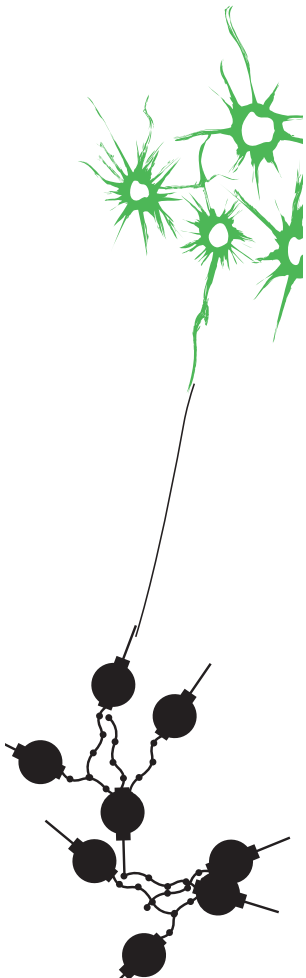

RAM ROBOTICS AND MECHATRONICS

LIVER RESPIRATORY MOTION ESTIMATION USING A-MODE ULTRASOUND AS SURROGATE SIGNAL AND LSTM NETWORKS AS CORRESPONDENCE MODEL

A. (Arend) Heijn

BSC ASSIGNMENT

Committee:
dr. ir. M. Abayazid
A. Cordon, MSc
dr. V. Kalpathy Venkiteswaran

July, 2024

017Ram2024
Robotics and Mechatronics
EEMCS
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

UNIVERSITY OF TWENTE. | TECHMED CENTRE

UNIVERSITY OF TWENTE. | DIGITAL SOCIETY INSTITUTE

Liver respiratory motion estimation using A-mode ultrasound as surrogate signal and LSTM networks as correspondence model

Arend Heijn
BSc report

***Abstract*—Purpose: Respiratory motion estimation of the liver using A-mode ultrasound as surrogate signal.**

Methods: Two LSTM networks of differing complexity have been made to function as motion model. The performance of these models was validated using a synthetic dataset. The best performing model architecture was additionally validated on data recorded from three human subjects. The ground truth was acquired from simultaneously recorded B-mode ultrasound data.

Results: The synthetic dataset had an MAER of 0.48 cm and 0.59 cm for the shallow and deep motion model respectively. Due to the better performance, the shallow model was further applied on the human subject data. The shallow model had an MAER of 0.83 cm, 0.18 cm and 0.54 cm for subject 1, 2 and 3 respectively.

Conclusion: Respiratory motion model performance differs significantly between subjects. The subjects with better model performance also had better surrogate signal quality. If the surrogate signal is of sufficient quality, the current methodology has the potential to outperform conventional biopsy protocol on tumours smaller than 1 cm.

I. Introduction

Liver cancer was responsible for 8.3% of all cancer related deaths globally in the year 2020 [1]. To successfully diagnose hepatocellular carcinoma, biopsies are of crucial importance [2, 3]. When performing these biopsies, the tumour experiences respiratory motion which complicates the procedure. Common solution is to ask the patient to perform a breath hold during insertion [4].

A. Problem statement

The use of a breath hold during needle insertion has a few limitations. For instance, the duration of the breath hold must be limited to a maximum of 30 seconds. The breath hold location also has a certain error, since the variability of breathing diminishes the repeatability of breath holds [5, 6].

The error of breath hold positions can be compensated for if the tumour is tracked real-time. Surgical robotics can be used to accurately follow the tumour if the location is known [7]. With surgical robotics, it is possible to take a biopsy consistently with higher accuracy than manual methods [7].

There are multiple pathways towards the required real-time liver displacement. A possibility is to measure a 'surrogate' signal which is easily measurable and highly correlated to the motion of interest. This surrogate signal can be used to compute an estimation of the motion of interest.

For instance, this has been done with a regression model using the position of external markers as surrogate signal to estimate the position of the liver [8]. The type of surrogate signal can vary greatly as long as it is correlated to the motion of interest. For instance, the use of A-mode ultrasound recordings as surrogate signal in human subjects to generate MRI images of the liver has also been studied [9]. With the outlook of applying these techniques in clinical practice, it is of importance to further explore the possibilities of using surrogate signals to estimate the liver position.

B. Research Objectives

The current paper will explore the prospect of using neural networks with Long Short Term Memory (LSTM) layers as motion model to estimate Superior-Inferior (SI) liver displacement. For this, the accurate liver displacement will be referred to as the ground truth. Two types of validation experiments will be conducted. Firstly, synthetic liver displacement and synthetic A-mode ultrasound is used as ground truth and surrogate signal respectively. The performance of a shallow model on the synthetic dataset will be compared to a deep model. Secondly, the model architecture with better performance on synthetic data will be validated in a study on 3 human subjects. A-mode ultrasound will be used to record the surrogate signal for the human study. Ultrasound is used because it has promising characteristics for application in motion models. Ultrasound is inexpensive, flexible, real-time, radiation free and has a high penetration depth [10].

II. Theoretical Background

A. Ultrasound

Ultrasound imaging utilises piezo electric crystals which vibrate in response to an electric potential and generate an electric potential in response to vibrations [11]. The generated sound waves attenuate and reflect based on acoustic absorption and impedance of the tissue respectively [12]. The reflected sound wave is thereafter transformed into an electric potential by the crystal. This electric potential is sent to a computer for analysis or post processing.

Multiple ultrasound modes like A-mode and B-mode ultrasound exist. A-mode ultrasound measures the properties of cells in a line perpendicular to the skin. B-mode ultrasound measures the properties of cells in a plane perpendicular to the skin [13].

The use of ultrasound comes with several advantages when working with motion models. Some of these advantages have been mentioned in section I-B. Ultrasound equipment is relatively inexpensive and already used for diagnostic purposes when hepatocellular carcinoma is suspected [14]. A disadvantage of using A-mode ultrasound is that the axis of imaging is in the anterior-posterior direction, which has a smaller respiratory compared to the SI direction [10].

