



## PAPER

## OPEN ACCESS

## RECEIVED

14 September 2023

## REVISED

1 February 2024

## ACCEPTED FOR PUBLICATION

25 March 2024

## PUBLISHED


2 April 2024

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



# Using positional tracking to improve abdominal ultrasound machine learning classification

Alistair Lawley<sup>1,\*</sup> , Rory Hampson<sup>1</sup>, Kevin Worrall<sup>2</sup> and Gordon Dobie<sup>1</sup><sup>1</sup> Department of Electronic and Electrical Engineering, University of Strathclyde, Royal College building, 204 George St, Glasgow G1 1XW, United Kingdom<sup>2</sup> Department of Engineering, University of Glasgow, James Watt South Building, University of Glasgow, Glasgow G12 8QQ, United Kingdom

\* Author to whom any correspondence should be addressed.

E-mail: [Alistair.Lawley@strath.ac.uk](mailto:Alistair.Lawley@strath.ac.uk)**Keywords:** machine learning, ultrasound, classification, infrared sensors

## Abstract

Diagnostic abdominal ultrasound screening and monitoring protocols are based around gathering a set of standard cross sectional images that ensure the coverage of relevant anatomical structures during the collection procedure. This allows clinicians to make diagnostic decisions with the best picture available from that modality. Currently, there is very little assistance provided to sonographers to ensure adherence to collection protocols, with previous studies suggesting that traditional image only machine learning classification can provide only limited assistance in supporting this task, for example it can be difficult to differentiate between multiple liver cross sections or those of the left and right kidney from image post collection. In this proof of concept, positional tracking information was added to the image input of a neural network to provide the additional context required to recognize six otherwise difficult to identify edge cases. In this paper optical and sensor based infrared tracking (IR) was used to track the position of an ultrasound probe during the collection of clinical cross sections on an abdominal phantom. Convolutional neural networks were then trained using both image-only and image with positional data, the classification accuracy results were then compared. The addition of positional information significantly improved average classification results from ~90% for image-only to 95% for optical IR position tracking and 93% for Sensor-based IR in common abdominal cross sections. While there is further work to be done, the addition of low-cost positional tracking to machine learning ultrasound classification will allow for significantly increased accuracy for identifying important diagnostic cross sections, with the potential to not only provide validation of adherence to protocol but also could provide navigation prompts to assist in user training and in ensuring adherence in capturing cross sections in future.

## 1. Introduction

Diagnostic ultrasound relies on the capture of cross-sectional images of anatomical structures within the body to provide a clinician with the requisite information to make a clinical decision. Capturing these anatomical cross sections is time consuming and requires a high level of user skill in anatomy and ultrasound operation [1, 2]. Machine learning has the potential to reduce the skill floor by assisting and automating ultrasound capture procedures, but to do so it must overcome the two fundamental difficulties: the differentiation of anatomical cross sections that are in close proximity and those that are visually similar. This is exemplified in previous studies [3, 4] showing that both experienced clinicians and neural networks [5] have substantial difficulty classifying abdominal cross sections where the anatomical structures were visually similar from image alone.

Machine learning has previously been used in the classification of 11 abdominal cross sections [3, 4] achieving respective accuracies of 77.9% and 82.2% using transfer learning. A classification study of 16

abdominal cross sections [5] achieved an accuracy 83.9%. The use of segmentation and landmarking [4, 6] was also shown to improve accuracy with models achieving 85.2% and 83.4% respectively, with increased accuracy possible if errors from similar cross sections were excluded. These studies show reduced accuracy where cross sections overlap or have visual similarities. Where a distinct dataset is used, that avoids these overlaps and visual similarities, accuracies of between 95.7% and 98.6% can be achieved [7]. This further highlights the limitations of using an image-only approach for abdominal cross sections, due to the lack of distinctive landmarks where there are overlapping classes within the imagery. Therefore, additional identifiers should be sought. Positional data has been previously used in medical ultrasound applications [8] such as 3D image reconstruction [9] and biopsy [10], but has not been utilised to assist machine learning in improving classification of diagnostic abdominal cross sections.

In order to test the efficacy of positional based tracking of an ultrasound probe for machine learning, two separate systems were tested: optical infrared tracking (IR) using a Vicon system, and an IR system based upon low-cost application specific integrated circuits (ASICs) IR sensors. Vicon has been shown to be highly accurate with within 2 mm [11], and is effective as a positional and registration reference measurement in other medical imaging applications [12, 13]. It also has shown to achieve high accuracies in motion capture, as part of complex automated classification processes such as respiratory tracking [14] and pose estimation [15]. The use of optical IR tracking would be difficult to implement within a clinical environment, due to the need for a large camera gantry, therefore a more mobile IR tracking system was designed based on a system used for full body tracking for virtual reality (VR). IR tracking has shown to be highly accurate at tracking while maintaining a low latency [16] with previous studies of similar positional systems being capable of tracking an ultrasound probe mounted to a robotic arm [17], spinal column tracking [18, 19], and tracking operator movements when applying machine learning to scanning the median nerve and radial artery [20].

This paper presents a proof-of-concept method to improve machine learning classification accuracy for abdominal scanning using positional information to augment image-based classification. This paper first compares image only machine learning classification to optical IR tracking within a Vicon system. Sensor-based IR tracking was then tested using a modified HTC VR tracking system. The use of the sensor-based IR tracking, while less accurate than Vicon, is to demonstrate the addition of positional tracking using a mobile sensor which would be more indicative of what could be used in a clinical environment. This paper does not seek to compare positional tracking precision, but the resultant output of the neural network classification using these tracking systems. This is to show how effective positional information is at improving classification of difficult to identify ultrasound cross sections and edge cases.

