $L2$ -norm (mean squared error) and cross-entropy (a measure of the difference between true and estimated distributions) (Gordon-Rodriguez, Loaiza-Ganem et al. 2020).

### 3.2.1 Optimisation

Optimisation is the process of finding a set of weights and biases that minimise the loss function. Gradient descent (Goodfellow, Bengio et al. 2016) is one of the most common optimisation algorithms in deep learning, where model parameters are updated by moving in the opposite direction of the gradient of the loss function with respect to the model parameters according to:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \alpha \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad (3.2)$$

where  $\boldsymbol{\theta}$  refers to the model parameters.  $\boldsymbol{\theta}_{t+1}$  and  $\boldsymbol{\theta}_t$  denotes the updated and current model parameters, respectively.  $t$  is the number of time steps where  $t \in \{1, 2, \dots, T\}$ .  $\mathcal{L}$  and  $\alpha$  are the loss function and learning rate, respectively. The learning rate is a hyperparameter defining the size of steps at each iteration while moving towards a minimum loss function.

Variants of gradient descent include:

- Batch gradient descent, where overall loss computation and weight update is done after all training samples complete the forward pass.

- Stochastic gradient descent, where weight update is done for each training sample, causes larger fluctuations in the loss function due to the high variance in per-sample weight update.
- Mini-batch gradient descent (Hinton, Srivastava et al. 2012), where the loss of a mini-batch (sub-sample) of training data is calculated in every update.

Gradient descent algorithms maintain a fixed learning rate for all weight updates during training. More efficient optimisation algorithms that can adaptively adjust the learning rate during training have been proposed. Some examples of these optimisation algorithms include Adadelta (Zeiler 2012), AdaGrad (Duchi, Hazan et al. 2011) and Adam (Kingma and Ba 2014). Adam is one of the most popular adaptive algorithms and is used to optimise the proposed deep learning models in this thesis. Adam updates model parameters by updating exponential moving averages of the loss gradient ( $m_t$ ) and the squared gradient ( $v_t$ ) on a current mini-batch. The equations representing the sequence of steps in Adam algorithm are as follows:

$$m_{t+1} = \beta_1 m_t + (1 - \beta_1) \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad (3.3)$$

$$v_{t+1} = \beta_2 v_t + (1 - \beta_2) \left( \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^2 \quad (3.4)$$

$$\hat{m}_{t+1} = \frac{m_{t+1}}{1 - \beta_1^t} \quad (3.5)$$

$$\hat{v}_{t+1} = \frac{v_{t+1}}{1 - \beta_2^t} \quad (3.6)$$

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \frac{\alpha}{\sqrt{\hat{v}_{t+1}} + \varepsilon} \hat{m}_{t+1} \quad (3.7)$$

where the subscripts  $t + 1$  and  $t$  refer to the updated and current estimates of the corresponding parameters, respectively.  $\beta_1, \beta_2 \in [0,1)$ , are hyperparameters controlling the exponential decay rates of  $m_t$  and  $v_t$ .  $\hat{m}_{t+1}$  and  $\hat{v}_{t+1}$  are bias-corrected estimates of  $m_{t+1}$  and  $v_{t+1}$ .  $\beta_1^t$  and  $\beta_2^t$  are  $\beta_1$  and  $\beta_2$  to the power of  $t$ , the number of time steps.  $\varepsilon$  is a very small number to prevent division by zero. The coefficient of  $\hat{m}_{t+1}$  in equation (3.7) is the adaptive learning rate. Good default settings of Adam hyperparameters for training deep neural networks are  $\alpha = 0.001$ ,  $\beta_1 = 0.09$ ,  $\beta_2 = 0.999$  and  $\varepsilon = 10^{-8}$ .

### 3.3 DEEP LEARNING IN MEDICAL IMAGE ANALYSIS

Deep learning approaches have achieved remarkable success in disease detection, characterisation, classifications, prognosis prediction and treatment response evaluation. The success of deep learning models in the medical imaging field is substantially defined by their focus on learning patterns from the input images rather than mathematical modelling (Zhou, Greenspan et al. 2020).

The state-of-the-art deep learning models for medical image analysis usually use variants of CNN (Krizhevsky, Sutskever et al. 2012) and U-Net (Ronneberger, Fischer et al. 2015) for semantic segmentation (Dolz, Gopinath et al. 2018, Li, Chen et al. 2018), classification (Habibalahi, Moghari et al. 2020, Papanastasopoulos, Samala et al. 2020, Rassadin 2020) and object detection (Shin, Roth et al. 2016, El Jurdi, Petitjean et al. 2020). More recently, several CNN-based variational autoencoders (VAE) (Kingma and Welling 2013) and generative adversarial networks (GANs) (Goodfellow, Pouget-Abadie et al. 2014) have opened new avenues for complex image synthesis tasks, including denoising (Biswas, Ghosh et al. 2020, Zhao, Zhou et al. 2020, Yin, Xia et al. 2021), cross-modality image synthesis (Hiasa, Otake et al. 2018, Ben-Cohen, Klang et al. 2019, Yu, Zhou et al. 2020, Li, Huang et al. 2021), image artifact reduction (Ghani and Karl 2019, Koike, Anetai et al. 2020) and segmentation (Evan, Iglesias et al. 2020, Zhang, Yu et al. 2021). For time-sequence data, recurrent neural networks (RNN) (Sherstinsky 2020) have shown impressive results for segmentation of different tissue structures such as abnormal, malignant or benign on images (Alom, Hasan et al. 2018, Bai, Suzuki et al. 2018, Chakravarty and Sivaswamy 2018), reconstruction of high-quality medical images from under-sampled data (Qin, Schlemper et al. 2018) and image quality enhancement (Rajeev, Samath et al. 2019, Lyu, Shan et al. 2021).

Stroke imaging can be a suitable application for deep learning techniques due to patient data and multi-model stroke image datasets that are of significance size (Hinman, Rost et al. 2017, Lee, Kim et al. 2017). Deep learning-based stroke research is diverse, including stroke detection (Chin, Lin et al. 2017, Olive-Gadea, Crespo et al. 2020), infarct/penumbra segmentation (Clèrigues, Valverde et al. 2020, Wang, Shou et al. 2020), treatment paradigms and tissue outcome prediction (Lucas, Kemmling et al. 2018, Nielsen, Hansen et al. 2018, Yu, Xie et al. 2020), onset time classification (Ho, Speier et al. 2017, Zhang, Polson et al. 2021), dose reduction and

denoising (Wu, Ren et al. 2020, Li, Zeng et al. 2021) and imaging protocol optimisation (van de Leemput, Prokop et al. 2019). In this thesis, variants of the U-Net, VAE, GAN and RNN network architectures have been used for CTP image denoising and optimisation of the CTP imaging protocol. In the following sections, we briefly outline the details of these architectures to provide context for Chapters 6 and 7. A detailed description of the VAE model can be found in Section 6.3.2.

### 3.4 U-NET

The U-Net (Ronneberger, Fischer et al. 2015) was originally developed for image segmentation. However, it has since been adapted for other applications, including denoising (Heinrich, Stille et al. 2018), tissue structural analysis (Kerfoot, Clough et al. 2018) and image registration (Li, Zhang et al. 2019).

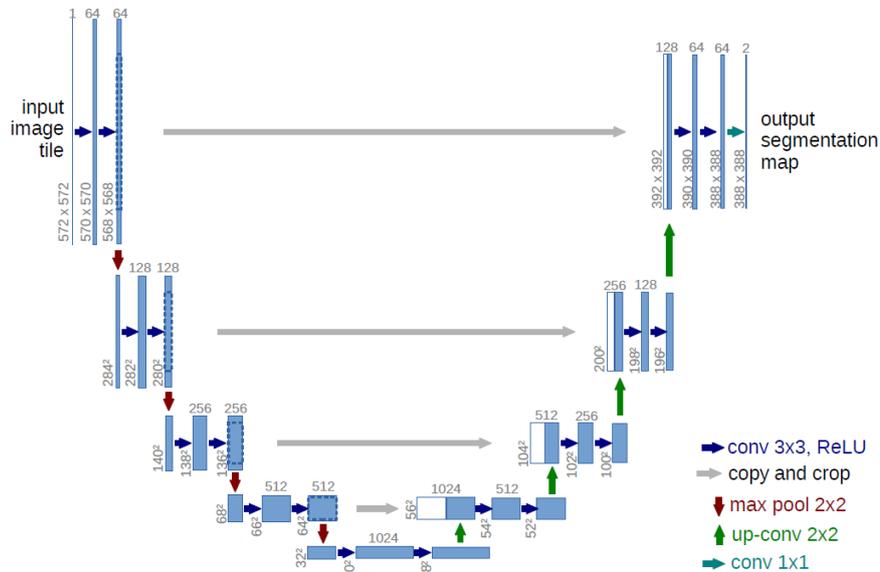


Figure 3.6 Typical architecture of a U-Net. The architecture comprises the contraction path or “encoder” to capture image context and the symmetric expansion path or “decoder” that enables precise localisation of the region of interest in the image. Blue and white boxes represent feature maps and copied feature maps, respectively. The arrows denote different operations. Image courtesy to (Ronneberger, Fischer et al. 2015)

Figure 3.6 shows the basic architecture of a U-Net. The architecture resembles a ‘U’ shape and comprises two paths. First, the contraction path or ‘encoder’ uses a typical CNN architecture comprising a stack of convolutional, activation and max-pooling layers to capture image context. Second, the symmetric expansion path or ‘decoder’ enables precise localisation of the image’s object or region of interest. At

each stage in the expansion path, the feature map is up-sampled and concatenated with feature maps from the corresponding layer in the contraction path via skip connections. Skip connections ensure the features learned at all scales while contracting the image are used to reconstruct it. The concatenation is then followed by a convolution and activation layer.

### 3.5 GENERATIVE ADVERSARIAL NETWORKS

The GAN model is designed for image synthesis tasks such as image segmentation (Zhu, Zhang et al. 2019), image denoising (Yang, Yan et al. 2018), image reconstruction (Zhang, Wang et al. 2018), intra-modality (e.g., MRI-to-MRI) (Yu, Zhou et al. 2019) and inter-modality (e.g., MRI-to-CT) (Pengjiang, Xu et al. 2020) image synthesis. The first GAN model was a generic architecture introduced in 2014 (Goodfellow, Pouget-Abadie et al. 2014). Figure 3.7 shows the generic GAN during training. It consists of a generator network,  $G$ , and a discriminator network,  $D$ , in competition. During training,  $G$  learns to map from random noise  $z \sim p(z)$  to realistic outputs  $G(z)$ , resembling the target image. At the same time,  $D$  tries to differentiate between synthesised images estimated by  $G$  and corresponding real target images  $x \sim p(x)$ . The generator and discriminator are trained alternately to minimise and maximise the objective function in turn. The loss function of the GAN,  $\mathcal{L}_{GAN}(G, D)$ , is defined as:

$$\begin{aligned} \operatorname{argmin}_G \max_D (\mathcal{L}_{GAN}(G, D)) = & \mathbb{E}_{x \sim p(x)} [\log (D(x))] \\ & + \mathbb{E}_{z \sim p(z)} [\log(1 - D(G(z)))] \end{aligned} \quad (3.8)$$

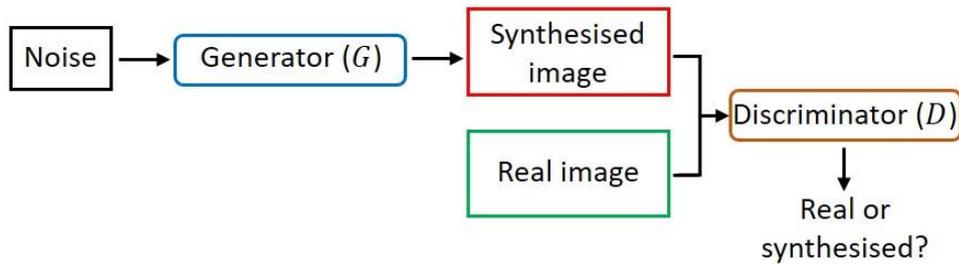


Figure 3.7 A generic GAN training framework.  $G$  learns to map from random noise to a synthesised image resembling the real image. At the same time,  $D$  tries to differentiate between the synthesised image and corresponding real image.

where  $G(\cdot)$  and  $D(\cdot)$  represent the outputs of  $G$  and  $D$ , respectively, and  $\mathbb{E}$  refers to the expected value.

The conditional GAN (cGAN) was first proposed in (Mirza and Osindero 2014) and conditions the learning process on specific input data rather than random noise to provide an additional guide for the synthesis process. Pix2pix (Isola, Zhu et al. 2017) is a variant of cGAN designed for image-to-image synthesis tasks. Figure 3.8 shows the pix2pix training framework.  $G$  takes in an input source image  $x \sim p_{data}(x)$  and learns how to generate a synthesised target image  $G(x)$  that resembles the corresponding true target image  $y \sim p_{data}(y)$ .  $D$  receives two pairs of images: the synthesised pair  $(x, G(x))$  and the real pair  $(x, y)$  and tries to differentiate them. The generator and discriminator compete to minimise and maximise the loss function, respectively.

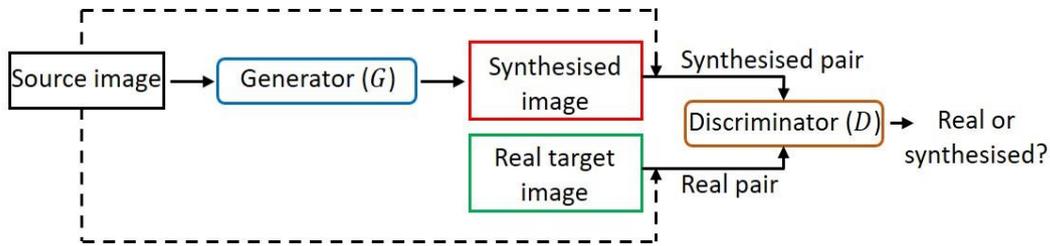


Figure 3.8 The pix2pix training framework.  $G$  learns to map from the source image to a synthesised image resembling the real target image.  $D$  receives two pairs of images: the synthesised pair, including the synthesised and source image; and the real pair, including the real target and source image, and tries to differentiate them.

There are many other variants of the GAN. Some examples are CycleGAN (Zhu, Park et al. 2017), which is used for unpaired image-to-image translation and super-resolution GAN (SRGAN) (Ledig, Theis et al. 2017), which is used to increase the resolution of images.

### 3.6 RECURRENT NEURAL NETWORKS

RNNs (Sherstinsky 2020) are a special type of neural network useful for processing sequential or time-series data. An RNN takes in the previous step output and current input to estimate the current outputs. Figure 3.9 shows there are two main differences between the traditional ANN and the RNN. Firstly, an ANN takes the input data simultaneously and passes them through the hidden layers to estimate the output. For the RNN, however, the data are input at different points in time. Thus, changing the order of the input data causes a significant change in the estimated output. Secondly, in an ANN, we assume that inputs and outputs are independent of each

other. By contrast, the output of an RNN depends on the previous elements within the data sequence.

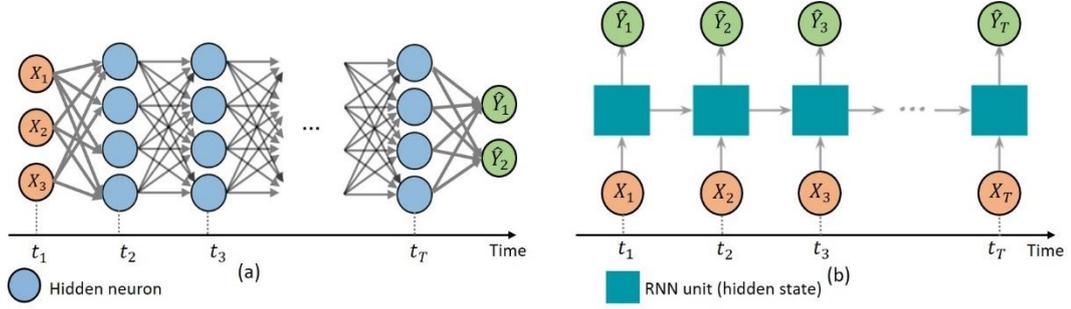


Figure 3.9 Comparison of an ANN and RNN at different time steps  $t_i$  ( $i \in \{1, 2, \dots, T\}$ ). (a) An ANN takes in the input data  $X_1, X_2$  and  $X_3$  at once and passes them through the hidden layers to estimate the outputs  $\hat{Y}_1$  and  $\hat{Y}_2$ . (b) An RNN takes in the input data  $X_i$  ( $i \in \{1, 2, \dots, T\}$ ) at different time steps  $t_i$  to estimate the output  $\hat{Y}_i$ . RNN units are connected via weight vectors.

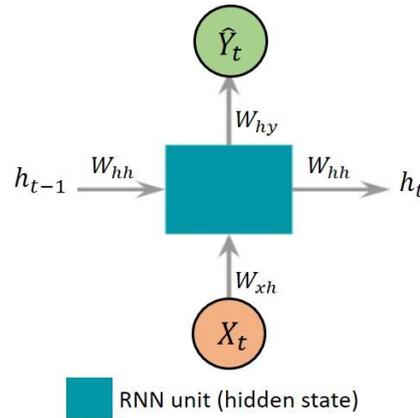


Figure 3.10 One RNN unit.  $X_t$  and  $\hat{Y}_t$  are the input and output at time step  $t$ .  $h_{t-1}$  and  $h_t$  are the outputs of the previous and current hidden states, respectively.  $W_{hh}$ ,  $W_{xh}$  and  $W_{hy}$  are the weight vectors connecting hidden states to each other, the input and the output, respectively.

Figure 3.10 shows one RNN unit (or hidden-state). At each unit, the model takes in two inputs:  $X_t$  ( $t \in \{1, 2, \dots, T\}$ ), which is the input value at time  $t$ , and  $h_{t-1}$  which is the output value from the previous unit at time  $t - 1$ . Two outputs are estimated at each unit at time  $t$ :  $h_t$ , which become the input to the next unit, and  $\hat{Y}_t$ , which is the final output of the current unit. The outputs at a single time step  $t$  are computed according to:

$$h_t = f(W_{xh}X_t + W_{hh}h_{t-1} + b_h) \quad (3.9)$$

$$\hat{Y}_t = f(W_{hy}h_t + b_y) \quad (3.10)$$

where  $W_{hh}$ ,  $W_{xh}$  and  $W_{hy}$  are the weight vectors connecting hidden states to each other, the input and the output, respectively,  $b_h$  and  $b_y$  are the biases, and  $f(\cdot)$  is the activation function. As equations 3.9 and 3.10 show, the current time step computation considers both the current input and the previous hidden layer output as context for prediction. Since the recurrent units hold the information about past values, it is sometimes referred to as internal memory.

The RNN loss function is computed at each time step, and the final loss accumulates all the time step losses. To update the network parameters during the training phase, an RNN uses backpropagation through time (BPTT). BPTT can cause problems, such as exploding and vanishing gradients (Grosse 2017). However, these can be addressed using gradient clipping (Pascanu, Mikolov et al. 2013) and gated recurrent units (GRU) (Chung, Gulcehre et al. 2014), respectively.

### **3.7 SUMMARY**

This chapter presented some general background and essential components of deep neural networks, followed by the basic architectures of the deep learning models used in Chapters 6 and 7 of this thesis. In Chapter 6, a recurrent VAE-GAN model is leveraged for CTP imaging protocol optimisation by reducing the total scan duration while avoiding the truncation of contrast concentration measurement. Chapter 7 follows by outlining the adaptation of a GAN model to a denoising task by generating normal-dose CTP images from low-dose images.

# Chapter 4: Head Movement During Cerebral CT Perfusion Imaging of Acute Ischaemic Stroke: Characterisation and Correlation with Patient Baseline Features

---

Published in *European Journal of Radiology*, 2021: p. 109979: **Dashtbani Moghari, M.**, et al., “Head Movement During Cerebral CT Perfusion Imaging of Acute Ischaemic Stroke: Characterisation and Correlation with Patient Baseline Features”.

## 4.1 ABSTRACT

**Purpose:** To quantitatively characterise head motion prevalence and severity and to identify patient-based risk factors for motion during cerebral CT perfusion (CTP) imaging of acute ischaemic stroke.

**Methods:** The head motion of 80 stroke patients undergoing CTP imaging was classified retrospectively into four categories of severity. Each motion category was then characterised quantitatively based on the average head movement with respect to the first frame for all studies. Statistical testing and principal component analysis (PCA) were then used to identify and analyse the relationship between motion severity and patient baseline features.

**Results:** 46/80 (58%) of patients showed negligible motion, 19/80 (24%) mild-to-moderate motion, and 15/80 (19%) considerable-to-extreme motion sufficient to affect diagnostic/therapeutic accuracy even with correction. The most prevalent movement was “nodding” with maximal translation/rotation in the sagittal/axial planes. There was a tendency for motion to worsen as scan proceeded and for faster motion to occur in the first 15 s. Statistical analyses showed that greater stroke severity (National Institutes of Health Stroke Scale (NIHSS)), older patient age and shorter time from stroke onset were predictive of increased head movement ( $p < 0.05$  Kruskal-Wallis). Using PCA, the combination of NIHSS and patient age was found to be highly predictive of head movement ( $p < 0.001$ ).

**Conclusions:** Quantitative methods were developed to characterise CTP studies impacted by motion and to anticipate patients at-risk of motion. NIHSS, age, and time from stroke onset function as good predictors of motion likelihood and could potentially be used pre-emptively in CTP scanning of acute stroke.

**Keywords:** CTP imaging, acute ischaemic stroke, head movement, motion risk factor

## 4.2 INTRODUCTION

A multimodal CT imaging regime, including non-contrast CT (NCCT) and CT Perfusion (CTP), is the most common approach for diagnosis, treatment decision support and monitoring of therapy response in stroke patients (Ledezma and Wintermark 2009, Morgan, Stephens et al. 2015, Heit and Wintermark 2016). In quantitative CTP analysis, haemodynamic parameters, including cerebral blood volume (CBV), cerebral blood flow (CBF), time-to-peak (TTP) or  $T_{max}$ , and mean transit time (MTT), are derived from CTP source data and, based on the estimated haemodynamic maps, the status of brain tissue and extent of abnormal perfusion areas such as infarct core (irreversibly damaged) and penumbra (at-risk) may be determined (Demeestere, Wouters et al. 2020, Moghari, Zhou et al. 2021). This analysis can, however, be compromised if head motion occurs during a CTP acquisition.

Quantitative CTP analysis assumes that the CTP frames (typically 30-40) collected over a 1-2 min acquisition are spatially aligned and free of motion artifacts, which may not be the case. Head movement can cause inconsistencies between frames (inter-frame motion) and fast motion may also cause inconsistencies in the raw data (intra-frame motion), resulting in streaking, distortion and blurring of reconstructed images (Popilock, Sandrasagaren et al. 2008, Yazdi and Beaulieu 2008). Basing diagnosis and treatment decisions on haemodynamic maps derived from motion-corrupted data is a potential problem (Fahmi, Riordan et al. 2014). Moreover, motion and its effects may go unnoticed, in which case the clinical implications are ignored. In the case of extreme motion that results in uninterpretable images, scans would usually be repeated. However, this means delay in treatment or repeated exposure to additional radiation dose, the latter being particularly significant in younger patients for whom radiation dose carries a greater long-term health risk (Wolterink, Leiner et al. 2017, Moghari, Zhou et al. 2019).

Since cerebral CTP imaging is motion-sensitive, understanding head movement characteristics and the patient-related risk factors affecting head motion is likely to provide insight into how its impact on image quality and analysis might be anticipated, mitigated or controlled. The aim of this study was to analyse the prevalence, severity and causes of head movement in stroke patients during CTP imaging. We develop a quantitative model describing the relationship between head movement and relevant baseline features such as stroke severity, patient age and gender, and time from stroke onset to scan. This model potentially has important implications for identifying motion-related risk factors prior to CTP imaging and informing image interpretation and treatment options.

### **4.3 MATERIALS AND METHODS**

#### **4.3.1 CTP acquisition and dataset**

Our dataset comprised CT perfusion studies of 80 consecutive patients (45 males [56%], 35 females [44%]) scanned at Westmead Hospital, Sydney, during 2018. All were positive stroke cases (no prior stroke) with right or left middle cerebral artery (MCA) occlusion. Average patient age was 69 yr (SD 16 yr, range 35-92 yr). Data collection and analysis was subject to a protocol approved by the Western Sydney Local Health District.

CTP scans were acquired using a Siemens Somatom Force CT scanner (dual source dual energy, 250 ms rotation time, 1120 projection views per rotation) in 4i cine mode at 70 kVp and 200 mAs. The first 25 CT volumes were acquired at 1.5 s intervals and the last 8 volumes at 3 s intervals, resulting in 33 brain volumes acquired over 1 minute. Each CT volume comprised  $22 \times 5$  mm-thick axial slices covering the full brain. Slices were reconstructed in a  $512 \times 512$  matrix with 0.43 mm in-plane resolution. A standard foam headrest was used to limit patient head movement during the acquisition.

#### **4.3.2 Head movement characterisation**

##### *Qualitative analysis*

Head motion from the 80 studies was initially characterised qualitatively by an experienced radiologist (NY), separating studies that were non-diagnostic (Category 3) or impacted considerably by motion (Category 2) from other scans based on visual inspection of the CTP images and haemodynamic maps. Category 2 was defined by

the presence of reconstructed frames with motion artifacts resulting from intra-frame motion or relative shifts/rotations between consecutive frames due to inter-frame motion. Figure 4.1a (frame 6) shows an example of artifact arising from intra-frame motion. Examples of inter-frame motion are shown in Figure 4.1a for the frame 5-6 transition, where an obvious rotation in the axial plane has occurred, and Figure 4.1d for the frame 6-8 transition, where changes are apparent in the nasal cavity due to a patient shift along the Z-axis. For Category 2, the radiologist deemed the intra- and inter-frame motion significant enough to cause uncertainty in interpretation and analysis, even after motion correction. Figure 4.2 shows an example of residual motion artifacts in the haemodynamic maps after applying frame-to-frame CTP registration for a study classified as Category 2. The remainder of studies were sub-categorised into two groups, Category 1 in which there was evidence of inter-frame motion but not considered clinically significant by the radiologist, and Category 0 in which there was no clear evidence of either inter- or intra-frame motion. The subjective criteria used for the categorisations are summarised in Table 4.1.

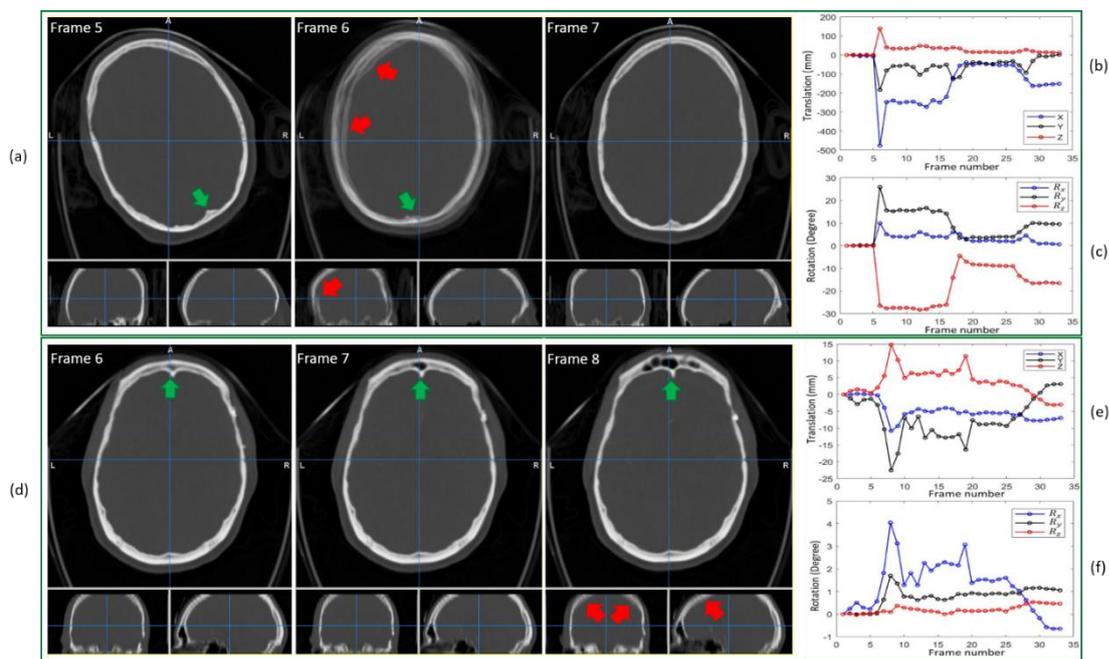


Figure 4.1 Evidence of motion and motion-induced artifacts over 3 consecutive CTP frames for a patient in Category 3 (top row) and Category 2 (bottom row). Each frame shows the axial (main panel), coronal (bottom-left panel) and sagittal (bottom-right panel) views of the middle slice of the brain. Frame-to-frame translations (b, e) and rotations (c, f) are shown to the right. Intra-frame motion artifacts are indicated by red arrows and inter-frame motion (relative rotation/shift of frames) is indicated by green arrows.

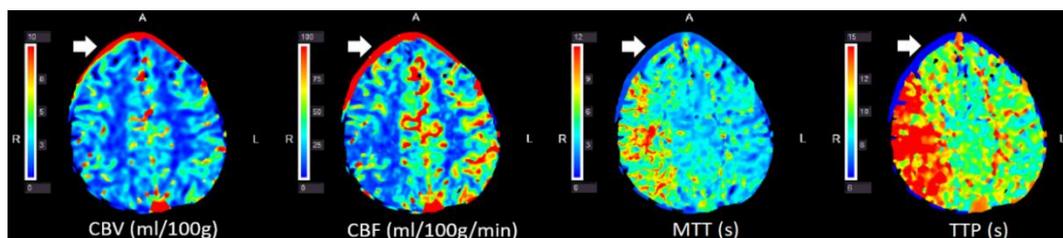


Figure 4.2 Example of residual motion artifacts (indicated by white arrows) in the haemodynamic maps after performing frame-to-frame motion correction. This example was from a study classified as Category 2 (“considerable” movement).

Table 4.1 Subjective head movement categorisation.

Category	Motion Severity	Salience	Clinical Implications
0	Negligible	No visible evidence of intra-frame or inter-frame motion.	None expected.
1	Mild-to-moderate	Evidence of inter-frame motion.	Motion is mostly corrected using frame-to-frame registration. Residual motion may still lead to errors in the quantitative analysis, but this is not considered to be clinically significant by the radiologist.
2	Considerable	Clear evidence of relative shifts between frames from inter-frame motion or motion artifacts from intra-frame motion.	Motion is partially corrected using frame-to-frame registration. The impact of residual motion on adding uncertainty in image interpretation or quantitative analysis is considered to be clinically significant by the radiologist.
3	Extreme	Evidence of severe motion artifacts.	Gross residual errors after motion correction preclude the extraction of meaningful information from images, rendering them non-diagnostic.

### *Quantitative analysis*

To quantify the severity of patient-specific head movement for a given patient, the following steps were used:

- (i) Each reconstructed CTP frame was rigidly registered to the first (reference) frame using Statistical Parametric Mapping software (SPM12) (FILgroup

2014). Head movement relative to the first frame was described by 6 motion degrees-of-freedom (DoF), including three translational components  $X$ ,  $Y$ , and  $Z$ , and three rotational components  $R_x$  (pitch),  $R_y$  (yaw) and  $R_z$  (roll), resulting in  $33 \times 6$  spatio-temporal motion parameters for each patient (Figure 4.1b,c and e,f). The coordinate system in which head motion was measured is shown in Figure 4.3. Default SPM parameter settings were used for the rigid frame-to-frame registrations (Ashburner, Barnes et al. 2014).

- (ii) For each DoF, the top 20% (7/33) of motion-affected frames were identified from step (i). This resulted in 42 identified frames per study, many of which were duplicated across the different DoF.
- (iii) The average magnitude of motion in each DoF was computed from the set of unique frames identified in step (ii). This set ranged in size from 10-24 frames for our 80 studies.

Steps (ii) and (iii) account for the fact that the motion may be limited to certain DoF and result in an estimate of the average motion of a study based on the frames most likely to impact image analysis. This is preferable to averaging motion across all frames which would bias motion estimates to the head pose (position and orientation) corresponding to the largest consistent group of frames.

After computing motion estimates for each study, quantitative characterisation of head movement for each of the four motion categories (Table 4.1) was performed by calculating the average and standard deviation (SD) of motion for all studies in each category. Additionally, we computed the average head movement of all patients at each frame to identify frame-to-frame and temporal trends of head motion in CTP studies.

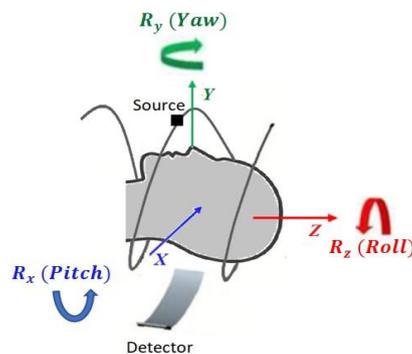


Figure 4.3 Rigid-body head movement and coordinate system definition. The X-ray source moves in a helical trajectory.  $Z$  is along the bed direction.

Table 4.2 Patient baseline features and grouping (designated Gp1 to Gp4) used in this study.

<b>Feature</b>	<b>Gp 1</b>	<b>Gp 2</b>	<b>Gp 3</b>	<b>Gp 4</b>
<b>Age (yr)</b>	30 – 50	51 – 75	≥ 76	-
<b>Gender</b>	Female	Male	-	-
<b>NIHSS<sup>(a)</sup></b>	1 – 4	5 – 14	15 – 24	25 – 42
<b>Time from stroke onset (hr)<sup>(b)</sup></b>	≤ 4.5	> 4.5	-	-

<sup>(a)</sup> National Institutes of Health Stroke Scale

<sup>(b)</sup> 4.5 hr is widely recognised as a suitable threshold for eligibility for thrombolytic drugs.

### 4.3.3 Relationship between head movement and patient baseline features

#### *Patient baseline feature selection and classification*

Four patient baseline features (Table 4.2) were considered for correlation with the prevalence and severity of head movement during CTP scans: two demographic features, patient age and gender, and two medical features, time from stroke onset and the National Institute of Health Stroke Scale (NIHSS). The NIHSS (Lyden 2017) is used to objectively quantify the impairment caused by stroke and ranges from 0 – 42, a higher score indicating greater stroke severity. All features were divided into 2 – 4 groups (designated Gp1 to Gp4). Gender (0 for female and 1 for male) and time from stroke onset (0 for < 4.5 hr and 1 for > 4.5 hr) were considered binary variables. The 4.5 hr cut-off for time from stroke onset is a commonly used threshold determining whether a patient is eligible for less invasive thrombolysis therapy (Zaheer, Robinson et al. 2011). Age and NIHSS were considered as categorical variables. We used three age groupings (30 – 50 yr, 51 – 75 yr, and ≥ 76 yr) and, based on (Brott, Adams Jr et al. 1989), four groups for NIHSS: NIHSS 1 – 4 (mild stroke), NIHSS 5 – 14 (mild-to-moderate stroke), NIHSS 15 – 24 (severe stroke), and NIHSS 25 – 42 (very severe stroke).