B. Respiratory Motion Models

Motion models are correspondence models with the purpose of estimating internal organ motion using a different external signal (a surrogate signal) as an input. In a review article, motion models are defined as any process which takes some surrogate as an input and produces a motion estimate as an output [5]. Motion models can be used when the motion of interest can not be measured directly or with can not be measured with sufficient temporal resolution [5]. A properly chosen surrogate signal should be highly correlated to the motion of interest and easy to measure [5].

A surrogate signal with a high temporal resolution will translate into estimated motion with a real-time resolution [5]. A high temporal resolution motion signal will help clinicians during biopsy procedures.

1) Types of Motion Models: Motion models can be divided into direct and indirect motion models. Direct motion models perform an operation directly on the (pre-processed) surrogate signal to estimate the motion of interest [5]. The current study will focus on making a direct motion model.

Indirect motion models have a set of internal parameters which contain information on the state of the ground truth and surrogate signal. When given the internal parameters, some function can make an estimation of the ground truth and surrogate signal [5]. This function can take many forms, such as polynomials or neural networks [5].

The internal parameters will contain the most important features of the surrogate signal if the surrogate signal can be estimated with low loss from the internal parameters. During application of an indirect model, the internal parameters are optimised to minimise the loss between the measured surrogate signal and the surrogate signal estimated from the internal parameters [5]. Since the ground truth is correlated to the surrogate signal, this will also optimise estimation of the ground truth using the internal parameters.

2) Neural Networks: The periodic characteristic of respiratory motion makes recurrent neural networks look promising for potential use in motion models. An example of a recurrent network applied as motion model is a study which combined convolutional and LSTM layers as motion model, with A-mode ultrasound as the surrogate signal [15].

A Simple Recurrent Network (SRN) receives data of past time points in addition to data on the current time point as input [16]. An overview of the inputs and outputs of an SRN node is displayed in Fig. 1.

LSTM networks are more complex recurrent networks than simple recurrent networks. They perform well on datasets with long time dependencies and combat backpropagation problems encountered when using other recurrent networks [16, 17].

LSTM networks make use of a memory cell which has a bidirectional regulation with an input, output and forget gate [16]. An overview of an LSTM node is displayed in Fig. 1. The input gate regulates the input of the memory cell. It is an activation function with as input a weighted summation of the node input, recurrent input and the content of the memory cell of previous time points. A block input is multiplied with the output of the input gate and the result enters the memory cell. The forget gate forms a feedback loop with the memory cell. This allows the forget gate to alter the memory cell based on the input, recurrent input and most notably the contents of the memory cell itself in previous time points. The output gate receives a weighted summation of the node input, recurrent input and the current memory cell as input. The output of the output gate is multiplied with values from the memory cell, forming the output of the node.

TABLE I
AN OVERVIEW OF THE USED A-MODE ULTRASOUND PARAMETERS

Parameter	Value
Voltage	Level 16
Pulse width	2,85 us
Sampling frequency	33,3 MHz
Analog filter	2-6 MHz
Pre-amplifier	+24 dB
Constant gain	20 dB
Delay	10 us
Pulse repetition frequency	30 Hz

III. MATERIALS AND METHODS

First, the materials used in the human subject experiments are explained and the used parameters are displayed. Afterwards, the architecture of the created correspondence model is described.

A. A-mode Ultrasound as Surrogate Signal

A-mode ultrasound data will be used as surrogate signal in the human subject study. An Optel Opbox 2.1 Mini Ultrasonic Box will be used as pulser/receiver [18]. An ultrasound transducer with a centre frequency of 3.5 MHz will be used. A Python GUI made by the Opbox manufacturer will be altered and used to record the surrogate signal over time.

The ultrasound parameters have been optimised by educated trial and error for the best Signal to Noise Ratio (SNR) of the liver peak. The final parameters used can be found in Table I.

B. B-mode Ultrasound as Ground Truth

The ground truth is extracted from B-mode ultrasound recordings. The data is recorded with a transducer with a centre frequency of 4 MHz. The depth is set to contain the entire liver plus approximately two centimetres of play. In practice, the imaging depth ranged from 19 cm to 21 cm. More gain is used at higher depth compared to lower depth. This is done to compensate for increased attenuation at higher