## 2. Method

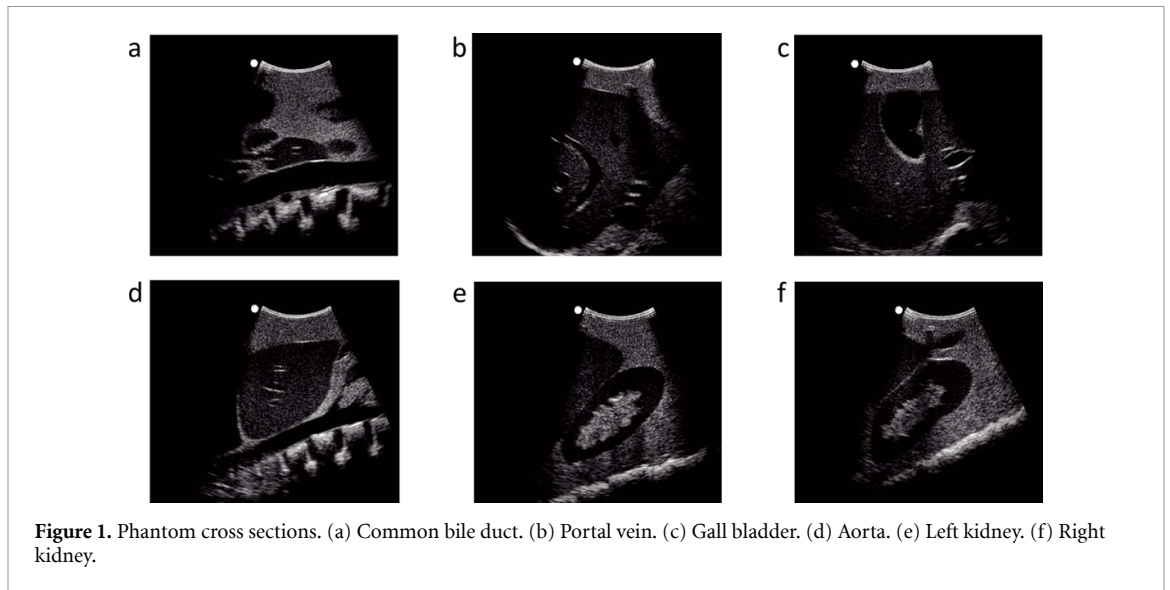
In order to make an effective comparison between image-only neural networks and those augmented with positional information, image and positional data was collected for six standard clinical abdominal cross sections and three normalisation points using an ultrasound abdominal phantom. This was performed within a laboratory environment using a medical ultrasound device and the optical IR or IR sensor positional tracking systems respectively. This data was then pre-processed into an image tensor and file containing classifier and raw coordinate output from the positional device to produce the dataset. The dataset was then split 80/20 at the session level to prevent data leakage and used to train a three-channel image only model and then a four-channel image and positional model. This model was validated using the unseen test set data and results outputted, the dataset was then re-split and the experiment repeated.

### 2.1. Dataset

A Kyoto Kagaku 'Echozy' ultrasound phantom (Kyoto Kagaku Co., Ltd, Japan) was scanned using a SonixTouch Q+ medical ultrasound system (SonixTouch, BK Ultrasound, USA) using a curved array, 5–2/60 ultrasound probe. These images were captured via HDMI cable using OpenCV [21] and were stored as .jpeg and .pt three-dimensional tensor files. Six cross sections were chosen as regions of interest (figure 1):

- (a) Right hypochondrium transverse approach for common bile duct.
- (b) Right intercostal approach sweeping through the liver to visualise the right portal vein.
- (c) Right hypochondrium longitudinal approach for the Gall Bladder.
- (d) Epigastric longitudinal approach sweeping through the aorta.
- (e) Transverse approach of the left kidney
- (f) Transverse approach of the right kidney.

These cross sections were chosen specifically based on classification error in previous studies [3–5, 7] and due to visual similarity, such as with the left and right kidneys and over lapping region of interest (ROI) such



**Table 1.** Phantom dataset size and composition.

Cross section	Optical	Infrared
Left kidney	3611	630
Right kidney	2629	563
Aorta	3335	725
Bile duct	3488	438
Gall bladder	2854	426
Portal vein	2697	628
Total images	<b>18 614</b>	<b>3410</b>

as with Gall bladder and Common Bile Duct. Complex sweep scans of aorta and portal veins that contain both visual similarities and overlapping anatomical structures were also chosen to provide added complexity to classification.

The optical IR dataset is made up of 137 sets of scans totalling 18 614 images, the IR dataset is made up of 22 sets of scans totalling 3410 images (table 1). Images were captured at a rate of 5 frames per second. Each set was performed as if scanning an individual patient with the sonographer using minor pressure and angle variation during the capture process while ensuring that the target ROI was visible and would adhere to standard clinical collection protocols. This was done to provide additional natural variation in the images.

## 2.2. Tracking system

Two methods of probe tracking were tested: Vicon optical IR tracking and sensor-based IR tracking. While the ultrasound and positional tracking systems were all capable a high rate of capture, a capture rate of 5 frames per second was used to prevent any de-synchronisation due to potential changes in system latency throughout the scanning process. As both Vicon optical IR and IR sensor tracker require line of sight and operate within the same frequency band, separate sessions were performed for each positional system to minimise any potential interference.

### 2.2.1. Vicon optical-based IR tracking

The Phantom was placed on a non-reflective surface within a fully calibrated Vicon optical measurement volume utilising a Vicon MX Giganet system [22] with 12 Vicon T160 cameras (16 MP, 18 mm focal length lens) mounted to a professional camera rig (figure 2). These cameras detect light reflected off tracking dots at a wavelength of  $\sim 850$  nm. Volume calibration was performed by placing the origin point on the floor 1 meter from the phantom ensuring that coordinates were as similar as possible between sessions. The ultrasound machine with the screen at its lowest position and laptop were placed at least 2 metres from the phantom and masked in the calibration setup to prevent interference with tracking. Vicon tracking markers were affixed to the probe, phantom and a Y frame that had been secured to the probe. The addition of the Y frame allowed for additional distance between tracking dots therefore increasing the sensitivity of the optical camera imagery and also ensuring line of sight could be maintained while the operator was positioning the probe. The Vicon API was used to stream the coordinates into python which was captured at a rate of 5

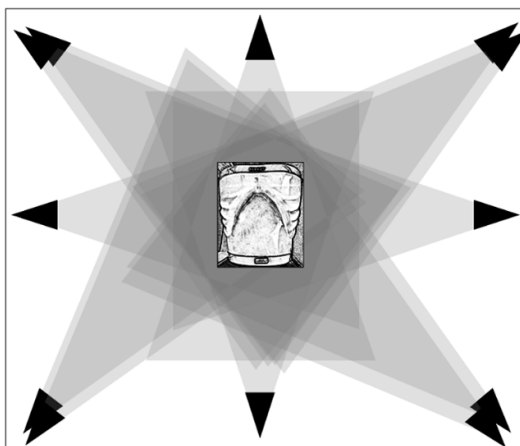


Figure 2. Optical tracking setup.