#### *Statistical analysis*

The Kruskal-Wallis test (McKight and Najab 2010), a non-parametric equivalent to the one-way analysis of variance (ANOVA) test, was used to identify statistically significant relationships between head motion categories and individual baseline features. A principal component analysis (PCA) (Abdi and Williams 2010) was also performed to evaluate possible relationships of a linear combination of the

baseline features with different head movement categories. Here we considered only the first principal component (PC1), which preserves the maximum variance of the features, and assessed the contribution of the different baseline features to PC1 using the PCA loadings- coefficients of the linear combination of the feature variables from which the principal components are constructed. The Kruskal-Wallis test was used to assess the statistical significance between PC1 and the motion categories. Covariance was used to assess the joint variability of individual baseline features and PC1 with the motion categories. Here, the sign of the covariance indicates the tendency (positive, negative) in the linear relationship between the variables. In all statistical analyses, the Shapiro–Wilk (SW) normality test (Yap and Sim 2011) was used to assess the data distribution.

#### 4.4 RESULTS

##### 4.4.1 Head movement characterisation

Figure 4.4 shows the results of the qualitative head motion categorisation. Approximately 60% (46/80) of patients had negligible motion, a quarter (19/80) exhibited mild-to-moderate motion, and nearly 20% had considerable or extreme motion.

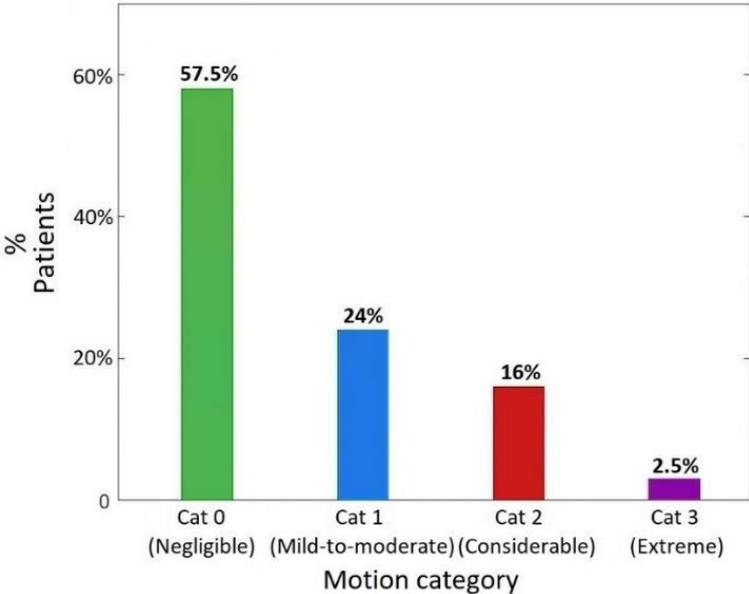


Figure 4.4 Percentage of stroke patients in each motion category (0-4) ( $N = 80$ ).

Table 4.3 Quantitative characterisation of head movement within the patient cohort for each motion category, using the method described in Section 4.3.2/Quantitative analysis. Values represent the mean motion amplitude with respect to the first frame. Errors represent  $\pm 1$  standard deviation.

Motion	$\bar{X}$ (mm)	$\bar{Y}$ (mm)	$\bar{Z}$ (mm)	$\bar{R}_x$ (deg)	$\bar{R}_y$ (deg)	$\bar{R}_z$ (deg)
<b>Cat 0</b>	$0.6 \pm 0.5$	$1.1 \pm 0.7$	$0.7 \pm 0.4$	$0.2 \pm 0.1$	$0.1 \pm 0.1$	$0.1 \pm 0.1$
<b>Cat 1</b>	$2.1 \pm 1.4$	$3.6 \pm 1.9$	$2.4 \pm 1.6$	$0.6 \pm 0.4$	$0.3 \pm 0.2$	$0.3 \pm 0.3$
<b>Cat 2</b>	$4.6 \pm 2.0$	$8.4 \pm 5.9$	$5.9 \pm 5.6$	$1.5 \pm 1.3$	$0.9 \pm 0.9$	$0.7 \pm 0.7$
<b>Cat 3</b>	$206.1 \pm 37.2$	$82.3 \pm 10.2$	$33.4 \pm 18.0$	$4.6 \pm 0.9$	$12.4 \pm 2.8$	$24.3 \pm 1.2$

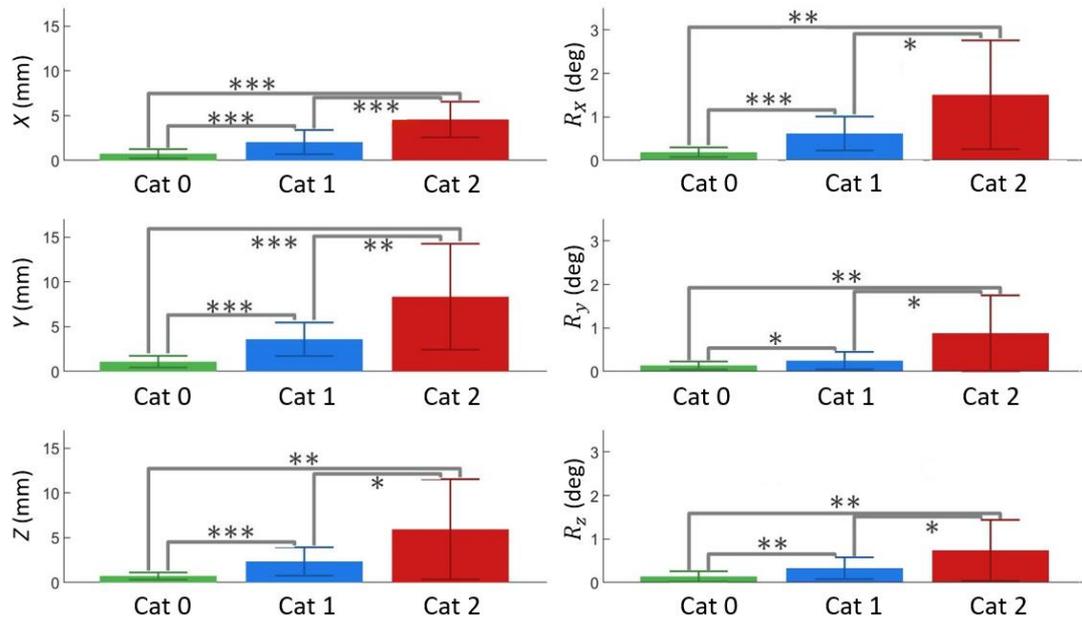


Figure 4.5 Statistical comparisons between average motion parameters measured for the different motion categories. Comparisons were performed using an unpaired Student's t-test. Significance is indicated by: \* ( $p<0.05$ ), \*\* ( $p<0.01$ ) and \*\*\* ( $p<0.001$ ). Error bars represent 1 standard deviation.

Table 4.3 summarises the quantitative results based on frame-to-frame registration for the different head motion categories. Category 0 was characterised by average translations  $\leq 1$  mm and average rotations  $\leq 0.2$  deg; Category 1 by average translations 2 – 4 mm and average rotations  $< 0.5$  deg; Category 2 by average translations 5 – 10 mm and average rotations  $\sim 1$  deg; and Category 3 by average translations several cm or more and average rotations  $> 5$  deg. The most sensitive DoFs for Categories 0 – 2 were translational motion along the Y-axis and rotation about the X-axis, reflecting a tendency of head ‘nodding’ motion. For the 2 patients in Category

3, motion was predominantly in the axial plane, i.e. translation along the  $X$ -axis and rotation about the  $Z$ -axis, reflecting head shifting and shaking. Figure 4.5 shows statistical comparisons between the three most populous motion categories. For all categories and all parameters, the differences were statistically significant.

Figure 4.6 illustrates the absolute translational (a-c) and rotational (d-f) head movement parameters as a function of CTP frame for 78/80 patients (the two patients with non-diagnostic images were excluded to improve visualisation). Translation and rotation range was limited to 0 – 20 mm and 0 – 5 deg, respectively. The median absolute motion versus time is shown in Figure 4.6g,h and show a clear trend of increasing motion in all DoF over the course of a scan.

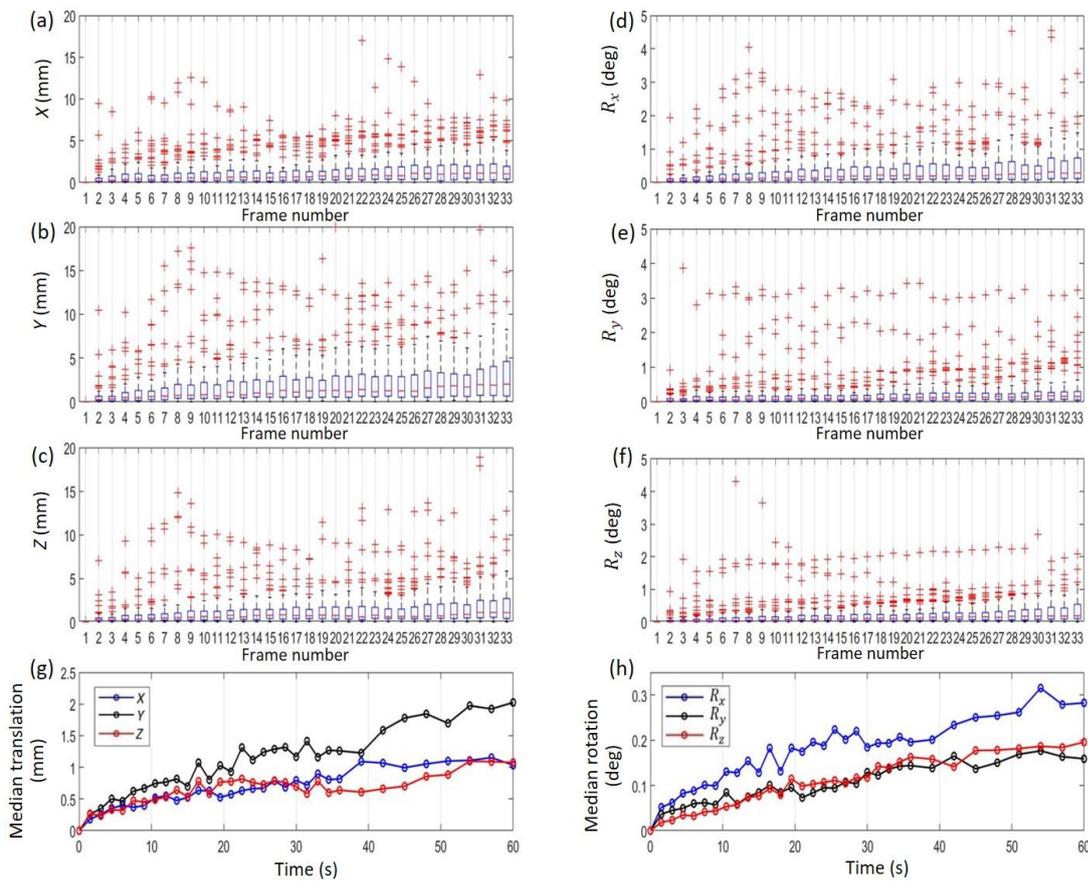


Figure 4.6 Box-whisker plots showing absolute translational (a-c) and rotational (d-f) head motion parameters with respect to the first frame, plotted as a function of CTP frame number for 78 patients. The bottom row shows the median absolute translational (g) and rotational (h) motion parameters against time.

#### 4.4.2 Head movement relationship with patient baseline features

The head movement characterisation results (Section 4.4.1) showed that extreme motion is rare (2/80). Therefore, for analysis of the relationship between head movement and patient features we combined Categories 2 (“considerable”) and 3 (“extreme”) into a single category.

The Kruskal-Wallis comparisons between individual features and the motion categories indicated a statistically significant relationship between head motion amplitude and all patient baseline features except gender: NIHSS ( $p = 0.0071$ ), time from stroke onset ( $p = 0.0151$ ), age ( $p = 0.0212$ ), gender ( $p = 0.2522$ ). The motion groups were related positively with NIHSS and age (positive covariance), and inversely with time from stroke onset (negative covariance).

The relationship between motion severity and a multi-component feature derived from PCA analysis is shown in Figure 4.7. Figure 4.7a shows the PC1 distribution of all studies in each motion category, indicating a highly significant positive relationship between PC1 and head motion ( $p = 0.0006$ ). Figure 4.7b shows the absolute value of the PC1 loading for each individual baseline feature. The raw PC1 loadings for age, gender, NIHSS and time from stroke onset were 0.46, -0.1, 0.87 and -0.18, respectively, indicating that NIHSS and age had the highest contribution to PC1.

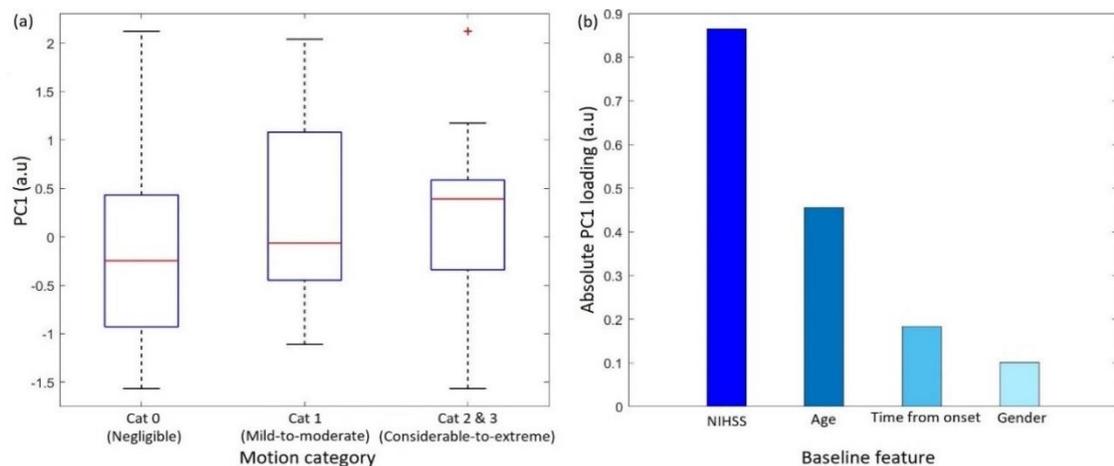


Figure 4.7 PCA-based correlation analysis using box-whisker plots. (a) First principal component (PC1) versus different motion categories; (b) absolute PC1 loading for different patient baseline features.

## 4.5 DISCUSSION

This study analysed the severity and potential correlates of head motion during ~60 second CTP scans of acute ischaemic stroke patients. Using statistical analyses,

head motion severity was correlated with patient baseline features in isolation and linear combination (PCA analysis) with the aim of identifying patient-related risk factors that tend to increase the likelihood of head motion, and which could potentially be used pre-emptively in CT scanning of stroke cases.

The qualitative classification of studies according to head motion was relatively coarse (4 categories), defined according to the perceived likelihood of residual motion-induced artifacts having a clinically significant impact on image interpretation and haemodynamic analysis after motion correction. This classification revealed that approximately 40% of patients had mild to extreme head movement during CTP scanning and that in half of those cases the residual artifacts had likely clinical implications. This reinforces the general consensus in the literature that clinically relevant head motion is common in CTP imaging despite the short scanning time. According to our study, motion correction could have a positive clinical impact on at least 20% of CTP studies.

The quantitative head motion analysis (Figure 4.5) confirmed the validity of the qualitative classification by indicating either significant or highly significant differences between all motion categories. Furthermore, on average, head movement occurs predominantly in the sagittal plane, mostly as translation, with greatest magnitude likely to be along the *Y* direction (Table 4.3). Average head rotation in all categories (excluding the extreme category) was < 2 deg, with the greatest average rotation amplitude about the *X*-axis (head “nodding”).

The temporal behaviour of motion (Figure 4.6g,h) indicates a net increase in motion as a scan progresses. More subtly, the data show that the greatest likelihood of motion is near the start (in the initial 15 s) of a scan, with motion tending to remain quite stable after this. The increased likelihood of motion at the start could be due to discomfort from contrast agent injection or associated with patient settling. The steeper slope over the first 15 s is indicative of relatively higher speed of head movement compared to later in the study. Based on the increasing interquartile range (Figure 4.6a-c and d-f) as the frames advance, we can further say that the motion patterns of patients tend to become more and more different as scanning proceeds. These data imply that methods to shorten scan duration are likely to reduce the overall amount and impact of head movement on CTP scans.

Analysing the relationship between patient baseline features and head motion categories revealed some interesting dependencies. The best predictor of motion severity was NIHSS ( $p < 0.01$ ) followed by time from stroke onset ( $p = 0.0151$ ) and age ( $p = 0.0212$ ). Gender was not predictive of motion ( $p = 0.2522$ ). Although NIHSS and time from stroke onset to scan have been recognised as important parameters for the clinical management of stroke patients, we are not aware of these parameters being explicitly linked to the likelihood of motion during scanning. When a linear combination of features was used via PCA analysis, the motion dependency was even stronger ( $p < 0.001$ ), indicating that PC1 is a better predictor of head motion severity than any individual feature considered. The relative contribution of each feature to the linear combination, characterised by the PCA loadings, implicated NIHSS and age as the 2 most important features, respectively. In both cases, the correlation was positive – the greater the NIHSS/age, the more severe the expected motion. Overall, these results suggest a multi-component feature derived using an approach like PCA may be a useful predictor of motion in CTP studies.

One obvious implication of our study is that older patients and patients with a higher NIHSS are more at risk of having motion-affected scans. Thus, measures should be taken to limit or correct for motion artifacts for this group in particular, and images from these studies should be interpreted with greater care. A more complete exploration of the potential uses of our prediction model will be considered in future work.

Several methods are used to reduce the amount of motion during CTP imaging. The most common approach is passive head restraint and improving patient comfort. Sedation can prevent motion but is not recommended in acute stroke imaging since it may cause haemodynamic effects which bias clinical assessment of neurological status (Odo, Crippa et al. 2016). Another option is faster CT scanners with faster gantry rotation or more detector rows and X-ray sources to reduce the probability of motion artifacts. In practice, it is difficult to fully eliminate the impact of head motion without some form of motion correction. CTP imaging software packages typically include basic frame-to-frame motion correction as default, where each CTP image frame is registered to a reference frame such as the first timeframe or another image such as the NCCT (Fahmi, Marquering et al. 2014). However, inter-frame head motion can imply underlying continuous (i.e. intra-frame) motion, which is not currently corrected for in

CTP data (Moghari, Noonan et al.). The development of retrospective intra-frame motion correction methods that are fast enough in the time-critical acute stroke context can potentially improve quantitative haemodynamic analysis.

Our dataset was limited to CTP studies of 80 consecutive positive stroke patients. Approximately 30% of patients presenting to the hospital with stroke-like symptoms are stroke mimics (Hand, Kwan et al. 2006) and usually undergo acute stroke CT protocols for diagnostic testing. The suggested analysis and predictive model should therefore be tested on a larger dataset including consecutive CTP studies of all suspected stroke cases (stroke mimics and positive stroke cases) to improve its specificity and practicality.

#### **4.6 CONCLUSION**

This study characterises the prevalence, severity and causes of head movement in CTP scans of acute stroke patients. Our results confirm that there are several clear factors (e.g. NIHSS, age, time from stroke onset to imaging) which function as good predictors of motion likelihood during CTP studies. The value of such prediction on CTP efficacy and on adjusting clinical workflows should be explored in future work.

# Chapter 5: Characterisation of the Intel RealSense D415 Stereo Depth Camera for Motion-Corrected CT perfusion Imaging

---

To be submitted to *Physics in Medicine and Biology*: **Dashtbani Moghari, M.**, et al., “Characterisation of the Intel RealSense D415 Stereo Depth Camera for Motion-Corrected CT perfusion Imaging”.

## 5.1 ABSTRACT

Despite using a protocol lasting only 1 min, head movement is problematic in cerebral CT perfusion (CTP) imaging of acute stroke, causing data inconsistencies that can compromise the quantitative haemodynamic modelling driving treatment decisions. Thus, retrospective motion correction is usually a default pre-processing step. Frame-to-frame registration is the most common form of retrospective correction but neglects the fact that motion is generally continuous, not discrete. External tracking devices provide continuous motion monitoring and thereby the opportunity to fully correct the acquired data for motion, both between frames and within frames. The aim of this study was to characterise the Intel D415 stereo depth camera, a compact low-cost markerless tracking device, in terms of its suitability for retrospective CTP motion correction. The characterisation consisted of (i) thermal stability and noise jitter performance for a realistic static head phantom; (ii) validation of static and dynamic pose measurement accuracy against ground-truth robotic motion; and (iii) adaptation of the Intel D415 to a clinical CT scanner and validation of dynamic human head pose measurement accuracy against the marker-based OptiTrack. The results showed that jitter was stable and thermally-induced pose drift was  $\leq 1.5$  mm and  $\leq 0.5^\circ$  during the first 10 - 20 min, after which it became stable. For static poses, the mean difference between the Intel D415 and commanded robot position was  $\leq 1.24 \pm 0.01$  mm and  $\leq 0.68 \pm 0.01^\circ$ . For dynamic poses measured while the phantom travelled smooth continuous trajectories with median speed  $0.031$  ms<sup>-1</sup> (range 0 -  $0.500$  ms<sup>-1</sup>), the mean root-mean-square-error (RMSE) was  $\leq 1.40 \pm 0.12$  mm and  $\leq 0.24 \pm 0.02^\circ$ . When tracking a simulated patient head trajectory derived from a clinical CTP scan, the mean RMSE was  $\leq 0.86 \pm 0.03$  mm and  $\leq 0.16 \pm 0.03^\circ$ . Tracking the head motion of a human volunteer inside a clinical CT scanner resulted in a mean RMSE between

the Intel D415 and OptiTrack of  $\leq 2.72 \pm 0.24$  mm and  $\leq 0.55 \pm 0.07^\circ$ . Overall, our results suggest that a single D415 tracking system can achieve promising pose estimation accuracy, though slightly worse than typical CT resolution for brain scans, including CTP. This error is likely to be reduced to a practical level by combining multiple devices and this should be investigated in future work.

**Keywords:** Computed tomography, CT perfusion imaging, head motion, motion tracking, The Intel D415

## 5.2 INTRODUCTION

Patient head movement during cerebral CT perfusion (CTP) imaging of suspected acute stroke cases introduces image artifacts that can compromise lesion characterisation and the ensuing treatment decisions for at least a fifth of CTP studies (Moghari, Young et al. 2021). Compared to a standard neurological CT scan that completes within a fraction of a second, the probability of head motion increases in CTP imaging due to the dynamic acquisition of 30-40 frames over a relatively long scanning time of ~1-2 minutes (Moghari, Young et al. 2021).

Head movement can cause CTP data inconsistencies in two different ways. The first is fast and discrete motion resulting in rigid misalignment between reconstructed frames, termed inter-frame motion corruption. The second is continuous motion resulting in data inconsistencies within individual frames, termed intra-frame motion corruption. Motion-corrupted frames may manifest visible artifacts like blurring, streaking, and ghosting (Yazdi and Beaulieu 2008, Boas and Fleischmann 2012), but the more serious impact is on the estimation of perfusion parameters derived from haemodynamic modelling (Fahmi, Marquering et al. 2014). These parameters, including cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT), and time-to-peak (TTP), are crucial for the lesion characterisation that drives treatment decision-making (Moghari, Zhou et al. 2021). To improve the performance of image-based stroke analysis, mitigating or correcting for patient head motion during imaging is, therefore, highly desirable.

The use of restraining devices such as a foam headrest can limit patient head movement during image acquisition but does not completely prevent it (Hanzelka, Dusek et al. 2013). More restrictive methods increase discomfort, reducing their acceptability by patients and potentially exacerbating movement (Montgomery,

Thielemans et al. 2006). An alternative approach to reduce the probability of motion artifacts is to use CT scanners with faster gantry rotation or more detector rows and/or X-ray sources in order to image the brain more rapidly (Boas and Fleischmann 2012, Kim, Nuyts et al. 2015). However, the timescale of the passage of contrast agent through the brain and the need to image the washout phase for modelling represents an important physiological constraint: although each frame can acquire faster using advanced hardware (reducing intra-frame motion), the total duration of the protocol – and thus the propensity for motion – remains unchanged.

Due to the limitations of motion prevention strategies, use of retrospective motion correction is a standard pre-processing step in the CTP image data analysis pipeline. By far the most common motion correction strategy is to rigidly register each individual CTP frame to a reference frame, for example the initial frame or another image such as a non-contrast CT (Moghari, Noonan et al. , Fahmi, Marquering et al. 2014). This data-driven frame-to-frame alignment is, however, only useful to correct inter-frame motion and cannot correct intra-frame motion caused by continuous head movement. This is where external motion tracking provides an advantage.

The rationale for external motion tracking in CTP imaging is two-fold: accuracy and continuity. Compared to data-driven techniques, external tracking can deliver more accurate motion estimates because they are independent of the motion-corrupted data. Continuous tracking is important because it is not reasonable to assume all motion is inter-frame. In fact, it is much more likely that individual frames are corrupted. Interestingly, all reports of motion-corrected CTP imaging focus on inter-frame motion, and the impact of intra-frame motion on the haemodynamic modelling of CTP data is not well understood.

Optical or infra-red (IR) motion tracking that relies on the detection of reflective discs or spheres attached to the head can provide highly accurate estimates of rigid-body head motion typically at 10-100 Hz over the course of a scan. This is contingent on markers remaining rigidly coupled to the head (Kim, Nuyts et al. 2015, Kim, Sun et al. 2016), which in practice is difficult to achieve using non-invasive attachment methods. Challenges related to the reliability and practicality of marker-based systems have motivated a push to markerless systems (Olesen, Sullivan et al. 2012, Kyme, Se et al. 2018). In a markerless approach, features used for object pose estimation are typically determined distinctive nature features or patterned (structured) illumination

(Olesen, Sullivan et al. 2012, Kyme, Aksoy et al. 2020). Computer vision techniques can then be used for stereo reconstruction of sparse landmarks or point clouds from which changing object pose may be estimated (Wilm, Olesen et al. 2011). Obtaining dense point clouds characterising the surface of an object at high frame rate has been made possible through a variety of consumer-grade devices called depth sensors such as Microsoft Kinect (Heß, Büther et al. 2015, Noonan, Howard et al. 2015, Silverstein and Snyder 2018) and Intel depth cameras (Bauer, Seitel et al. 2013, Lindsay, Mukherjee et al. 2015, Bier, Ravikumar et al. 2018, Siena, Byrom et al. 2018).

Low-cost and compact depth sensors are an attractive option for integrating into medical imaging applications provided they meet the accuracy requirements. The Intel RealSense depth sensors were introduced onto the market in 2015 and were quickly deployed in applications such as security (Bock 2018), posture and gesture recognition (Arachchi, Hakim et al. 2018, Chiang and Fan 2018, Liao, Li et al. 2018) and healthcare (Siena, Byrom et al. 2018, Aoki, Suzuki et al. 2019). The D400-series, introduced in 2018, feature hardware-accelerated stereo vision calculations and support the use of unstructured light. To date, these sensors have not been thoroughly characterised for CT applications such as motion-corrected CTP imaging.

The aim of this study was to investigate the feasibility of applying the Intel RealSense D415 stereo depth camera for accurate and reliable motion estimation to enable CT/CTP motion correction. We characterise the performance of this device with respect to thermal stability and jitter, static and dynamic six degree-of-freedom (DoF) pose accuracy, and adaptability to the clinical setting.

## **5.3 MATERIAL AND METHODS**

### **5.3.1 Intel RealSense D415**

The Intel D415 is a stereo-optical system consisting of right and left IR cameras (imagers) and an RGB sensor, allowing to capture both depth and colour information of the scene, and an optional IR projector (Figure 5.1). The IR projector enhances depth estimation for low-textured objects by projecting a structured pattern onto the scene. The scene is captured by the imagers and the raw data sent to an onboard vision processor for depth map estimation. The depth map is generated via stereo matching, where the disparity (horizontal displacement) between a pair of corresponding pixels

in the left and right images is calculated for each pixel. Thus, each pixel of the depth map represents the distance to a region of the surface of the scene captured in the FoV.

The Intel D415 uses camera sensors of  $1280 \times 720$  pixels with pixel size  $1.4 \mu\text{m} \times 1.4 \mu\text{m}$ , has a weight of 72 g and dimensions  $99 \times 20 \times 23$  mm (length  $\times$  depth  $\times$  height). Key technical specifications are summarised in Table 5.1.



Figure 5.1 Intel RealSense D415 Stereo Depth Camera.

Table 5.1 Specifications of the Intel RealSense D415.

Component	Specification
Horizontal field of view (FoV)	$65^\circ \pm 2^\circ$
Vertical FoV	$40^\circ \pm 1^\circ$
Diagonal FoV	$72^\circ \pm 2^\circ$
Depth map resolution	Up to $1280 \times 720$ pixels
Frame rate	Up to 90 fps
Minimum working distance	0.16 m
Maximum range	$\sim 10$ m
Connector	USB Type-C

### 5.3.2 Pose tracking

The pose tracking software is an implementation of the signed distance function (SDF) tracker algorithm (Canelhas, Stoyanov et al. 2013), which uses a GPU parallelised iterative closest point approach to register sequential depth frames from the Intel Realsense onto a static volume. Depth frames are converted into a 3D point cloud where each point represents the  $X$ ,  $Y$  and  $Z$  position of part of the surface of the object being imaged. When the tracking software is initialised, the point cloud associated with the first frame is transformed into a SDF volume in which each voxel contains the distance to the nearest surface, implicitly describing the scene in a

volumetric form. Subsequent depth frames are converted to point clouds before being registered to the SDF by sampling the SDF values at each point's  $x$ ,  $y$  and  $z$  locations in the 3D GPU memory and iteratively optimising the rigid transform aligning the current frame to the SDF volume space.

With the depth camera fixed, any motion of the object in the tracking volume is measured as a relative transformation of the tracking volume, in an arbitrary world frame, with respect to the coordinate frame of the Intel depth camera (Figure 5.2a). Real-time pose tracking rates ( $\geq 10$  Hz) could be achieved on a wide range of GPU-enabled hardware. The software was developed in C++ and OpenGL for cross compatibility between Windows and Linux based systems.

### **5.3.3 The Intel D415 performance characterisation: robotic phantom**

#### *Experimental setup*

The repeatability, stability and static and dynamic 6 DoF pose estimation accuracy of the Intel D415 was characterised using a head phantom manipulated by the 6 DoF UR3 robot (Universal Robots, Odense Denmark). The UR3 is highly flexible, having  $\pm 360^\circ$  rotation in all joints, and has pose repeatability of  $\pm 0.1$  mm and  $\pm 0.01^\circ$  (Jiang, Luo et al. 2017, Alnaghy, Kyme et al. 2019) and positional accuracy  $< 0.35$  mm (Kirschner, Schlotzhauer et al. 2017).

The experimental setup is shown in Figure 5.2. A realistic rubber human face mask phantom was rigidly mounted to the UR3 tool centre point (TCP) and viewed using a single Intel D415 rigidly fixed on a tripod at a nominal working distance of 0.25 m. The Intel D415 was connected to a powered USB3 hub interfacing with the host computer in order to stabilise the frame rate and avoid frame drop-out. Neither the Intel D415 IR projector or the RGB camera images were used in any of our experiments.

The UR3 pose and associated timestamp were extracted in real-time at 125 Hz from the robot controller MODBUS server via a dedicated port on the existing TCP/IP connection while the Intel D415 simultaneously recorded the pose of the phantom in its native coordinate frame (Figure 5.2a) at 30 fps. Default pre-sets were used for the Intel D415 and the depth frame resolution was set to  $1280 \times 720$  pixels.

For all of the robotic experiments we computed the median speed and speed range for the motion trajectories tested. The reported speed corresponded to the

phantom nose-tip, a point 100 mm along the +Z direction of the phantom coordinate frame (Figure 5.2a). Speed of this point was computed from the MODBUS data using an offset transformation as described in (Alnaghy, Kyme et al. 2019).

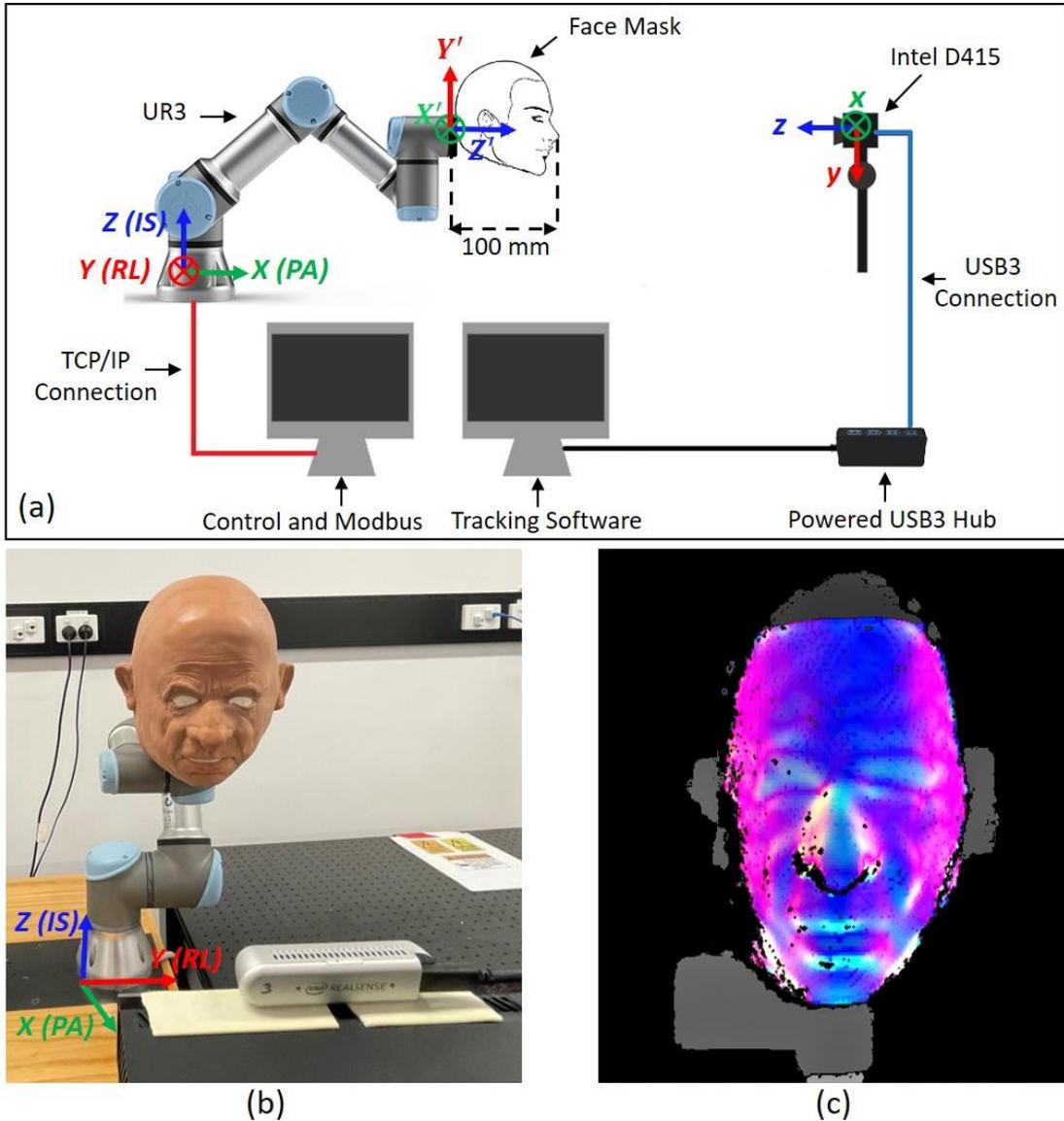


Figure 5.2 Robotic motion tracking, (a) Schematic of the experimental setup to control the Intel D415 and robotic phantom. The moving phantom frame coincided with the default UR3 tool frame ( $X'Y'Z'$ ). The robot base frame ( $XYZ$ ) is defined based on posterior to Anterior (PA), right to left (RL), and inferior to superior (IS).  $xyz$  is the Intel D415 coordinate frame. (b) Rubber human face mask phantom mounted to the UR3 and viewed by the Intel D415. (c) Depth map generated by the Intel D415 using a region-of-interest confined to the phantom face. The coordinate frame shown in (b) is the robot base frame. Right/left in (a) is out/into the page.