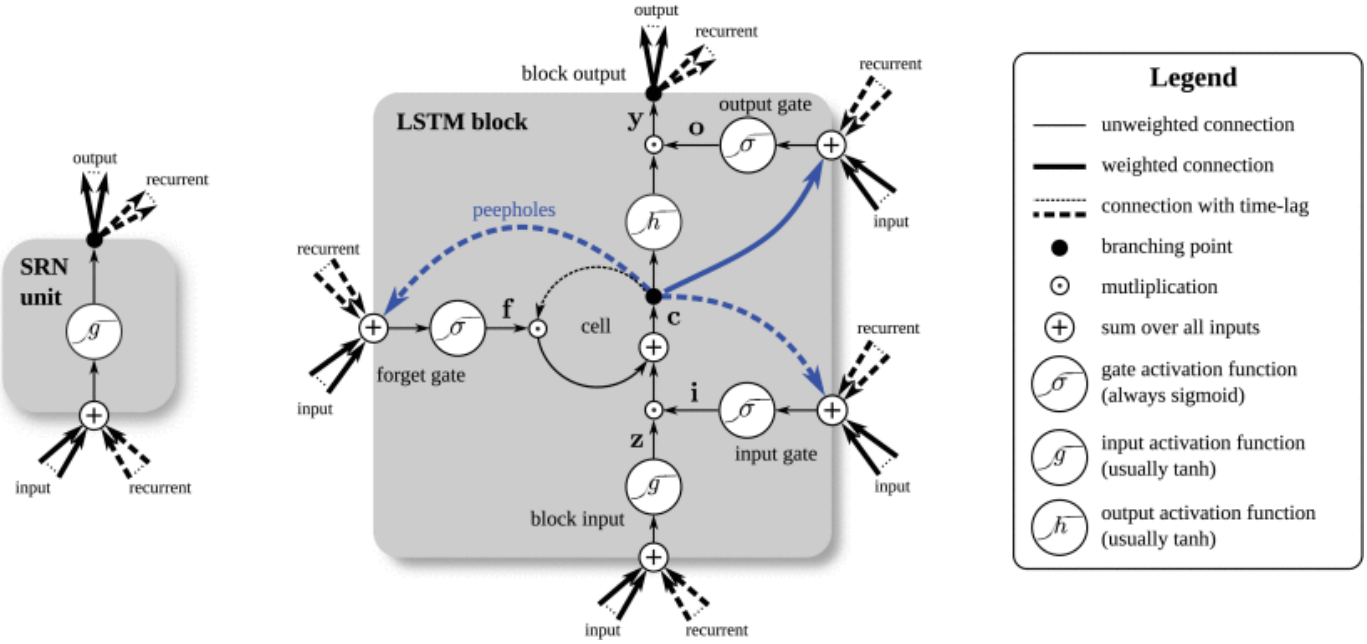

Fig. 1. A detailed schematic of a Simple Recurrent Network (SRN) unit compared to an LSTM block. The figure is created by Greff *et al* [16].

imaging depths. The recordings are made with a frequency of 13 or 15 Hz. Two assumptions have been made in the current research. Firstly, the ground truth is assumed to be accurate and valid. If there was suspicion of an invalid ground truth, the liver displacement was extracted from the B-mode ultrasound recordings with a different method. Secondly, the interference of A-mode and B-mode ultrasound within the body is assumed to be negligible. The degree of interference has been investigated by observing the A-mode ultrasound signal with and without the B-mode transducer on the body. No visual difference has been observed, but it should nevertheless be noted as an assumption.

C. LSTM Correspondence Model

Two neural networks have been made with differing complexity to investigate which model can estimate the ground truth with the lowest Mean Absolute Error (MAE). LSTM networks will be used as correspondence model. Respiratory motion sequences contain long time dependencies due to the periodic nature of the movement. LSTM networks are well suited for sequences with long time dependencies [16]. The created models will be trained, validated and tested on data measured from the same subject.

1) Model architecture: The input sequence of both models contains a window of data which consists of the most recent time point and the previous 29 time points. This means at 15 Hz, the model inputs consist of data of the last 2 seconds. In practice, noise overpowered the overall trend of the signal if the window size was lower than two seconds. A window larger than 2 seconds resulted in reduced performance due to the lesser amount of training windows created.

The output sequence of both models contain a ground truth estimation for every corresponding input sequence time point. The models are trained in batches of size 30 and uses an L1 loss function with an AdamW optimizer. Both models will be trained for a maximum of 100 epochs, after which the model of the epoch with the lowest loss will be used for testing. The learning rate is optimised every time a model is trained. On datasets with a high quality surrogate signal, data was split into 60% for training, 20% for validation and 20% for testing. On datasets where lower quality surrogate signal parts have to be cropped away, this split is approximated while dividing the high quality surrogate signal parts over training, validation and testing.

The most recent time point of the output window displays the estimated ground truth in the present. In a real time scenario, the accuracy of the estimated ground truth in the present would matter more than estimations in the past. To investigate this, real time plots will be made using the last time point of every output window.

As stated earlier, two neural networks have been made with differing complexity. The more shallow LSTM neural network made consists of four layers, and is visible in Fig. 2. Firstly, there are two LSTM layers which consist of 64 nodes, with a dropout of 0.15 after every LSTM layer. The last LSTM layer generates an output for every time point. The output for the last (most recent) timepoint in the window is inserted into a ReLu layer. Lastly, the data passes through a dense layer which estimates a window of the current and some past ground truth values.

The second model investigated is more complex. This model is similar to the shallow model aside from the amount of LSTM layers. The deep model has a total of four LSTM

layers. The first two LSTM layers consist of 64 nodes, and the second two consist of 32 nodes. All LSTM layers have a dropout of 0.15. The last timepoint in the output window of the LSTM layers is inserted into a ReLu layer. Lastly, the data passes through a dense layer which generates the output window.

D. Model performance metrics

The performance of the models will be measured with multiple metrics. These metrics will be used to evaluate all models created for use on either human or synthetic data. For the synthetic data and human subject experiments, the MAER is of most interest. The goal of this study is to examine the performance of the model in a real time scenario, which is best reflected by the MAER.

Firstly, the Mean Absolute Error of the Windows created by the models is abbreviated as MAEW. This is the MAE of the output window of the model, so this error is optimised during training.

Secondly, the Mean Absolute Error for the Real time plot is abbreviated as MAER. This is the MAE of the real time plot made with the last time point of every window. The MAER gives an indication for the expected performance in a real time scenario.

Lastly, the Mean Absolute Error Standard Deviation is abbreviated as MAERSTD. This is the standard deviation of the MAER. A high MAERSTD indicates a high degree of dispersion in the data.

IV. Validation

A. Simulated Experiments

To validate whether the created correspondence model is able to correlate a perfectly synchronised and well correlated A-mode ultrasound signal to the liver displacement, synthetic data has been created. The synthetic data will also help determine which model should be applied on the human subject data.

1) Synthetic Ground Truth: The synthetic ground truth data is made to resemble the liver displacement with the same type and degree of noise. The signal consists of few sinusoidal signals with different frequencies summed with random noise. The final ground truth signal can be seen in orange in subfigure b. and c. of Fig. 4.