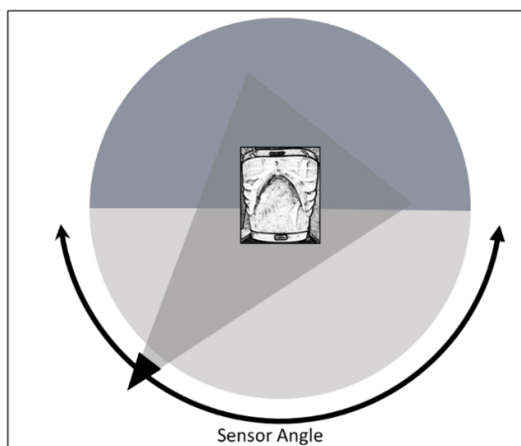


Figure 3. IR tracking setup.

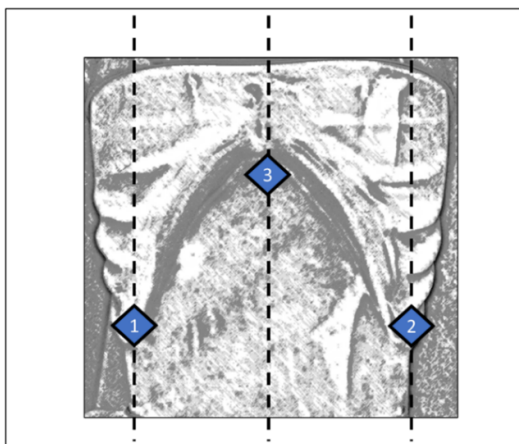
frames per second via a Wi-Fi connection from the laptop capturing the ultrasound images to a computer running the Vicon optical tracking development kit.

### 2.2.2. Infrared sensor tracking

This positional system is a modified setup based on those used for full body tracking for VR [23]. The system itself had been modified to track a single HTC VIVE (3.0) tracker [16, 24] which was affixed to the ultrasound probe using a strap and hot glue. A Steam VR base station (2.0) [25] was attached via a mounting strap to the ultrasound cart which was positioned anteroinferior to the Phantom ensuring clear line of sight (figure 3). The base station produces pulses of infrared light at a wavelength of  $\sim 850$  nm which is then detected by simple ASIC IR sensors on the VIVE tracker. VIVE tracker has previously been shown to be accurate to within  $0.68 \pm 0.32$  cm translationally and  $1.64 \pm 0.18^\circ$  rotationally [18] in comparison to the Vicon tracking system. The Base station was moved after each collection set to mimic moving to a new patient or clinical space. Note that anterosuperior scans were performed but excluded as they provided conflicting reversed positional data, this data could have been used if the angle of the phantom was tracked during the IR experiment, or a second base station used to provide additional point of reference. Software requirements for a headset and VR stage were bypassed by using a modified system profile and using developer options within the Steam VR software. OpenXR [26] was used to extract the coordinates from the VR runtime with a modified API used to stream the coordinates into python which was captured at a rate of 5 frames per second using a USB cable to the Steam VR base station.

### 2.2.3. Phantom coordinate normalisation

In order for the positional data to be used to effectively track the ultrasound probes movement it is necessary to normalise coordinates provided to the neural network so that they are of similar scale. In order to test



**Figure 4.** Abdominal normalisation points. (1) 9–10 on right midclavicular line, (2) 9–10 on left midclavicular line, (3) horizontally positioned on xiphoid notch.

normalisation methods, scans of three fixed points on the phantom were taken before each set of scans was performed as shown in figure 4:

- On the right midclavicular line, between the right 9th and 10th ribs.
- The probe is positioned on the xiphoid notch along the midsternal line with the probe positioned anteriorly.
- On the left midclavicular line, between the left 9th and 10th ribs.

These anatomical points on the ribcage, are less subject to variation due to patient positioning or disease process, are not subject to patient dignity concerns, and can be precisely and consistently pinpointed by a clinician. Use of soft tissue landmarks such as the umbilicus would be impractical in cases with abdominal distension where these features would be subject to greater variation. These defined points on the abdomen were used to normalise the coordinates for each axis, where multiple points are used, a simple mean is used to provide a single normalisation point. This normalisation point was then applied during the conversion to positional tensor.

Post-capture normalisation for the optical data was not required, as the optical IR positional data was automatically calibrated to a point within the measurement volume during each collection session, meaning that differences in coordinates between scan sessions was very small. However, the IR sensor base station was moved after each cycle of data collection to represent moving between patients and potential changes in clinical area.

In order to evaluate the amount of normalisation required prior to input into the neural networks four sets of normalisation data was produced by setting a new zero point:

- No normalisation—using the original captured values.
- 1-point normalisation from point one on the anatomical right side of the phantom. This locates a single point on the abdomen within the tracked volume.
- 2-point normalisation using a simple mean of points 1 and 2 on the phantom. These measurements would allow for the sizing of the abdomen along a single dimension.
- 3-point normalisation using a combined simple mean of all three normalisation points. This would allow for the two-dimensional sizing of the abdomen.

Training of the neural network was performed for image only and for each of the normalisation point using the same dataset split so that results could be compared.

### 2.3. Machine learning implementation

All training and testing was performed on a 64bit version of Windows 10, using a Intel core i9 and Nvidia 40 series GPU using python [27] (version 11.4) and the CUDA toolkit (version 11.7). The SciPy metrics library was used to analyse model output. A pre-trained ResNet-50 [28] convolutional neural network from the torchvision library was used as the basis for study, with weights based on ImageNet challenge dataset [29]. The final layer of this network is adjusted to output 6 classes. Image-only method uses the default 3 channel neural network. For the positional study, the neural network was modified to accept a 4th channel for the inclusion of the positional data.

As the dataset consists of scan of a single phantom, overfitting is a concern, as such the experiment was repeated 50 times for each normalisation type to provide an average training response, over a maximum of 5 epoch using early stopping [30] and a small batch size of 64 to promote better generalisation [31]. Training used a learning rate of  $1.00 \times 10^{-04}$  using the ADAM optimizer [32]. Training and testing methodology was identical for both 3 and 4 channel versions of the network.

The dataset images were converted into tensors with 3 channels of size  $330 \times 370$  pixels. The optical dataset was split 80/20 into training and test sets for each experimental run using a stratified random split to ensure a balanced training set, with both the 3 and 4 channels networks trained and validated using this split, so that a direct comparison between image-only and positional tracking could be performed. For the positional experiments the IR dataset was split 50/50 between training and testing. The positional data was converted into a tensor, the normalisation sum performed and input into the network alongside the 3 image channels. Training was repeated for each normalisation state on the same data split to ensure comparison could be performed. The datasets were split at session level to prevent bias due to data leakage.