### ***System cross-calibration and synchronisation***

To directly compare the commanded UR3 poses with the Intel D415 measurements, we required (i) a cross-calibration of the two systems so that Intel D415 pose measurements could be reported in the robot frame, and (ii) temporal synchronisation of the pose measurements from the two devices.

To cross-calibrate the UR3 and Intel D415, 35 discrete poses of the phantom were measured simultaneously using the two devices. The calibration poses were chosen to sample all translational DoF (right–left (*RL*), inferior–superior (*IS*), and posterior–anterior (*PA*)) and rotational DoF (*rRL*, *rIS*, and *rPA*) (Figure 5.2a). At each calibration pose of the phantom, the Intel D415 and UR3 sampled the pose multiple times and pose averaging (Stavdahl, Bondhus et al. 2005) was used to obtain estimates with reduced noise and improved stability. By default, the UR3 expresses three-dimensional (3D) rotations using axis-angle format in which a 3-element vector describes the axis of rotation and its norm describes the magnitude of rotation. This format was converted to a rotation matrix and then combined with the position vector to represent the pose as a  $4 \times 4$  homogeneous transformation matrix,  $T$ , according to:

$$T = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \quad (5.1)$$

where  $\mathbf{R}$  and  $\mathbf{t}$  are a  $3 \times 3$  rotation matrix and  $3 \times 1$  translation vector, respectively.

The cross-calibration matrix  $\mathbf{X}$  was obtained using the hand-eye method in (Daniilidis 1999).  $\mathbf{X}$  relates pose changes in the robot and Intel D415 frames according to (Shiu and Ahmad 1987):

$$\Delta \mathbf{A}_{ij}^r = \mathbf{A}_i \mathbf{A}_j^{-1} \quad (5.2)$$

$$\Delta \mathbf{B}_{ij}^I = \mathbf{B}_i^{-1} \mathbf{B}_j \quad (5.3)$$

$$\Delta \mathbf{A}_{ij}^r \mathbf{X} = \mathbf{X} \Delta \mathbf{B}_{ij}^I \quad (5.4)$$

where  $\Delta \mathbf{A}_{ij}^r$  and  $\Delta \mathbf{B}_{ij}^I$  are homogeneous transformation matrices describing pose change or movement between pose  $i$  and  $j$  in the UR3 (*r*) and the Intel D415 (*I*) coordinate systems, respectively.  $\mathbf{A}$  and  $\mathbf{B}$  are homogeneous transformation matrices describing poses in the UR3 and the Intel D415 coordinate systems, respectively. In our experiments, we considered  $i = 1$  and  $j \in \{1, 2, \dots, n\}$  where  $n$  is the total number of

poses. According to equation (4), a pose change in the UR3 coordinate system is obtained by:

$$\Delta \mathbf{A}_{ij}^r = \mathbf{X} \Delta \mathbf{B}_{ij}^l \mathbf{X}^{-1} \quad (5.5)$$

The synchronisation of the UR3 and Intel D415 pose measurements was achieved by manually aligning the respective motion traces, using the first transition of the phantom from stationary to moving and the last transition of the phantom from moving to stationary as reference points. Since the rate of pose measurement was different for the Intel D415 (30 fps) and UR3 MODBUS (125 Hz), the UR3 measurements were down-sampled to 30 fps before performing cross-calibration or pose comparisons.

### *Stability assessment*

To assess the thermal dependency and noise stability (jitter) of the Intel D415, the head phantom (Figure 5.2) was tracked stationary for 70 minutes, repeated three times at the same pose. The amplitude and variance of the noise and the thermal drift were computed as a function of time from the raw Intel D415 pose estimates.

### *Static and dynamic pose accuracy*

#### *Static performance*

The robot was programmed to translate the phantom  $\pm 40$  mm along each axis ( $RL$ ,  $IS$  and  $PA$ ) and  $\pm 10^\circ$  around each axis ( $rRL$ ,  $rIS$ , and  $rPA$ ), stopping for 3 s at each pose. The 12 discrete poses recorded by the Intel D415 were compared with the commanded UR3 poses after applying the cross-calibration and averaging over the samples collected during the 3 s per pose. The experiment was repeated five times to assess the variance.

#### *Dynamic performance*

The dynamic performance of the Intel D415 was assessed in three experiments: (1) tracking the 1D pose-pose transition  $\pm 40$  mm along each axis ( $RL$ ,  $IS$  and  $PA$ ) and  $\pm 10^\circ$  around each axis ( $rRL$ ,  $rIS$ , and  $rPA$ ) at median speed of  $0.262 \text{ ms}^{-1}$  (range 0 -  $0.717 \text{ ms}^{-1}$ ); (2) tracking a continuous motion trajectory through 35 distinct poses at three different median speeds:  $0.031 \text{ ms}^{-1}$  (range 0 -  $0.501 \text{ ms}^{-1}$ ),  $0.022 \text{ ms}^{-1}$  (range 0 -  $0.261 \text{ ms}^{-1}$ ), and  $0.006 \text{ ms}^{-1}$  (range 0 -  $0.061 \text{ ms}^{-1}$ ); and (3) tracking a simulated patient head motion sequence derived from a clinical CTP scan. This particular motion was assessed by the reporting radiologist as having considerable

impact on the CTP images and causing residual artifacts in the haemodynamic maps even after standard frame-to-frame motion correction.

The patient motion track was obtained by rigidly registering each of the 33 reconstructed CTP frames to the first frame using SPM12.  $X$ ,  $Y$ , and  $Z$  of the SPM12 coordinate system correspond to  $RL$ ,  $PA$  and  $IS$ , respectively. To simulate the patient motion track, we matched the UR3 and SPM12 coordinate systems and programmed the UR3 to visit the 33 poses, as reported by SPM12, within 60 s. This corresponded to a median speed of  $0.020 \text{ ms}^{-1}$  (range 0 -  $0.093 \text{ ms}^{-1}$ ).

For all three experiments, poses recorded by the Intel D415 were compared to the UR3 poses after applying the cross-calibration. The average root-mean-square-error (RMSE) and mean difference for each DoF were used to assess the pose accuracy. All experiments were repeated five times to assess the variance.

### **5.3.4 The Intel D415 performance characterisation: clinical environment**

To evaluate the performance of the Intel D415 in a clinical setting, the device was adapted to a clinical CT scanner to track both the phantom and human head motion. The Intel D415 was mounted to the gantry of a Siemens Biograph mCT scanner using a custom 3D-printed attachment facilitating pan/tilt positioning and connected to a powered USB3 hub interfacing with the host computer. Performance was measured relative to the OptiTrack IR motion tracking system (NaturalPoint, Inc.). Three synchronised OptiTrack cameras ( $1280 \times 1024$  resolution, 100 fps) were clamped to the wall at the rear of the CT scanner. All cameras had direct line-of-sight to the head inside the gantry at a nominal working distance of 2 m. The OptiTrack multi-camera configuration was calibrated using the standard calibration wand procedure before collecting measurements. The experimental setup is shown in Figure 5.3.

#### ***System cross-calibration and synchronisation***

To directly compare the Intel D415 pose estimates against the OptiTrack, the two systems were cross-calibrated using the hand-eye method as described in Section 5.3.3/System cross-calibration and synchronisation. The realistic rubber face mask was positioned on the CT bed, directly facing the Intel D415 and centrally located in its FoV. The phantom had five IR reflective markers affixed to the scalp and all three OptiTrack cameras had line-of-sight to every marker to maximise OptiTrack tracking

accuracy. The phantom was moved manually to 35 discrete calibration poses within the FoV of both tracking systems at 5-10 s intervals and simultaneously tracked by both devices. Pose averaging (Stavdahl, Bondhus et al. 2005) was used to obtain estimates with reduced noise before computing  $\mathbf{X}$  in equation (5.4).

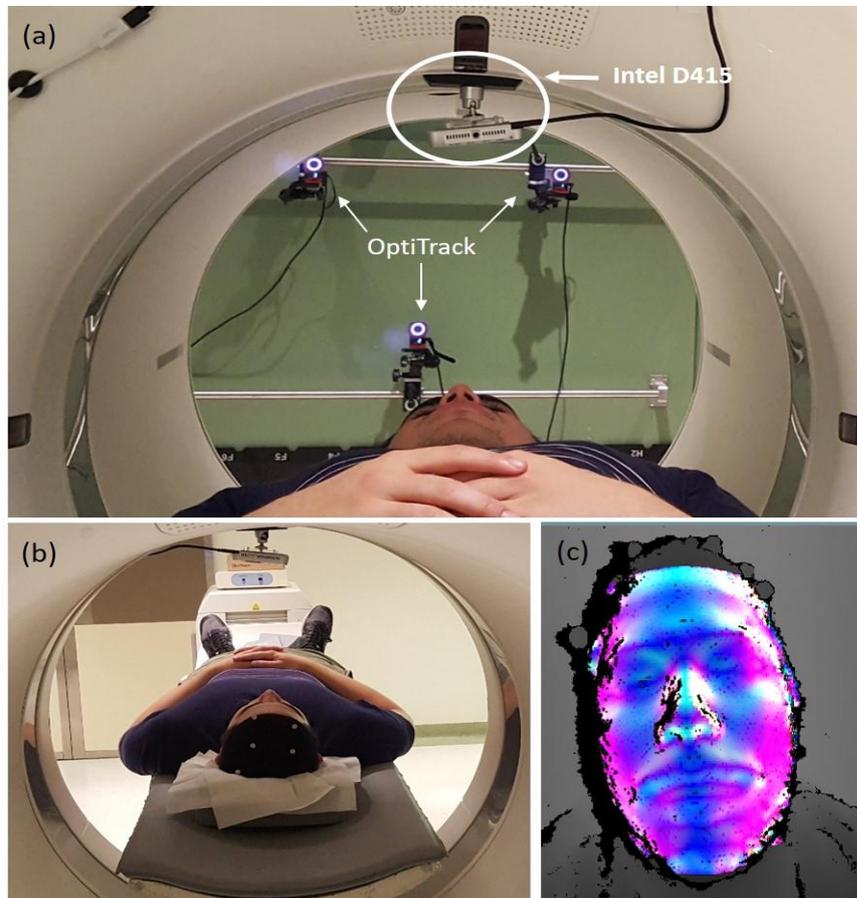


Figure 5.3 Evaluating the Intel D415 performance in a clinical CT scanner. (a) Experimental setup showing head tracking of a volunteer using the mounted Intel D415 (white circle) and simultaneously using the IR-based OptiTrack system with 3 cameras; (b) IR-reflective markers affixed to the OptiTrack head cap (shown here for the human volunteer, but the same applied to the head phantom test); (c) depth map generated by the Intel D415 using a region-of-interest confined to the human volunteer face.

### *Phantom experiments*

The head phantom was positioned inside the CT scanner and moved continuously by hand in three separate 60-s trials while being tracked by the two systems. The Intel D415 pose measurements were compared with the OptiTrack estimates after applying the cross-calibration.

## *Human volunteer experiments*

A human volunteer wore a tightly fitted fabric head cap affixed with five IR reflective markers defining a rigid-body registered with the OptiTrack system (Figure 5.3b). The volunteer lay in the scanner and performed three 60-s motion sequences while their head motion was tracked by the OptiTrack and Intel D415 systems simultaneously. The volunteer was instructed to perform slow and continuous head motion in each of the three trials. The Intel D415 head motion trajectories were compared against the OptiTrack estimates after applying the cross-calibration.

## 5.4 RESULTS

### 5.4.1 The Intel D415 performance characterisation: robotic phantom

#### *Stability assessment*

The Intel D415 tracking of a static phantom over 70 minutes (Figure 5.4) showed a gradual drift in pose measurements as the sensor's internal temperature increased from start-up. The pose measurements became stable after 10 - 20 minutes with the stabilisation of the internal temperature. The average internal temperature change over three trials was  $12.66 \pm 2.08$  °C. The absolute pose drift was  $\leq 1.5$  mm and  $\leq 0.5^\circ$  with respect to translational and rotational DoFs, respectively. The largest jitter amplitude for translational measurements was 0.5 mm, observed for the *RL* and *IS* axes. The jitter amplitude for rotations was approximately  $0.1^\circ$  about all three axes.

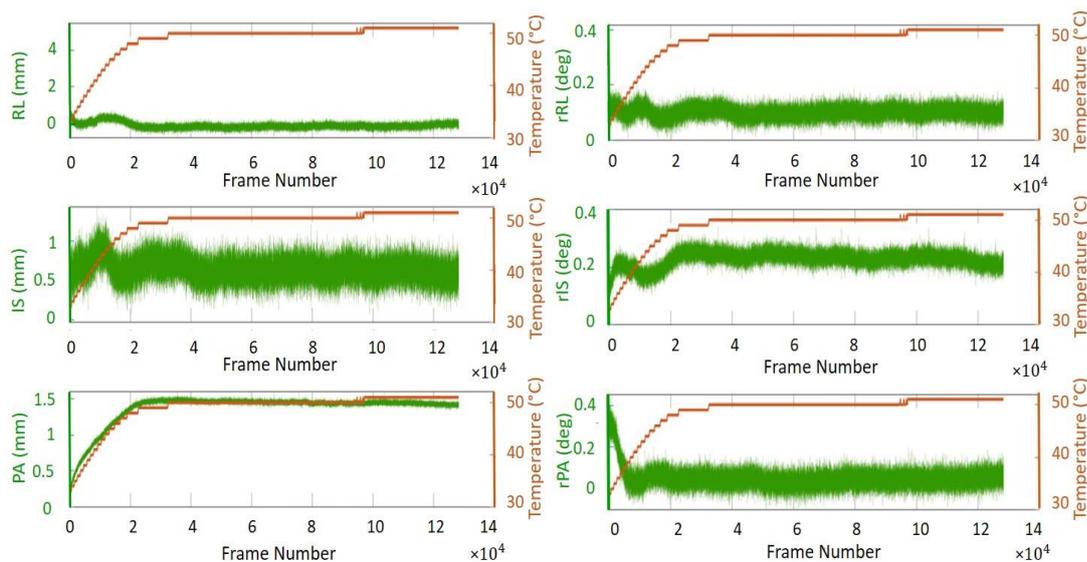


Figure 5.4 A representative thermal stability of the Intel D415 pose measurements after start-up (green) and the corresponding internal temperature change (orange) over 70 min (approximately 130,000 frames).

Table 5.2 1D static and dynamic pose accuracy of the Intel D415 compared to the UR3 robot. Measurements are reported as mean  $\pm$  SD.

Direction	Commanded UR3 position	Static accuracy <sup>(a)</sup> (mean difference)	Dynamic accuracy <sup>(b)</sup> (mean RMSE)	Dynamic accuracy <sup>(b)</sup> (mean difference)
<i>RL</i> (mm)	40	1.24 $\pm$ 0.01	1.08 $\pm$ 0.63	0.98 $\pm$ 0.74
	-40	0.56 $\pm$ 0.07	0.78 $\pm$ 0.31	0.06 $\pm$ 0.79
<i>IS</i> (mm)	40	-0.71 $\pm$ 0.02	1.26 $\pm$ 0.70	-1.05 $\pm$ 0.65
	-40	0.23 $\pm$ 0.03	0.77 $\pm$ 0.39	0.29 $\pm$ 0.63
<i>PA</i> (mm)	40	0.67 $\pm$ 0.05	0.60 $\pm$ 0.09	0.50 $\pm$ 0.08
	-40	0.05 $\pm$ 0.04	0.4 $\pm$ 0.13	0.32 $\pm$ 0.17
<i>rRL</i> ( $^{\circ}$ )	10	-0.10 $\pm$ 0.01	0.05 $\pm$ 0.01	-0.02 $\pm$ 0.01
	-10	0.01 $\pm$ 0.01	0.07 $\pm$ 0.01	-0.06 $\pm$ 0.01
<i>rIS</i> ( $^{\circ}$ )	10	0.68 $\pm$ 0.01	0.37 $\pm$ 0.04	0.25 $\pm$ 0.07
	-10	-0.02 $\pm$ 0.01	0.67 $\pm$ 0.09	0.57 $\pm$ 0.08
<i>rPA</i> ( $^{\circ}$ )	10	-0.05 $\pm$ 0.01	0.04 $\pm$ 0.01	-0.02 $\pm$ 0.02
	-10	0.01 $\pm$ 0.01	0.04 $\pm$ 0.01	-0.03 $\pm$ 0.02

(a) Static localisation accuracy is expressed as the difference between the commanded UR3 position (ground-truth) and average pose reported by the Intel D415.

(b) Dynamic accuracy is expressed as the average of RMSE and mean difference between UR3 poses and Intel poses along the motion trajectories.

### *Static performance*

Table 5.2 shows the static localisation accuracy of the Intel D415 reported as the difference between the commanded UR3 pose and mean pose reported by the Intel D415 for six translational and six rotational poses, averaged over five trials. The static pose accuracy of the Intel D415 was  $\leq 1.24 \pm 0.01$  mm and  $\leq 0.68 \pm 0.01^{\circ}$  with respect to translational and rotational axes, respectively.

### *Dynamic performance*

Table 5.2 shows the dynamic accuracy and repeatability of the Intel D415 measurements against UR3 for 1D transition along six translational and about six rotational positions. The average RMSE was  $\leq 1.26 \pm 0.70$  mm and  $\leq 0.67 \pm 0.09^{\circ}$  and the absolute mean difference was  $\leq 1.05 \pm 0.65$  mm and  $\leq 0.57 \pm 0.08^{\circ}$ .

Table 5.3 Dynamic pose accuracy of the Intel D415 compared to the UR3 robot. Measurements are reported as mean  $\pm$  SD.

Direction	Metric	Dynamic accuracy <sup>(a)</sup> ( $S_1$ ) <sup>(b)</sup>	Dynamic accuracy <sup>(a)</sup> ( $S_2$ ) <sup>(b)</sup>	Dynamic accuracy <sup>(a)</sup> ( $S_3$ ) <sup>(b)</sup>	Dynamic accuracy <sup>(a)</sup> (Patient motion) <sup>(c)</sup>
<i>RL</i> (mm)	Mean RMSE $\pm$ SD	$1.40 \pm 0.12$	$1.38 \pm 0.02$	$1.25 \pm 0.01$	$0.53 \pm 0.04$
<i>IS</i> (mm)		$1.13 \pm 0.27$	$0.91 \pm 0.04$	$0.85 \pm 0.02$	$0.86 \pm 0.03$
<i>PA</i> (mm)		$0.55 \pm 0.12$	$0.50 \pm 0.04$	$0.45 \pm 0.04$	$0.77 \pm 0.07$
<i>rRL</i> ( $^\circ$ )		$0.12 \pm 0.02$	$0.10 \pm 0.00$	$0.09 \pm 0.00$	$0.11 \pm 0.01$
<i>rIS</i> ( $^\circ$ )		$0.22 \pm 0.01$	$0.22 \pm 0.01$	$0.20 \pm 0.00$	$0.16 \pm 0.03$
<i>rPA</i> ( $^\circ$ )		$0.24 \pm 0.02$	$0.23 \pm 0.00$	$0.14 \pm 0.00$	$0.03 \pm 0.01$
<i>RL</i> (mm)	Mean difference $\pm$ SD	$-0.24 \pm 0.01$	$0.23 \pm 0.01$	$0.14 \pm 0.01$	$0.40 \pm 0.05$
<i>IS</i> (mm)		$-0.11 \pm 0.02$	$0.10 \pm 0.01$	$0.10 \pm 0.01$	$0.19 \pm 0.01$
<i>PA</i> (mm)		$-0.38 \pm 0.02$	$-0.4 \pm 0.04$	$-0.35 \pm 0.03$	$0.49 \pm 0.01$
<i>rRL</i> ( $^\circ$ )		$-0.04 \pm 0.00$	$0.01 \pm 0.00$	$0.01 \pm 0.00$	$-0.09 \pm 0.01$
<i>rIS</i> ( $^\circ$ )		$0.06 \pm 0.00$	$-0.01 \pm 0.00$	$-0.01 \pm 0.00$	$0.07 \pm 0.01$
<i>rPA</i> ( $^\circ$ )		$0.13 \pm 0.00$	$0.14 \pm 0.00$	$0.06 \pm 0.00$	$-0.01 \pm 0.00$

(a) Dynamic accuracy is expressed as the average of RMSE and mean difference between UR3 poses (ground-truth) and Intel poses for a continuous trajectory.

(b) The arbitrary continuous trajectory travelled at a median speed of  $S_i, i \in \{1, 2, 3\}$ :  $S_1 = 0.031 \text{ ms}^{-1}$  (range 0 -  $0.501 \text{ ms}^{-1}$ ),  $S_2 = 0.022 \text{ ms}^{-1}$  (range 0 -  $0.261 \text{ ms}^{-1}$ ), and  $S_3 = 0.006 \text{ ms}^{-1}$  (range 0 -  $0.061 \text{ ms}^{-1}$ ).

(c) Simulated head motion trajectory of a patient during a 60 s CTP scan travelled at a median speed of  $0.020 \text{ ms}^{-1}$  (range 0 -  $0.093 \text{ ms}^{-1}$ ).

Table 5.3 shows the dynamic localisation accuracy of the Intel D415 measurements against the UR3 for the arbitrary continuous trajectory consisting of 35 distinct poses, for three different maximum speeds. The mean RMSE was  $\leq 1.40 \pm 0.12 \text{ mm}$  and  $\leq 0.24 \pm 0.02^\circ$  at the median speed of  $0.031 \text{ ms}^{-1}$  (range 0 -  $0.501 \text{ ms}^{-1}$ ),  $\leq 1.38 \pm 0.02 \text{ mm}$  and  $\leq 0.23 \pm 0.00^\circ$  at the median speed of  $0.022 \text{ ms}^{-1}$  (range 0 -  $0.261 \text{ ms}^{-1}$ ), and  $\leq 1.25 \pm 0.01 \text{ mm}$  and  $\leq 0.20 \pm 0.00^\circ$  at the median speed of  $0.006 \text{ ms}^{-1}$  (range 0 -  $0.061 \text{ ms}^{-1}$ ). Figure 5.5 shows a comparison of the Intel D415 motion estimates against the UR3 for the arbitrary dynamic path travelled at a median speed of  $0.022 \text{ ms}^{-1}$  (range 0 -  $0.261 \text{ ms}^{-1}$ ).

The dynamic localisation accuracy of the Intel D415 for a simulated patient head movement is shown in Table 5.3 (right column). The mean RMSE was  $\leq 0.86 \pm 0.03$  mm and  $\leq 0.16 \pm 0.03^\circ$  and absolute mean difference was  $\leq 0.49 \pm 0.01$  mm and  $\leq 0.09 \pm 0.01$ . The pose measurements from each device are compared in Figure 5.6 for each DoF.

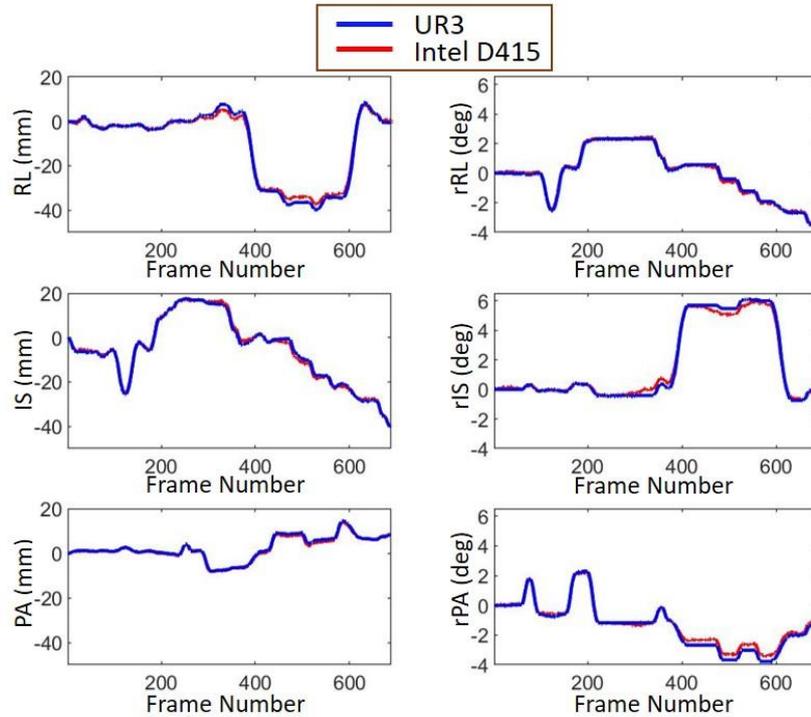


Figure 5.5 Comparison of motion estimates for the Intel D415 (red) and UR3 (blue) for an arbitrary continuous motion trajectory traversed at a median speed of  $0.022 \text{ ms}^{-1}$  (range  $0 - 0.261 \text{ ms}^{-1}$ ).

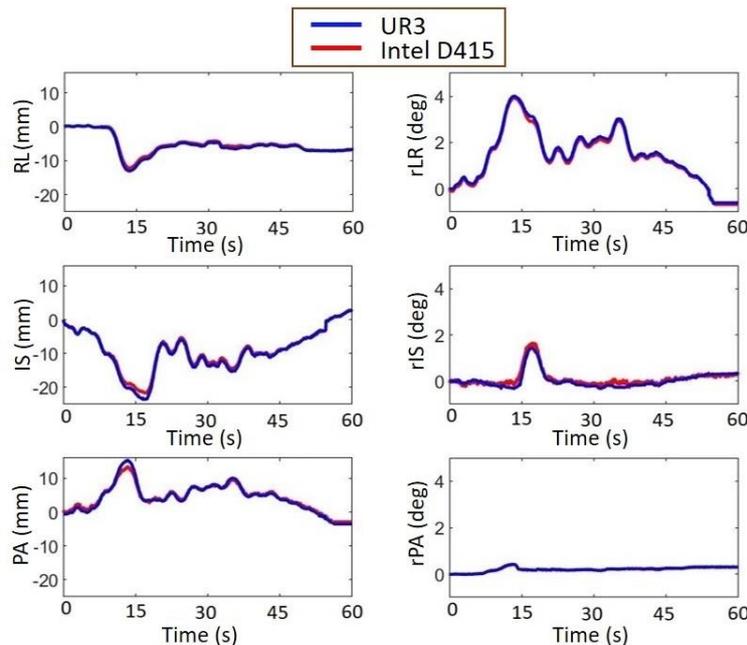


Figure 5.6 Comparison of motion estimates for the Intel D415 (red) and UR3 (blue) for a simulated head motion trajectory of a patient during a 60 s CTP scan.

### 5.4.2 The Intel D415 performance characterisation: clinical environment

Table 5.4 shows the dynamic localisation accuracy of the Intel D415 against the OptiTrack for the phantom and volunteer experiments in a clinical setting. For the head phantom, the mean RMSE was  $\leq 3.74 \pm 2.39$  mm and  $\leq 0.91 \pm 0.66^\circ$  and absolute mean difference was  $\leq 1.13 \pm 3.77$  mm and  $\leq 0.26 \pm 0.82^\circ$ . The pose measurements from each device are compared in Figure 5.7 for each DoF.

For volunteer head movement, the mean RMSE was  $\leq 2.72 \pm 0.24$  mm and  $\leq 0.55 \pm 0.07^\circ$  and absolute mean difference was  $\leq 1.12 \pm 2.34$  mm and  $\leq 0.17 \pm 0.26^\circ$ . The pose measurements from each device are compared in Figure 5.8 for each DoF.

Table 5.4 Dynamic pose accuracy of the Intel D415 compared to the OptiTrack for the phantom and volunteer motion inside a clinical CT scanner. Measurements are reported as mean  $\pm$  SD.

	Metric	<i>RL</i> (mm)	<i>IS</i> (mm)	<i>PA</i> (mm)	<i>rRL</i> ( $^\circ$ )	<i>rIS</i> ( $^\circ$ )	<i>rPA</i> ( $^\circ$ )
Dynamic accuracy <sup>(a)</sup> (Phantom motion)	Mean RMSE $\pm$ SD	$3.74 \pm 2.39$	$3.53 \pm 2.99$	$2.24 \pm 1.34$	$0.91 \pm 0.66$	$0.76 \pm 0.59$	$0.68 \pm 0.47$
	Mean difference $\pm$ SD	$1.09 \pm 3.16$	$1.13 \pm 3.77$	$-0.66 \pm 2.58$	$-0.19 \pm 1.53$	$0.05 \pm 1.11$	$-0.26 \pm 0.82$
Dynamic accuracy <sup>(a)</sup> (Volunteer motion)	Mean RMSE $\pm$ SD	$2.72 \pm 0.24$	$1.70 \pm 0.41$	$0.81 \pm 0.14$	$0.32 \pm 0.08$	$0.55 \pm 0.07$	$0.47 \pm 0.40$
	Mean difference $\pm$ SD	$0.34 \pm 0.47$	$-1.12 \pm 2.34$	$-0.36 \pm 0.67$	$-0.17 \pm 0.26$	$-0.02 \pm 0.06$	$-0.06 \pm 0.53$

(a) Dynamic accuracy is expressed as the average of RMSE and mean difference between OptiTrack poses (ground-truth) and Intel poses for a continuous trajectory.

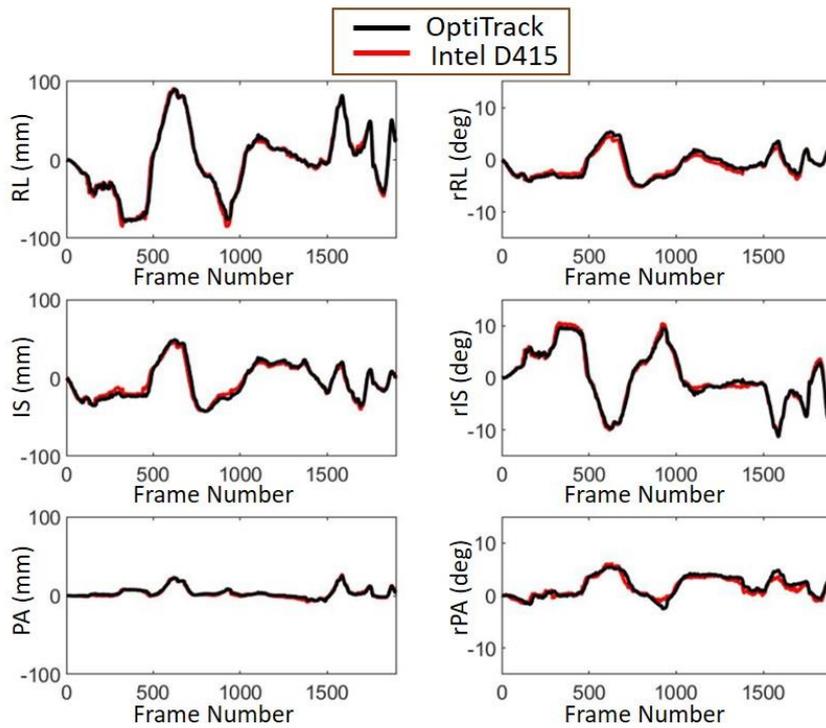


Figure 5.7 Comparison of motion estimates for the Intel D415 (red) and OptiTrack (black) when tracking a head phantom inside a clinical CT scanner for 60 s.

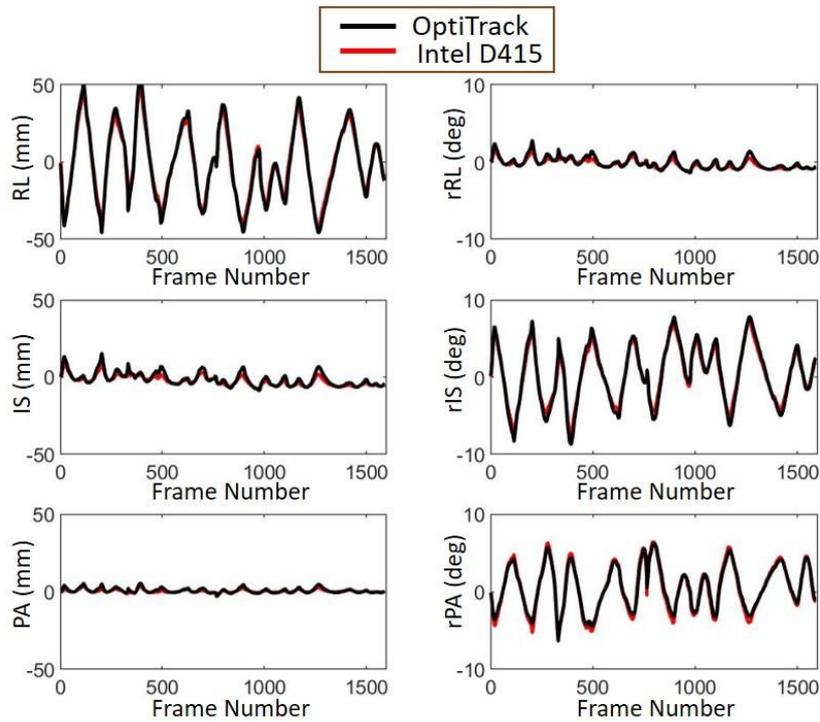


Figure 5.8 Comparison of motion estimates for the Intel D415 (red) and OptiTrack (black) when tracking human head movement for 60 s inside a clinical CT scanner.