2) Synthetic Surrogate Signal: The synthetic surrogate signal is made to resemble the unprocessed A-mode ultrasound data with the same types of noise. For every time point, the synthetic surrogate has a graph of data which contains three peaks and random noise. Of the three peaks, one moves in the same frequency as the ground truth. The two other peaks move at a different frequency and can be thought of as different signals like the movement caused by the heartbeats. The final synthetic surrogate signal created can be seen in subfigure a of Fig. 4. The SNR of this data can be improved with several processing steps like a low pass filter on the depth dimension as done on human subject data explained in section IV-B2. However the goal of the synthetic data is to test whether the model performs in a highly correlated but noisy environment, so further preprocessing would defeat the purpose.

B. Clinical Study with Human Subjects

To validate whether the created correspondence model is applicable to a human subject scenario, the ground truth and surrogate signal were recorded in three human subjects. The study was evaluated by an ethics committee, and the subjects signed a consent form.

The A-mode ultrasound transducer is placed at a specific place on the abdomen, which is further explored in section A. The subject is instructed to perform a breath hold of 10 seconds, followed by 60 seconds of breathing and a second breath hold of 10 seconds.

To avoid motion artefacts, a belt was made to hold the A-mode transducer. This belt was only used on subject 1, as in some situations the transducer lost contact with the skin using the belt. The B-mode ultrasound transducer is held by the examiner and should be kept as stationary as possible. To avoid artefacts, the examiner should note sufficient ultrasonic gel is applied. The subject is coached to breathe so the surrogate signal quality is maximised.

1) Ground Truth Data: To extract SI liver displacement from ultrasound images, several processing steps are applied. The B-mode ultrasound images are cropped to the region of interest and contrast enhanced by a MATLAB script written by a previous researcher at RaM. Segmentation is done to find the liver border. After segmentation, the average horizontal displacement of the largest connected area is used as liver displacement for that time point. Lastly, the ground truth is smoothed and the breath holds are cropped away, leaving only the normal breathing data.

2) Surrogate Data: To increase the SNR of the surrogate signal, several processing steps are applied. An overview of the processing steps and the data after each processing step can be seen in Fig. 3.

Firstly, the breath holds are cropped away. This step is very important for synchronisation, and must be done accurately for maximum motion model performance. If the temporal resolution is low and the model performance is poor, the cause is likely synchronisation. In such a case, this step is repeated multiple times by shifting the region of interest to find a suitable crop. Afterwards, the data is interpolated over the temporal axis to match the sample frequency of the B-mode ultrasound by fitting to a cubic spline.

Secondly, the data is cropped along the distance axis to contain only distances where peaks can be seen moving with respiratory motion.

Thirdly, the absolute value is taken to make Fourier filtering possible. The data is first filtered over the time dimension, and afterwards over the depth dimension. Both filters are subject specific low pass filters.

Fourthly, normalisation is applied and every value below a subject specific threshold is set to 0. The threshold is optimised for every subject individually and ranges from 0.15 to 0.25. Finally, for every time point a weighted middle point

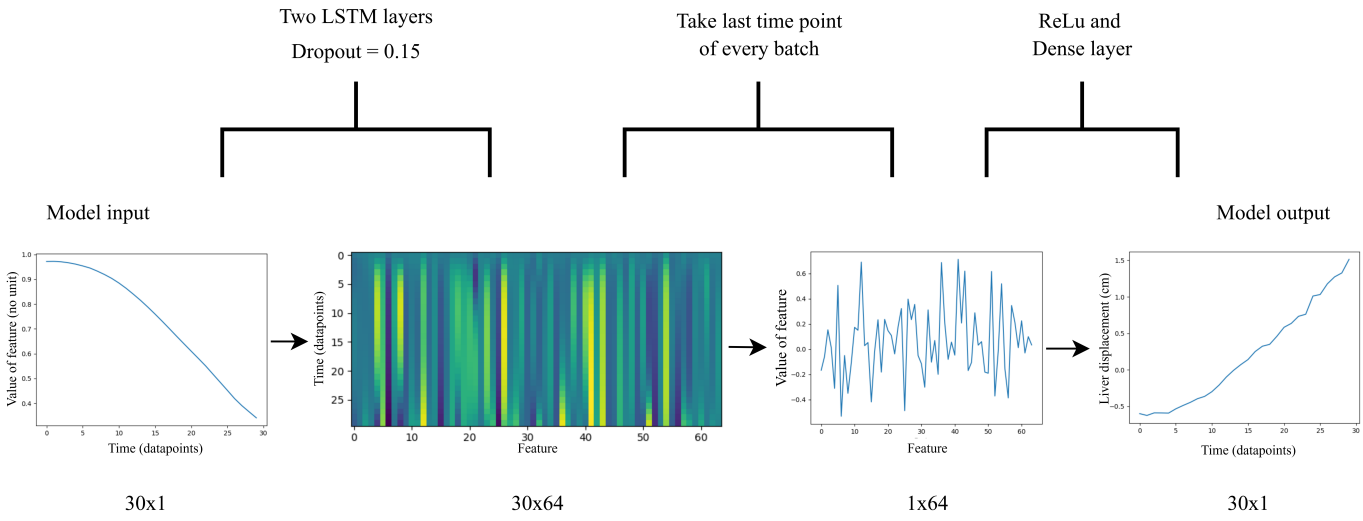

Fig. 2. An overview of the architecture of the shallow model created. From left to right the model input, intermediate features of the model and final output are seen. Dimensions are visible below the plot, and are time x feature in that order.