### 3. Results

Neural networks trained using the optical IR dataset produced average accuracies of 91.47% for image-only based training and 95.75% with the addition of positional data, an average improvement of 4.3% (table 2). The largest accuracy improvements can be seen in classification of the bile duct (6.4%) and portal vein (7.8%). The highest performing image-only network achieved an accuracy of 96.34% with the largest error in the classification of Gall Bladder and Bile Duct. The highest performing optical IR network achieved an overall accuracy of 98.84%, with errors in aorta, bile duct and gall bladder classification. When examining accuracy variance as seen in figure 5, networks trained with positional data, achieved an average reduction in variance of 23% overall.

When statistically comparing image-only and optical IR tracked results by performing a twin tailed *T*-Test with the assumption of heteroscedastic variance (table 2), the optical IR tracking results proved to be statically significant with an averaged *P* value of 0.0482. When the results are analysed on a class by class basis, results are shown to be highly significant achieving *P*-values  $< 0.003$ , however there is insufficient statistical significance when comparing aorta classification results (*p*-value 0.2436).

Using figure 5 to compare the accuracy all 50 trained networks, the deviation in class accuracy was substantially higher in image only trained network in comparison to those using optical IR tracking data. Misclassification of gall bladder, bile duct, aorta and portal vein was the largest cause of deviation for both image-only and optically tracked networks. Misclassification of left and right kidney was reduced to less than 3% in optically tracked networks with the average network achieving above 99% accuracy for the kidney cross section images.

When examining the networks with the highest accuracy using a confusion matrix (figure 6), the largest source of error for both image-only and optically tracked network is between bile duct and gall bladder, this error is present throughout both network types, and was consistent across all 100 networks. If we compare the image-only networks to its optically tracked network trained on the same dataset split, the optically tracked network improves upon the image only accuracy result by an average of 4% with optically tracked always improving on its image only counterpart.

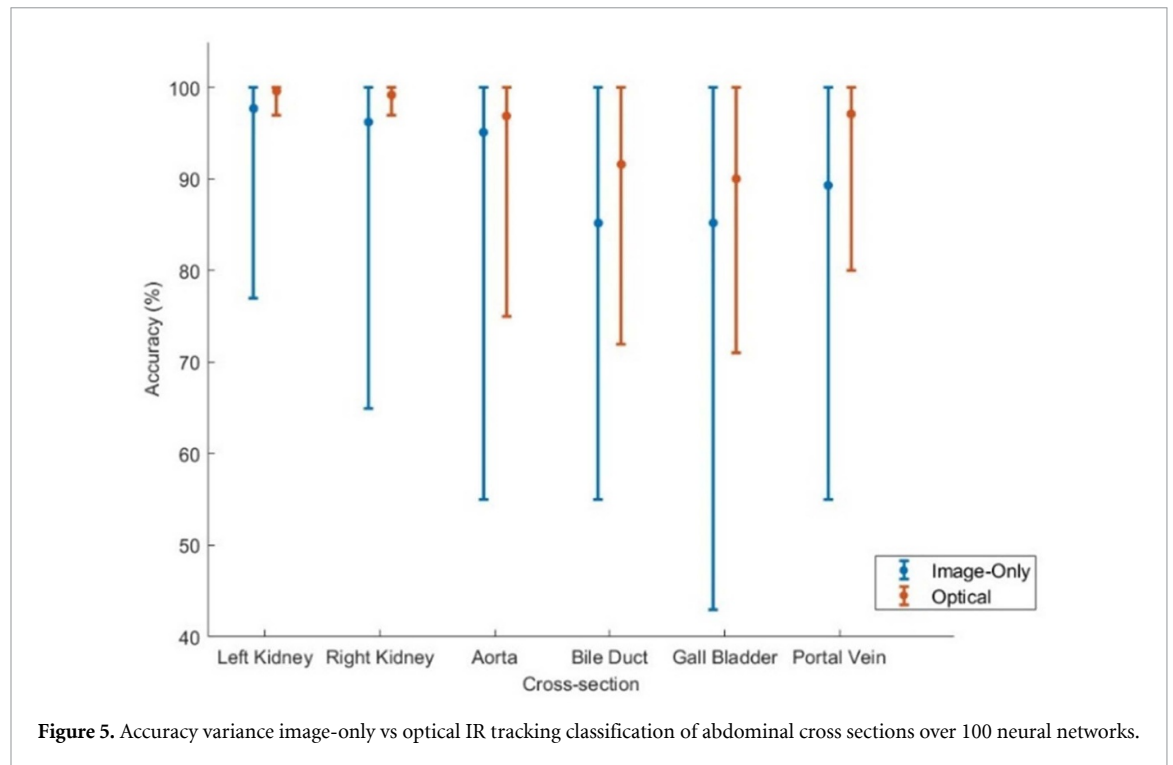
The inclusion of the IR data into the training and test sets (table 3) saw an average accuracy of 89.70% for image only classification. Positional accuracy achieved average accuracy of 92.71% without any form of normalisation, 93.61% for single point normalisation, 93.73% for two-point normalisation, and 93.28% for three-point normalisation respectively. When compared against image-only accuracy results, networks trained on non-calibrated positional data achieved an average improvement of  $\sim 3\%$  improvement, with calibrated positional data achieving  $\sim 4\%$  improvement in cross section classification. Common bile duct and gall bladder classification were the largest sources of error in both image-only and positional tracked networks. The maximum achieved network accuracy was 97.5% for image only, 98.5% for no normalisation, 98.7% one point of normalisation, 98.3% for two points of normalisation and 97.5% for 3 points of normalisation.

While there is an overall improvement in classification accuracy when using IR sensor tracking with normalisation, a single factor ANOVA test showed that there is insufficient statistical significance (*F*-value 0.3521, *P*-value 0.7038) in the results to distinguish between 1, 2 and 3 points of normalisation. This is likely due to a limitation of this study as there is insufficient difference in the size of the abdominal cavity to confirm efficacy of normalisation.