## 5.5 DISCUSSION

This study reports a characterisation of the Intel D415 depth sensor performance to evaluate its potential use in head motion tracking during CT/CTP imaging. Thermal stability and measurement jitter were assessed over 70-min trials and static and dynamic pose estimation accuracy was validated against ground-truth robotic motion in benchtop experiments and against a multi-camera stereo-optical tracking system in the clinical environment. The Intel D415 stability, static and dynamic localisation accuracy, was assessed through robotic motion tracking experiments and in a clinical setting against OptiTrack.

Although thermally-induced pose drift is observed (Figure 5.4), the impact is negligible over the 1-min scanning time in CTP imaging. Nevertheless, the accuracy of pose measurements cannot be assumed until thermal stability has been reached approximately 10 - 20 min post power-up. This is also when cross-calibration should be performed.

Measurement jitter remained stable over 70-min trials and was also unaffected by changes in the Intel D415 internal temperature. Jitter is likely due to errors in stereo matching and uncertainty in the iterative process used to calculate the position and orientation of the object. A potential limitation is that the jitter may depend on the shape of the object and may be slightly different for a different object.

In robotic motion testing, the static pose accuracy of the Intel D415 appeared to be DoF-dependent, and that the DoFs with poorest accuracy were shifts along *RL* (mean difference  $1.24 \pm 0.01$  mm) and head roll about *IS* (mean difference  $0.68 \pm 0.01^\circ$ ) axes. Motion along these DoFs is more likely to result in occlusion near the nose, a crucial area of high topological change that enables robust depth-map alignment. Thus, occlusions of this region result in larger pose uncertainty and creating a gap in the depth map. Occlusion occurs when pixels cannot be seen by both the left and right imagers, which precludes solving for the depth and creates gap. A possible solution to reduce the likelihood of occlusion is to use multiple depth sensors to track the object from multiple views. Use of multiple sensors is also likely to improve pose estimation accuracy. This will be investigated in future work.

The dynamic accuracy of the Intel D415 (mean RMSE) for the arbitrary dynamic path travelled at the median speed of  $0.031 \text{ ms}^{-1}$  (range  $0 - 0.501 \text{ ms}^{-1}$ ) was  $\leq 1.40$

$\pm 0.12$  mm and  $\leq 0.24 \pm 0.02^\circ$ . The dynamic accuracy was slightly better for slower motion compared to faster motion (Table 5.3). The Intel D415 showed sub-millimetre accuracy in tracking the simulated patient motion trace (mean RMSE  $\leq 0.86 \pm 0.03$  mm and  $\leq 0.16 \pm 0.03^\circ$ ). However, a conservative accuracy estimate (mean RMSE) based on all experiments involving the robotically controlled mask (Table 5.2 and Table 5.3) was  $< 2$  mm and  $< 1^\circ$  for translations and rotations, respectively.

Compared to the robotic motion testing, the dynamic accuracy of the Intel D415 was poorer when validating it against the OptiTrack inside the CT scanner (Table 5.4). Possible reasons for this include: (i) less control over motion speed and range compared to robotic motion, and increased likelihood of occlusion or the object leaving the FoV, and (ii) increased likelihood of error in the OptiTrack (ground-truth) measurement due to obstruction of the IR markers or deformation of the head cap, causing relative movement of markers, can produce an artificially high positional error.

Typically, CTP volumes comprise slices with 5 mm thickness and 0.43 mm in-plane resolution. Our results showed that the pose estimation accuracy of the Intel D415 was lower than a typical CTP image resolution. This implies that although the Intel D415 will aid motion correction, there will be residual error. Since the pose accuracy of the Intel D415 appeared to be DoF-dependent, an optimised multiple sensor setup to track the object from multiple views can potentially reduce the impact of object position by improving the stereo matching. This will be considered in our future work.

Continuous motion monitoring via external tracking devices provides the opportunity to fully correct the acquired data for both inter- and intra-frame motions. Although data-driven approaches for comprehensive motion correction exist (Sun, Kim et al. 2016), their practicality for the clinical CT applications is limited due to the high computational cost and their limited accuracy due to the type of motion to be estimated or the noise in the data. Motion correction can be more accurate when the motion is estimated independently from the raw data via an external tracking device. External tracking system could also provide useful initialisations or to fill in gaps when the data-driven estimation is unreliable.

## 5.6 CONCLUSION

This study characterised the Intel D415 RealSense depth sensor as a markerless motion tracking device for head pose estimation at close range ( $\leq 0.25$  m) for potential use in motion corrected CTP imaging. The Intel D415 conservative accuracy estimate (mean RMSE) for the experiments involved the robotically controlled head phantom was  $< 2$  mm and  $< 1^\circ$ , and for volunteer motion tracking inside a clinical CT scanner against OptiTrack was  $< 3$  mm and  $< 1^\circ$  for translations and rotations, respectively.

# Chapter 6: Reducing scan duration and radiation dose in cerebral CT perfusion imaging of acute stroke using a recurrent neural network

---

To be submitted to *NeuroImage*: **Dashtbani Moghari, M.**, et al., “Reducing scan duration and radiation dose in cerebral CT perfusion imaging of acute stroke using a recurrent neural network”.

## 6.1 ABSTRACT

Cerebral CT perfusion (CTP) imaging is one of the most common methods used to diagnose acute ischaemic stroke and support treatment decisions. Wherever possible, shortening the duration of the CTP acquisition is desirable to reduce the accumulated radiation dose and the risk of patient head movement. In this study, we present a novel application of a stochastic adversarial video prediction (SAVP) approach to reduce the acquisition time of CTP imaging. A variational autoencoder (VAE) and generative adversarial networks (GAN) VAE-GANs were implemented in a recurrent framework to predict the last 8 (24 s), 13 (31.5 s) and 18 (39 s) image frames of the CTP acquisition from the first 25 (36 s), 20 (28.5 s) and 15 (21 s) acquired frames, respectively. The supervised model was trained using 65 stroke cases and tested on 10 unseen cases. Predicted frames were assessed against ground truth (measured frames) in terms of image quality of CTP data and haemodynamic maps, bolus shape characteristics and volumetric analysis of lesions. In the most challenging case predicting the last 18 CTP frames, the bolus area and full-width-half-maximum exhibited an average percentage difference of  $< 4\% \pm 4\%$  compared to the ground-truth. The best image quality of predicted haemodynamic maps was obtained for cerebral blood volume (CBV) followed (in order) by cerebral blood flow (CBF), mean transit time (MTT) and time to peak (TTP). Average volumetric error of the lesion was overestimated by 7.58 mL (15%), 20.78 mL (28%), and 28.36 mL (22%) for the infarct, penumbra and hypo-perfused regions, respectively, and the corresponding spatial agreement for these regions was 67%, 76% and 83% (mean dice coefficient). This study suggests that a recurrent VAE-GANs could potentially be used as a practical method to predict a portion of CTP frames from truncated acquisitions.

Predicting the last 18 frames can preserve the majority of clinical content of the images while simultaneously reducing the scan duration and radiation dose by 65% and 54.5%, respectively. Further work is required to investigate the generalisability of the approach across datasets and scanners and the impact of the proposed method on treatment decision-making.

**Keywords:** CT perfusion imaging, scan duration, dose reduction, deep learning, acute ischaemic stroke

## 6.2 INTRODUCTION

A multimodal CT imaging regime, including non-contrast CT (NCCT), CT perfusion (CTP) and CT angiography (CTA), is commonly used for diagnosis and determining the best treatment options for acute ischaemic stroke patients (Ledezma and Wintermark 2009, Morgan, Stephens et al. 2015, Heit and Wintermark 2016). In quantitative CTP analysis, haemodynamic parameters such as cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT) and time to peak (TTP) are derived from CTP source data for each voxel of the brain using mathematical models (Lin, Bivard et al. 2013). Based on estimated perfusion maps, the voxel-wise status of the brain tissue can be determined - specifically, the extent of hypo-perfused regions including irreversibly damaged and irrecoverable tissue (infarct core) and potentially salvageable tissue (penumbra) (Demeestere, Wouters et al. 2020).

A CTP imaging protocol involves the rapid acquisition of successive volumetric frames, each covering all/part of the brain, over ~1-2 minutes after contrast agent administration. There are two important limitations associated with this protocol. Firstly, the acquisition delivers a radiation dose of 5-6 mSv to the patient (Manniesing, Oei et al. 2015), not excessive on its own, but large when considered alongside the dose from NCCT, CTA and other potential follow-up CT scans post treatment. Therefore, reducing the radiation dose from CTP imaging without compromising the image quality or the accuracy of haemodynamic modelling is highly desirable, especially for younger adults and paediatric patients who are more likely to be impacted by the long-term stochastic effects of ionising radiation (Wolterink, Leiner et al. 2017, Moghari, Zhou et al. 2019). The second limitation is the risk of patient head movement between frames, especially during the terminal phase of the scan (Moghari, Noonan et al. , Hanzelka, Dusek et al. 2013), resulting in streaks, distortion

and blurring of reconstructed images (Popilock, Sandrasagaren et al. 2008, Yazdi and Beaulieu 2008). All such errors can impact the downstream haemodynamic modelling. An obvious motion mitigation strategy is reducing the total scan time, however this truncates the late contrast concentration measurements important for image-based stroke analysis (Mangla, Ekholm et al. 2014, Copen, Deipolyi et al. 2015, Kasasbeh, Christensen et al. 2016).

Several paradigms have been investigated to reduce the total radiation dose in CTP imaging. The most intuitive approach is to reduce the dose per frame while preserving the total number of frames and the total scan time. Many methods have been reported to denoise low-dose CT images, including frames from CTP imaging. Related work can be classified into sinogram domain filtration methods (Wang, Lu et al. 2005, Karimi, Deman et al. 2016), statistical iterative reconstruction methods (Kim, Kim et al. 2015, Lee, Kim et al. 2019), and image post-processing techniques including traditional filter-based methods (Mendrik, Vonken et al. 2010, Mendrik, Vonken et al. 2011, Pisana, Henzler et al. 2017) and more recent deep learning-based methods (Wolterink, Leiner et al. 2017, Moghari, Zhou et al. 2019, Moghari, Zhou et al. 2021). The first two methods are not practical in CTP imaging of acute stroke due to challenges such as access to the raw sinogram data from commercial scanners, high computational cost, and time delays between acquisition and reconstruction. By contrast, deep learning-based post-processing techniques are very promising for the restoration of low-dose CTP frames (Kadimesetty, Gutta et al. 2018, Liu and Fang 2018, Moghari, Zhou et al. 2021).

A second way to reduce radiation dose in CTP imaging is to collect fewer frames by increasing the frame-to-frame time interval, and to estimate the missing (down-sampled) frames. Missing frames can be estimated using traditional or CNN-based interpolation of an image time series (Xiao, Liu et al. 2019, Zhu, Tong et al. 2020); this is generally based only on the two neighbouring frames rather additional images in the sequence and may be unreliable if these frames are degraded by intra- or inter-frame motion.

Although both of these paradigms are effective in reducing the radiation dose, the total scan time – and thus the likelihood of motion – is unchanged. A third approach to dose reduction in CTP is to reduce the number of frames by reducing the total scan time. This has the advantage of simultaneously reducing the radiation dose and the

likelihood of head motion, specifically motion that would occur in the late frames. In this study, we present a novel application of a stochastic video prediction (SVP) technique to demonstrate the feasibility of this approach. Despite many interesting applications of SVP in video analysis (Kumar, Babaeizadeh et al. 2019, Villegas, Pathak et al. 2019, Franceschi, Delasalles et al. 2020), to date it has not been applied to CTP data. Most SVP applications are limited to 3D spatiotemporal video data (2D + time) and not designed to handle 4D dynamic volumetric data (3D + time) like we encounter in CTP imaging. We describe and test a SVP approach in a recurrent framework (Medsker and Jain 2001) to predict the last 8 (24 s), 13 (31.5 s) or 18 (39 s) CTP image frames from a sequence of initial 25 (36 s), 20 (28.5 s), or 15 (21 s) acquired frames, respectively. Unlike the frame interpolation techniques, our approach bases the prediction of later frames on a sequence of dynamic data rather than two adjacent frames, increasing the model efficiency in the presence of data inconsistency due to motion.

## **6.3 MATERIALS AND METHODS**

### **6.3.1 CT perfusion data and data pre-processing**

The retrospective dataset comprised CTP studies from 75 patients (42 males [56%], 33 females [44%]) scanned consecutively in 2018 at Westmead Hospital, Sydney. All studies were acute stroke cases with occlusion in the right or left middle cerebral artery. The average age of the patients was 71 yr (SD 15 yr, range 35-92 yr). Data collection and analysis were performed in accordance with a human ethics protocol approved by the Western Sydney Local Health District.

The CTP images were acquired using a dual-source, dual-energy Siemens Somatom Force CT scanner in 4i cine mode with 1-min acquisition at 70 kVp and 200 mAs. Source rotation time was 250 ms with 1120 projection views per rotation. A standard foam headrest was used to limit patient head movement during the acquisition. The CT dose index (CTDI<sub>vol</sub>) and dose-length product of the CTP scans was 159.84 mGy and 2398.0 mGy.cm, respectively. At the start of the scan, ~45 mL of non-ionic iodinated contrast agent was administered intravenously at 7 mL/s via a power injector with a 5 s delay. 33 brain volumes were acquired over the 1-min scan at 1.5 s intervals for the first 25 CT volumes and 3 s intervals for the last 8 volumes. To cover the full brain, each CT volume comprised 22 axial slices with 5 mm

thickness. Slices were reconstructed in a  $512 \times 512$  matrix with 0.43 mm in-plane resolution.

All studies were pre-processed by removing the background and skull from the reconstructed image slices and scaling the intensity of brain voxels linearly from -1 to 1.

### 6.3.2 Model components and architecture

We adapted the stochastic adversarial video prediction (SAVP) model proposed in (Lee, Zhang et al. 2018). The model combines a variational autoencoder (VAE) (Kingma and Welling 2013) and GANs (Goodfellow, Pouget-Abadie et al. 2014) in a recurrent framework (Sherstinsky 2020) that allows previous outputs (predictions) to be fed back to the model as inputs. In the following sub-sections, we first describe the principle of VAE and GAN models and then the specific architecture of the proposed VAE-GANs (Lee, Zhang et al. 2018) model for application to CTP analysis.

#### *Variational auto encoder (VAE)*

A VAE consists of two connected neural networks, the encoder and decoder. The encoder takes an input data sample  $x$  and compresses (encodes) it into a more compact representation  $z$ , known as the latent variable, in a lower dimensional space called the latent space. In latent space, similar data points are spatially closer together, forming clusters. The decoder learns to take the latent representation and reconstruct (decode) it back to the original data space. To avoid discontinuities in the latent space (i.e. gaps between the clusters), the posterior distribution  $q(z|x)$  is calculated by assigning a mean,  $\mu$ , and standard deviation,  $\sigma$ , to each random variable in the latent space. This stochastic generation of variables introduces local variation, resulting in a smooth latent space within and around the clusters. The encoder, however, can learn very different  $\mu$  and  $\sigma$  values for the different classes (clusters), thus introducing discontinuity between them. Ideally, different classes should be as close to each other as possible while still being distinct, allowing for smooth sampling and hence efficient decoding to the data space. This proximity and differentiation requirement is enforced using Kullback–Leibler divergence ( $D_{KL}$ ) in the VAE loss function.  $D_{KL}$  measures the divergence between the posterior  $q(z|x)$  and prior  $p(z)$  distributions, and minimizing  $D_{KL}$  optimizes  $\mu$  and  $\sigma$  by forcing these distributions to be closer. The VAE loss ( $\mathcal{L}_{VAE}$ ) is thus given by:

$$\mathcal{L}_{VAE} = \mathcal{L}_\ell + D_{KL} \quad (6.1)$$

where  $\mathcal{L}_\ell$  and  $D_{KL}$  refer to the reconstruction error and Kullback–Leibler divergence, respectively, and are defined according to:

$$\mathcal{L}_\ell = -\mathbb{E}_{q(z|x)}[\log p(x|z)] \quad (6.2)$$

$$D_{KL} = D_{KL}(q(z|x)||p(z)) = \sum_z q(z|x) \log \frac{q(z|x)}{p(z)} \quad (6.3)$$

where  $\mathbb{E}$  refers to maximum likelihood estimation.

### ***Generative adversarial network (GAN)***

The GAN consists of a generator network,  $G$ , and discriminator network,  $D$ , in competition. The generator learns to map (decode) the latent variables  $z$  to the data space while the discriminator, which is simply a classifier, tries to distinguish real data  $x \sim p(x)$  from generated data  $\hat{x} \sim p(\hat{x})$  estimated by  $G$ . The objective of the GAN is to determine the binary classifier that optimally distinguishes between the real and generated data and simultaneously enables  $G$  to fit the true data distribution.

The generator and discriminator are trained alternately to minimize and maximize the objective function in turn. The objective function of the GAN,  $\mathcal{L}_{GAN}(G, D)$ , is defined as:

$$\operatorname{argmin}_G \operatorname{max}_D (\mathcal{L}_{GAN}(G, D)) = \mathbb{E}_{x \sim p(x)} [\log(D(x))] + \mathbb{E}_{\hat{x} \sim p(\hat{x})} [\log(1 - D(\hat{x}))] \quad (6.4)$$

### ***Model architecture***

Our VAE-GANs model was designed to predict the late 8, 13, or 18 CTP frames from the sequence of 25, 20, or 15 initial frames, respectively. The GAN is intended to produce predictions that are perceptually realistic, whereas the VAE component functions to improve the prediction diversity, which is important for CTP data due to the diversity of shape of TACs for different tissue types. Thus, the combination of the VAE and GAN is designed to deliver predictions that are simultaneously realistic and that sample a range of possible futures.

The model architecture is shown in Figure 6.1. During the training phase (Figure 6.1a), the recurrent generators ( $G$ ) are conditioned on the previous frame  $\tilde{x}_{t-1}$  and random latent code  $z_{t-1}$  to predict the next frame  $\hat{x}_t$ . Here the previous frame  $\tilde{x}_{t-1}$  could either be a ground-truth frame  $x_{t-1}$  (as for the initial frame), or the last prediction

$\hat{x}_{t-1}$ . As shown in Figure 6.1a, the latent code  $z_{t-1}$  is sampled from two distributions: (1) a single posterior distribution  $q(z_{t-1}|x_{t-1:t})$  estimated by an interface encoder network ( $E$ ), a feed-forward network encoding two ground-truth adjacent frames ( $x_{t-1}, x_t$ ) denoted by  $x_{t-1:t}$ ; (2) the prior distribution  $p(z_{t-1})$ . The proposed VAE-GANs model jointly optimises the VAE and GAN losses during the training. The model objective,  $\mathcal{L}_{VAE-GANs}$ , is defined as:

$$\begin{aligned} \operatorname{argmin}_{G,E} \max_{D_{GAN}, D_{VAE}} (\mathcal{L}_{VAE-GANs}) = \\ \lambda_1 \mathcal{L}_1(G, E) + \lambda_{KL} D_{KL}(E) + \mathcal{L}_{GAN}(G, D_{GAN}) + \mathcal{L}_{GAN}^{VAE}(G, E, D_{VAE}) \end{aligned} \quad (6.5)$$

where  $\mathcal{L}_1$  is the  $L1$  penalty between the predicted  $\hat{x}_t$  frame and ground-truth  $x_t$  frame,  $\mathcal{L}_{GAN}$  is the objective of the GAN with discriminator  $D_{GAN}$  and latents sampled from  $p(z_{t-1})$ ,  $\mathcal{L}_{GAN}^{VAE}$  is similar to  $\mathcal{L}_{GAN}$  except that it uses the latents sampled from  $q(z_{t-1}|x_{t-1:t})$  and has a separate discriminator  $D_{VAE}$  (Figure 6.1a).  $\lambda_1$  and  $\lambda_{KL}$  are hyperparameters chosen by evaluating similarity metrics on the validation set during training.

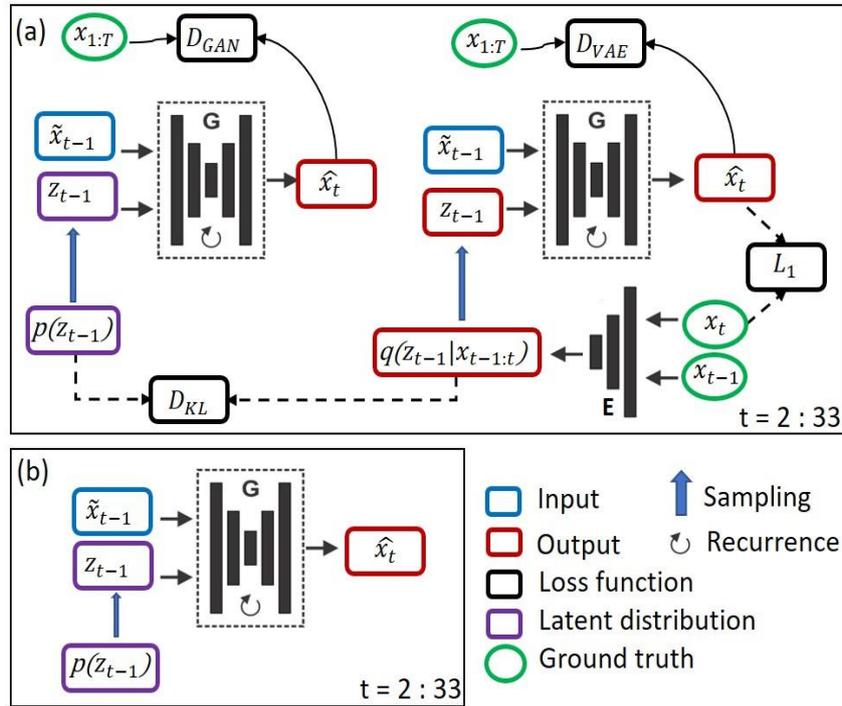


Figure 6.1 Block diagram of the VAE-GANs model to predict late CTP frames ( $\hat{x}_t$ ) from the sequence of previous (i.e. measured) frames ( $x_t$ ) shown in the training phase (a) and the testing phase (b), where  $E, G, D, p$  and  $q$  refer to the encoder, generator, discriminator, and prior and posterior distributions, respectively,  $L_1$  is the  $L1$  norm, and  $D_{KL}$  is the Kullback–Leibler divergence between  $p$  and  $q$  (see main text for details).

During the testing phase (Figure 6.1b), the generator takes in the previous frame  $\tilde{x}_{t-1}$  and random latent code  $z_{t-1}$  sampled from a prior distribution  $p(z_{t-1})$  to synthesise the next frame  $\hat{x}_t$ . The process iterates at each successive time step with the synthesised frames being fed back into  $G$ .

### 6.3.3 Network training and hyperparameter settings

The network was implemented in TensorFlow (Abadi, Barham et al. 2016) and trained on a dedicated workstation with a NVIDIA GeForce RTX 2080 Ti GPU. We performed 650000 iterations of the Adam optimizer (Kingma and Ba 2014) to train the model. The learning rate was decayed to zero linearly for the last 20000 iterations. For the GAN models, an optimizer with  $\beta_1 = 0.5$ ,  $\beta_2 = 0.999$  and a learning rate of 0.0002 was used. The estimation error term  $\lambda_1$  was set to 100 since this resulted in the best similarity performance on the validation set. For the VAE model an optimizer with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  and a learning rate of 0.001 was used.

The network was trained using 65 CTP studies (47,190 2D slices) and tested using 10 studies. The training dataset was augmented to a virtual size of 195 studies using randomly applied rotation and scaling.

### 6.3.4 Model validation and performance assessment

The model described in the previous sections allows us to predict frames  $\hat{x}_t, t \in \{t_{cut}, \dots, 33\}$  given measured frames  $x_t, t \in \{1, \dots, t_{cut} - 1\}$ . We tested the performance of this model for three increasingly challenging prediction tasks, corresponding respectively to  $t_{cut} = 26$ ,  $t_{cut} = 21$  and  $t_{cut} = 16$ , i.e. predicting 8, 13 and 18 frames, respectively.

#### *Image quality metrics*

The image quality of slices in predicted frames was compared to ground-truth frames using three metrics (Yu, Zhou et al. 2019): peak signal-to-noise ratio (PSNR), root mean squared error (RMSE) and structural similarity index (SSIM). PSNR is defined as:

$$PSNR(\hat{x}_t, x_t) = 10 \log \left( \frac{NR^2}{\|\hat{x}_t - x_t\|_2^2} \right) \quad (6.6)$$

where  $R$  and  $N$  denote the maximum dynamic range of the measured and predicted image, and the total number of voxels in the brain region, respectively. PSNR

is a relative image quality estimate usually expressed in decibels; higher PNSR can be indicative of better quality of the predicted set. RMSE is defined as:

$$RMSE(\hat{x}_t, x_t) = \sqrt{\frac{\sum_N (\hat{x}_t - x_t)^2}{N}} \quad (6.7)$$

and quantifies the discrepancy per voxel between the ground-truth and predicted images. Lower RMSE (close to zero) can indicate higher quality of estimated images. SSIM is defined as:

$$SSIM(\hat{x}_t, x_t) = \frac{(2\mu_{x_t}\mu_{\hat{x}_t} + c_1)(2\sigma_{x_t\hat{x}_t} + c_2)}{(\mu_{x_t}^2 + \mu_{\hat{x}_t}^2 + c_1)(\sigma_{x_t}^2 + \sigma_{\hat{x}_t}^2 + c_2)} \quad (6.8)$$

where  $\mu_{\hat{x}_t}$ ,  $\mu_{x_t}$  and  $\sigma_{\hat{x}_t}$ ,  $\sigma_{x_t}$  are the mean and variance, respectively, of  $\hat{x}_t$  and  $x_t$ ,  $\sigma_{x_t\hat{x}_t}$  is the covariance of  $\hat{x}_t$  and  $x_t$ ,  $c_1 = (0.01R)^2$  and  $c_2 = (0.03R)^2$  (Wang, Simoncelli et al. 2003). Since both PSNR and RMSE are based on the mean-squared-error (MSE) between the ground-truth and predicted set, they are susceptible to bias from over-smoothing. SSIM is a useful complement to PSNR and RMSE and measures the perceived alteration in structural information between two images. SSIM ranges from 0 to 1, with a higher SSIM indicating greater similarity between the two images. The metrics were calculated for each test subset individually and then averaged. The same quantitative metrics were used to evaluate the associated perfusion maps.

### ***Bolus shape analysis***

Since the shape of the contrast bolus affects the evaluation of the tissue status, we compared the characteristics of the contrast bolus passage for the predicted and ground-truth images to aid assessment of the approach – in particular, its performance in preserving the clinical content of images. The venous output function (VOF) was used to represent the bolus shape for each patient in the test group (Kasasbeh, Christensen et al. 2016). VOF is a straightforward signal to determine due to its large dimension and is less susceptible to the partial volume effect compared to the arterial input function (AIF).

The VOF was localised semi-automatically by searching for a voxel with the highest area under the time attenuation curve (TAC) within manually defined square regions of interest (ROI) placed on the straight or sagittal sinus. The same ROIs were placed on the corresponding predicted and ground-truth images. We fitted the first-

pass bolus in the VOF graph to a gamma-variate curve (Bennink, Oosterbroek et al. 2015, Kasasbeh, Christensen et al. 2016) to obtain a robust estimate of the area under the curve (AUC), bolus peak height ( $C_{max}$ ), and VOF width defined as the full-width-at-half-maximum (FWHM) of the fitted gamma-variate curve. We also compared the average VOF obtained from predicted and ground-truth images for patients in the test group. As bolus arrival time was different for each patient, all VOFs were aligned to their time-to-peak before averaging (Bennink, Oosterbroek et al. 2015).

### ***Infarct and penumbra size and location***

Since the size and extent of the infarct and penumbra is vital for stroke physicians to determine the best treatment options for patients, we compared the spatial agreement of these regions derived from the ground-truth and predicted haemodynamic maps.

The size of the infarct and penumbra was computed by thresholding the perfusion parameters in the predicted and ground-truth haemodynamic maps using the thresholds described in (Yu, Han et al. 2016): the penumbral region was defined by a delay  $\geq 3$  s relative to the normal hemisphere and infarct core as the sub-region of the penumbra with rCBF  $\leq 30\%$ . The delay in the pathological hemisphere was expressed as the difference between the TTP values of each voxel in the ipsilateral hemisphere from the mean TTP of the contralateral hemisphere. rCBF denotes the percentage ratio of the CBF in the ipsilateral hemisphere to the mean CBF in the contralateral hemisphere. Infarct/penumbra volumes were calculated from the predicted and ground-truth haemodynamic maps and the average and relative lesion size estimation error, respectively  $A$  and  $A_{rel}$ , were computed as (Moghari, Zhou et al. 2021):

$$A = \frac{\sum_n A_{\text{Test}} - A_{\text{GT}}}{n} \quad (6.9a)$$

$$A_{rel} = \frac{\sum_n \left( \frac{A_{\text{Test}} - A_{\text{GT}}}{A_{\text{GT}}} \right)}{n} \quad (6.9b)$$

where  $A_{\text{GT}}$  and  $A_{\text{Test}}$  show the lesion size in the ground-truth and test (predicted) images, respectively, and  $n$  is the number of test cases.

Spatial agreement for the lesion was quantified using the dice coefficient (F1-score), defined as:

$$\text{Dice coefficient} = 2 \left[ \left( \frac{1}{\text{Precision}} \right) + \left( \frac{1}{\text{Sensitivity}} \right) \right]^{-1} \quad (6.10)$$

where precision and sensitivity are defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6.11)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (6.12)$$

$TP$  denotes true positive voxels (correctly classified infarct/penumbra tissue),  $FP$  denotes false positive voxels (healthy tissue misclassified as infarct/penumbra) and  $FN$  denotes false negative voxels (infarct/penumbra tissue misclassified as healthy) (Figure 6.2). The highest possible value of the dice coefficient is 1, which indicates 100% spatial agreement between lesion estimates in the predicted and ground-truth data.

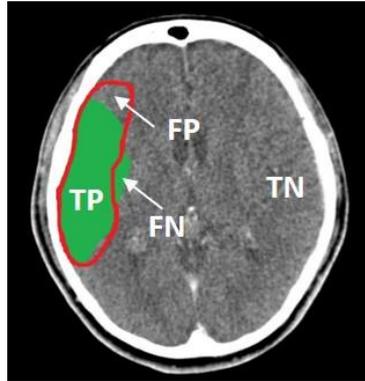


Figure 6.2 A typical lesion summary map. The red border indicates the lesion estimate obtained using the proposed VAE-GANs maps and the green region indicates the lesion estimate obtained using the ground truth haemodynamic maps. The false positive ( $FP$ , healthy tissue misclassified as abnormal), false negative ( $FN$ , abnormally perfused tissue misclassified as healthy), true positive ( $TP$ , correctly classified abnormally perfused tissue) and true negative ( $TN$ , correctly classified normally perfused tissue) regions are shown in the image for this example.

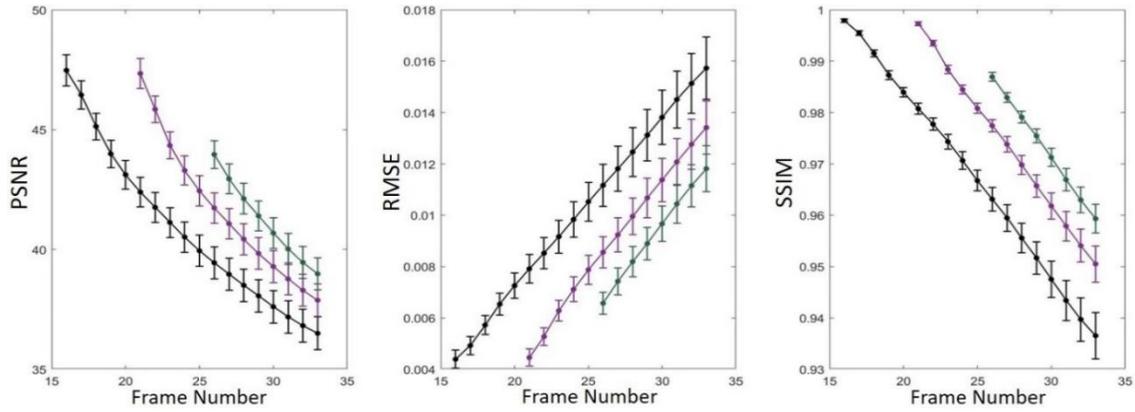


Figure 6.3 Average PSNR, SSIM, and RMSE (mean  $\pm$  SD) of the predicted CTP frames computed for 10 test studies. In each plot the green curve shows predicted CTP frames 26-33 estimated from frames 1-25, the purple curve represents predicted CTP frames 21-33 estimated from frames 1-20, and the black curve shows the metric for predicted CTP frames 16-33 estimated from frames 1-15.

## 6.4 RESULTS

### 6.4.1 Analysis of CTP images

Figure 6.3 shows the PSNR, RMSE and SSIM image quality results for the predicted CTP frames. These data indicate that image quality degraded monotonically from early predicted frames to late frames. The range of average PSNR, RMSE, and SSIM was 47.48-36.50 dB, 0.004-0.015, and 0.997-0.937, respectively. The degradation was approximately linear with frame number for SSIM and RMSE. Moreover, for all three metrics the standard deviation increased over these frames, suggesting that performance of the method becomes more variable the more distant the predicted frames are from measured data. Figure 6.4 shows some representative examples of how the image quality degradation manifests, including increased blurring of late predicted frames.

### 6.4.2 Analysis of bolus shape

Figure 6.5 and Table 6.1 show the results of bolus shape analysis for the most challenging scenario of predicting the last 18 frames (typically, the entire downslope of TAC) given the initial 15 frames. The average percentage difference between the VOF derived from predicted versus ground-truth images was  $< 4 \pm 4\%$ . The lowest percentage difference was observed for AUC followed (in order) by  $C_{max}$  and FWHM.