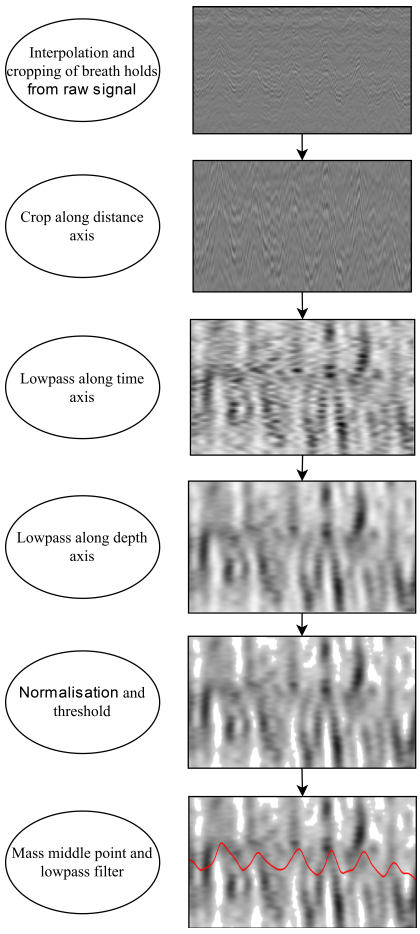

Fig. 3. An overview of the architecture of the shallow model created. Dimensions are batch x time x feature in that order.

is calculated. This is done by taking the depth of every pixel, multiplied by the intensity of that pixel and divided by the total intensity of all pixels. This weighted middle point is filtered by another low pass filter and normalised again.

From here on, two pathways can be followed to acquire the final surrogate signal. The first pathway is for recorded data with a high correlation to the ground truth. The dataset is divided into training, validation and testing parts. As many overlapping windows as possible with a size of 30 are created. The training, validation and testing parts stay separate, so no single window spans two different datasets.

The second pathway is for recorded data with only few breathing cycles clearly correlated to the ground truth data. The data is temporally cropped to contain only the part which clearly shows a liver peak moving with respiration. The cropped regions are split into training, validation and testing segments. Every segment is windowed the same way as mentioned in the other pathway, with no window spanning two segments to preserve continuity. These windows form the input to the correspondence model.

V. Experimental Results

A. Synthetic data

The output and ground truth of the shallow and deep model on the created synthetic data are displayed in Fig. 4. The estimations mostly follow the ground truth, however at some points the error is clearly visible. In Table II the performance metrics are displayed. The MAER of the shallow and deep model are 0.48 cm and 0.59 cm, respectively. The MAERSTD is larger than the MAER for both models, suggesting a large spread in errors. All performance metrics suggest the shallow model performs better. Therefore the shallow motion model will be used to correspond the human subject surrogate signal to the ground truth.

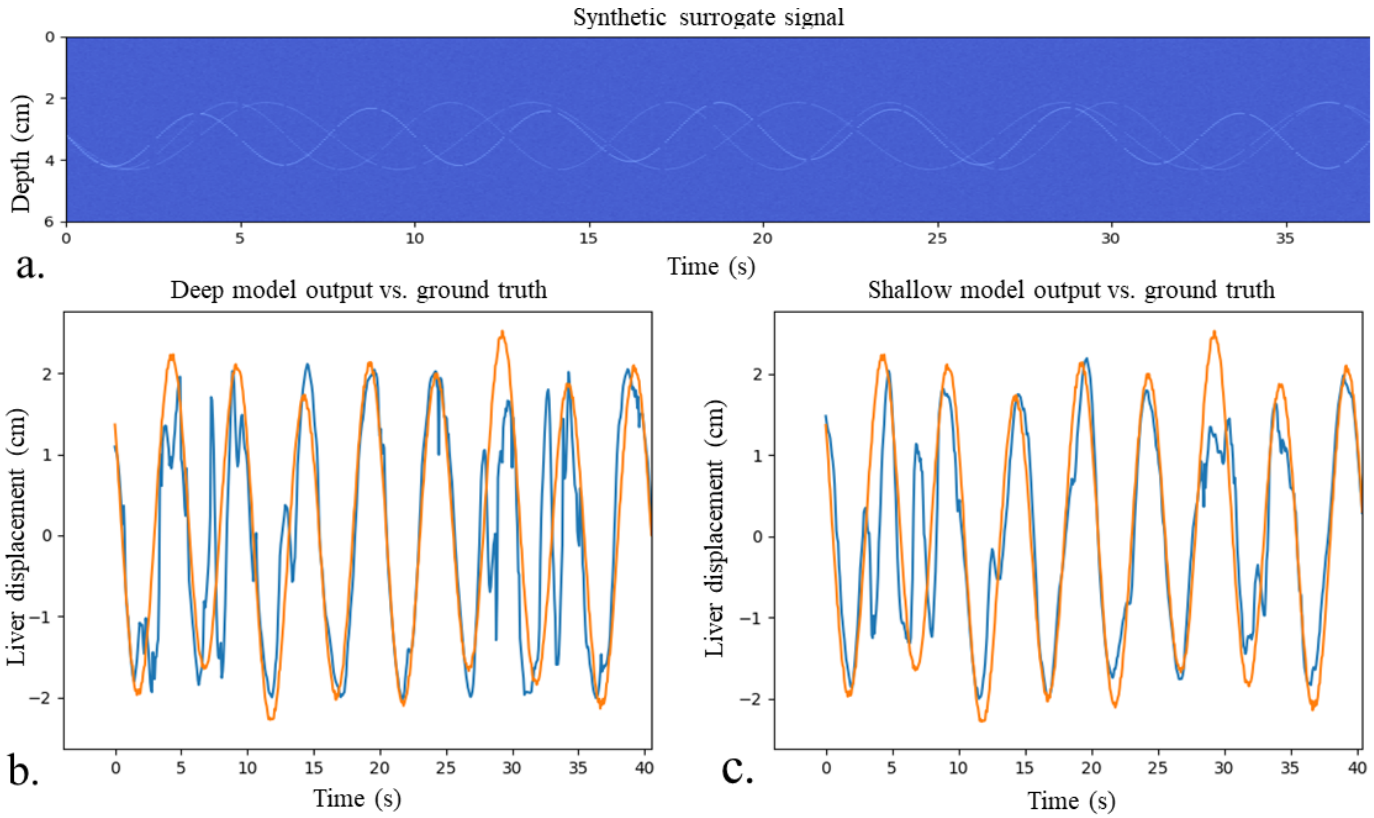

Fig. 4. a. The first 40 seconds of the created synthetic data used for testing. b. A plot of the deep model test performance on synthetic data. The synthetic ground truth is visible in orange with the corresponding model output visible in blue. c. A plot of the shallow model test performance on synthetic data. The synthetic ground truth is visible in orange with the corresponding model output visible in blue.