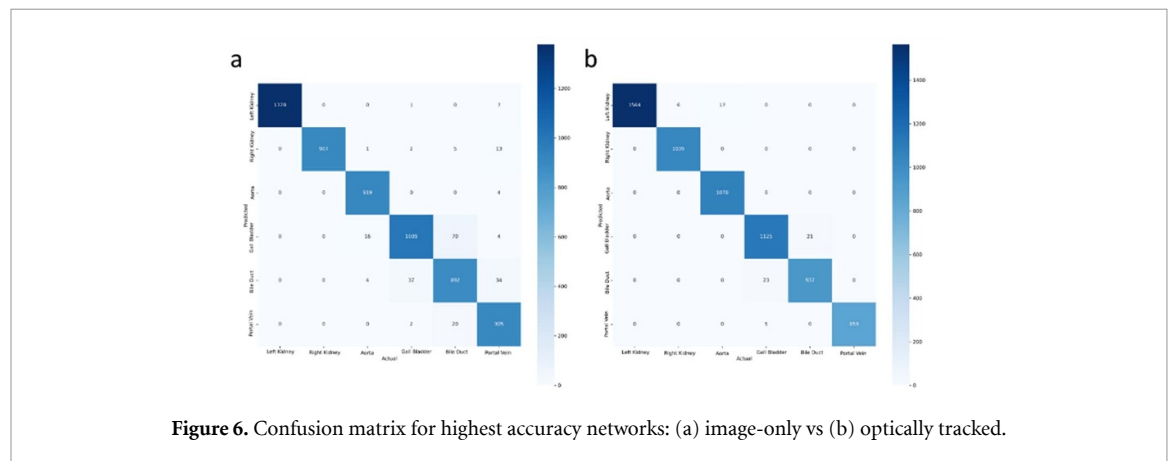
A comparison of the accuracy of the 50 trained networks (figure 7) shows that while overall accuracy was improved, there was increased training variance in comparison to optical IR. Despite an increase in overall accuracy, networks trained with positional data with no normalisation saw an increase in training variance by

**Table 2.** Optical IR tracking average accuracy results of 50 neural networks.

	Image only	Optical IR tracking	Accuracy improvement	Training variance	Comparative <i>P</i> value
Left kidney	97.7%	99.6%	1.9%	19.6%	0.0104
Right kidney	96.2%	99.2%	3.0%	29.7%	0.0033
Aorta	95.1%	96.9%	1.8%	14.7%	0.2436
Bile duct	85.2%	91.6%	6.4%	15.4%	0.0019
Gall bladder	85.2%	90.0%	4.8%	35.1%	0.0298
Portal vein	89.3%	97.1%	7.8%	23.6%	0.0003
Average	<b>91.5%</b>	<b>95.8%</b>	<b>4.3%</b>	<b>23.0%</b>	<b>0.0482</b>



**Figure 5.** Accuracy variance image-only vs optical IR tracking classification of abdominal cross sections over 100 neural networks.



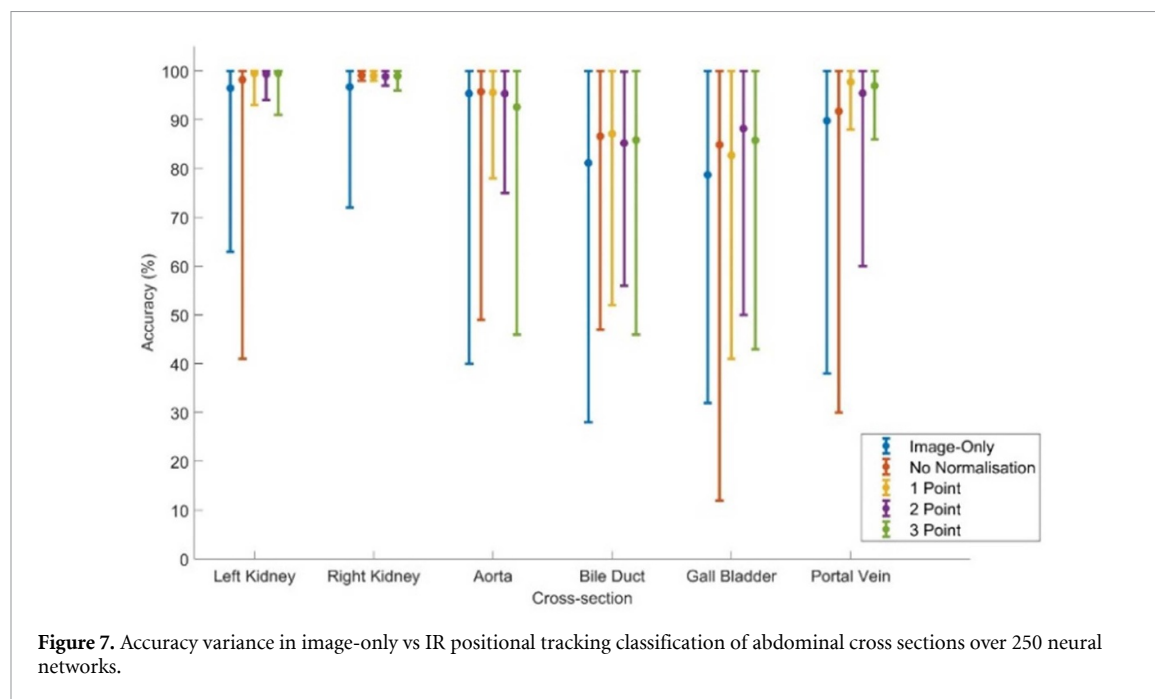
**Figure 6.** Confusion matrix for highest accuracy networks: (a) image-only vs (b) optically tracked.

5.3%, compared to improvements of 21.7% for one-point normalisation, 18.6% for two-point, and 15.4% improvement for three-point normalisation. Accuracy values were also lower than image-only results for 12 out of the 50 no normalisation networks, as this also occurred in a number of calibrated networks this is most likely due to training variance. This notably did not occur in the more accurately calibrated optically tracked networks.

When comparing the confusion matrix for the IR tracking networks (figure 8), the gall bladder and bile duct are the most confused classifications. This result is directly comparable to that seen in figure 6.

**Table 3.** IR positional tracking average accuracy results of 50 neural networks.

	Image only	No normalisation	1-point normalisation	2-point normalisation	3-point normalisation
Left kidney	96.5%	98.2%	99.5%	99.3%	99.5%
Right kidney	96.7%	99.1%	99.0%	98.9%	99.0%
Aorta	95.4%	95.7%	95.6%	95.3%	92.6%
Bile duct	81.1%	86.6%	87.1%	85.2%	85.8%
Gall bladder	78.7%	84.9%	82.7%	88.2%	85.8%
Portal vein	89.8%	91.7%	97.7%	95.4%	97.0%
Average	89.7%	92.7%	93.6%	93.7%	93.3%



## 4. Discussion

This paper demonstrates the use of positional data to improve classification of abdominal ultrasound cross sections on an ultrasound phantom using both optical IR and Sensor-based IR tracking systems. On average neural networks trained on optical IR tracking data provided the highest accuracy network, followed by IR Sensor tracking, and standard image-only classification.