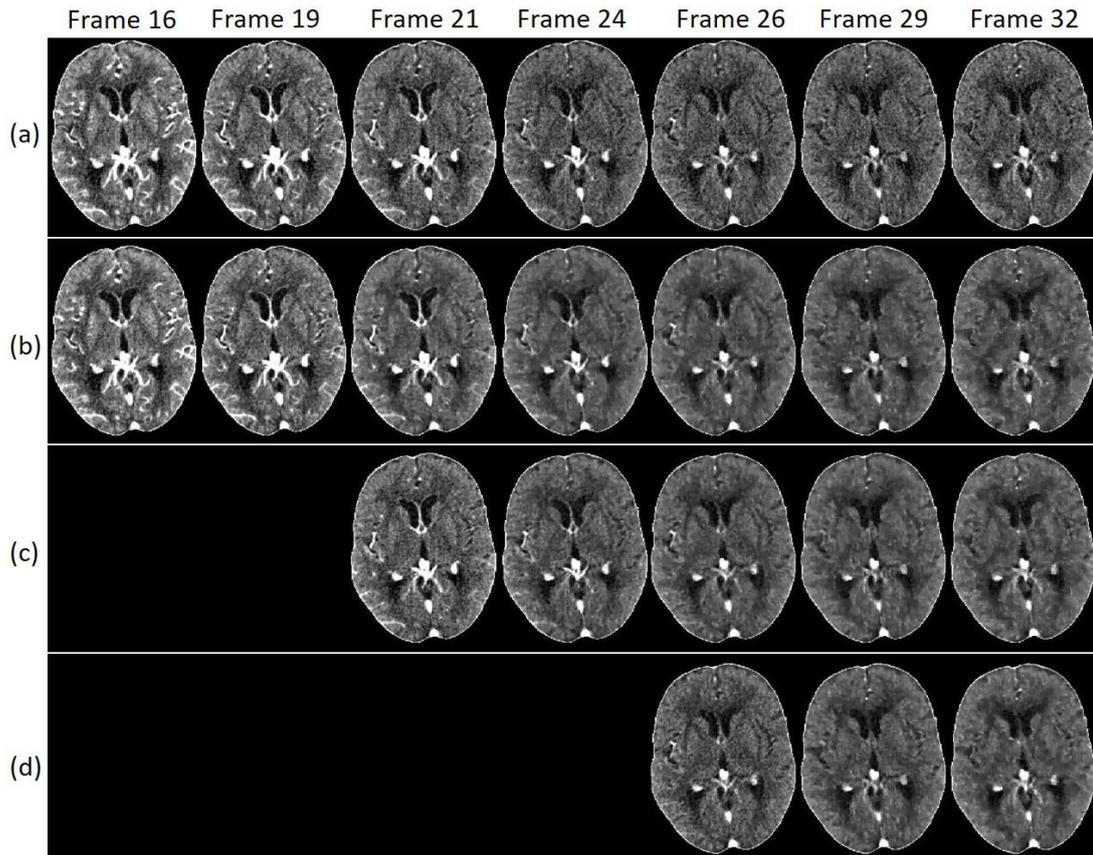


Figure 6.4 Examples of image quality degradation from early predicted frames to late frames. Row (a) shows ground-truth CTP frames. Rows (b)-(d) show the corresponding predicted frames estimated from frames 1-15 (b), frames 1-20 (c), and frames 1-25 (d).

Table 6.1 Comparison of the fitted VOF characteristics (AUC, FWHM and  $C_{max}$ ) in the ground-truth and predicted CTP images for the case of predicting frames 16-33 using frames 1-15, computed for 10 test cases. Values represent the mean  $\pm$  1 standard deviation.

	AUC	FWHM	$C_{max}$
<b>Absolute difference</b>	165.00 $\pm$ 153.76	0.66 $\pm$ 0.59	9.71 $\pm$ 10.27
<b>Percentage difference<sup>(a)</sup></b>	1.70 $\pm$ 1.36	3.74 $\pm$ 3.25	2.70 $\pm$ 3.10

(a) The percentage ratio of the absolute difference to the ground-truth values

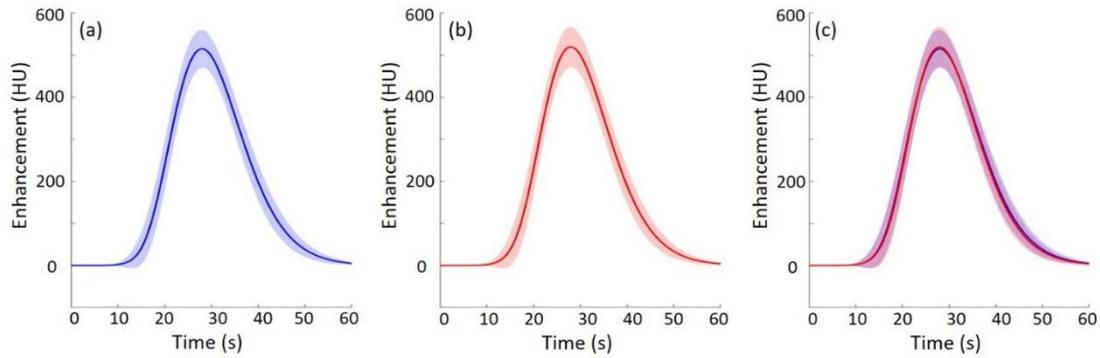


Figure 6.5 The mean fitted venous output function (VOF) for the case of predicting frames 16-33 using frames 1-15, computed for 10 test cases: (a) ground-truth, (b) predicted, and (c) overlay of (a) and (b). The shaded regions indicate  $\pm 1$  SD. The peak enhancement of all VOFs were aligned prior to averaging.

### 6.4.3 Haemodynamics and lesion analysis

Figure 6.6 compares the predicted and ground-truth haemodynamic maps for CBV, CBF, MTT, and TTP when predicting frames 16-33 from frames 1-15. Differences are difficult to discern visually for CBV and CBF, whereas for MTT and TTP it is clear that predicted maps tended to overestimate.

The quantitative comparison is summarised in Table 6.2 and Table 6.3. The best agreement in image quality (Table 6.2) was observed for CBV followed (in order) by CBF, MTT and TTP.

Evaluation of the spatial agreement of the lesion volume in the real and predicted images (Table 6.3) showed the average precision, sensitivity, and dice coefficient of the infarct and penumbra were greater than 60%, 70% and 65%, respectively, with the penumbra better estimated than infarct. The total hypo-perfused region showed greater average precision ( $> 75\%$ ), sensitivity ( $> 90\%$ ) and dice coefficient ( $> 80\%$ ) compared to either infarct or penumbra. The lesion size error metrics ( $A$  and  $A_{rel}$  in Table 6.3) indicated an overestimation of the average lesion volume by around 8 mL (15%) for the infarct, 21 mL (28%) for the penumbra, and 28 mL (22%) for the hypo-perfused regions. For all measurements, the SD was less than 15 mL and 18%.

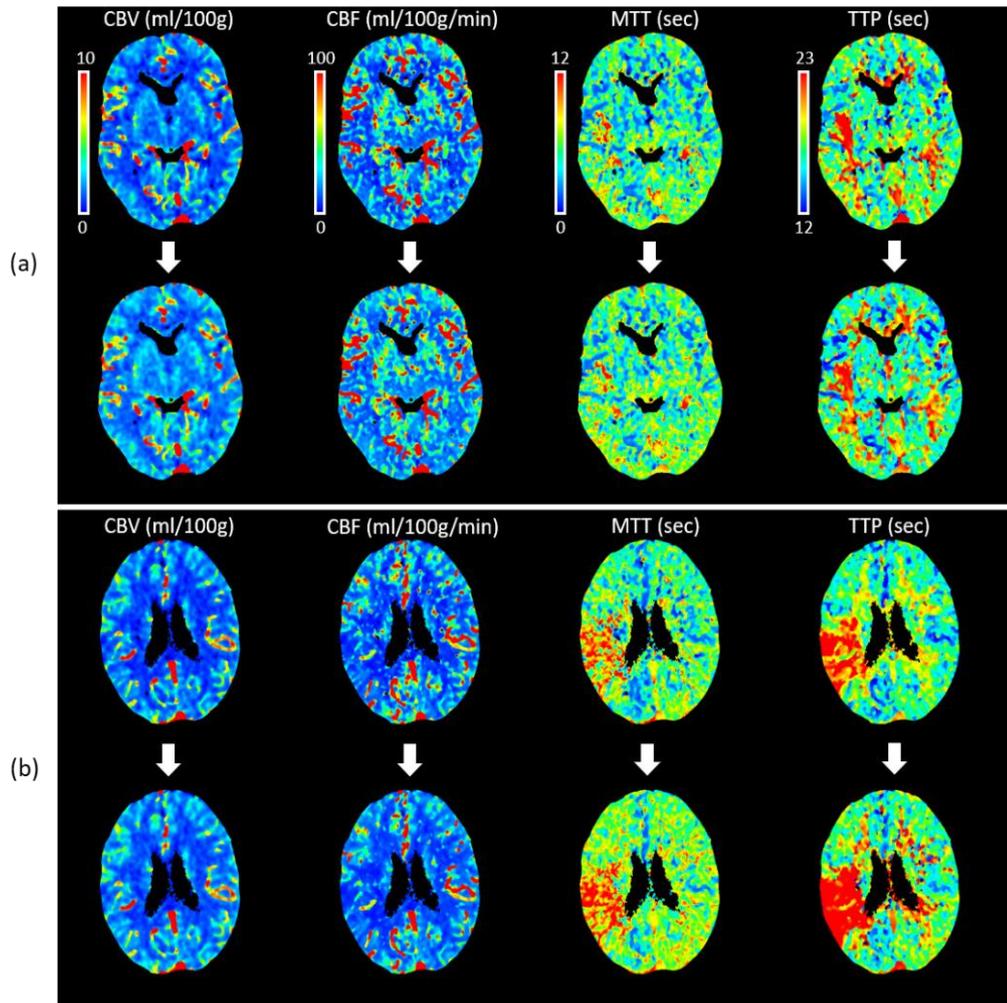


Figure 6.6 Ground-truth hemodynamic maps (top row) and the corresponding predicted maps using VAE-GANs (bottom row) for the case of predicting frames 16-33 using frames 1-15, shown for two patients (a) and (b). Columns from left to right show CBV, CBF, MTT and TTP maps.

Table 6.2 PSNR, RMSE and SSIM (mean  $\pm$  SD) of haemodynamic maps computed for the case of predicting frames 16-33 from frames 1-15, averaged across 10 test studies.

	PSNR	SSIM	RMSE
<b>CBV</b>	$33.42 \pm 7.47$	$0.97 \pm 0.03$	$2.06 \pm 0.78$
<b>CBF</b>	$26.97 \pm 5.35$	$0.94 \pm 0.03$	$3.1 \pm 0.50$
<b>MTT</b>	$21.06 \pm 3.47$	$0.87 \pm 0.03$	$3.96 \pm 0.45$
<b>TTP</b>	$18.14 \pm 4.65$	$0.83 \pm 0.05$	$4.45 \pm 0.72$

Table 6.3 Lesion size characterisation (mean  $\pm$  SD) computed for the case of predicting frames 16-33 from frames 1-15, averaged across the 10 test studies.

	<b>Precision</b>	<b>Sensitivity</b>	<b>Dice</b>	<b><math>A^a</math>(mL)</b>	<b><math>A_{rel}^b</math></b>
<b>Infarct</b>				7.58 $\pm$ 7.92	
	0.62 $\pm$ 0.21	0.72 $\pm$ 0.26	0.67 $\pm$ 0.23	20.78	$\pm$ 0.15 $\pm$ 0.14
<b>Penumbra</b>	0.68 $\pm$ 0.12	0.86 $\pm$ 0.07	0.76 $\pm$ 0.09	12.89	0.28 $\pm$ 0.18
<b>Hypo-perfused</b>	0.76 $\pm$ 0.08	0.92 $\pm$ 0.09	0.83 $\pm$ 0.07	28.36	$\pm$ 0.22 $\pm$ 0.11
				14.11	

(a) See equation (6.9a)

(b) See equation (6.9b)

## 6.5 DISCUSSION

In this study we conditioned a SAVP approach using the sequence of the first 25 (36 s), 20 (28.5 s), and 15 (21 s) reconstructed frames of a CTP study to predict the last 8 (24 s), 13 (31.5 s), and 18 (39 s) frames in order to reduce both the scan duration and the radiation dose. Feasibility of the method was assessed based on the image quality of the CTP images and haemodynamic maps, bolus shape and volumetric lesion characterisation.

The tendency of the image quality of predicted frames to degrade approximately linearly with frame number (Figure 6.3) is what one would expect from a recurrent model that estimates each frame based on all previous frames. The cumulative error reduces for the models making predictions based on a higher number of initial ground-truth frames.

The haemodynamics results (Table 6.2) indicated that CBV is the easiest parameter to predict reliably into the future, and TTP is the most difficult. A likely reason for this is that CBV is less susceptible to the TAC noise compared to TTP since it is calculated from the AUC of impulse response function (IRF). By contrast, the TTP is estimated time to a single value of the TAC, which can be more affected by noise.

In terms of bolus shape, predicting half of the frames resulted in less than 4% difference on average compared to using all frames (Table 6.1). However, the most important comparison from a clinical perspective (lesion analysis, Table 6.3) indicated

systematic overestimation of the infarct/penumbra in the predicted images. Since the infarct/penumbra volumes are calculated by thresholding the perfusion parameters, optimising the threshold on the predicted haemodynamic maps could potentially reduce the overestimation. It was beyond the scope of this study to find the optimised threshold or determine if such observed lesion overestimations were clinically significant. Infarct/penumbra have very noisy TACs (Figure 6.7c,d), with the temporal changes in voxel values as low as 10-20 HU. By contrast, the TACs for regions of healthy brain (Figure 6.7a,b) exhibit a well-defined shape and temporal enhancement from the passage of contrast agent. Therefore, the prediction and differentiation of the voxels in the infarct and penumbra regions is a more challenging task compared to the healthy regions. Our results showed the lesion characterisation in terms of “hypo-perfused” (i.e. combined infarct and penumbra) was better than either infarct or penumbra alone across all metrics. This indicates that although our method results in a loss of accuracy in delineating these two regions, it remains quite robust in identifying regions of reduced perfusion.

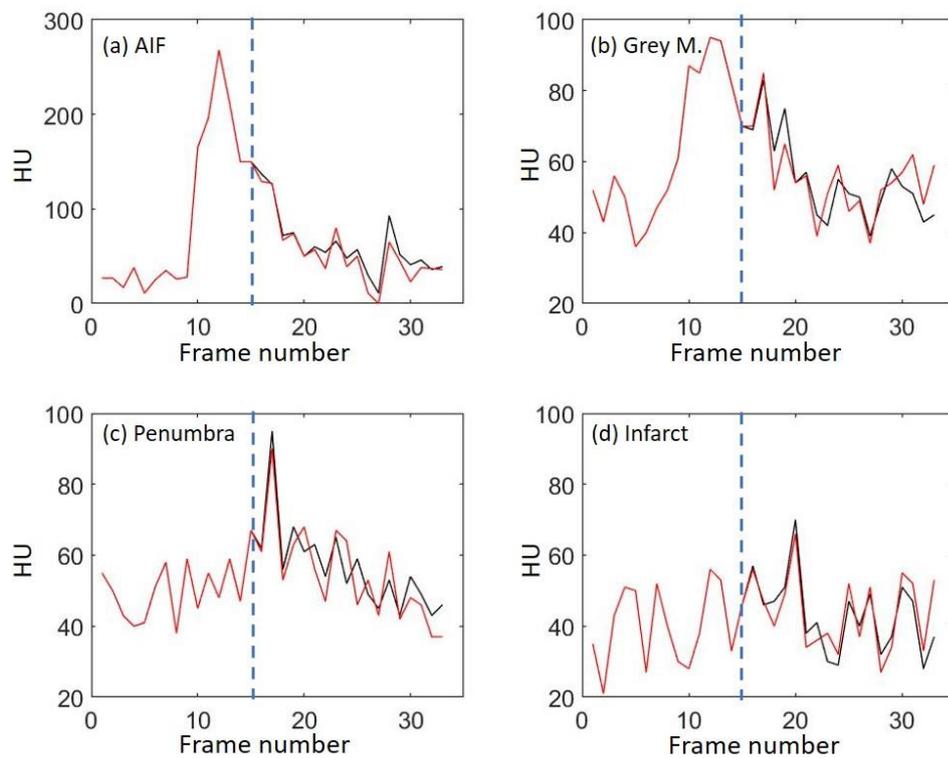


Figure 6.7 Examples of the time-attenuation curves (TAC) of a representative voxel in a region including (a) the arterial input function (AIF), (b) normal grey matter, (c) penumbra, and (d) infarct. In each plot, the red and black curves represent the ground-truth TAC and the TAC estimated using the VAE-GANs, respectively. Left of the vertical dashed line shows the model input.

There are several potential advantages of using the proposed VAE-GANs approach in CTP imaging. Firstly, it can be used to reduce the scan duration, thereby reducing the likelihood of patient head movement during the terminal phase of the scan in addition to the radiation dose. For example, prediction of the last 18 (39 s) CTP frames from the first 15 (21 s) frames reduces the scan duration and radiation dose by around 62% and 55%, respectively. Secondly, the model was fully automated and able to predict the late frames of the whole brain volume in around 30 s in the testing phase. This is in accordance with the requirements for a practical approach in time-critical acute ischaemic stroke management. A further potential application of the method is to replace motion-corrupted frames with higher quality predicted frames in a standard CTP protocol.

Although our results suggest that the proposed VAE-GANs is promising as a potential practical method to reduce scan duration in CTP imaging, there are some important limitations which should temper the conclusions. Firstly, the size and diversity of the training set could affect the model performance. Our training set was relatively small (65 studies, tripled using data augmentation methods) and performance should therefore be tested for a much larger cohort of CTP studies in which the lesion size distribution is greater. Secondly, the clinical content of the predicted images must be further assessed to determine how treatment decisions would be impacted by reducing scan time. Thirdly, the generalisability of the SAVP approach in the CTP application should be evaluated beyond the single scanner and single protocol tested in this study.

## **6.6 CONCLUSION**

In this study, we introduced and assessed a novel application of a deep learning approach to predict late CTP image frames from the early frames. The method has important potential implications for reducing the radiation dose and simultaneously reducing the probability of patient head movement in the terminal phase of the scan. Further clinical evaluation is needed to fully assess the utility of the method in practice, determining if the approach can match the clinical outcomes of analyses based on standard CTP protocols and assessing the generalisability of the method across a more expansive training/testing set of individuals and scanners.

# Chapter 7: Efficient Radiation Dose Reduction in Whole-Brain CT Perfusion Imaging Using a 3D GAN: Performance and Clinical Feasibility

---

Published in *Physics in Medicine & Biology*, 2021. 66(7): p. 075008: **Dashtbani Moghari, M.**, et al., “Efficient radiation dose reduction in whole-brain CT perfusion imaging using a 3D GAN: performance and clinical feasibility”.

## 7.1 ABSTRACT

Dose reduction in cerebral CT perfusion (CTP) imaging is desirable but is accompanied by an increase in noise that can compromise the image quality and the accuracy of image-based haemodynamic modelling used for clinical decision support in acute ischaemic stroke. The few reported methods aimed at denoising low-dose CTP images lack practicality by considering only small sections of the brain or being computationally expensive. Moreover, the prediction of infarct and penumbra size and location – the chief means of decision support for treatment options – from denoised data has not been explored using these approaches. In this work, we present the first application of a 3D generative adversarial network (3D GAN) for predicting normal-dose CTP data from low-dose CTP data. Feasibility of the approach was tested using real data from 30 acute ischaemic stroke patients in conjunction with low dose simulation. The 3D GAN model was applied to  $64^3$  voxel patches extracted from two different configurations of the CTP data– frame-based and stacked. The method led to whole-brain denoised data being generated for haemodynamic modelling within 90 seconds. Accuracy of the method was evaluated using standard image quality metrics and the extent to which the clinical content and lesion characteristics of the denoised CTP data were preserved. Results showed an average improvement of 5.15-5.32 dB PSNR and 0.025-0.033 SSIM for CTP images and 2.66-3.95 dB PSNR and 0.036-0.067 SSIM for functional maps at 50% and 25% of normal dose using GAN model in conjunction with a stacked data regime for image synthesis. Consequently, the average lesion volumetric error reduced significantly (p-value < 0.05) by 18-29% and dice coefficient improved significantly by 15-22%. We conclude that GAN-based

denoising is a promising practical approach for reducing radiation dose in CTP studies and improving lesion characterisation.

**Keywords:** CT perfusion imaging, acute ischemic stroke, dose reduction, denoising, deep learning, generative adversarial network

## 7.2 INTRODUCTION

Stroke is the second leading cause of death in the world (World Health Organization 2018, May 24) and results from a loss of blood supply to parts of the brain, potentially leading to impairment in cognition, language and motor control. Approximately 85% of stroke cases are ischaemic, caused by arterial blockage from a clot. A multimodal CT imaging regime, including non-contrast CT (NCCT), CT perfusion (CTP) and CT angiography (CTA), is the most common protocol for diagnosing acute ischaemic stroke and determining the best treatment options (Heit and Wintermark 2016). To ascertain the functional status of brain tissue and detect areas of abnormal perfusion, parameters including cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT), and time-to-peak (TTP) or  $T_{max}$  are derived from CTP data. CBV and CBF quantify, respectively, the blood volume (mL/100 g) and blood flow (mL/100 g/min) through brain tissue. MTT (s) reflects the average time for blood to transit through a region of interest. TTP and  $T_{max}$  (s) are measures of relative time of peak enhancement for brain tissue voxels. Based on the estimated perfusion maps of these parameters, the status of the brain tissue and the extent of the hypo-perfused (abnormally low perfusion) regions such as infarct core (severely hypo-perfused or dead tissue) and penumbra (potentially salvageable hypo-perfused tissue) can be obtained. This differentiation is important because of recent clinical advances in the acute treatment of stroke patients.

A CTP imaging protocol typically involves the rapid acquisition of several frames of the brain volume over ~1 minute after contrast agent administration. It therefore carries a relatively high radiation dose of around 5-6 mSv (Manniesing, Oei et al. 2015). The ability to reduce CTP radiation dose without compromising image quality and the accuracy of image-based stroke modelling is highly desirable, especially for younger adults and paediatric patients who are more vulnerable to the harmful stochastic effects of ionising radiation (Wolterink, Leiner et al. 2017).

However, in CTP imaging there is a trade-off between radiation dose and image quality. Reducing radiation dose increases noise in the image volumes from which the haemodynamic parameters must be computed. A variety of methods have been proposed for de-noising low-dose CT images while maintaining important physiological information. Related work can be classified into sinogram domain filtration, statistical iterative reconstruction (SIR) and image post-processing techniques.

In sinogram filtering approaches, the raw data are smoothed prior to image reconstruction using, for example, a bilateral or statistical nonlinear filter (Wang, Lu et al. 2005). It is, however, not always possible to obtain well-formatted raw projection data from commercial scanners, and edge blurring and resolution loss are also common drawbacks of this filtering approach. SIR methods accurately model the statistical properties of the raw data in the reconstruction and also optionally incorporate prior information using methods such as nonlocal means (Li, Yu et al. 2014), total variation (Sidky and Pan 2008) and dictionary learning (Xu, Yu et al. 2012). They have been shown to have advantages over analytical reconstruction for low-dose CT (Willeminck, Leiner et al. 2013, Geyer, Schoepf et al. 2015). Similar to sinogram filtering methods, SIR methods require access to the raw sinogram data, which is not always possible. Moreover, high computational cost and time delays between acquisition and reconstruction limit the practicality of SIR methods.

An alternative method for enhancing the quality of low-dose CT is post-processing or denoising of reconstructed images. Image-based denoising is particularly challenging in low-dose CTP images of ischaemic stroke patients since the voxel enhancement over time of dead/salvageable brain tissue can be very low, and thus masked by noise (Pisana, Henzler et al. 2017). Traditional image-based denoising methods used for CTP images and/or the haemodynamic maps derived from these images include filter-based approaches such as k-means clustering guided bilateral filtering (Pisana, Henzler et al. 2017) and time-intensity profile similarity (TIPS) filters (Mendrik, Vonken et al. 2010, Mendrik, Vonken et al. 2011). Potential drawbacks of filter-based approaches include uneven performance improvements, over-smoothing, edge blurring and other filter-induced artifacts.

Recently, deep learning-based methods have shown promising performance compared to traditional solutions for challenging medical imaging tasks, including

denoising. Several neural network architectures have been reported for image restoration and low-dose CT denoising in 2D and 3D (Wolterink, Leiner et al. 2017, Shan, Zhang et al. 2018). The problem of denoising 4D data, such as CTP data, is more challenging since both the spatial and temporal information must be considered within the network.

To the best of our knowledge, there are very few reports of deep learning-based methods for low-dose CTP. In (Kadimesetty, Gutta et al. 2018) and (Liu and Fang 2018) a 2D denoising CNN was used, ignoring 3D spatial information from adjacent slices within a frame and temporal information across CTP frames. In (Xiao, Liu et al. 2019), super-resolution and denoising was achieved by inputting three different 2D cross-sections of spatio-temporal CTP data to three independent super resolution denoising networks. The output of the three networks was merged before a final layer to output the denoised image sequence. Although the method demonstrated that deep learning-based denoising is feasible in low-dose CTP, application was limited to only one slice along the  $z$ -axis for each patient and whole brain lesion analysis was ignored. Moreover, the volumetric accuracy of the infarct and penumbral regions, which is of great importance for clinical decision support for the type of treatment, was not investigated.

In this work we chose a 3D generative adversarial network (GAN) architecture for denoising CTP data because of the excellent performance these models have demonstrated in a variety of image estimation tasks (Sixt, Wild et al. 2018, Arslan and Seke 2019, Yu, Zhou et al. 2019, Chen, Lin et al. 2020, Zhu, Zhang et al. 2020), including CT-related tasks such as metal artefact reduction (Ghani and Karl 2019, Koike, Anetai et al. 2020) and image denoising (Wolterink, Leiner et al. 2017, Yang, Yan et al. 2018, Huang, Chen et al. 2020). In (Xiao, Peters et al. 2020), integration of transfer learning with a 2D GAN was implemented for multimodal CT image super resolution. To date, however, 3D GANs have not been reported for denoising 4D contrast-enhanced data acquired in CTP. The GAN model consists of a generator (e.g. a traditional CNN-based model) to synthesise normal-dose images from low-dose images and a discriminator to distinguish between the synthesized images and true images. The generator and the discriminator compete to achieve a Nash equilibrium at which each component can no longer improve on its objective (Fedus, Rosca et al. 2017). The contributions of this paper are summarised as follows:

- i. We demonstrate the first application of a 3D GAN model for dose reduction in whole-brain 4D CTP of acute ischaemic stroke patients through prediction of normal-dose data from low-dose data. The U-net like generator allows much more efficient estimation of normal-dose images of the full brain compared to voxel-wise estimation methods.
- ii. We demonstrate practical and effective ways to exploit the spatio-temporal relationships in 4D CTP data using a 3D GAN architecture.
- iii. We present the first report of comprehensive stroke lesion characterisation based on synthesized normal-dose CTP images for the whole brain volume. This is achieved by using standard image quality metrics in addition to quantifying the extent to which the clinical content and lesion characteristics of synthesized images is preserved.

## **7.3 MATERIAL AND METHODS**

### **7.3.1 Dataset, data acquisition and pre-processing**

The dataset comprised retrospective CTP perfusion studies for 30 patients (21 males, 9 females) scanned in 2018 at the Department of Radiology, Westmead Hospital, Sydney. All studies were selected as positive stroke cases having occlusion in the right or left middle cerebral artery (MCA). The average age of the patients was 66 yr (SD 11.36 yr, range 43-86 yr). The study was carried out in accordance with a human research ethics protocol approved by Western Sydney Local Health District (ID: LNR/18/WMEAD/19).

The CTP scans were acquired using a Siemens Somatom Force CT scanner (dual source dual energy; 250ms rotation time, 1120 projection views per rotation, equating to a projection acquisition rate of 4,480 projections per second). Approximately 45 mL of non-ionic iodinated contrast was administered intravenously at 7 mL/s using a power injector with a 5 s delay. Scanning was performed in 4i cine mode with 1-min acquisition at 70 kVp and 200 mAs. The CT dose index (CTDI<sub>vol</sub>) and dose length product (DLP) of the CT perfusion scans were 159.84 mGy and 2398.0 mGy-cm, respectively. The first 25 CT volumes were acquired at 1.5 s intervals and the last 8 volumes at 3 s intervals, resulting in 33 brain volumes acquired over 1 minute. Each CT volume comprised 22 axial slices with 5 mm thickness covering the full brain volume. Slices were reconstructed in a 512 x 512 matrix with 0.43 mm in-plane

resolution. The total number of voxels for each patient was therefore 512 x 512 x 22 x 33.

Before model training, the skull and background were cropped from the raw images using the brain extraction tool (Jenkinson, Beckmann et al. 2012) and the CT numbers of brain voxels, ranging from -1024 to 1200 HU, were linearly scaled between -1 and 1.

### 7.3.2 Simulating low-dose CTP images

Tube current (mAs) is an important parameter affecting the quality of reconstructed CTP images. For a fixed scan duration, reducing the tube current reduces the radiation dose at the expense of increasing the noise in CTP images. Low-dose CTP images were simulated using the method described in (Britten, Crotty et al. 2004), where spatially correlated statistical noise is added directly to reconstructed images. The noise level in reconstructed CT images can be characterised by the standard deviation ( $\sigma$ ) of the CT number inside a uniform area such as the ventricle in the brain. If  $I_1$  represents the tube current normally used for the CTP imaging protocol and  $I_2$  represents some arbitrary lower tube current, the noise level  $\sigma_2$  associated with the new exposure can be estimated from:

$$\frac{\sigma_2}{\sigma_1} = \sqrt{\frac{I_2}{I_1}} \quad (7.1)$$

The difference in noise,  $\sigma_{diff}$ , between the two tube currents is calculated from:

$$\sigma_2^2 = \sigma_1^2 + \sigma_{diff}^2 \quad (7.2)$$

and allows us to simulate the higher noise level associated with tube current  $I_2$ . Since noise in reconstructed CT data is spatially correlated, we must also match the spectral properties of the simulated noise to real data (Britten, Crotty et al. 2004). To do this, a uniform cylindrical water phantom was imaged at different tube currents on the same scanner used for CTP studies and the noise autocorrelation function was calculated from the water phantom data. We then used a filter kernel derived from an 11 x 11 window centred on the peak of the autocorrelation function to convolve white Gaussian noise images prior to scaling them to the noise level  $\sigma_{diff}$  in equation 7.2. The final simulated low-dose CTP images, corresponding to the lower tube current  $I_2$ ,

were obtained by adding the noise images to the normal-dose CTP images acquired at tube current  $I_1$ .

In this study the reference tube current ( $I_1$ ) was 200 mAs and noise level associated with it was  $8.3 \pm 0.62$  HU, based on the average standard deviation of CT number inside the lateral ventricle for 30 patients imaged using our standard CTP protocol. We simulated low-dose images with two increased levels of noise corresponding to tube currents of 100 mAs and 45 mAs, respectively. In the remainder of the paper we refer to these as nominally 50% and 25% of the reference tube current.

### 7.3.3 GAN model architecture

The 3D GAN model (Wang, Yu et al. 2018) consists of a generator and a discriminator in adversarial competition. In this study, we conditioned the model to take in low-dose 3D CTP image patches instead of a random noise vector. During model training (Figure 7.1), the generator learns a mapping from the input  $x \sim P_{low}(x)$  (distribution of low-dose CTP image) to the output  $y \sim P_{normal}(y)$  (distribution of true normal-dose or “ground truth” image (GT)). Thus, the generator synthesizes  $G(x)$  – what we refer to as the synthesized or estimated normal-dose image – that resembles  $y$ . The discriminator takes in two pairs of images, the synthesized image pair  $(x, G(x))$  and the corresponding real image pair  $(x, y)$  and attempts to determine the authenticity of  $G(x)$ .

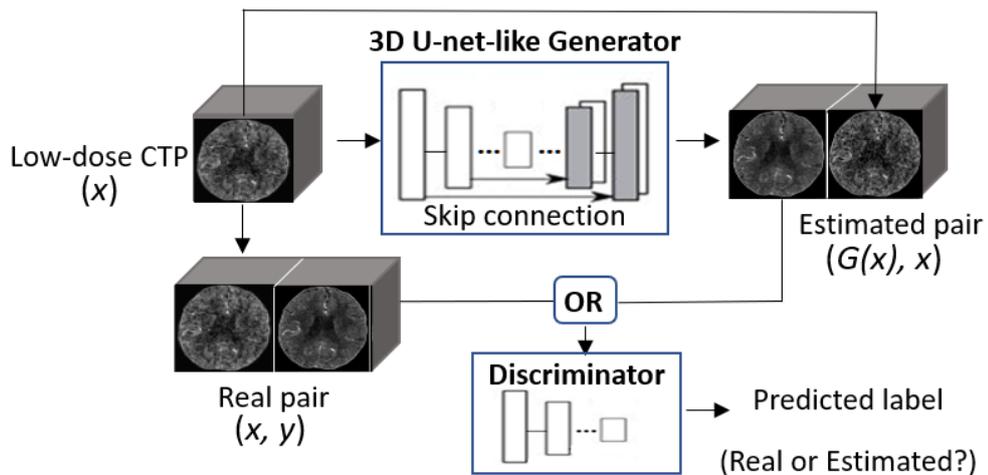


Figure 7.1 Overview of the 3D GAN training framework to estimate normal-dose CTP images from corresponding low-dose images.

The generator and discriminator are trained alternately to minimize and maximize the objective function in turn. To train the generator network, in addition to the discriminator feedback an  $L1$  norm penalty was used to preserve voxel-wise intensity similarity between the synthesised and real images and to suppress image blurring. Therefore, the objective ( $V$ ) of the 3D GAN was defined as:

$$\begin{aligned}
& \operatorname{argmin}_G \operatorname{max}_D (V_{GAN}(G, D) + \lambda V_{L1}(G)) & (7.3) \\
& = E_{x \sim P_{low}(x), y \sim P_{normal}(y)} [\log(D(x, y))] \\
& + E_{x \sim P_{low}(x)} [\log(1 - D(x, G(x)))] \\
& + \lambda E_{x \sim P_{low}(x), y \sim P_{normal}(y)} [y - G(x)]_{L1}
\end{aligned}$$

where  $G$  and  $D$  refer to the generator and discriminator, respectively, and  $E$  refers to maximum likelihood estimation. The first two terms of equation (7.3) represent the objective of the GAN model and the third term is the  $L1$  penalty.

### ***Generator architecture***

Since the low-dose and normal-dose CTP images derive from the same modality and thus share similar low-level information, a 3D U-net architecture is well suited as the generator network. The generator was a symmetric 3D U-net architecture with 6 down-convolutional (encoder) and 6 up-convolutional (decoder) layers with skip connections between each convolutional block (Figure 7.1).

Given the limited number of patient studies, limited GPU memory and the required training batch-size, we extracted large 3D image patches of size  $64 \times 64 \times 64$  voxels from 4D CTP images to train the model. In the encoder path,  $4 \times 4 \times 4$  filters were applied with a stride of 2 to the image patches, reducing the size of feature maps from  $64^3$  to  $1 \times 1 \times 1$ . A leaky ReLu activation function with -0.2 slope was used to introduce nonlinearity at each layer. In the decoder path,  $2 \times 2 \times 2$  filters were applied with a stride of 2, followed by a ReLu at each layer. Batch normalization was applied in each convolutional layer to improve learning (Garbin, Zhu et al. 2020).

### ***Discriminator architecture***

The discriminator inputs a real/synthesized pair of 3D image patches to determine the authenticity of the synthesized patch. It consisted of four convolutional layers where  $4 \times 4 \times 4$  filters with stride 2 were applied to image patches at each layer along with a leaky ReLu activation function with slope of 0.2. After the last layer, a

kernel of size  $1 \times 1 \times 1$  and a sigmoid layer was used to output a classification of 0 (real) or 1 (fake) for the synthesised CTP image patches.