TABLE II
THE PERFORMANCE METRICS FOR THE DEEP AND SHALLOW MODEL ON SYNTHETIC DATA

Model size	MAEW (cm)	MAER (cm)	MAERSTD (cm)
Shallow	0.45	0.48	0.53
Deep	0.54	0.59	0.61

TABLE III
THE PERFORMANCE METRICS FOR THE SHALLOW MODEL ON DATA FROM HUMAN SUBJECTS

Subject	MAEW (cm)	MAER (cm)	MAERSTD (cm)
1	0.89	0.83	0.68
2	0.18	0.18	0.17
3	0.44	0.54	0.30
Overall	0.50 ± 0.36	0.52 ± 0.33	0.38 ± 0.27

B. Human subject study

The shallow model is applied on datasets acquired from three human subjects because it performed the best on synthetic data. The performance metrics are seen in Table III. The ground truth of subject 1 and 2 were recorded at 15 Hz with an imaging depth of 19 cm, with 13 Hz and 21 cm used for subject 3.

The processed surrogate signal and model output compared to the ground truth of subject 2 is displayed in Fig. 5.

The model performance differs significantly per subject. The average MAER of all subjects is $0.52 \pm 0.33cm$, with an average MAERSTD of 0.38 ± 0.27. The lowest MAER achieved is 0.18 cm from subject 2.

The surrogate signal of subject 1 and 3 contained large time intervals with low correlation to respiratory motion. Analysing these subjects like normal results in an MAER larger than 1 cm. For these subjects, the pathway for low quality surrogate signals explained in section IV-B2 was followed. The surrogate signal for subject 2 did not contain uncorrelated time intervals. Because of this, it was possible to use the pathway for high quality surrogate signals explained in section IV-B2 for subject 2.

VI. DISCUSSION

A. Discussion of results

The MAER of the shallow and deep model on synthetic data are 0.48 cm and 0.59 cm. This is a large error when considering the data is perfectly synchronised and perfectly correlated ignoring the noisy time points. During testing, it was found that a synthetic surrogate signal which was less complex had an MAER in the order of millimetres. In the used surrogate signal, some time points have been replaced

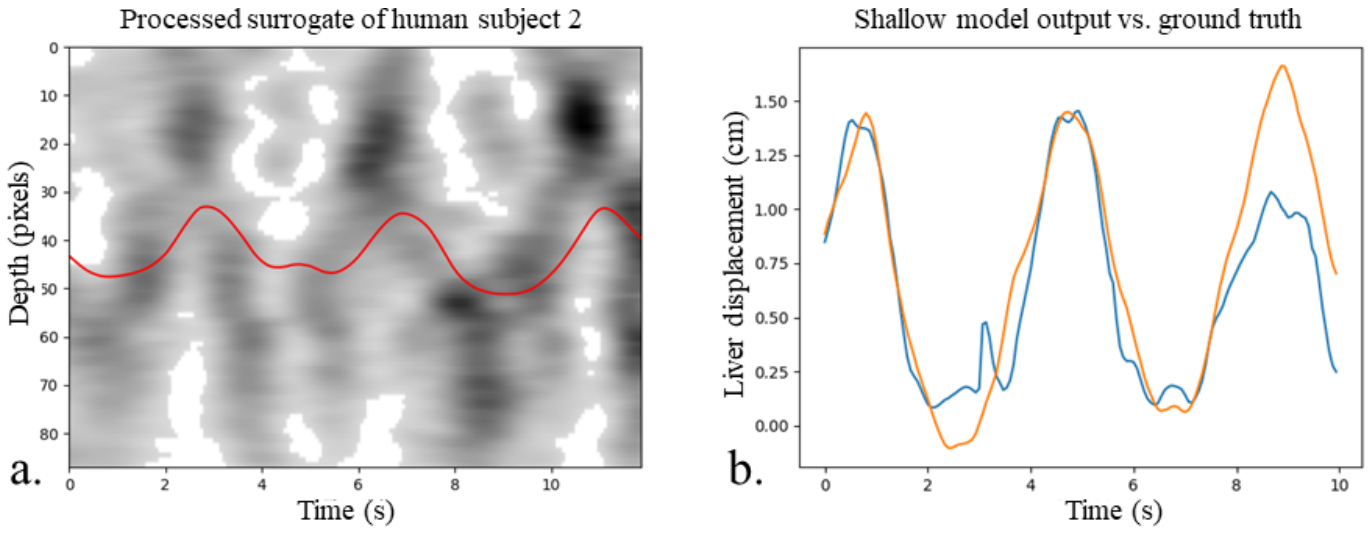

Fig. 5. a. A plot of the processed human subject surrogate signal used for testing. The filtered ultrasound data is visible in the background as grayscale image with the red line representing the weighted middle point of every time point filtered by a low pass filter. b. A real-time test plot consisting of the last value of every window produced by the model. The testing domain of the ground truth is visible in orange with the corresponding model output in blue.

by white noise. While a recurrent network should be able to look at past time points which do not consist of noise, this has introduced some 'nonsense' time points in the training data. This likely is the cause of the high MAER on the synthetic data.