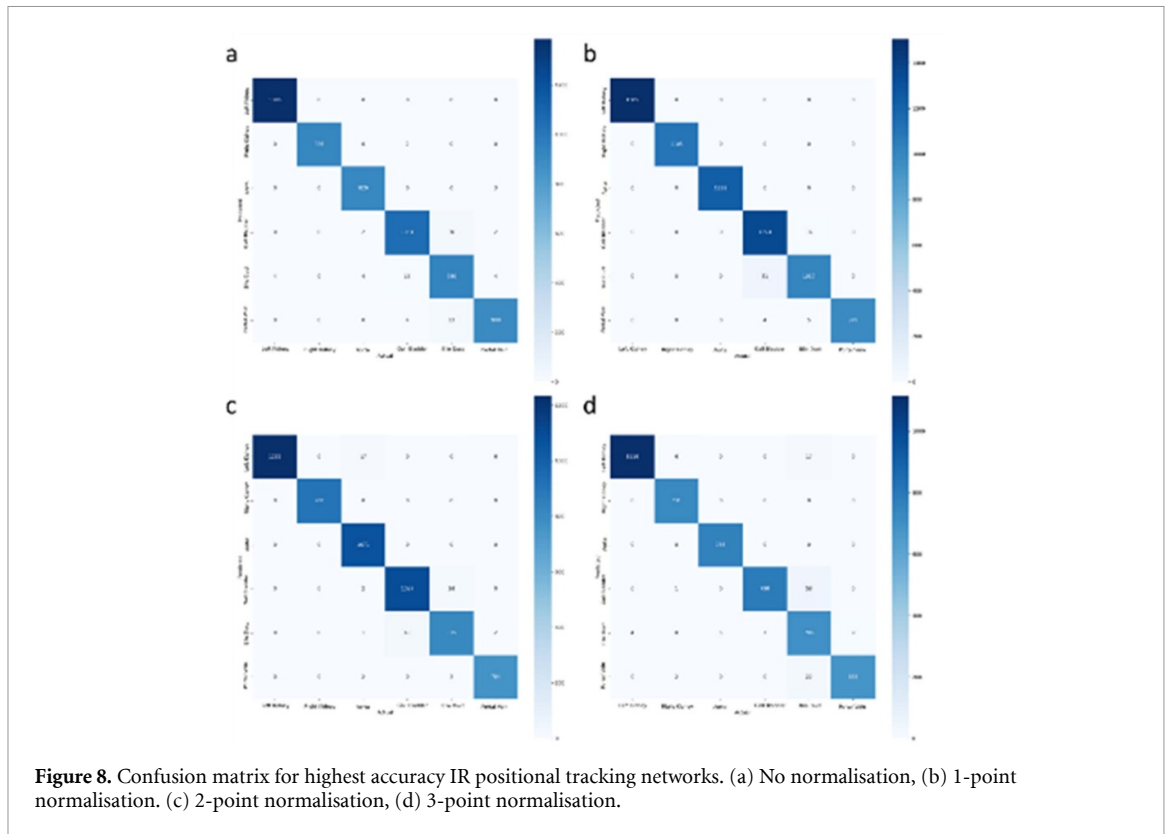
### 4.1. Study limitations

While the phantom was designed to provide an accurate representation of ultrasound cross sections for clinical training purposes, it is an idealised representation of a human abdomen and cannot fully represent the difficulties usually encountered during image acquisition in ultrasound scanning such as:

- Shadowing is limited as phantom materials do not have the same density range that would reduce amplitude and obscure ROI.
- Attenuation changes from different tissue thicknesses or densities present in the anterior abdominal wall are not represented in the phantom.
- Image artefacts common in ultrasound such as those from the digestive tract are not present in the phantom, as these artefacts, such as gas pockets are not represented in the phantom.
- The phantom is of fixed size and as such cannot represent different sized abdominal cavities, however this does not reduce clinical applicability as while the scale would change, in a high proportion of cases the position of cross sectional landmarks would remain the same in relation to one another, and as such can still be used for positional identification.

The use of a single subject (the phantom) has caused image-based accuracy results to be inflated despite use of a holdout test set and variation in probe position during collection, there is still a high level of similarity





**Figure 8.** Confusion matrix for highest accuracy IR positional tracking networks. (a) No normalisation, (b) 1-point normalisation. (c) 2-point normalisation, (d) 3-point normalisation.

due to the structures of the phantom being identical in all scans. This does not however prevent comparison of image-only and positional performance as the same subject and image set is used therefore any bias is present in both control and experimental tests. In this case, overfitting would favour an image-based solution, however, results clearly indicate that positional tracking has provided a significant improvement to abdominal cross section classification, strongly suggesting that positional tracking merits further study.

Figures 5 and 7 show significant accuracy variance, in order to reduce overfitting strict early stopping was applied during training, which has led to underfitting of some classifiers, resulting in increased variance than would normally be seen. While this is a study limitation, the reduced variance in networks augmented with positional data is highly indicative that this is improving classification accuracy.

The use of normalisation did increase overall accuracy by  $\sim 4\%$ , but there is very little difference between 1, 2 and 3 points of normalisation. This is due to a limitation of the experimental setup, because the phantom is a fixed size, once a fixed point on the abdomen is located, no additional variation in abdomen shape or volume is required to be taken into account. In a human trial the abdomen could potentially have much greater levels of variation and therefore the requirement of additional normalisation points should not be discounted in future experimental trials. The networks using positional data with no normalisation had the most variance in accuracy result, achieving the worst performing network at 73.3%, but still outperforming the image-only networks on average. This is likely partially due to the error rotatory angle data  $[a, b, c]$  being much smaller than that of positional  $[x, y, z]$  data. This would be particularly useful in the recognition between left and right kidney, which maintained accuracy comparable to calibrated trained networks despite providing positional  $[x, y, z]$  values that were likely incompatible with those already seen by the network during training.

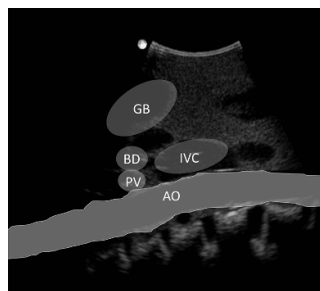
#### 4.2. Accuracy

While accuracy has been used as the main metric throughout this paper, examining the harmonic mean for the highest accuracy neural networks (table 4) confirms high precision and recall for all methodologies used. This is due to the limited subject matter available with using only one phantom. Despite using a single phantom, overfitting has been sufficiently reduced using variation in the image capture technique, early stopping, small batch size and experimental repetition to provide indicative results, there is a distinct correlation between the use of positional information when training a neural network and the improvement of classification result.