### 7.3.4 4D data handling

We compared two methods to leverage the 4D nature of CTP data within the 3D GAN model (Figure 7.2). The first method (Figure 7.2a) involved concatenating the 33 volumetric frames for a patient and extracting the  $64 \times 64 \times 64$  patches from the concatenated data set ( $512 \times 512 \times 726$ ) with a stride of 64. This provided 768 ( $8 \times 8 \times 12$  in  $x, y$  and  $z$ , respectively) image patches for each training/testing sample. Due to the choice of stride, there was no overlap between patches. Zero-padding was used for the last 64 patches along the  $z$  direction. We refer to this data arrangement as ‘frame-based’ and to the generator trained using this arrangement as  $G_{xyz}$ .

The second method (Figure 7.2b) involved concatenating the time sequence for each 2D brain slice and is based on the fact that we are ultimately interested in tracking a voxel time series in the brain to calculate haemodynamic parameters. These slice-based time sequences were further concatenated to form a 3D data set of dimensions  $512 \times 512 \times 726$ , from which  $64 \times 64 \times 64$  spatio-temporal image patches were extracted. We refer to this data arrangement as ‘stacked’ and to the generator trained on this data configuration as  $G_{xyt}$ .

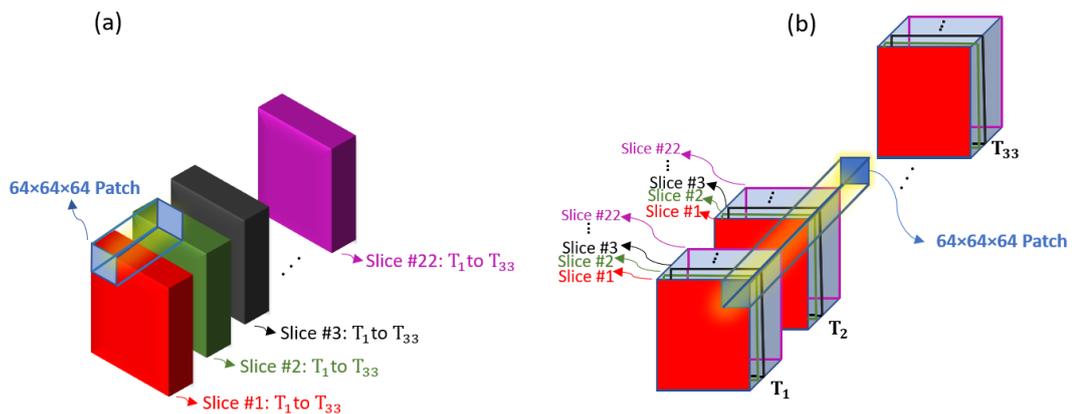


Figure 7.2 Extraction of 3D input patches from 4D CTP data using the frame-based (a) and stacked (b) regimes.

Conceptually, the CT number fluctuation in the plane reflects spatial information within a slice in both the frame-based and stacked data configurations. The CT number fluctuation along the 3rd dimension of the frame-based configuration prioritises the spatial information from adjacent slices within a frame; for the stacked configuration it prioritises the temporal information related to wash-in/wash-out of contrast agent across CTP frames. The GAN network was trained and tested on the two data arrangements separately to compare the performance difference with respect to this trade-off.

### **7.3.5 Model training and testing**

The network was implemented in PyTorch (Paszke, Gross et al. 2019) and trained on a dedicated workstation with a NVIDIA GeForce RTX 2080 Ti GPU. We performed 150 epochs to train the network for each cross-validation split. For the first 100 epochs a fixed learning rate of 0.0002 was applied; this was linearly reduced to zero for the next 50 epochs. The Adam solver with a batch size of 4 was used to train all networks. The estimation error term  $\lambda$  was set to 300 as it resulted in the best estimation performance. The 3D GAN model proposed in (Wang, Yu et al. 2018) was adapted for denoising low-dose CTP images.

A five-fold cross-validation strategy was used to train and test the network under our different experimental conditions (noise level, 4D data regimen). For each cross-validation split, subjects were divided into five subsets, four for training (24 patients) and one for testing (6 patients). Based on the method described in Section 7.3.4, the total number of training and testing patches was 18,432 ( $768 \times 24$ ) and 4,608 ( $768 \times 6$ ) for the frame-based and stacked data regimes, respectively. All estimated image patches for a given regime were merged together to form the complete estimated CTP volume. Vitrea software (Vital Images, Minneapolis, Minnesota, USA) was used to generate haemodynamic maps from the true normal-dose, simulated low-dose and GAN-denoised CTP images. Vitrea uses a delay insensitive singular value deconvolution (SVD+) algorithm (Angel 2010) to produce functional haemodynamic maps with frame-to-frame motion correction included.

### 7.3.6 Model validation and performance assessment

#### *Image quality metrics*

The quality of GAN-denoised images and the associated perfusion maps derived from them was assessed against the ground-truth using two widely used image quality metrics (Yu, Zhou et al. 2019): peak signal-to-noise ratio (PNSR) and structural similarity index (SSIM). The metrics are defined as follows:

$$PSNR(y, G(x)) = 10 \log \left( \frac{NR^2}{\|y - G(x)\|_2^2} \right) \quad (7.4a)$$

where  $R$  is the dynamic range and  $N$  denotes the total number of voxels in the image.

$$SSIM(y, G(x)) = \frac{(2\mu_y\mu_{G(x)} + c_1)(2\sigma_{yG(x)} + c_2)}{(\mu_y^2 + \mu_{G(x)}^2 + c_1)(\sigma_y^2 + \sigma_{G(x)}^2 + c_2)} \quad (7.4b)$$

where  $\mu_y$ ,  $\mu_{G(x)}$  and  $\sigma_y$ ,  $\sigma_{G(x)}$  are the mean and variance, respectively, of  $y$  and  $G(x)$ ,  $\sigma_{yG(x)}$  is the covariance of  $y$  and  $G(x)$ ,  $c_1 = (0.01R)^2$  and  $c_2 = (0.03R)^2$  (Wang, Simoncelli et al. 2003).

PSNR is a relative image quality estimate based on a logarithmic decibel scale; higher PNSR can be indicative of better synthesis performance. However, since PSNR is based on the mean squared error between the true and GAN-denoised images it is susceptible to bias from over-smoothing. By contrast, SSIM quantifies the perceived alteration in structural information between two images, including edges, and is a useful complement to PSNR. SSIM varies between 0 and 1 and higher SSIM implies greater similarity between two images. The metrics were computed for each test subset individually and then averaged.

#### *Comparison with alternative denoising methods*

We note that a 3D U-net generator with skip connections is a key component of the proposed GAN model. Compared to a 3D U-net alone, the adversarial training (discriminator feedback) of the GAN increases the estimation performance. Thus, it is redundant to compare the performance of the proposed GAN model with a 3D CNN (U-net) and instead we chose to compare the performance with a block-matching and 4D filtering (BM4D) (Maggioni, Katkovnik et al. 2012) and also a residual encoder-

decoder convolutional neural network (RED-CNN) (Chen, Zhang et al. 2017). These approaches have been widely used in noise reduction applications and have shown impressive denoising performance in CT and MRI (Maggioni, Katkovnik et al. 2012, Shi, Hu et al. 2016, Chen, Zhang et al. 2017, Shan, Zhang et al. 2018, Ma, Feng et al. 2019).

The BM4D algorithm is an extension of the BM3D filter for volumetric data in which mutually similar 3D patches of the image are grouped into a 4D array and jointly filtered in a transform domain. We implemented BM4D to denoise each individual CTP frame at the two increased noise levels. RED-CNN is a 2D deep learning model incorporating a symmetric convolutional and deconvolutional neural network aided by residual learning (He, Zhang et al. 2016) with shortcut connections. To train the RED-CNN, input image patches of size  $64 \times 64$  were extracted from the original images and the loss was minimised using the Adam optimiser. Both BM4D and RED-CNN methods were implemented based on the image pre-processing and hyper-parameter settings in the original papers.

### ***Expert quality rating***

The diagnostic quality of the haemodynamic maps derived from GAN-denoised CTP images was assessed using a blinded expert ranking. Two experts (one radiologist and one stroke physician) independently ranked the CBV, CBF and TTP maps obtained from the noisy (simulated low-dose) and denoised (BM4D, RED-CNN, frame-based GAN and stacked GAN) data relative to the true normal dose (GT) maps. This blinded comparison was done for a randomly-selected slice from each of 5 randomly-selected test studies, and at 2 dose levels (50% and 25% of the normal dose), resulting in a total of 150 assessments ( $5 \text{ slices} \times 5 \text{ methods} \times 3 \text{ maps} \times 2 \text{ dose levels}$ ). Slices were chosen from the slice range 6-20 to avoid the top and bottom of the brain. For each assessment, the experts compared the labelled GT map against the unlabelled noisy and denoised options presented in random order. The ranking was based on structure preservation, lesion preservation and overall quality using a subjective six-point scale (0 = non-diagnostic and 5 = equivalent to GT).

### ***Infarct and penumbra size and location***

Ultimately, it is the size and location of the infarct core and penumbra which are of primary clinical importance. Stroke physicians use this information, derived from the haemodynamic summary maps, as a decision support for determining the best

treatment option. Therefore, a key part of our evaluation was to compare the infarct, penumbra, and hypo-perfused (infarct + penumbra) region size and location derived from true normal-dose, simulated low-dose and GAN-denoised CTP data.

Infarct/penumbra volumes are usually determined by thresholding the perfusion parameters in haemodynamic maps. The optimal choice of perfusion parameters and thresholds is an ongoing topic of research (Wintermark, Flanders et al. 2006, Yu, Han et al. 2016, Chen, Bivard et al. 2019). We applied a previously established threshold [42] to the perfusion maps generated by Vitrea in order to investigate the lesion size agreement between low-dose/denoised images and the corresponding true images. According to (Yu, Han et al. 2016), the penumbral region is characterised by delay  $\geq 3$  s relative to the contralateral hemisphere and the infarct core is the sub-region of the penumbra for which  $rCBF \leq 30\%$ .  $rCBF$  denotes relative CBF and is defined as the percentage ratio of the local CBF value in the ipsilateral hemisphere to the mean CBF in the contralateral hemisphere. Delay was calculated by subtracting the TTP value for each voxel from the mean TTP of the contralateral normal hemisphere. We computed infarct/penumbra summary maps from true normal-dose, simulated low-dose and GAN-denoised low-dose CTP images. The average absolute and relative lesion size estimation error, respectively  $E_{abs}$  and  $E_{rel}$ , were calculated as:

$$E_{abs} = \frac{\sum_N |A_{GT} - A_{Test}|}{N} \quad (7.5a)$$

$$E_{rel} = \frac{\sum_N \left( \frac{|A_{GT} - A_{Test}|}{A_{GT}} \right)}{N} \quad (7.5b)$$

where  $A_{GT}$  is the lesion size in the GT image,  $A_{Test}$  is the lesion size in the test image (low-dose or GAN-denoised), and  $N$  is the number of cases.

We also compared the dice coefficient (F1-score) to assess the spatial agreement of the estimated lesion with the GT. The dice coefficient is defined as:

$$\text{Dice coefficient} = 2 \left[ \left( \frac{1}{\text{Precision}} \right) + \left( \frac{1}{\text{Sensitivity}} \right) \right]^{-1} \quad (7.6a)$$

where precision and sensitivity are given by:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7.6b)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (7.6c)$$

Here  $TP$  denotes true positive (correctly classified infarct/penumbra tissue),  $FP$  denotes false positive (healthy tissue misclassified as infarct/penumbra) and  $FN$  denotes false negative (infarct/penumbra tissue misclassified as healthy) (Figure 7.3). The highest possible dice coefficient is 1 and indicates perfect spatial agreement between lesion estimates in the denoised and GT data.

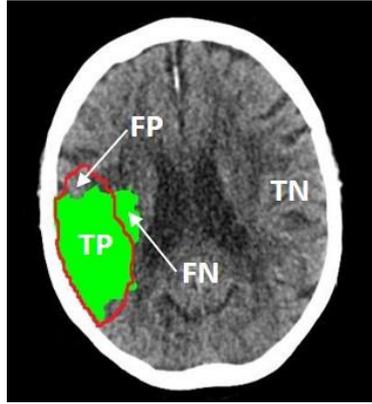


Figure 7.3 An example lesion summary map. The red border indicates the lesion volume obtained using denoised haemodynamic maps and the green region using normal-dose haemodynamic maps (GT). Four regions including false positive ( $FP$ , healthy tissue misclassified as abnormal), false negative ( $FN$ , abnormally perfused tissue misclassified as healthy), true positive ( $TP$ , correctly classified abnormally perfused tissue) and true negative ( $TN$ , correctly classified normally perfused tissue) are shown in the image.

## 7.4 RESULTS

### 7.4.1 Analysis of CTP images

Table 7.1 shows the PSNR and SSIM results for the simulated low-dose and GAN-denoised CTP images relative to the true normal-dose CTP images. Results for the BM4D and RED-CNN denoising approaches are shown for comparison. The 3D GAN using the stacked data regime ( $G_{xyt}$ ) outperformed the BM4D, RED-CNN and  $G_{xyz}$  methods for both metrics and both increased noise levels. There was a statistically significant difference between the SSIM metric computed for the  $G_{xyt}$  and RED-CNN methods ( $p < 0.005$ ), indicating an important improvement in the preservation of structural information in CTP images using the proposed approach. Compared to the low-dose case with no denoising, the stacked regime improved PSNR and SSIM by

5.32 dB and 0.025, respectively, at 50% dose and 5.15 dB and 0.033, respectively, at 25% dose.

Table 7.1 PSNR and SSIM (mean  $\pm$  SD calculated for 30 patients) for denoised CTP images at 50% and 25% of the reference dose. Bold font indicates the best performance.

Method	50% dose		25% dose	
	PSNR	SSIM	PSNR	SSIM
Low-dose	37.07 $\pm$ 4.64	0.962 $\pm$ 0.008	32.14 $\pm$ 4.89	0.931 $\pm$ 0.016
BM4D	38.08 $\pm$ 2.13	0.960 $\pm$ 0.009	32.93 $\pm$ 2.53	0.937 $\pm$ 0.011
RED-CNN	41.82 $\pm$ 2.34	0.971 $\pm$ 0.005	36.58 $\pm$ 2.59	0.949 $\pm$ 0.008
3D GAN ( $G_{xyz}$ )	41.18 $\pm$ 3.40	0.978 $\pm$ 0.009	36.04 $\pm$ 3.54	0.958 $\pm$ 0.014
3D GAN ( $G_{xyt}$ )	<b>42.39 <math>\pm</math> 2.37</b>	<b>0.987 <math>\pm</math> 0.006</b>	<b>37.29 <math>\pm</math> 2.61</b>	<b>0.964 <math>\pm</math> 0.012</b>

#### 7.4.2 Analysis of haemodynamic maps

Figure 7.4 shows an example of raw CTP images and haemodynamic maps for the GT and denoised data. Qualitatively, maps generated from the GAN with the stacked data regime showed better contrast and better overall agreement with the normal-dose maps compared to the frame-based GAN, BM4D and RED-CNN-based approaches.

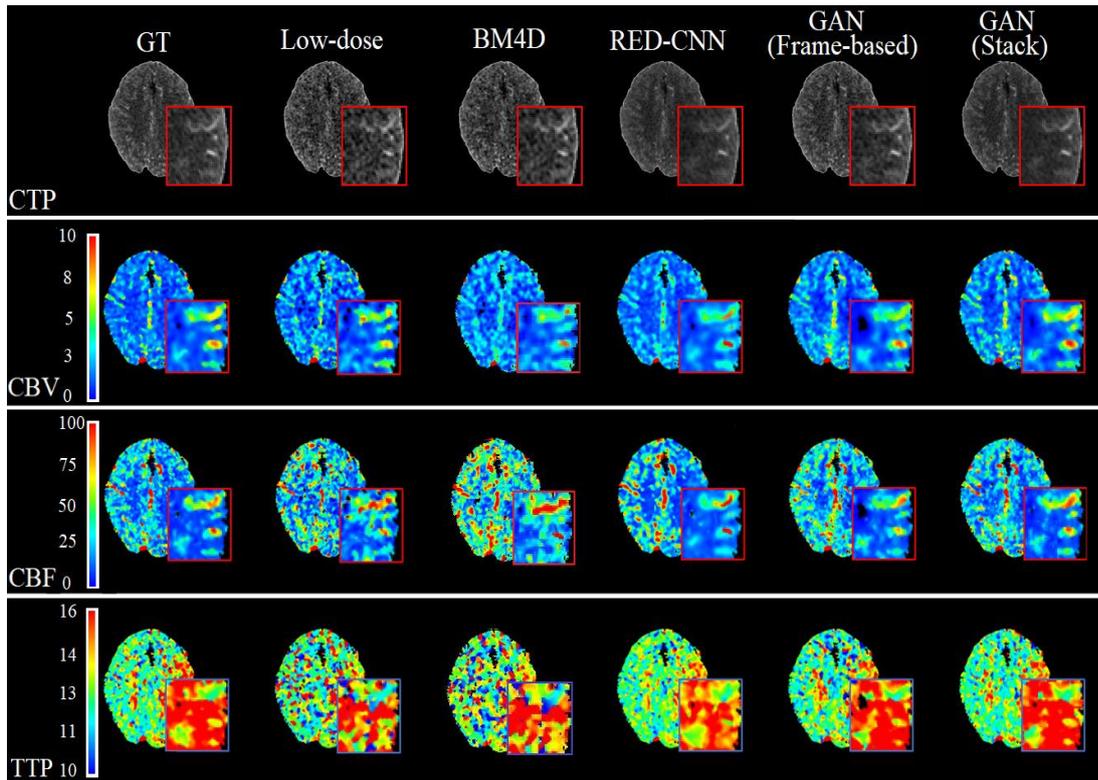


Figure 7.4 CTP images (top row) and haemodynamic maps (CBV, CBF, TTP) obtained using different denoising approaches (columns) at 25% dose.

The qualitative results were confirmed by the quantitative assessment (Table 7.2). The GAN network outperformed the other methods for all haemodynamic maps at 50% / 25% dose, improving PSNR by 2.66 / 2.78 dB (CBV), 3.95/ 3.56 dB (CBF) and 3.76 / 2.78 dB (TTP), and improving SSIM by 0.036 / 0.049 (CBV), 0.049/ 0.047 CBF) and 0.053 / 0.067 (TTP).

Table 7.2 PSNR and SSIM (mean  $\pm$  SD calculated for 30 patients) for haemodynamic maps obtained using different denoising methods applied to CTP images acquired at 50% and 25% of normal dose. Bold font indicates the best performance.

Method	Dose	CBV		CBF		TTP	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Low-dose		30.36 $\pm$	0.935 $\pm$	25.63	0.868 $\pm$	24.1 $\pm$	0.831 $\pm$
		2.78	0.018	$\pm$ 5.11	0.033	3.11	0.025
BM4D		30.72 $\pm$	0.941 $\pm$	26.49	0.870 $\pm$	25.01	0.836 $\pm$
		2.05	0.016	$\pm$ 3.11	0.042	$\pm$ 2.25	0.032
RED-CNN	50%	31.75 $\pm$	0.950 $\pm$	27.63	0.880 $\pm$	26.53	0.845 $\pm$
		2.01	0.012	$\pm$ 3.31	0.039	$\pm$ 2.11	0.022
3D GAN ( $G_{xyz}$ )		31.93 $\pm$	0.961 $\pm$	28.2 $\pm$	0.893 $\pm$	26.51	0.862 $\pm$
		2.83	0.011	3.48	0.046	$\pm$ 3.23	0.036
3D GAN ( $G_{xyt}$ )		<b>33.02 <math>\pm</math></b>	<b>0.971 <math>\pm</math></b>	<b>29.58</b>	<b>0.917 <math>\pm</math></b>	<b>27.86</b>	<b>0.884 <math>\pm</math></b>
		<b>0.78</b>	<b>0.009</b>	$\pm$ <b>2.71</b>	<b>0.032</b>	$\pm$ <b>2.01</b>	<b>0.019</b>
Low-dose		26.18 $\pm$	0.891 $\pm$	21.78	0.815 $\pm$	20.69	0.766 $\pm$
		2.99	0.025	$\pm$ 4.91	0.052	$\pm$ 3.49	0.034
BM4D		26.61 $\pm$	0.894 $\pm$	22.21	0.819 $\pm$	21.26	0.772 $\pm$
		2.08	0.026	$\pm$ 3.22	0.058	$\pm$ 2.46	0.033
RED-CNN	25%	27.21 $\pm$	0.907 $\pm$	23.16	0.831 $\pm$	22.45	0.784 $\pm$
		2.31	0.023	$\pm$ 3.28	0.055	$\pm$ 2.31	0.031
3D GAN ( $G_{xyz}$ )		27.82 $\pm$	0.918 $\pm$	23.98	0.848 $\pm$	22.88	0.798 $\pm$
		2.53	0.021	$\pm$ 4.27	0.061	$\pm$ 3.19	0.038
3D GAN ( $G_{xyt}$ )		<b>28.96 <math>\pm</math></b>	<b>0.940 <math>\pm</math></b>	<b>25.43</b>	<b>0.862 <math>\pm</math></b>	<b>23.56</b>	<b>0.833 <math>\pm</math></b>
		<b>1.52</b>	<b>0.015</b>	$\pm$ <b>2.87</b>	<b>0.043</b>	$\pm$ <b>2.10</b>	<b>0.034</b>

### 7.4.3 Expert quality rating

The mean quality ratings of the experts are shown in Figure 7.5. In nearly all cases, average scores for structure, lesion and quality were higher and had lower standard deviation for haemodynamic maps computed from  $G_{xyt}$  images. The scores on the haemodynamic maps computed from  $G_{xyt}$  outputs had statistically significant

difference with the low-dose images. Overall, the average lowest score was for perfusion maps calculated from low-dose and BM4D denoised images.

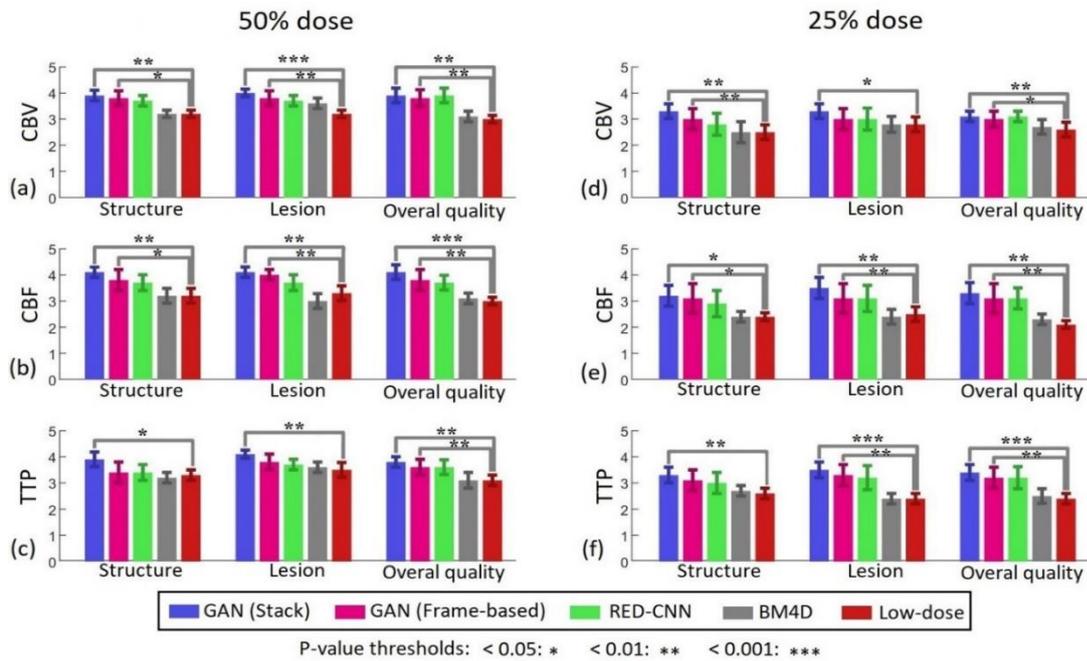


Figure 7.5 Expert quality rating (mean  $\pm$  SD) for haemodynamic maps obtained using different methods at 50% (a-c) and 25% (d-f) dose. Statistically significant scores with respect to the low-dose haemodynamic maps are indicated by \*, \*\*, and \*\*\* for GAN denoised images.

#### 7.4.4 Infarct and penumbra size and location

Figure 7.6, Figure 7.7, and Figure 7.8 show infarct, penumbral and hypo-perfused volumes obtained at 50% and 25% dose versus the corresponding volume obtained using true normal-dose data (GT). Several features are clear from these results: (i) low-dose images led to overestimation of lesion size; (ii) the GAN using a stacked data regime provided the best agreement of the infarct/penumbral/hypo-perfused volumes and the GT for both lower doses; and (iii) the GAN using a frame-based regime led to the infarct/penumbral volumes being over/under-estimated but the hypo-perfused volume being in close agreement with the GT.

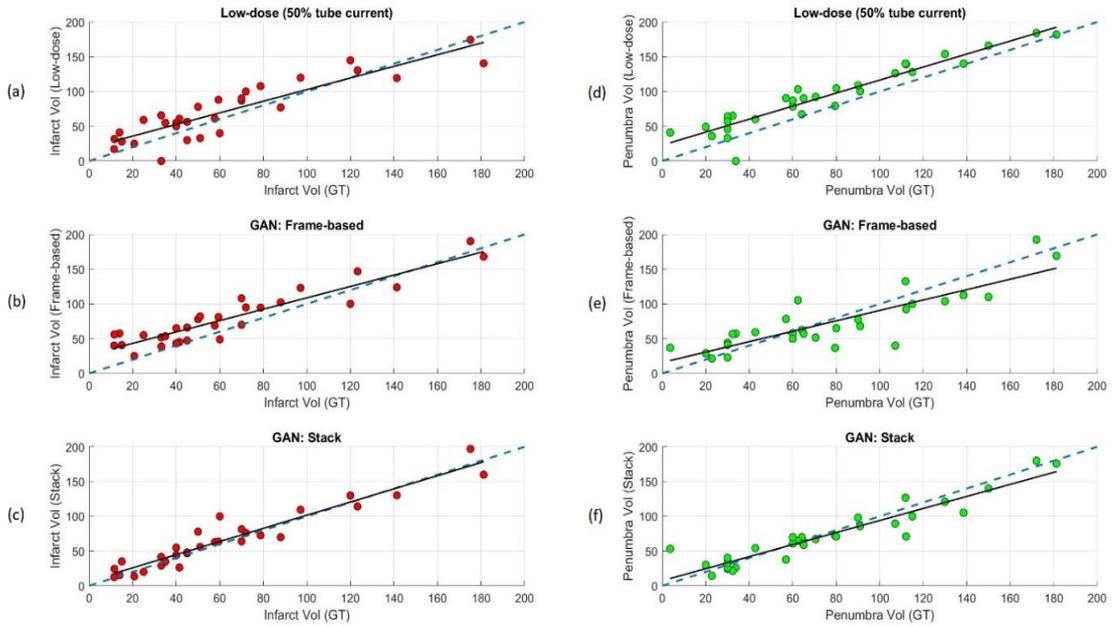


Figure 7.6 Estimated volume (mL) of infarct (a-c) and penumbra (d-f) from data obtained at 50% dose versus the volume from normal-dose (GT) images, shown for 30 acute stroke cases. Results are shown for low-dose CTP (a, d),  $G_{xyz}$  (b, e) and  $G_{xyt}$  (c, f) data. The dashed line shows 100% agreement with the GT.

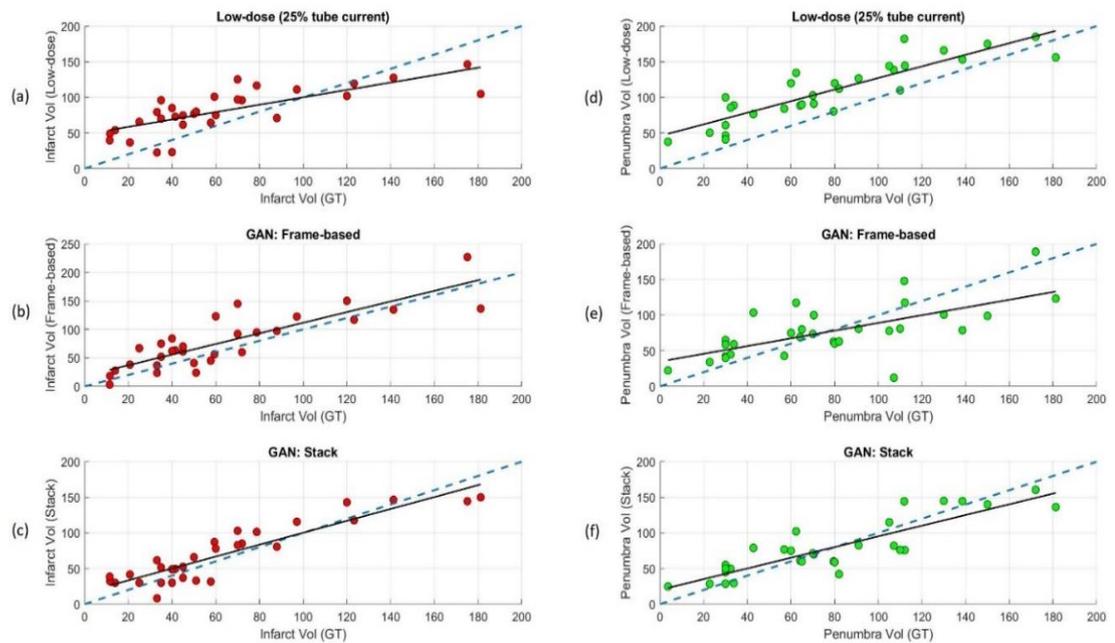


Figure 7.7 Estimated volume (mL) of infarct (a-c) and penumbra (d-f) from data obtained at 25% dose versus the volume from normal-dose (GT) images, shown for 30 acute stroke cases. Results are shown for low-dose CTP (a, d),  $G_{xyz}$  (b, e) and  $G_{xyt}$  (c, f) data. The dashed line shows 100% agreement with the GT.

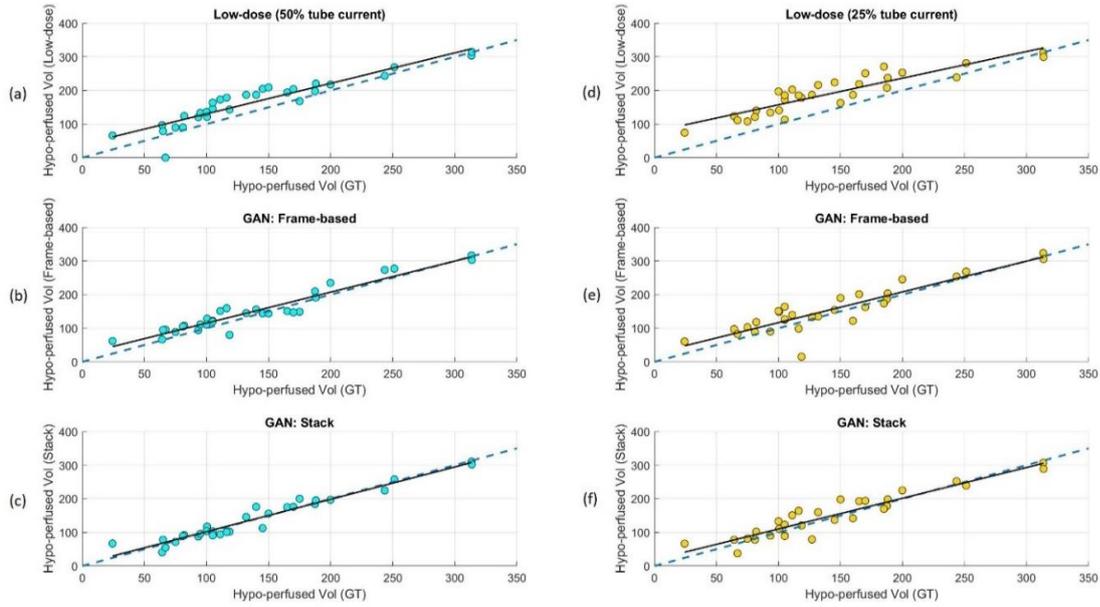


Figure 7.8 Estimated volume (mL) of hypo-perfused (infarct + penumbra) regions at 50% (a-c) and 25% (d-f) dose versus the volume in normal-dose (GT) images, shown for 30 acute stroke cases. Results are shown for low-dose CTP images (a, d),  $G_{xyz}$  (b, e) and  $G_{xyt}$  (c, f) data. The dashed line shows 100% agreement with the GT.

Table 7.3 and Table 7.4 show the lesion size/location error metrics in more detail for 50% and 25% dose, respectively. Denoising using the stacked data regime reduced the absolute volumetric error for the infarct, penumbral and hypo-perfused regions at 50% dose by 8.96 mL (20%), 8.55 mL (29%) and 19.42 mL (18%), respectively (Table 7.3). The corresponding reduction in absolute volumetric error at 25% dose was 14.23 mL (29%), 15.41 mL (26%) and 30.19 mL (28%). Although the frame-based regime did not significantly reduce the size estimation error of the infarct/penumbra volumes, it did reduce the hypo-perfused volume error by 12.28 mL (12%) and 25.74 mL (22%) at 50% and 25% dose, respectively. Compared to the low-dose images with no denoising, the stacked data regime improved the dice coefficient at 50% / 25% dose by 15% / 21% (infarct), 20% / 22% (penumbra) and 17% / 19% (hypo-perfused). Moreover, it resulted in statistically significant improvement for the infarct, penumbral and hypo-perfused regions for all metrics at 50% and 25% dose.

Overall, Table 7.3 and Table 7.4 show that lesion characterisation is consistently better, across all our chosen metrics, for the stacked regime compared to the frame-based regime.

Table 7.3 Lesion size estimation error and dice coefficient (mean  $\pm$  SD calculated for 30 cases) at 50% dose. Bold font indicates the best performance. Statistically significant differences ( $p < 0.05$ ) with respect to the low-dose data are indicated by \*.