In a real-time clinical scenario on human subjects, a biopsy would give a false negative if the absolute error of the estimated liver displacement is smaller than half the diameter of a tumour. The smallest tumour for which biopsy is recommended during diagnosis has a diameter of 1 cm [14]. The estimated true positive chance with the shallow model for a tumour with a radius of 1 cm is showed in Table IV. The true positive rates are 48 %, 95 % and 59 % for subject 1, 2 and 3 respectively. For this analysis, the tumour is assumed spherical and inaccuracies from other sources are neglected. A study by Durand et al found percutaneous liver biopsies to have a true positive rate of 89% and a false negative rate of 10% [19]. The false negative group had a tumour size of $0.6 \pm 0.3cm$, while the majority of the subjects had a tumour bigger than 3 cm. The current study fails to reliably have a better true positive rate than the conventional biopsy methodology on tumours with a diameter of 0.6 cm. The estimated true positive rate for subject 2 is the highest, at 79%. This suggests the current methodology will need improvement before clinical testing. Below, suggestions are made for future research to more consistently achieve high model performance.

As seen in Table IV, the average displacement between consecutive ground truth maxima and minima is 5.4, 1.3 and 2.1 cm for subject 1, 2 and 3 respectively. Subject 1 also has the largest MAER and subject 2 the smallest MAER. This suggests a positive correlation between ground truth amplitude and MAER. Due to this, the performance of the model should partly be interpreted in combination with the amplitude of the ground truth. However due to the small sample size, no definitive conclusions can be drawn for the relationship between MAER and ground truth amplitude.

B. Discussion of surrogate signal

The quality of the surrogate signal likely has had a negative impact on the performance of the created motion models. The SNR of the border of the liver was visibly lower on A-mode imaging compared to B-mode imaging. Besides that, the surrogate shows movements not seen in the ground truth. This diminishes the correlation between the surrogate and ground truth, leading to reduced model performance.

To keep the surrogate correlated to the ground truth, the datasets of subjects 1 and 3 had to be cropped temporally to the regions which showed a clear liver peak moving with respiration. Consequently, both subjects only had 2 full breathing cycles of training data. The data recorded from subject 2 had higher quality and could be used without cropping breathing data. The model performance on subject 2 is likely the best due to the higher amount of training data left after cropping.

The poor surrogate signal quality also resulted in a low temporal accuracy of the start and end breath holds. This caused some problems with synchronisation. However, synchronisation using the peaks and valleys of the high quality breathing cycles was found to be accurate enough.

No transducer placement has been found which captures a surrogate signal with sufficient quality for every human subject. Optimisation of the used parameters improved the visibility of the liver peak to a small extent.

To reduce motion artefacts in the ultrasound recording, a belt was made to hold the A-mode ultrasound transducer stationary. However the belt was found to apply uneven pressure, causing contact loss during some part of the breathing cycle for some subjects. For this reason the ultrasound belt has only been used on subject 1, with the transducer being held by the examiner for subject 2 and 3.

Two processing steps have not been implemented in the current study, but are thought to further increase the clarity of respiration in the fully processed surrogate signal. These processing steps should be considered in future research.

Firstly, the threshold step which results in the grayscale image seen in Fig. 5 could be replaced by a more complex masking. The masking should set all low intensity zones which are not contributing to a sinus-like trend to zero.

Secondly, there are moments where the ground truth observes a change in displacement while the surrogate signal does not. In these moments some parts of the breathing cycle will resemble the data recorded moments before. This problem will be referred to as self-resemblance. As an example, an extreme scenario of self-resemblance with synthetic data not used in this study has been plotted in Fig 6. When using this self-resembling signal as input to some motion model, half of all ground truth phases will correspond to the input of 0, likely resulting in reduced performance. Fitting the surrogate data to a less self-resembling curve will more accurately contain information about the phase of breathing. Note that the respiratory frequency is not constant, unlike what is assumed in the synthetic example given. The human subject data could not be fit to a simple sinusoid with this method due to the changing respiratory frequency.

Aside from the mentioned processing steps, future work might focus on processing steps applied in B-mode ultrasound to implement them on the raw A-mode ultrasound signal. This is thought to help the liver peak SNR due to the observed quality difference between A-mode and B-mode ultrasound. Though it will not solve the underlying problem, recording more data or data augmentation should also be considered.

C. Discussion of the motion model

The created motion model optimises the MAEW. In future work, the motion model should be changed to optimise MAER. This performance metric is more relevant to a real-time scenario.

In the current study, the motion model is trained and tested on datasets from the same subject. The motion model performance is not tested on a different subject than was trained on. A general model applicable to multiple subjects would have large merit in a clinical situation. Such a model would need to be trained on a big dataset and the surrogate would need to be recorded from the same position every time. After training, the model can be applied to multiple subjects without the need for B-mode ultrasound at all. This would take away the B-mode ultrasound costs and increase accessibility. The inter subject performance of the made motion model is expected to be low due to the small amount of training data and high variability of breathing between subjects.

VII. Conclusion

The current study has successfully created two neural network motion models with LSTM layers which estimate the liver position using A-mode ultrasound as surrogate signal.