When comparing the image-only accuracy results between tables 2 and 3, there is a significant drop in the average classification accuracy of the gall bladder and bile duct between networks trained on the optical

**Table 4.** F-1 score for highest accuracy neural networks from optical and IR experiments.

	Image only	Optical tracking	IR No normalisation	IR 1-point	IR 2-point	IR 3-point
Left kidney	1	0.99	1	1	0.99	0.99
Right kidney	0.99	1	1	1	1	0.99
Aorta	0.99	0.99	1	1	0.99	1
Bile duct	0.94	0.98	0.97	0.96	0.97	0.96
Gall bladder	0.91	0.98	0.96	0.95	0.96	0.92
Portal vein	0.96	1	0.98	1	0.99	0.98

**Figure 9.** Bile duct cross section with gall bladder present in image, gall bladder (GB), bile duct (BD), inferior vena cava (IVC), portal vein (PV), aorta (AO).

set alone vs those trained with additional images from the IR sensor dataset. This is not however reflected by the level of accuracy achieved by the highest performing image-only networks with the optical image-only achieving 96.34% compared to 97.5% with the addition of images from the IR dataset. There was also significantly more variance in training result in the IR image-only networks, with the lowest network result being 79.5%, which is 1.8% lower than achieved by the worst performing optical image-only network. It is important to note that image-only results would be likely be lower with a larger sample size, imagery would also lack the same level of clarity in human trials, where body shape and differences in thickness and density would cause changes to attenuation properties that that would have to be considered.

#### 4.3. Accuracy variance

While use of positional tracking significantly improved overall accuracy and reduced training variance, the causes for this variance should be examined to ensure the validity of the results. Training variance was a significant factor in poor performance. Networks trained on the same dataset split produced similar class accuracy results with networks trained using positional data achieved consistently higher accuracy than their image-only counterparts trained on the same dataset split. Use of a small batch size produced more weight updates per epoch allowing faster convergence but also added additional training noise causing variance. Training was restrained to a maximum of 5 epoch to reduce overfitting but this has also led to significant variance in image-only and positional models with the lowest accuracy due to poor training performance.

#### 4.4. Dataset variance and overlap

As seen in the results for both optical IR and IR sensor experiments, the largest source of error for all models was between gall bladder and common bile duct. As this was consistent across both network types it is important to rule out an error within the dataset itself. Analysing the images where this error had occurred revealed that in adjusting the probe position to add variation to the dataset, a number of the images capture both gall bladder and bile duct (figure 9). These images still contain the target features but also cover the other anatomical structure as well. This overlapping visual information is the exact type of edge case that was targeted during cross section selection and exists within a number of real clinical protocols. There was a significant improvement in accuracy suggesting that probe angle information is making a significant difference in the classification of cross sections where the target ROI overlaps.

#### 4.5. Clinical practicality

Optical IR tracking was the most precise position tracking used, providing a 99% accurate neural network with the average accuracy result being ~95%, but this was achieved using an expensive Vicon measurement volume with an external calibration software, and required additional hardware attached to the ultrasound probe to maintain line of sight with the camera rig. A fixed camera set up would significantly reduce the mobility of the ultrasound system, which is one of the core benefits of ultrasound over other medical

imaging technologies, a mobile camera setup would require a substantial amount of space to ensure good visibility of the patient from multiple angles, as well as precise calibration, potentially blocking access in a busy clinical area. While excellent at validating positional data as a viable method of improving neural network recognition of abdominal cross section views, optical IR tracking is not clinically practical outside of specialist facilities. Sensor-based IR tracking while being less accurate overall, still achieved ~98% accuracy using simple normalisation techniques with an average of ~93%, higher than that achieved by image-only. The IR tracking system still required line of sight but was compact with the base station able to be easily mounted to the screen arm, although in many tight clinical areas it could still be a challenge to ensure that the base station is far enough back to fully cover the ROI and could potentially be blocked by the patient or clinician during the scan causing no positional data to be collected. The HTC VIVE tracker was small enough not to interfere with scanning, initially the strap holding the tracker in place would slip from the probes ergonomic design, but this problem was easily solved with adhesive which ensured a tracker position on the probe was maintained throughout data collection. During collection of IR scans, the position of the base station relative to the phantom was initially a full 360 degree loop with the base station moved after each set of scans, but due to the use of a single base station, and the fact no tracker was attached to the phantom, it was not possible to fully localise the positional data, scans taken with the base station in an anterosuperior position were mirrored in comparison to the optical IR tracking data. As such all anterosuperior scans were excluded from the IR dataset instead of manually adjusting these values and potentially adding additional human error to the training set. Ideally a sensor system that does not require line of sight such as electromagnetic sensors should be used to track the probe although these too would have to be carefully considered to ensure they do not interfere with other medical equipment.

## 5. Conclusion

This paper highlights the potential of positional sensor information as an additional data source when training neural networks on diagnostic cross sections that may be hard to differentiate using image alone. Optical IR positional tracking was highly accurate and substantially increased classification accuracy. Mobile sensor-based IR tracking provided a less accurate, but more practical example of applying positional information to machine learning for clinical use cases but also highlighted a number of difficulties that would need to be overcome before such technologies could be used. Contextualising cross section imagery that are in close proximity, or where there is a high level of visual similarity is no longer a challenge when the position of the probe is known in relation to other scans.

The collection and use of use of positional information as part of an ultrasound scan will allow a neural network to know the position of the probe relative to the patient, opening up many exciting opportunities for future research. Immediate future work should focus on increasing the sample size using a cadaver study to further test data and normalisation requirements. Electromagnetic sensors will be tested as a method of probe tracking as this technology does not require line of sight. Neural networks that can localise the probe to the position on the abdomen can provide feedback to the sonographer to assist in the positioning and fine tuning of the probe for the collection of potentially higher quality ultrasound cross sections that fully capture the required anatomical structures as mandated in the clinical protocol. It would also allow a more experienced user to sweep the probe over the ROI with the neural network selecting and potentially annotating the required cross sections automatically, speeding up scan times and reducing workload.

## Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://doi.org/10.15129/6151ae14-8acf-4bed-a1b6-d9367aeec218>.