Method	Infarct			Penumbra			Hypo-perfused		
	$E_{abs}^a$	$E_{rel}^b$	Dice coeff	$E_{abs}$	$E_{rel}$	Dice coeff	$E_{abs}$	$E_{rel}$	Dice coeff
Low-dose	19.40	0.43	0.51	20.29	0.60	0.55	32.32	0.33	0.70
	$\pm$ 10.1	$\pm$ 0.22	$\pm$ 0.15	$\pm$ 11.48	$\pm$ 0.38	$\pm$ 0.11	$\pm$ 20.04	$\pm$ 0.33	$\pm$ 0.10
GAN ( $G_{xyz}$ )	18.90	0.43	0.57*	19.87	0.59	0.59*	20.04*	0.21*	0.78*
	$\pm$ 11.65	$\pm$ 0.25	$\pm$ 0.19	$\pm$ 14.24	$\pm$ 0.47	$\pm$ 0.13	$\pm$ 12.37	$\pm$ 0.28	$\pm$ 0.12
GAN ( $G_{xyt}$ )	<b>10.44*</b>	<b>0.23*</b>	<b>0.66*</b>	<b>11.74*</b>	<b>0.31*</b>	<b>0.75*</b>	<b>12.90*</b>	<b>0.15*</b>	<b>0.87*</b>
	$\pm$ <b>8.98</b>	$\pm$ <b>0.19</b>	$\pm$ <b>0.16</b>	$\pm$ <b>10.13</b>	$\pm$ <b>0.20</b>	$\pm$ <b>0.09</b>	$\pm$ <b>10.40</b>	$\pm$ <b>0.16</b>	$\pm$ <b>0.08</b>

a) See equation (7.5a)

b) See equation (7.5b)

Table 7.4 Lesion size estimation error and dice coefficient (mean  $\pm$  SD calculated for 30 cases) at 25% dose. Bold font indicates the best performance. Statistically significant differences ( $p < 0.05$ ) with respect to the low-dose data are indicated by \*.

Method	Infarct			Penumbra			Hypo-perfused		
	$E_{abs}^a$	$E_{rel}^b$	Dice coeff	$E_{abs}$	$E_{rel}$	Dice coeff	$E_{abs}$	$E_{rel}$	Dice coeff
Low-dose	30.99	0.70	0.40	32.84	0.72	0.42	50.18	0.48	0.63
	$\pm$ 17.8	$\pm$ 0.62	$\pm$ 0.14	$\pm$ 17.2	$\pm$ 0.87	$\pm$ 0.12	$\pm$ 28.11	$\pm$ 0.40	$\pm$ 0.13
GAN ( $G_{xyz}$ )	23.98	0.49	0.50*	26.64	0.57	0.47	24.44*	0.26*	0.71*
	$\pm$ 19.67	$\pm$ 0.40	$\pm$ 0.17	$\pm$ 20.84	$\pm$ 0.62	$\pm$ 0.11	$\pm$ 22.05	$\pm$ 0.31	$\pm$ 0.13
GAN ( $G_{xyt}$ )	<b>16.76*</b>	<b>0.41*</b>	<b>0.61*</b>	<b>17.43*</b>	<b>0.46*</b>	<b>0.64*</b>	<b>19.99*</b>	<b>0.20*</b>	<b>0.82*</b>
	$\pm$ <b>11.13</b>	$\pm$ <b>0.41</b>	$\pm$ <b>0.16</b>	$\pm$ <b>16.33</b>	$\pm$ <b>0.54</b>	$\pm$ <b>0.13</b>	$\pm$ <b>13.79</b>	$\pm$ <b>0.21</b>	$\pm$ <b>0.11</b>

a) See equation (7.5a)

b) See equation (7.5b)

## 7.5 DISCUSSION

In this study we have demonstrated the feasibility of using a 3D GAN model to perform efficient synthesis of normal-dose whole-brain CTP data sets from low-dose data sets and the utility of the synthesised data for haemodynamic modelling. Feasibility was demonstrated using a variety of assessments, including image quality metrics, an expert quality rating, and volumetric lesion characterisation accuracy. The study also demonstrates the use of different data packaging regimes to leverage both the spatial and temporal (i.e. 4D) information in CTP data within a 3D GAN architecture.

Our results show that the BM4D, RED-CNN and GAN approaches all provide quality improvement on low-dose CTP data. It indicates that lower-dimensional denoising methods can provide an advantage in higher-dimensional data. Until fully 4D approaches become feasible for CTP, iterative application of lower dimensional methods remains an important option to consider for comparison with new approaches.

Detecting temporal changes in contrast is vital in CTP analysis because the time-attenuation curve (TAC) of each voxel is used to model the haemodynamics. It also implicitly differentiates structures (e.g. arteries, veins, bone, parenchyma). Our results showed that GAN-based denoising outperformed more traditional denoising approaches across all qualitative and quantitative metrics considered for haemodynamic maps. Moreover, the results obtained across all metrics show that GAN-based denoising using a stacked data regime outperforms the frame-based regime. Despite some irregularity in traversing the spatial and temporal domains due to the chosen patch size, prioritising temporal information in the 3rd dimension of the stacked configuration resulted in the best restoration of tissue haemodynamic information. More complex (and potentially less efficient) implementations using irregular patch sizes are possible and would potentially change the performance.

At both 50% and 25% dose, GAN-based denoising using the stacked regime resulted in an average improvement of 5.15-5.32 dB PSNR and 0.025-0.033 SSIM for CTP images and 2.66-3.95 dB PSNR and 0.036-0.067 SSIM for functional maps. The improvement in the image quality metrics in turn led to an average lesion volumetric error reduction of 18-29% and a dice coefficient improvement of 15-22%. Therefore, GAN-based denoising of CTP data in conjunction with an appropriate data packaging appears to be a promising denoising approach for low-dose CTP images.

GAN-based denoising exhibited better average performance in reducing volumetric error and improving dice coefficient for hypo-perfused region compared to the infarct and penumbra regions within the hypo-perfused volume. The temporal change in voxel values within the hypo-perfused volume can be as low as 10-20 HU which can be masked by noise at low dose, making the restoration and differentiation of infarct and penumbra challenging. The change in voxel values is much higher for healthy tissue than for the hypo-perfused region due to the passage of contrast agent. Therefore, healthy tissue analysis is more resistant to noise and the synthesized voxel values and resulting TACs of the healthy and hypo-perfused regions can be distinguished with higher accuracy compared to the synthesized infarct and penumbra voxels.

Treatment of stroke with intravenous and endovascular reperfusion is time-critical, with disability related to time after stroke onset (Emberson, Lees et al. 2014, Bourcier, Goyal et al. 2019), so it is important that any additional processing does not significantly increase time to treatment. The GAN-based denoising demonstrated in this study was fully automated and able to process the whole brain volume in less than 2 minutes. This is consistent with the requirements for a practical method for acute stroke.

Although our results suggest that GAN-based image synthesis may provide a practical method to reduce dose in CTP, there are several important limitations of the study which should temper the conclusions. Firstly, model performance is dependent on and limited by the size and diversity of the training set. Our training set was relatively small (30 patients) (though amplified using a patch-based approach) and performance should therefore be tested for a much larger cohort of patients and lesion size distribution. Moreover, the generalisability of the GAN-based approach should be assessed beyond the single scanner / single protocol tested in this study. Secondly, low-dose CTP images were generated by simulating reduced tube current and, although this involved an established simulation approach (Britten, Crotty et al. 2004), comparing GAN performance for the true and simulated low-dose images is needed. This could be achieved, for example, by down-sampling the normal-dose CTP projection data prior to reconstruction. Finally, our expert quality ranking was based on a qualitative assessment of estimated haemodynamic maps. This should be

extended in future to include the impact on treatment decisions and clinical outcome of using synthesized images.

## **7.6 CONCLUSION**

This study demonstrates the feasibility of denoising low-dose CTP data sets using a 3D GAN for application in ischaemic stroke analysis. The method incorporates pre-processing aimed at leveraging the full spatio-temporal (4D) information of CTP data within the 3D GAN architecture. Results show that the method outperformed other CT denoising approaches in terms of traditional image quality-based metrics and in preserving critical image content from which clinical outcome measures are derived via haemodynamic modelling. The method has important potential implications for dose reduction in CTP imaging, especially for younger patients. Further work is required to determine if the approach can match the clinical outcomes of analyses based on normal-dose CTP data, and to assess the generalisability of the method across a more expansive testing set of individuals and scanners.

# Chapter 8: SUMMARY, CONCLUSIONS & FURTHER WORK

---

## 8.1 SUMMARY AND CONCLUSIONS

Cerebral CTP imaging provides an important decision-support for physicians in diagnosis and treatment planning for acute ischaemic stroke. However, current CTP imaging protocols and quantitative stroke analysis based on these data have key shortcomings that limit their safety and accuracy. The aim of this thesis was to identify those limitations and develop potential solutions to address them.

Through the review of quantitative CTP imaging for acute ischaemic stroke (Chapter 2), motion artifacts and radiation dose were identified as ongoing challenges related to quantitative CTP imaging. Therefore, the specific aims of this thesis were:

- i. To investigate the prevalence, severity and dependencies of head motion during CTP imaging;
- ii. To investigate the suitability of compact, markerless, consumer-grade motion tracking devices for use in motion-corrected CTP imaging;
- iii. To develop and validate an accurate and reliable deep learning-based image processing approach to reduce CTP imaging scan time, thereby simultaneously reducing both the radiation dose and the likelihood of motion;
- iv. To investigate the feasibility of deep learning-based noise reduction in CTP imaging to reduce radiation dose without compromising the accuracy of haemodynamic modelling.

To address Aim 1 we characterised the prevalence, severity, patterns and dependencies of head movement during CTP imaging in 80 acute stroke cases (Chapter 4). By classifying motion into four clinically relevant categories of severity and tracking this over time, several key findings emerged:

- i. Motion was common in CTP imaging despite the short scanning time with ~40% of patients exhibiting mild to extreme motion;

- ii. 20% of CTP studies had residual motion artifacts with likely clinical implications even after performing frame-to-frame motion correction;
- iii. Point (ii) provides evidence for the importance of intra-frame motion correction, something which has not been reported on for CTP imaging;
- iv. There was a net increase in motion as the scan progressed, suggesting that shortening the scan duration could reduce the overall amount and impact of head movement (Chapter 6).

To gain further insight into the dependencies of the motion categories, PCA and statistical testing was used to model the relationship between patient baseline features and head motion severity (Chapter 4). The results showed that the combination of NIHSS and patient age are highly predictive of head movement ( $p < 0.001$ ). Specifically, older patients and patients with a higher NIHSS are more at risk of having motion-affected scans. Thus, measures should be taken to limit or correct for motion artifacts for this group in particular, and images from these studies should be interpreted with greater care.

In summary, the Chapter 4 study provides important justification for the development and implementation of motion correction methods in CTP imaging, including the previously unexplored area of intra-frame motion, and also suggests that modelling can be used to inform pre-emptive strategies for motion mitigation. A limitation of our study was our dataset, which was limited to CTP studies of positive stroke patients. The suggested analysis and predictive model should therefore be tested on a larger dataset including consecutive CTP studies of all suspected stroke cases (stroke mimics and positive stroke cases) to improve its specificity and practicality.

Building on the key finding from Chapter 4 that intra-frame motion may be as important, if not more important, than inter-frame motion, Chapter 5 addressed Aim 2 by characterising a compact, low-cost markerless device for tracking continuous head motion. The suitability of the Intel RealSense D415 depth sensor for estimating motion during CTP imaging was tested based on thermal stability and noise jitter performance, static and dynamic pose measurement accuracy, and adaptation and testing within a clinical CT scanner. The results suggested that the device is capable of sub-millimetre and sub 0.5-deg accuracy for rigid-body motion, but that values of several millimetres and 0.5-1 deg are more likely in practice for tracking the human head. Although it

would still provide some correction for CT, there would undoubtedly be considerable residual error since the intrinsic spatial in-plane resolution in CTP imaging is 0.43 mm. Nevertheless, we postulate that using multiple calibrated Intel D415 sensors simultaneously will likely provide improved accuracy and is something that should be investigated. Furthermore, a device such as the Intel D415 could function very well to augment data-driven correction by providing good initialisations of motion and filling in gaps where data-driven estimation fails.

One implication from the motion analysis in Chapter 4 was that shortening a CTP acquisition could have an important role in reducing the likelihood of motion. Indeed, reducing the duration of the scan can also have a second important benefit in reducing the radiation dose. In Chapter 6, we addressed Aim 3 by developing and validating a novel application of a SVAP approach to reduce the total CTP acquisition time, thereby simultaneously reducing the radiation dose and the likelihood of head motion during the scan. SVAP was adapted to predict late CTP image frames from a sequence of early frames, thus avoiding the truncation of the TAC. Our results indicated that image quality degraded monotonically from early predicted frames to late frames. In the most challenging case predicting the last 18 CTP frames, the bolus shape characteristics exhibited an average percentage difference of  $< 4 \pm 4\%$  compared to the ground-truth. Average volumetric error of the hypo-perfused region was overestimated by 28.36 mL (22%) and a systematic overestimation of the infarct/penumbra was observed in the predicted images. Since the infarct/penumbra volumes are calculated by thresholding the perfusion parameters, optimising the threshold on the predicted haemodynamic maps could potentially reduce the overestimation. The results showed that predicting the last 18 frames can preserve the majority of clinical content of the images while simultaneously reducing the scan duration and radiation dose by 65% and 54.5%, respectively.

Although the SAVP frame prediction approach (Chapter 6) shows promise, a more conventional approach to dose reduction is the denoising of low-dose data. In Chapter 7 we addressed Aim 4 by proposing a 3D GAN model for predicting normal-dose CTP data from low-dose CTP data. The results showed that prioritising temporal information in adapting 4D CTP data to the 3D GAN model resulted in better restoration of tissue haemodynamic information. The average lesion volumetric error reduced significantly by 18 - 29% and dice coefficient improved significantly by 15 -

22% at 50% and 25% of normal radiation dose using GAN model. The GAN-based denoising demonstrated processing the whole brain volume in less than 90 seconds, showing promise for reducing radiation dose in CTP studies in the emergency setting of acute stroke.

In summary, the main contributions of this thesis are:

- Characterisation of head motion in CTP imaging to motivate and inform the development of motion correction strategies and pre-emptively mitigate motion;
- Characterisation of the Intel RealSense D415 depth sensor as candidate device to support comprehensive – that is, inter-frame and intra-frame – motion-corrected CT/CTP imaging;
- Demonstration of the feasibility to predict clinically useful late-time CTP imaging frames using state-of-the-art deep neural networks, and to leverage this for motion and dose reduction;
- Demonstration of the first application of a 3D GAN model for dose reduction in whole-brain 4D CTP of acute ischaemic stroke patients through prediction of normal-dose data from low-dose data;
- Demonstration of practical and effective ways to exploit the spatio-temporal relationships in 4D CTP data within lower-dimensional models.

## **8.2 FURTHER WORK**

There is considerable scope for further work. Six areas for further investigation include:

### **(1) Motion dependency simulation study**

To investigate the type and timing of patient head movement resulting in the most significant impact on haemodynamic maps, a systematic simulation study should be performed. The suggested approach is to add translational and rotational movement to single and multiple frames of a 33-frame reference (motion-free) scan, and compare the lesion calculated from the motion-affected and reference scans. Based on the discrepancy of these comparisons, the most influential types of motion and the most susceptible frames and frame combinations can be identified.

## **(2) Intra-frame motion simulation study**

Although this thesis suggests that intra-frame motion is important, the extent to which intra-frame motion impacts CTP analysis and therefore should be corrected is still not known. By simulating head motion patterns derived from motion-corrupted CTP scans previously classified in Chapter 4 of this thesis as low-to-moderate, considerable, and extreme, the impact of intra-frame motion correction on lesion classification can be assessed. The recently proposed data-driven CT motion correction method (Sun, Kim et al. 2016) can be applied to the simulated motion-affected CTP data and the head movement threshold for which intra-frame motion correction is effective can be identified.

## **(3) Intra-frame motion correction**

The previously developed data-driven CT motion correction method (Sun, Kim et al. 2016) can be applied to real motion-affected CTP data of acute stroke patients. The approach concurrently estimates head motion and the motion-corrected image from the raw CT projection data. The data obtained will not only enable an evaluation of the performance of intra-frame motion correction for clinical CTP datasets, it will also reveal the type and extent of intra-frame motion encountered clinically, which has been unknown to date.

## **(4) Intra-frame motion correction and external tracking**

Motion correction may be more accurate when it is based on motion estimates that are independent of the raw data. An optimised setup of multiple Intel D415 sensors (Chapter 5) can potentially provide accurate, continuous and real-time motion monitoring, which can be input into the motion correction algorithms. Therefore, the next steps can be: (i) optimisation of the camera setup inside the clinical CT scanner, (ii) synchronisation of the motion tracking device with the CT acquisition, and (iii) the use of the motion data recorded by the sensors for comprehensive motion correction.

## **(5) Deep learning approach to replace motion-corrupted frames**

Our study in Chapter 6 showed that that a SAVP approach can be used to estimate late CTP image frames from the early frames. A further potential application of this method is to replace motion-corrupted frames with higher quality predicted frames in a standard CTP protocol. The diagnostic quality of the haemodynamic maps derived from the measured (motion-corrupted) and predicted CTP data can then be

assessed using an expert ranking. The quantitative assessment can be performed by comparing the lesion characteristics derived from the measured and predicted CTP data with the lesions on follow-up NCCT or MRI data at 24 hours. For the patients with complete reperfusion, the baseline (acute phase) infarct core normally has the same size and pattern with the infarct core on the follow-up images. For the patients with no reperfusion, the baseline hypo-perfused region usually completely turns to infarct core on the follow-up images. Thus, by co-registering the follow-up images to the baseline predicted and measured data, the percentage improvement in the baseline lesion characteristic while using the SAVP approach can be characterised.

#### **(6) Misclassified infarct/penumbra characterisation**

To improve quantitative CTP data analysis, it is important to identify the misclassified lesions and exclude them at the baseline. To characterise the misclassified infarct core at baseline, the final infarct can be segmented on follow-up NCCT or MRI as a reference. Baseline CTP and follow-up image data can then be co-registered to allow for the classification of ischaemic lesion agreement and misclassified ischaemic core (i.e. core identified on CTP, not in reference). Absolute and relative CTP parameters (CBV, CBF, TTP, and MTT), their combinations, and other quantitative metrics (Dashtbani Moghari 2016) can be calculated for both misclassified CTP ischaemic core and ischaemic lesion agreement regions and compared. To characterise the misclassified penumbra, the CTP data of patients not eligible for any type of treatment can be selected and penumbral tissue segmented on CTP images. For patients not receiving any treatment, it is expected that all penumbral tissue becomes infarct core on the follow-up NCCT or MRI. The baseline CTP and follow-up images can be co-registered to allow for the classification of the agreement between baseline hypo-perfused tissue and follow-up infarct core and misclassified penumbra (i.e. penumbra identified on CTP, not in reference). The absolute and relative haemodynamic parameters and their combinations can be extracted from both misclassified and reference penumbral regions for comparison.

## References

---

- (Aug 2021). "D400 Series Visual Presets." from <https://github.com/IntelRealSense/librealsense/wiki/D400-Series-Visual-Presets>.
- . "Ischaemic Stroke." Retrieved 1 Sep 2021, 2021, from <https://www.stroke.org.uk/what-is-stroke/types-of-stroke/ischaemic-stroke>.
- (2021, 25 Jun 2021). "Deaths in Australia." Retrieved 1 Sep 2021, 2021, from <https://www.aihw.gov.au/reports/life-expectancy-death/deaths-in-australia/contents/leading-causes-of-death>.
- Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving and M. Isard (2016). Tensorflow: A system for large-scale machine learning. 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16).
- Abdi, H. and L. J. Williams (2010). "Principal component analysis." Wiley interdisciplinary reviews: computational statistics 2(4): 433-459.
- Alnaghy, S., A. Kyme, V. Caillet, D. T. Nguyen, R. O'Brien, J. T. Booth and P. J. Keall (2019). "A six-degree-of-freedom robotic motion system for quality assurance of real-time image-guided radiotherapy." Physics in Medicine & Biology 64(10): 105021.
- Alom, M. Z., M. Hasan, C. Yakopcic, T. M. Taha and V. K. Asari (2018). "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation." arXiv preprint arXiv:1802.06955.
- Angel, E. (2010). "SVD+ Dynamic Volume CT: Delay insensitive Brain Perfusion Analysis." White Paper Toshiba America Medical Systems I (ed).
- Aoki, H., A. Suzuki and T. Shiga (2019). Study on non-contact heart beat measurement method by using depth sensor. World Congress on Medical Physics and Biomedical Engineering 2018, Springer.
- Arachchi, S. K., N. L. Hakim, H.-H. Hsu, S. V. Klimenko and T. K. Shih (2018). Real-time static and dynamic gesture recognition using mixed space features for 3D virtual world's interactions. 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), IEEE.
- Arslan, A. T. and E. Seke (2019). "Face depth estimation with conditional generative adversarial networks." IEEE Access 7: 23222-23231.
- Ashburner, J., G. Barnes, C. Chen, J. Daunizeau, G. Flandin, K. Friston, S. Kiebel, J. Kilner, V. Litvak and R. Moran (2014). "SPM12 manual." Wellcome Trust Centre for Neuroimaging, London, UK: 2464.
- Australian Stroke Foundation. (2017). "Types of stroke." Retrieved Apr, 2018, from <https://strokefoundation.org.au/About-Stroke/Types-of-stroke>.
- Bai, W., H. Suzuki, C. Qin, G. Tarroni, O. Oktay, P. M. Matthews and D. Rueckert (2018). Recurrent neural networks for aortic image sequence segmentation with sparse annotations. International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer.

- Balaji, A., P. Babu and K. Vanitha (2015). "Analysis of CT perfusion parameters in the setting of cerebral ischemic stroke." Panacea Journal of Medical Sciences **5**(3): 130-136.
- Bang, O. Y., J. L. Saver, B. H. Buck, J. R. Alger, S. Starkman, B. Ovbiagele, D. Kim, R. Jahan, G. R. Duckwiler and S. R. Yoon (2008). "Impact of collateral flow on tissue fate in acute ischaemic stroke." Journal of Neurology, Neurosurgery & Psychiatry **79**(6): 625-629.
- Bauer, S., A. Seitel, H. Hofmann, T. Blum, J. Wasza, M. Balda, H.-P. Meinzer, N. Navab, J. Hornegger and L. Maier-Hein (2013). Real-time range imaging in health care: a survey. Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications, Springer: 228-254.
- Ben-Cohen, A., E. Klang, S. P. Raskin, S. Soffer, S. Ben-Haim, E. Konen, M. M. Amitai and H. Greenspan (2019). "Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection." Engineering Applications of Artificial Intelligence **78**: 186-194.
- Benjamin, E. J., S. S. Virani, C. W. Callaway, A. M. Chamberlain, A. R. Chang, S. Cheng, S. E. Chiuve, M. Cushman, F. N. Delling and R. Deo (2018). "Heart disease and stroke statistics—2018 update: a report from the American Heart Association." Circulation **137**(12): e67-e492.
- Bennink, E., J. Oosterbroek, A. D. Horsch, J. W. Dankbaar, B. K. Velthuis, M. A. Viergever and H. W. de Jong (2015). "Influence of thin slice reconstruction on CT brain perfusion analysis." Plos one **10**(9): e0137766.
- Bhaskar, S., P. Stanwell, D. Cordato, J. Attia and C. Levi (2018). "Reperfusion therapy in acute ischemic stroke: dawn of a new era?" BMC neurology **18**(1): 8.
- Bier, B., N. Ravikumar, M. Unberath, M. Levenston, G. Gold, R. Fahrig and A. Maier (2018). "Range imaging for motion compensation in c-arm cone-beam ct of knees under weight-bearing conditions." Journal of Imaging **4**(1): 13.
- Biswas, B., S. K. Ghosh and A. Ghosh (2020). DVAE: deep variational auto-encoders for denoising retinal fundus image. Hybrid Machine Intelligence for Medical Image Analysis, Springer: 257-273.
- Bivard, A., C. Levi, N. Spratt and M. Parsons (2013). "Perfusion CT in acute stroke: a comprehensive analysis of infarct and penumbra." Radiology **267**(2): 543-550.
- Bivard, A., N. Spratt, C. Levi and M. Parsons (2011). "Perfusion computer tomography: imaging and clinical validation in acute ischaemic stroke." Brain **134**(11): 3408-3416.
- Boas, F. E. and D. Fleischmann (2012). "CT artifacts: causes and reduction techniques." Imaging in medicine **4**(2): 229-240.
- Bock, R. D. (2018). Low-cost 3D security camera. Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything, International Society for Optics and Photonics.
- Bourcier, R., M. Goyal, D. S. Liebeskind, K. W. Muir, H. Desal, A. H. Siddiqui, D. W. Dippel, C. B. Majoie, W. H. Van Zwam and T. G. Jovin (2019). "Association of time from stroke onset to groin puncture with quality of reperfusion after mechanical thrombectomy: a meta-analysis of individual patient data from 7 randomized clinical trials." JAMA neurology **76**(4): 405-411.

- Britten, A., M. Crotty, H. Kiremidjian, A. Grundy and E. Adam (2004). "The addition of computer simulated noise to investigate radiation dose and image quality in images with spatial correlation of statistical noise: an example application to X-ray CT of the brain." The British journal of radiology **77**(916): 323-328.
- Brott, T., H. P. Adams Jr, C. P. Olinger, J. R. Marler, W. G. Barsan, J. Biller, J. Spilker, R. Holleran, R. Eberle and V. Hertzberg (1989). "Measurements of acute cerebral infarction: a clinical examination scale." Stroke **20**(7): 864-870.
- Canelhas, D. R., T. Stoyanov and A. J. Lilienthal (2013). SDF tracker: A parallel algorithm for on-line pose estimation and scene reconstruction from depth images. 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE.
- Chakravarty, A. and J. Sivaswamy (2018). "RACE-net: a recurrent neural network for biomedical image segmentation." IEEE journal of biomedical and health informatics **23**(3): 1151-1162.
- Chen, C., A. Bivard, L. Lin, C. R. Levi, N. J. Spratt and M. W. Parsons (2019). "Thresholds for infarction vary between gray matter and white matter in acute ischemic stroke: a CT perfusion study." Journal of Cerebral Blood Flow & Metabolism **39**(3): 536-546.
- Chen, H., Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou and G. Wang (2017). "Low-dose CT with a residual encoder-decoder convolutional neural network." IEEE transactions on medical imaging **36**(12): 2524-2535.
- Chen, L., S.-Y. Lin, Y. Xie, Y.-Y. Lin, W. Fan and X. Xie (2020). DGGAN: Depth-image Guided Generative Adversarial Networks for Disentangling RGB and Depth Images in 3D Hand Pose Estimation. The IEEE Winter Conference on Applications of Computer Vision.
- Chiang, T. and C.-P. Fan (2018). 3D depth information based 2D low-complexity hand posture and gesture recognition design for human computer interactions. 2018 3rd International Conference on Computer and Communication Systems (ICCCS), IEEE.
- Chin, C.-L., B.-J. Lin, G.-R. Wu, T.-C. Weng, C.-S. Yang, R.-C. Su and Y.-J. Pan (2017). An automated early ischemic stroke detection system using CNN deep learning algorithm. 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), IEEE.
- Choi, K., M. Vania and S. Kim (2019). Semi-supervised learning for low-dose ct image restoration with hierarchical deep generative adversarial network (hd-gan). 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE.
- Chung, J., C. Gulcehre, K. Cho and Y. Bengio (2014). "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555.
- Clèrigues, A., S. Valverde, J. Bernal, J. Freixenet, A. Oliver and X. Lladó (2020). "Acute and sub-acute stroke lesion segmentation from multimodal MRI." Computer methods and programs in biomedicine **194**: 105521.
- Copen, W., A. Deipolyi, P. Schaefer, L. Schwamm, R. González and O. Wu (2015). "Exposing hidden truncation-related errors in acute stroke perfusion imaging." American Journal of Neuroradiology **36**(4): 638-645.

- Cuenod, C. and D. Balvay (2013). "Perfusion and vascular permeability: basic concepts and measurement in DCE-CT and DCE-MRI." Diagnostic and interventional imaging **94**(12): 1187-1204.
- Daniilidis, K. (1999). "Hand-eye calibration using dual quaternions." The International Journal of Robotics Research **18**(3): 286-298.
- Dashtbani Moghari, M. (2016). "Optimised methylene-blue detection and quantification using conventional Raman spectroscopy."
- Demeestere, J., A. Wouters, S. Christensen, R. Lemmens and M. G. Lansberg (2020). "Review of Perfusion Imaging in Acute Ischemic Stroke: From Time to Tissue." Stroke **51**(3): 1017-1024.
- Demuth, H.-U., R. M. Dijkhuizen, T. D. Farr, M. Gelderblom, K. Horsburgh, C. Iadecola, D. D. Mcleod, D. Michalski, T. H. Murphy and J. Orbe (2017). "Recent progress in translational research on neurovascular and neurodegenerative disorders." Restorative neurology and neuroscience **35**(1): 87-103.
- Dolz, J., K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers and I. B. Ayed (2018). "HyperDense-Net: a hyper-densely connected CNN for multi-modal image segmentation." IEEE transactions on medical imaging **38**(5): 1116-1126.
- Duchi, J., E. Hazan and Y. Singer (2011). "Adaptive subgradient methods for online learning and stochastic optimization." Journal of machine learning research **12**(7).
- El Jurdi, R., C. Petitjean, P. Honeine and F. Abdallah (2020). "Bb-unet: U-net with bounding box prior." IEEE Journal of Selected Topics in Signal Processing **14**(6): 1189-1198.
- Emberson, J., K. R. Lees, P. Lyden, L. Blackwell, G. Albers, E. Bluhmki, T. Brott, G. Cohen, S. Davis and G. Donnan (2014). "Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials." The Lancet **384**(9958): 1929-1935.
- Engedal, T. S., N. Hjort, K. D. Hougaard, C. Z. Simonsen, G. Andersen, I. K. Mikkelsen, J. K. Boldsen, S. F. Eskildsen, M. B. Hansen and H. Angleys (2017). "Transit time homogenization in ischemic stroke—A novel biomarker of penumbral microvascular failure?" Journal of Cerebral Blood Flow & Metabolism: 0271678X17721666.
- Evan, M. Y., J. E. Iglesias, A. V. Dalca and M. R. Sabuncu (2020). An Auto-Encoder Strategy for Adaptive Image Segmentation. Medical Imaging with Deep Learning, PMLR.
- Fahmi, F., H. A. Marquering, J. Borst, G. J. Streekstra, L. F. Beenen, J. M. Niesten, B. K. Velthuis, C. B. Majoie and E. VanBavel (2014). "3D movement correction of CT brain perfusion image data of patients with acute ischemic stroke." Neuroradiology **56**(6): 445-452.
- Fahmi, F., A. Riordan, L. F. Beenen, G. J. Streekstra, N. Janssen, H. de Jong, C. Majoie, E. van Bavel and H. A. Marquering (2014). "The effect of head movement on CT perfusion summary maps: simulations with CT hybrid phantom data." Medical & biological engineering & computing **52**(2): 141-147.

Fedus, W., M. Rosca, B. Lakshminarayanan, A. M. Dai, S. Mohamed and I. Goodfellow (2017). "Many paths to equilibrium: GANs do not need to decrease a divergence at every step." arXiv preprint arXiv:1710.08446.

Fernández-López, D., J. Faustino, R. Daneman, L. Zhou, S. Y. Lee, N. Derugin, M. F. Wendland and Z. S. Vexler (2012). "Blood–brain barrier permeability is increased after acute adult stroke but not neonatal stroke in the rat." Journal of Neuroscience **32**(28): 9588-9600.

FILgroup. (2014). from <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>.

Franceschi, J.-Y., E. Delasalles, M. Chen, S. Lamprier and P. Gallinari (2020). Stochastic latent residual video prediction. International Conference on Machine Learning, PMLR.

Gao, S. and R. Wang (2013). "Takayasu arteritis presenting with massive cerebral ischemic infarction in a 35-year-old woman: a case report." Journal of Medical Case Reports **7**(1): 1-5.

Garbin, C., X. Zhu and O. Marques (2020). "Dropout vs. batch normalization: an empirical study of their impact to deep learning." Multimedia Tools and Applications: 1-39.

Geyer, L. L., U. J. Schoepf, F. G. Meinel, J. W. Nance Jr, G. Bastarrika, J. A. Leipsic, N. S. Paul, M. Rengo, A. Laghi and C. N. De Cecco (2015). "State of the art: iterative CT reconstruction techniques." Radiology **276**(2): 339-357.

Ghani, M. U. and W. C. Karl (2019). "Fast enhanced CT metal artifact reduction using data domain deep learning." IEEE Transactions on Computational Imaging **6**: 181-193.

Gomez, C. R. (1993). Time is brain!, WB Saunders.

González, R. G. (2013). "Current state of acute stroke imaging." Stroke **44**(11): 3260-3264.

Goodfellow, I., Y. Bengio and A. Courville (2016). Deep learning, MIT press.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio (2014). "Generative adversarial nets." Advances in neural information processing systems **27**.

Gordon-Rodriguez, E., G. Loaiza-Ganem, G. Pleiss and J. P. Cunningham (2020). "Uses and abuses of the cross-entropy loss: case studies in modern deep learning."

Grosse, R. (2017). "Lecture 15: Exploding and vanishing gradients." University of Toronto Computer Science.

Habibalahi, A., M. D. Moghari, J. M. Campbell, A. G. Anwer, S. B. Mahbub, M. Gosnell, S. Saad, C. Pollock and E. M. Goldys (2020). "Non-invasive real-time imaging of reactive oxygen species (ROS) using auto-fluorescence multispectral imaging technique: A novel tool for redox biology." Redox biology **34**: 101561.

Hand, P. J., J. Kwan, R. I. Lindley, M. S. Dennis and J. M. Wardlaw (2006). "Distinguishing between stroke and mimic at the bedside: the brain attack study." Stroke **37**(3): 769-775.

Hanzelka, T., J. Dusek, F. Ocasek, J. Kucera, J. Sedy, J. Benes, G. Pavlikova and R. Foltan (2013). "Movement of the patient and the cone beam computed tomography

scanner: objectives and possible solutions." Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology **116**(6): 769-773.

He, K., X. Zhang, S. Ren and J. Sun (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition.

Heinrich, M. P., M. Stille and T. M. Buzug (2018). "Residual U-net convolutional neural network architecture for low-dose CT denoising." Current Directions in Biomedical Engineering **4**(1): 297-300.

Heiss, W.-D., J. Sobesky and V. Hesselmann (2004). "Identifying thresholds for penumbra and irreversible tissue damage." Stroke **35**(11 suppl 1): 2671-2674.

Heit, J. J. and M. Wintermark (2016). "Perfusion computed tomography for the evaluation of acute ischemic stroke: strengths and pitfalls." Stroke **47**(4): 1153-1158.

Heß, M., F. Büther, F. Gigengack, M. Dawood and K. P. Schäfers (2015). "A dual-Kinect approach to determine torso surface motion for respiratory motion correction in PET." Medical physics **42**(5): 2276-2286.

Hiasa, Y., Y. Otake, M. Takao, T. Matsuoka, K. Takashima, A. Carass, J. L. Prince, N. Sugano and Y. Sato (2018). Cross-modality image synthesis from unpaired data using CycleGAN. International workshop on simulation and synthesis in medical imaging, Springer.

Hinman, J. D., N. S. Rost, T. W. Leung, J. Montaner, K. W. Muir, S. Brown, J. F. Arenillas, E. Feldmann and D. S. Liebeskind (2017). "Principles of precision medicine in stroke." Journal of Neurology, Neurosurgery & Psychiatry **88**(1): 54-61.

Hinton, G., N. Srivastava and K. Swersky (2012). "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent." Cited on **14**(8): 2.

Ho, K. C., F. Scalzo, K. V. Sarma, S. El-Saden and C. W. Arnold (2016). A temporal deep learning approach for MR perfusion parameter estimation in stroke. Pattern Recognition (ICPR), 2016 23rd International Conference on, IEEE.

Ho, K. C., W. Speier, S. El-Saden and C. W. Arnold (2017). Classifying acute ischemic stroke onset time using deep imaging features. AMIA Annual Symposium Proceedings, American Medical Informatics Association.

Hoffmann, A., G. Zhu and M. Wintermark (2012). "Advanced neuroimaging in stroke patients: prediction of tissue fate and hemorrhagic transformation." Expert review of cardiovascular therapy **10**(4): 515-524.

Hom, J., J. Dankbaar, T. Schneider, S.-C. Cheng, J. Bredno and M. Wintermark (2009). "Optimal duration of acquisition for dynamic perfusion CT assessment of blood-brain barrier permeability using the Patlak model." American Journal of Neuroradiology **30**(7): 1366-1370.

Horsch, A. D., J. W. Dankbaar, T. van Seeters, J. M. Niesten, M. J. Luitse, P. C. Vos, I. C. van der Schaaf, G.-J. Biessels, Y. van der Graaf and L. J. Kappelle (2016). "Relation between stroke severity, patient characteristics and CT-perfusion derived blood-brain barrier permeability measurements in acute ischemic stroke." Clinical neuroradiology **26**(4): 415-421.

Huang, Z., Z. Chen, Q. Zhang, G. Quan, M. Ji, C. Zhang, Y. Yang, X. Liu, D. Liang and H. Zheng (2020). "CaGAN: a cycle-consistent generative adversarial network with

attention for low-dose CT imaging." IEEE Transactions on Computational Imaging **6**: 1203-1218.

Isola, P., J.-Y. Zhu, T. Zhou and A. A. Efros (2017). Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition.

Jenkinson, M., C. F. Beckmann, T. E. Behrens, M. W. Woolrich and S. M. Smith (2012). "Fsl." Neuroimage **62**(2): 782-790.

Jia, B., F. Scalzo, E. Agbayani, G. W. Woolf, L. Liu, Z. Miao and D. S. Liebeskind (2016). "Multimodal CT techniques for cerebrovascular and hemodynamic evaluation of ischemic stroke: occlusion, collaterals, and perfusion." Expert review of neurotherapeutics **16**(5): 515-525.

Jiang, G., M. Luo, K. Bai and S. Chen (2017). "A precise positioning method for a puncture robot based on a PSO-optimized BP neural network algorithm." Applied Sciences **7**(10): 969.

Jung, S., M. Gilgen, J. Slotboom, M. El-Koussy, C. Zubler, C. Kiefer, R. Luedi, M.-L. Mono, M. R. Heldner and A. Weck (2013). "Factors that determine penumbral tissue loss in acute ischaemic stroke." Brain **136**(12): 3554-3560.

Kadimesetty, V. S., S. Gutta, S. Ganapathy and P. K. Yalavarthy (2018). "Convolutional Neural Network-Based Robust Denoising of Low-Dose Computed Tomography Perfusion Maps." IEEE Transactions on Radiation and Plasma Medical Sciences **3**(2): 137-152.

Karimi, D., P. Deman, R. Ward and N. Ford (2016). "A sinogram denoising algorithm for low-dose computed tomography." BMC medical imaging **16**(1): 1-14.

Kasasbeh, A. S., S. Christensen, M. Straka, N. Mishra, M. Mlynash, R. Bammer, G. W. Albers and M. G. Lansberg (2016). "Optimal computed tomographic perfusion scan duration for assessment of acute stroke lesion volumes." Stroke **47**(12): 2966-2971.

Kerfoot, E., J. Clough, I. Oksuz, J. Lee, A. P. King and J. A. Schnabel (2018). Left-ventricle quantification using residual U-Net. International Workshop on Statistical Atlases and Computational Models of the Heart, Springer.

Kim, J., J. Nuyts, A. Kyme, Z. Kuncic and R. Fulton (2015). "A rigid motion correction method for helical computed tomography (CT)." Physics in Medicine & Biology **60**(5): 2047.

Kim, J., T. Sun, A. Alcheikh, Z. Kuncic, J. Nuyts and R. Fulton (2016). "Correction for human head motion in helical x-ray CT." Physics in Medicine & Biology **61**(4): 1416.

Kim, Y., Y. K. Kim, B. E. Lee, S. J. Lee, Y. J. Ryu, J. H. Lee and J. H. Chang (2015). "Ultra-low-dose CT of the thorax using iterative reconstruction: evaluation of image quality and radiation dose reduction." American journal of roentgenology **204**(6): 1197-1202.

Kingma, D. P. and J. Ba (2014). "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980.

Kingma, D. P. and M. Welling (2013). "Auto-encoding variational bayes." arXiv preprint arXiv:1312.6114.

- Kirschner, D., A. Schlotzhauer, M. Brandstötter and M. Hofbaur (2017). Validation of relevant parameters of sensitive manipulators for human-robot collaboration. International Conference on Robotics in Alpe-Adria Danube Region, Springer.
- Klein, S., M. Staring, K. Murphy, M. A. Viergever and J. P. Pluim (2010). "Elastix: a toolbox for intensity-based medical image registration." IEEE transactions on medical imaging **29**(1): 196-205.
- Koike, Y., Y. Anetai, H. Takegawa, S. Ohira, S. Nakamura and N. Tanigawa (2020). "Deep learning-based metal artifact reduction using cycle-consistent adversarial network for intensity-modulated head and neck radiation therapy treatment planning." Physica Medica **78**: 8-14.
- Konstas, A., G. Goldmakher, T.-Y. Lee and M. Lev (2009). "Theoretic basis and technical implementations of CT perfusion in acute ischemic stroke, part 1: theoretic basis." American Journal of Neuroradiology **30**(4): 662-668.
- Krizhevsky, A., I. Sutskever and G. E. Hinton (2012). "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems **25**: 1097-1105.
- Kudo, K., S. Christensen, M. Sasaki, L. Østergaard, H. Shirato, K. Ogasawara, M. Wintermark and S. Warach (2013). "Accuracy and reliability assessment of CT and MR perfusion analysis software using a digital phantom." Radiology **267**(1): 201-211.
- Kumar, M., M. Babaeizadeh, D. Erhan, C. Finn, S. Levine, L. Dinh and D. Kingma (2019). "Videoflow: A conditional flow-based model for stochastic video generation." arXiv preprint arXiv:1903.01434.
- Kurz, K., G. Ringstad, A. Odland, R. Advani, E. Farbu and M. Kurz (2016). "Radiological imaging in acute ischaemic stroke." European journal of neurology **23**(S1): 8-17.
- Kyme, A. Z., M. Aksoy, D. L. Henry, R. Bammer and J. Maclaren (2020). "Marker-free optical stereo motion tracking for in-bore MRI and PET-MRI application." Medical physics **47**(8): 3321-3331.
- Kyme, A. Z. and R. R. Fulton (2021). "Motion estimation and correction in SPECT, PET and CT." Physics in Medicine & Biology.
- Kyme, A. Z., S. Se, S. R. Meikle and R. R. Fulton (2018). "Markerless motion estimation for motion-compensated clinical brain imaging." Physics in Medicine & Biology **63**(10): 105018.
- Ledezma, C. J. and M. Wintermark (2009). "Multimodal CT in stroke imaging: new concepts." Radiologic clinics of North America **47**(1): 109-116.
- Ledig, C., L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz and Z. Wang (2017). Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the IEEE conference on computer vision and pattern recognition.
- Lee, A. X., R. Zhang, F. Ebert, P. Abbeel, C. Finn and S. Levine (2018). "Stochastic adversarial video prediction." arXiv preprint arXiv:1804.01523.
- Lee, E.-J., Y.-H. Kim, N. Kim and D.-W. Kang (2017). "Deep into the brain: artificial intelligence in stroke imaging." Journal of stroke **19**(3): 277.

- Lee, N. K., S. Kim, S. B. Hong, T. U. Kim, H. Ryu, J. W. Lee and J. Y. Kim (2019). "Low-dose CT with the adaptive statistical iterative reconstruction V technique in abdominal organ injury: comparison with routine-dose CT with filtered back projection." *American Journal of Roentgenology* **213**(3): 659-666.
- Li, F., W. Huang, M. Luo, P. Zhang and Y. Zha (2021). "A new VAE-GAN model to synthesize arterial spin labeling images from structural MRI." *Displays*: 102079.
- Li, G., L. Zhang, S. Hu, D. Fu and M. Liu (2019). *Adversarial Network with Dual U-net Model and Multiresolution Loss Computation for Medical Images Registration*. 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE.
- Li, S., D. Zeng, Z. Bian, D. Li, M. Zhu, J. Huang and J. Ma (2021). "Learning non-local perfusion textures for high-quality computed tomography perfusion imaging." *Physics in Medicine & Biology* **66**(11): 115007.
- Li, X., H. Chen, X. Qi, Q. Dou, C.-W. Fu and P.-A. Heng (2018). "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes." *IEEE transactions on medical imaging* **37**(12): 2663-2674.
- Li, Z., L. Yu, J. D. Trzasko, D. S. Lake, D. J. Blezek, J. G. Fletcher, C. H. McCollough and A. Manduca (2014). "Adaptive nonlocal means filtering based on local noise level for CT denoising." *Medical physics* **41**(1).
- Liao, B., J. Li, Z. Ju and G. Ouyang (2018). *Hand gesture recognition with generalized hough transform and DC-CNN using realsense*. 2018 Eighth International Conference on Information Science and Technology (ICIST), IEEE.
- Lin, L., A. Bivard and M. W. Parsons (2013). "Perfusion patterns of ischemic stroke on computed tomography perfusion." *Journal of stroke* **15**(3): 164.
- Lindsay, C., J. M. Mukherjee, K. Johnson, P. Olivier, X. Song, L. Shao and M. A. King (2015). *Marker-less multi-frame motion tracking and compensation in PET-brain imaging*. Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging, International Society for Optics and Photonics.
- Liu, F., C. Cuevas, A. A. Moss, O. Kolokythas, T. J. Dubinsky and P. E. Kinahan (2008). "Gas bubble motion artifact in MDCT." *American Journal of Roentgenology* **190**(2): 294-299.
- Liu, P. and R. Fang (2018). *SDCNet: Smoothed dense-convolution network for restoring low-dose cerebral CT perfusion*. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE.
- Lucas, C., A. Kemmling, N. Bouteldja, L. F. Aulmann, A. Madany Mamlouk and M. P. Heinrich (2018). "Learning to predict ischemic stroke growth on acute CT perfusion data by interpolating low-dimensional shape representations." *Frontiers in neurology* **9**: 989.
- Lyden, P. (2017). "Using the national institutes of health stroke scale: a cautionary tale." *Stroke* **48**(2): 513-519.
- Lyu, Q., H. Shan, Y. Xie, A. C. Kwan, Y. Otaki, K. Kuronuma, D. Li and G. Wang (2021). "Cine cardiac mri motion artifact reduction using a recurrent neural network." *IEEE Transactions on Medical Imaging*.

- Ma, Y., P. Feng, P. He, Z. Long and B. Wei (2019). Low-dose CT with a deep convolutional neural network blocks model using mean squared error loss and structural similar loss. Eleventh International Conference on Information Optics and Photonics (CIOP 2019), International Society for Optics and Photonics.
- Maggioni, M., V. Katkovnik, K. Egiazarian and A. Foi (2012). "Nonlocal transform-domain filter for volumetric data denoising and reconstruction." IEEE transactions on image processing **22**(1): 119-133.
- Mangla, R., S. Ekholm, B. S. Jahromi, J. Almast, M. Mangla and P.-L. Westesson (2014). "CT perfusion in acute stroke: know the mimics, potential pitfalls, artifacts, and technical errors." Emergency radiology **21**(1): 49-65.
- Manniesing, R., M. T. Oei, B. van Ginneken and M. Prokop (2015). "Quantitative dose dependency analysis of whole-brain CT perfusion imaging." Radiology **278**(1): 190-197.
- McKight, P. E. and J. Najab (2010). "Kruskal-wallis test." The corsini encyclopedia of psychology: 1-1.
- Medsker, L. R. and L. Jain (2001). "Recurrent neural networks." Design and Applications **5**.
- Meier, P. and K. L. Zierler (1954). "On the theory of the indicator-dilution method for measurement of blood flow and volume." Journal of applied physiology **6**(12): 731-744.
- Mendrik, A., E.-j. Vonken, J.-W. Dankbaar, M. Prokop and B. van Ginneken (2010). Noise filtering in thin-slice 4D cerebral CT perfusion scans. Medical Imaging 2010: Image Processing, International Society for Optics and Photonics.
- Mendrik, A. M., E.-j. Vonken, B. van Ginneken, H. W. de Jong, A. Riordan, T. van Seeters, E. J. Smit, M. A. Viergever and M. Prokop (2011). "TIPS bilateral noise reduction in 4D CT perfusion scans produces high-quality cerebral blood flow maps." Physics in medicine and biology **56**(13): 3857.
- Mies, G., S. Ishimaru, Y. Xie, K. Seo and K.-A. Hossmann (1991). "Ischemic thresholds of cerebral protein synthesis and energy state following middle cerebral artery occlusion in rat." Journal of Cerebral Blood Flow & Metabolism **11**(5): 753-761.
- Mirza, M. and S. Osindero (2014). "Conditional generative adversarial nets." arXiv preprint arXiv:1411.1784.
- Mo Kim, S., M. A. Haider, M. Milosevic, D. A. Jaffray and I. W. Yeung (2011). "A method for patient dose reduction in dynamic contrast enhanced CT study." Medical physics **38**(9): 5094-5103.
- Moghari, M. D., P. Noonan, D. L. Henry, R. Fulton, N. Young, K. Moore and A. Kyme Characterization of the Intel RealSense D415 Stereo Depth Camera for Motion-Corrected CT Imaging. 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE.
- Moghari, M. D., P. Noonan, D. L. Henry, R. Fulton, N. Young, K. Moore and A. Kyme (2019). Characterization of the Intel RealSense D415 Stereo Depth Camera for Motion-Corrected CT Imaging. 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE.

- Moghari, M. D., N. Young, R. R. Fulton, A. Evans and A. Z. Kyme (2021). "Head Movement During Cerebral CT Perfusion Imaging of Acute Ischaemic Stroke: Characterisation and Correlation with Patient Baseline Features." European Journal of Radiology: 109979.
- Moghari, M. D., N. Young, K. Moore, R. R. Fulton, A. Evans and A. Z. Kyme (2021). "Head movement during cerebral CT perfusion imaging of acute ischaemic stroke: Characterisation and correlation with patient baseline features." European Journal of Radiology **144**: 109979.
- Moghari, M. D., L. Zhou, B. Yu, K. Moore, N. Young, R. Fulton and A. Kyme (2019). Estimation of full-dose 4D CT perfusion images from low-dose images using conditional generative adversarial networks. 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE.
- Moghari, M. D., L. Zhou, B. Yu, N. Young, K. Moore, A. Evans, R. R. Fulton and A. Z. Kyme (2021). "Efficient radiation dose reduction in whole-brain CT perfusion imaging using a 3D GAN: performance and clinical feasibility." Physics in Medicine & Biology **66**(7): 075008.
- Montgomery, A. J., K. Thielemans, M. A. Mehta, F. Turkheimer, S. Mustafovic and P. M. Grasby (2006). "Correction of head movement on PET studies: comparison of methods." Journal of Nuclear Medicine **47**(12): 1936-1944.
- Morgan, C. D., M. Stephens, S. L. Zuckerman, M. S. Waitara, P. J. Morone, M. C. Dewan and J. Mocco (2015). "Physiologic imaging in acute stroke: Patient selection." Interventional Neuroradiology **21**(4): 499-510.
- Mukherjee, J. M., P. H. Pretorius, K. L. Johnson, B. F. Hutton and M. A. King (2010). Comparison of data-driven and external-surrogate based motion estimation strategies in cardiac SPECT imaging. IEEE Nuclear Science Symposium & Medical Imaging Conference, IEEE.
- Nielsen, A., M. B. Hansen, A. Tietze and K. Mouridsen (2018). "Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning." Stroke **49**(6): 1394-1401.
- Noonan, P., J. Howard, W. Hallett and R. Gunn (2015). "Repurposing the Microsoft Kinect for Windows v2 for external head motion tracking for brain PET." Physics in Medicine & Biology **60**(22): 8753.
- Oddo, M., I. A. Crippa, S. Mehta, D. Menon, J.-F. Payen, F. S. Taccone and G. Citerio (2016). "Optimizing sedation in patients with acute brain injury." Critical Care **20**(1): 1-11.
- Olesen, O. V., J. M. Sullivan, T. Mulnix, R. R. Paulsen, L. Hojgaard, B. Roed, R. E. Carson, E. D. Morris and R. Larsen (2012). "List-mode PET motion correction using markerless head tracking: proof-of-concept with scans of human subject." IEEE transactions on medical imaging **32**(2): 200-209.
- Olive-Gadea, M., C. Crespo, C. Granes, M. Hernandez-Perez, N. Pérez de la Ossa, C. Laredo, X. Urra, J. Carlos Soler, A. Soler and P. Puyalto (2020). "Deep learning based software to identify large vessel occlusion on noncontrast computed tomography." Stroke **51**(10): 3133-3137.

- Østergaard, L., A. Gjedde, O. Wu, C. Gyldensted, B. Rosen and G. Sorensen (2001). Microscopic flow heterogeneity: its determination by PWI and role in cerebral oxygen metabolism. Proceedings of the 9th Annual Meeting of ISMRM, Glasgow, Scotland.
- Paciaroni, M., V. Caso and G. Agnelli (2009). "The concept of ischemic penumbra in acute stroke and therapeutic opportunities." European neurology **61**(6): 321-330.
- Papanastasopoulos, Z., R. K. Samala, H.-P. Chan, L. Hadjiiski, C. Paramagul, M. A. Helvie and C. H. Neal (2020). Explainable AI for medical imaging: deep-learning CNN ensemble for classification of estrogen receptor status from breast MRI. Medical Imaging 2020: Computer-Aided Diagnosis, International Society for Optics and Photonics.
- Pascanu, R., T. Mikolov and Y. Bengio (2013). On the difficulty of training recurrent neural networks. International conference on machine learning, PMLR.
- Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein and L. Antiga (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems.
- Pengjiang, Q., K. Xu, T. Wang, Z. Qiankun, H. Yang, B. Atallah, Z. Junqing, T. Bryan and M. R. F Jr (2020). "Estimating CT from MR abdominal images using novel generative adversarial networks." Journal of Grid Computing **18**(2): 211-226.
- Peretz, S., D. Orion, D. Last, Y. Mardor, Y. Kimmel, S. Yehezky, E. Lotan, Z. e. Itsekson-Hayosh, S. Koton and D. Guez (2017). "Incorporation of relative cerebral blood flow into CT perfusion maps reduces false 'at risk' penumbra." Journal of neurointerventional surgery: neurintsurg-2017-013268.
- Pisana, F., T. Henzler, S. Schönberg, E. Klotz, B. Schmidt and M. Kachelrieß (2017). "Noise reduction and functional maps image quality improvement in dynamic CT perfusion using a new k-means clustering guided bilateral filter (KMGB)." Medical physics **44**(7): 3464-3482.
- Popilock, R., K. Sandrasagaren, L. Harris and K. A. Kaser (2008). "CT artifact recognition for the nuclear technologist." Journal of nuclear medicine technology **36**(2): 79-81.
- Qin, C., J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal and D. Rueckert (2018). "Convolutional recurrent neural networks for dynamic MR image reconstruction." IEEE transactions on medical imaging **38**(1): 280-290.
- Rajeev, R., J. A. Samath and N. Karthikeyan (2019). "An intelligent recurrent neural network with long short-term memory (LSTM) BASED batch normalization for medical image denoising." Journal of medical systems **43**(8): 1-10.
- Rassadin, A. (2020). Deep Residual 3D U-Net for Joint Segmentation and Texture Classification of Nodules in Lung. International Conference on Image Analysis and Recognition, Springer.
- Reid, E., D. Graham, M. R. Lopez-Gonzalez, W. M. Holmes, I. M. Macrae and C. McCabe (2012). "Penumbra detection using PWI/DWI mismatch MRI in a rat stroke model with and without comorbidity: comparison of methods." Journal of Cerebral Blood Flow & Metabolism **32**(9): 1765-1777.
- Riordan, A. J., M. Prokop, M. A. Viergever, J. W. Dankbaar, E. J. Smit and H. W. de Jong (2011). "Validation of CT brain perfusion methods using a realistic dynamic head phantom." Medical physics **38**(6Part1): 3212-3221.

- Ritchie, H. and M. Roser (2018). "Causes of death." Our World in Data.
- Ronneberger, O., P. Fischer and T. Brox (2015). U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, Springer.
- Saad, N., S. A. Abu-Bakar, A. Muda, S. Muda and A. Syafeeza (2014). "Automatic Brain Lesion Detection and Classification Based on Diffusion-Weighted Imaging using Adaptive Thresholding and a Rule-Based Classifier." International Journal of Engineering and Technology (IJET) **6**(6).
- Saad, N. M., S. Bakar, A. S. Muda and M. M. Mokji (2015). "Review of brain lesion detection and classification using neuroimaging analysis techniques." Jurnal Teknologi **74**(6): 73-85.
- Saita, K., M. Chen, N. J. Spratt, M. J. Porritt, G. T. Liberatore, C. R. Levi, G. A. Donnan, U. Ackermann, H. J. Tochon-Danguy and J. I. Sachinidis (2004). "Imaging the ischemic penumbra with 18F-fluoromisonidazole in a rat model of ischemic stroke." Stroke **35**(4): 975-980.
- Saver, J. L. (2006). "Time is brain—quantified." Stroke **37**(1): 263-266.
- Scalzo, F., J. R. Alger, X. Hu, J. L. Saver, K. A. Dani, K. W. Muir, A. M. Demchuk, S. B. Coutts, M. Luby and S. Warach (2013). "Multi-center prediction of hemorrhagic transformation in acute ischemic stroke using permeability imaging features." Magnetic resonance imaging **31**(6): 961-969.
- Scalzo, F., Q. Hao, J. R. Alger, X. Hu and D. S. Liebeskind (2012). "Regional prediction of tissue fate in acute ischemic stroke." Annals of biomedical engineering **40**(10): 2177-2187.
- Shan, H., Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong and G. Wang (2018). "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network." IEEE transactions on medical imaging **37**(6): 1522-1534.
- Sherstinsky, A. (2020). "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network." Physica D: Nonlinear Phenomena **404**: 132306.
- Shi, L., Y. Hu, Y. Chen, X. Yin, H. Shu, L. Luo and J.-L. Coatrieux (2016). "Improving low-dose cardiac CT images based on 3D sparse representation." Scientific reports **6**: 22804.
- Shin, H.-C., H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura and R. M. Summers (2016). "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning." IEEE transactions on medical imaging **35**(5): 1285-1298.
- Shiu, Y. C. and S. Ahmad (1987). "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form  $AX = XB$ ."
- Sidky, E. Y. and X. Pan (2008). "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization." Physics in Medicine & Biology **53**(17): 4777.

- Siena, F. L., B. Byrom, P. Watts and P. Breedon (2018). "Utilising the intel realsense camera for measuring health outcomes in clinical research." Journal of medical systems **42**(3): 1-10.
- Silverstein, E. and M. Snyder (2018). "Comparative analysis of respiratory motion tracking using Microsoft Kinect v2 sensor." Journal of applied clinical medical physics **19**(3): 193-204.
- Sixt, L., B. Wild and T. Landgraf (2018). "Rendergan: Generating realistic labeled data." Frontiers in Robotics and AI **5**: 66.
- Stavdahl, O., A. K. Bondhus, K. Y. Pettersen and K. E. Malvig (2005). "Optimal statistical operators for 3-dimensional rotational data: geometric interpretations and application to prosthesis kinematics." Robotica **23**(3): 283-292.
- Sun, T., R. Clackdoyle, J.-H. Kim, R. Fulton and J. Nuyts (2017). "Estimation of local data-insufficiency in motion-corrected helical CT." IEEE Transactions on Radiation and Plasma Medical Sciences **1**(4): 346-357.
- Sun, T., J. H. Kim, R. Fulton and J. Nuyts (2016). "An iterative projection-based motion estimation and compensation scheme for head x-ray CT." Medical physics **43**(10): 5705-5716.
- Tsai, J. P. and G. W. Albers (2017). Current Imaging Strategies for Patient Selection in Acute Ischemic Stroke Trials. Neuroprotective Therapy for Stroke and Ischemic Disease, Springer: 751-774.
- van de Leemput, S. C., M. Prokop, B. van Ginneken and R. Manniesing (2019). "Stacked bidirectional convolutional LSTMs for deriving 3D non-contrast ct from spatiotemporal 4D CT." IEEE transactions on medical imaging **39**(4): 985-996.
- Verro, P., L. N. Tanenbaum, N. M. Borden, S. Sen and N. Eshkar (2002). "CT angiography in acute ischemic stroke: preliminary results." Stroke **33**(1): 276-278.
- Villegas, R., A. Pathak, H. Kannan, D. Erhan, Q. V. Le and H. Lee (2019). "High fidelity video prediction with large stochastic recurrent neural networks." arXiv preprint arXiv:1911.01655.
- Wang, H., M. Naghavi, C. Allen, R. M. Barber, Z. A. Bhutta, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen and M. M. Coates (2016). "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015." The lancet **388**(10053): 1459-1544.
- Wang, J., H. Lu, T. Li and Z. Liang (2005). Sinogram noise reduction for low-dose CT by statistics-based nonlinear filters. Medical Imaging 2005: Image Processing, International Society for Optics and Photonics.
- Wang, K., Q. Shou, S. J. Ma, D. Liebeskind, X. J. Qiao, J. Saver, N. Salamon, H. Kim, Y. Yu and Y. Xie (2020). "Deep learning detection of penumbral tissue on arterial spin labeling in stroke." Stroke **51**(2): 489-497.
- Wang, X.-C., P.-Y. Gao, Y. Lin, L. Ma, Guan-ruiLiu, J. Xue, B.-B. Sui, C.-J. Wang and Y.-J. Wang (2009). "Clinical value of computed tomography perfusion source images in acute stroke." Neurological research **31**(10): 1079-1083.

Wang, Y., B. Yu, L. Wang, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen and L. Zhou (2018). "3D conditional generative adversarial networks for high-quality PET image estimation at low dose." Neuroimage **174**: 550-562.

Wang, Z., E. P. Simoncelli and A. C. Bovik (2003). Multiscale structural similarity for image quality assessment. The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, Ieee.

Wiesmann, M., S. Berg, G. Bohner, R. Klingebiel, V. Schöpf, B. Stoeckelhuber, I. Yousry, J. Linn and U. Missler (2008). "Dose reduction in dynamic perfusion CT of the brain: effects of the scan frequency on measurements of cerebral blood flow, cerebral blood volume, and mean transit time." European radiology **18**(12): 2967-2974.

Willeminck, M. J., T. Leiner, P. A. de Jong, L. M. de Heer, R. A. Nievelstein, A. M. Schilham and R. P. Budde (2013). "Iterative reconstruction techniques for computed tomography part 2: initial results in dose reduction and image quality." European radiology **23**(6): 1632-1642.

Wilm, J., O. V. Olesen, R. R. Paulsen, L. Højgaard, B. Roed and R. Larsen (2011). Real time surface registration for PET motion tracking. Scandinavian Conference on Image Analysis, Springer.

Wintermark, M., A. E. Flanders, B. Velthuis, R. Meuli, M. Van Leeuwen, D. Goldsher, C. Pineda, J. Serena, I. v. d. Schaaf and A. Waaijer (2006). "Perfusion-CT assessment of infarct core and penumbra: receiver operating characteristic curve analysis in 130 patients suspected of acute hemispheric stroke." Stroke **37**(4): 979-985.

Wintermark, M., M. Sesay, E. Barbier, K. Borbély, W. P. Dillon, J. D. Eastwood, T. C. Glenn, C. B. Grandin, S. Pedraza and J.-F. Soustiel (2005). "Comparative overview of brain perfusion imaging techniques." Stroke **36**(9): e83-e99.

Wolterink, J. M., T. Leiner, M. A. Viergever and I. Išgum (2017). "Generative adversarial networks for noise reduction in low-dose CT." IEEE transactions on medical imaging **36**(12): 2536-2545.

World Health Organization. (2018, May 24). "The top 10 causes of death." from <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>.

Wu, D., H. Ren and Q. Li (2020). "Self-supervised dynamic CT perfusion image denoising with deep neural networks." IEEE Transactions on Radiation and Plasma Medical Sciences **5**(3): 350-361.

Xiao, Y., P. Liu, Y. Liang, S. Stolte, P. Sanelli, A. Gupta, J. Ivanidze and R. Fang (2019). "STIR-Net: Deep Spatial-Temporal Image Restoration Net for Radiation Reduction in CT Perfusion." Frontiers in neurology **10**.

Xiao, Y., K. R. Peters, W. C. Fox, J. H. Rees, D. A. Rajderkar, M. M. Arreola, I. Barreto, W. E. Bolch and R. Fang (2020). Transfer-Gan: Multimodal Ct Image Super-Resolution Via Transfer Generative Adversarial Networks. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE.

Xu, Q., H. Yu, X. Mou, L. Zhang, J. Hsieh and G. Wang (2012). "Low-dose X-ray CT reconstruction via dictionary learning." IEEE transactions on medical imaging **31**(9): 1682-1697.

- Yang, F., S. S. B. Bal, Y.-F. Sung and G.-S. Peng (2020). "Mathematical Framework of Deconvolution Algorithms for Quantification of Perfusion Parameters." Acta Neurologica Taiwanica **29**(3): 79-85.
- Yang, Q., P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun and G. Wang (2018). "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss." IEEE transactions on medical imaging **37**(6): 1348-1357.
- Yap, B. W. and C. H. Sim (2011). "Comparisons of various types of normality tests." Journal of Statistical Computation and Simulation **81**(12): 2141-2155.
- Yazdi, M. and L. Beaulieu (2008). "Artifacts in spiral X-ray CT scanners: problems and solutions." International Journal of Biological and Medical Sciences **4**(3): 135-139.
- Yin, Z., K. Xia, Z. He, J. Zhang, S. Wang and B. Zu (2021). "Unpaired image denoising via wasserstein GAN in low-dose CT image with multi-perceptual loss and fidelity loss." Symmetry **13**(1): 126.
- Yu, B., L. Zhou, L. Wang, Y. Shi, J. Fripp and P. Bourgeat (2019). "Ea-GANs: Edge-aware Generative Adversarial Networks for Cross-modality MR Image Synthesis." IEEE transactions on medical imaging.
- Yu, B., L. Zhou, L. Wang, Y. Shi, J. Fripp and P. Bourgeat (2020). "Sample-adaptive gans: Linking global and local mappings for cross-modality mr image synthesis." IEEE transactions on medical imaging **39**(7): 2339-2350.
- Yu, Y., Q. Han, X. Ding, Q. Chen, K. Ye, S. Zhang, S. Yan, B. C. Campbell, M. W. Parsons and S. Wang (2016). "Defining core and penumbra in ischemic stroke: a voxel-and volume-based analysis of whole brain CT perfusion." Scientific reports **6**: 20932.
- Yu, Y., Y. Xie, T. Thamm, E. Gong, J. Ouyang, C. Huang, S. Christensen, M. P. Marks, M. G. Lansberg and G. W. Albers (2020). "Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging." JAMA network open **3**(3): e200772-e200772.
- Zaheer, Z., T. Robinson and A. K. Mistri (2011). "Thrombolysis in acute ischaemic stroke: an update." Therapeutic advances in chronic disease **2**(2): 119-131.
- Zeiler, M. D. (2012). "Adadelat: an adaptive learning rate method." arXiv preprint arXiv:1212.5701.
- Zhang, H., J. S. Polson, K. Nael, N. Salamon, B. Yoo, S. El-Saden, F. Scalzo, W. Speier and C. W. Arnold (2021). "Intra-domain task-adaptive transfer learning to determine acute ischemic stroke onset time." Computerized Medical Imaging and Graphics **90**: 101926.
- Zhang, J., L. Yu, D. Chen, W. Pan, C. Shi, Y. Niu, X. Yao, X. Xu and Y. Cheng (2021). "Dense GAN and Multi-layer Attention based Lesion Segmentation Method for COVID-19 CT Images." Biomedical Signal Processing and Control: 102901.
- Zhang, P., F. Wang, W. Xu and Y. Li (2018). Multi-channel generative adversarial network for parallel magnetic resonance image reconstruction in k-space. International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer.

- Zhao, C., T. Martin, X. Shao, J. R. Alger, V. Duddalwar and D. J. Wang (2020). "Low dose CT perfusion with K-space weighted image average (KWIA)." IEEE transactions on medical imaging **39**(12): 3879-3890.
- Zhao, K., L. Zhou, S. Gao, X. Wang, Y. Wang, X. Zhao, H. Wang, K. Liu, Y. Zhu and H. Ye (2020). "Study of low-dose PET image recovery using supervised learning with CycleGAN." Plos one **15**(9): e0238455.
- Zhou, S. K., H. Greenspan, C. Davatzikos, J. S. Duncan, B. van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert and R. M. Summers (2020). "A review of deep learning in medical imaging: Image traits, technology trends, case studies with progress highlights, and future promises." arXiv preprint arXiv:2008.09104.
- Zhu, F. (2013). "Brain perfusion imaging: performance and accuracy."
- Zhu, F., D. R. Gonzalez, T. Carpenter, M. Atkinson and J. Wardlaw (2013). "Lesion area detection using source image correlation coefficient for CT perfusion imaging." IEEE journal of biomedical and health informatics **17**(5): 950-958.
- Zhu, H., D. Tong, L. Zhang, S. Wang, W. Wu, H. Tang, Y. Chen, L. Luo, J. Zhu and B. Li (2020). "Temporally downsampled cerebral CT perfusion image restoration using deep residual learning." International journal of computer assisted radiology and surgery **15**(2): 193-201.
- Zhu, J.-Y., T. Park, P. Isola and A. A. Efros (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on computer vision.
- Zhu, X., L. Zhang, L. Zhang, X. Liu, Y. Shen and S. Zhao (2020). "GAN-Based Image Super-Resolution with a Novel Quality Loss." Mathematical Problems in Engineering **2020**.
- Zhu, X., X. Zhang, X.-Y. Zhang, Z. Xue and L. Wang (2019). "A novel framework for semantic segmentation with generative adversarial network." Journal of Visual Communication and Image Representation **58**: 532-543.