## Funding

This work was supported by a UK Engineering and Physical Sciences Research Council (EPSRC) Future Ultrasonic Engineering Center for Doctoral Training (FUSE CDT) under Grants EP/S023879/1 and 2296317.

## Conflict of interest

The authors declare no conflicts of interest.

## ORCID iD

Alistair Lawley  <https://orcid.org/0000-0002-0903-1116>

## References

- [1] Schott C K et al 2021 Retention of point-of-care ultrasound skills among practicing physicians: findings of the VA National POCUS Training Program *Am. J. Med.* **134** 391–399.e8
- [2] Mulder T A, van de Velde T, Dokter E, Boekstijn B, Olgers T J, Bauer M P and Hierck B P 2023 Unravelling the skillset of point-of-care ultrasound: a systematic review *Ultrasound J.* **15** 1–13
- [3] Cheng P M and Malhi H S 2017 Transfer learning with convolutional neural networks for classification of abdominal ultrasound images *J. Digit. Imag.* **30** 234–43
- [4] Xu Z et al 2018 Less is more: simultaneous view classification and landmark detection for abdominal ultrasound images *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st Int. Conf., Proc., Part II 11 (Granada, Spain, 16–20 September 2018)* (Springer) pp 711–9
- [5] Lawley A, Hampson R, Worrall K and Dobie G 2024 Analysis of neural networks for routine classification of sixteen ultrasound upper abdominal cross sections *Abdom Radiol.* **49** 651–61
- [6] Dandan L, Huanhuan M, Yu J and Yi S 2020 A multi-model organ segmentation method based on abdominal ultrasound image *2020 15th IEEE Int. Conf. on Signal Processing (ICSP)* vol 1 (IEEE) pp 505–10
- [7] Reddy D S, Rajalakshmi P and Mateen M 2021 A deep learning based approach for classification of abdominal organs using ultrasound images *Biocybernetics Biomed. Eng.* **41** 779–91
- [8] Peng C, Cai Q, Chen M and Jiang X 2022 Recent advances in tracking devices for biomedical ultrasound imaging applications *Micromachines* **13** 1855
- [9] Mohamed F and Siang C V 2019 A survey on 3D ultrasound reconstruction techniques *Artif. Intell. Appl. Med. Biol.* 73–92
- [10] Beigi P, Salcudean S E, Ng G C and Rohling R 2021 Enhancement of needle visualization and localization in ultrasound *Int. J. Comput. Assist. Radiol. Surg.* **16** 169–78
- [11] Summan R, Pierce S G, Macleod C N, Dobie G, Gears T, Lester W, Pritchett P and Smyth P 2015 Spatial calibration of large volume photogrammetry based metrology systems *Measurement* **68** 189–200
- [12] Hampson R, West G and Dobie G 2023 Tactile, orientation, and optical sensor fusion for tactile breast image mosaicking *IEEE Sens. J.* **23** 5315–24
- [13] Rana S, Hampson R and Dobie G 2019 Breast cancer: model reconstruction and image registration from segmented deformed image using visual and force based analysis *IEEE Trans. Med. Imag.* **39** 1295–305
- [14] Mok G S, Zhang Q, Sun J, Zhang D, Pretorius P H and King M A 2019 Preliminary investigation of auto-classification of respiratory trace using convolutional neural network for adaptive respiratory gated myocardial perfusion SPECT *2019 IEEE Nuclear Science Symp. and Medical Imaging Conf. (NSS/MIC)* (IEEE) pp 1–3
- [15] Vafadar S, Gajny L, Boëssé M and Skalli W 2019 Evaluation of CNN-based human pose estimation for body segment lengths assessment *VipIMAGE 2019: Proc. VII ECCOMAS Thematic Conf. on Computational Vision and Medical Image Processing (Porto, Portugal, 16–18 October 2019)* (Springer) pp 179–87
- [16] Caserman P, Garcia-Agundez A, Konrad R, Göbel S and Steinmetz R 2019 Real-time body tracking in virtual reality using a Vive tracker *Virtual Real.* **23** 155–68
- [17] Ameler T et al 2019 A comparative evaluation of steamvr tracking and the optitrack system for medical device tracking *2019 41st Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 1465–70
- [18] Van der Veen S M, Bordeleau M, Pidcoe P E, France C R and Thomas J S 2019 Agreement analysis between vive and vicon systems to monitor lumbar postural changes *Sensors* **19** 3632
- [19] Meszaros-Beller L, Antico M, Fontanarosa D and Pivonka P Assessment of spinal curvatures in static postures using localized 3D ultrasound: a proof-of-concept study
- [20] Marharjan J, Mitchell B R, Chan V W and Kim E 2020 Guided ultrasound imaging using a deep regression network *Proc. SPIE* **11319** 28–36
- [21] Bradski G 2000 The openCV library *Dr Dobb's J.: Softw. Tools Prof. Program.* **25** 120–3
- [22] Vicon *Vicon award winning motion tracking systems* (available at: [www.vicon.com/](http://www.vicon.com/)) (Accessed 7 August 2023)
- [23] Caserman P, Garcia-Agundez A and Göbel S 2019 A survey of full-body motion reconstruction in immersive virtual reality applications *IEEE Trans. Vis. Comput. Graphics* **26** 3089–108
- [24] HTC *Introducing VIVE Tracker (3.0)* (available at: [www.vive.com/uk/accessory/tracker3/](http://www.vive.com/uk/accessory/tracker3/)) (Accessed 7 August 2023)
- [25] HTC *SteamVR Base Station 2.0* (available at: [www.vive.com/uk/accessory/base-station2/](http://www.vive.com/uk/accessory/base-station2/)) (Accessed 7 August 2023)
- [26] K. Group *OpenXR Overview* (available at: [www.khronos.org/openxr/](http://www.khronos.org/openxr/)) (Accessed 7 08 2023)
- [27] Paszke A et al 2017 Automatic differentiation in pytorch
- [28] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8
- [29] Russakovsky O et al 2015 Imagenet large scale visual recognition challenge *Int. J. Comput. Vis.* **115** 211–52
- [30] Caruana R, Lawrence S and Giles C 2000 Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping *Advances in Neural Information Processing Systems* vol 13 pp 402–8
- [31] Keskar N S, Nocedal J, Tang P T P, Mudigere D and Smelyanskiy M 2017 On large-batch training for deep learning: generalization gap and sharp minima *5th Int. Conf. on Learning Representations, ICLR 2017*
- [32] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)