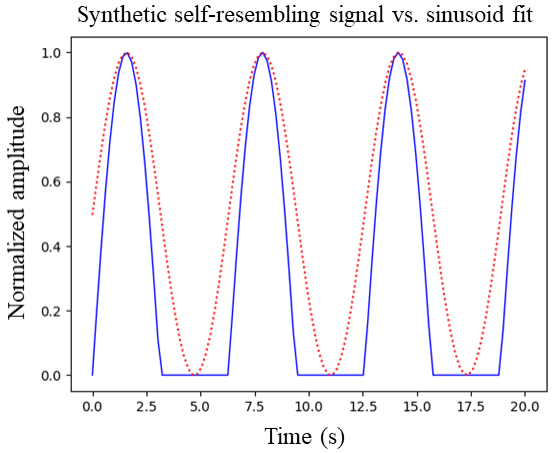

Fig. 6. An example of periodic data with self resemblance in blue. A sinusoidal signal fitted to this signal is seen in red

TABLE IV
The Estimated True Positive (ETP) chance of a biopsy procedure for a given tumour diameter and average ground truth amplitude of the human subjects

Subject	1.0 cm ETP (%)	0.6 cm ETP (%)	Ground truth amplitude (cm)
1	48	26	5.4
2	95	79	1.3
3	59	18	2.1
Overall	67 ± 25	41 ± 27	2.9 ± 2.2

A synthetic dataset was made to validate the model performance on perfectly synchronised, well correlated data. On this dataset, the shallow model and the deep model had an MAER of 0.48 cm and 0.59 cm respectively.

The more shallow model performed better on the synthetic dataset, so this model was applied on human subject data. For human subject 1, 2 and 3, an MAER of 0.83 cm, 0.18 cm and 0.54 cm was found respectively. Model performance was found to differ significantly per human subject. Differences in surrogate signal quality are thought to be a plausible explanation for the divergent model performance. An estimation of the true positive rate in a situation using the created shallow model reveals the methodology still needs improvement before usage in a clinical situation.

Appendix

The following section is meant to clarify A-mode ultrasound usage for future researchers. In the current study, the OpBox 2.1 one dimensional ultrasound was found to record very noisy data. For future researchers, several points could be of importance to work with and improve on the current methodology. These will be explained below.

The A-mode ultrasound connects via USB to the laptop that is used. To make a suitable connection, a driver from the manufacturer is installed. Python is chosen from a subsection of OpBox 2.1 compatible languages to analyse the data. Access to the data in real-time is important for finding a suitable transducer placement. To achieve this, the GUI provided by

the manufacturer, which already showed data in real-time, was edited to record data over time as well.

When aiming to take A-mode ultrasound measurements, the first thing to be done is to find a suitable placement for the ultrasound transducer. This will be done within a field of interest on the abdomen starting at the ribs and stops at 2 cm down from the ribs. The transducer is placed at a position within the field of interest and the real-time plotted graph is observed to see whether the liver peak is present, moves with respiration and does not disappear during part of the respiratory cycle. If the signal is inadequate, this process is repeated for different positions until it is. Holding the transducer at a slight angle to image a bigger part of the liver located under the ribs is found to help significantly for some subjects. Research has been done to find a transducer placement which always results in an adequate signal, but this placement has not been found. Take care to use enough ultrasonic gel for all measurements done, as otherwise the SNR will diminish greatly.

To minimise motion artefacts on the A-mode ultrasound, a belt has been made which attaches to a 3d printed transducer attachment made by a previous researcher [9]. However, the transducer was found to lose contact with the skin during some part of the breathing cycle for some subjects. Consequently, the belt is only used for subject 1.

At first, the ultrasound parameters used in a similar research by a previous researcher were used [9]. These parameters did not produce the desired result, as the liver peak had a very low SNR and would regularly disappear completely during a part of the breathing cycle. To resolve this issue, all relevant ultrasound parameters were changed systematically to see the effect of every individual parameter. The parameter values seen in Table I were chosen because they resulted in the best SNR of the liver peak. All surrogate measurements were done with a 3.5 MHz transducer.

These changes increased the SNR enough to reveal the liver peak in most subjects. However, the liver peak would still disappear completely at some period in the respiratory cycle for most subjects. It has not yet been possible to define a proper protocol which always records a high quality surrogate signal. The problems seems to be very subject specific.

REFERENCES

[1] Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249, 2021.

[2] Jordi Bruix and Morris Sherman. Management of hepatocellular carcinoma: An update. *Hepatology*, 53(3):1020–1022, 2011.

[3] European Association for the Study of the Liver (Josep M. Llovet) and European Organisation for Research and Treatment of Cancer (Michel Ducreux). Easl–eortc clinical practice guidelines: Management of hepatocellular carcinoma. *Journal of Hepatology*, 56(4):908–943, 2012.

[4] Valeer Desmet and Johan Fevery. Liver biopsy. *Baillière's Clinical Gastroenterology*, 9(4):811–828, 1995. Investigations in Hepatology.

[5] Jamie R. McClelland. Respiratory motion models: A review. *Medical Image Analysis*, 17(1):19–42, 2013.

[6] Kurt Von Siebenthal, Gábor Székely, U Gamper, Peter Boesiger, A Lomax, and Philippe Cattin. 4d mr imaging of respiratory organ motion its variability. *Physics in medicine and biology*, 52:1547–64, February 2007.

[7] Françoise Jeanette Siepel, Bogdan Mihai Maris, Marcel Klaas Welleweerd, Vincent Groenhuis, Paolo Fiorini, and Stefano Stramigioli. Needle and biopsy robots: a review. *Current Robotics Reports*, March 2021. Springer deal.

[8] Shamel Fahmi. Respiratory motion estimation of the liver with abdominal motion as a surrogate : a supervised learning approach, August 2017.

[9] Gerard W. Veenstra. Generating high frame rate mr images using surrogate signals, August 2019.

[10] Yuwen Wu, Zhisen Wang, Yuyi Chu, Renyuan Peng, Haoran Peng, Hongbo Yang, Kai Guo, and Juzhong Zhang. Current research status of respiratory motion for thorax and abdominal treatment: A systematic review. *Biomimetics*, 9(3), 2024.

[11] Franz Wudy, Christoph Stock, and Heiner J. Gores. Measurement methods — electrochemical: Quartz microbalance. In Jürgen Garche, editor, *Encyclopedia of Electrochemical Power Sources*, pages 660–672. Elsevier, Amsterdam, 2009.

[12] Timothy G. Leighton. What is ultrasound? *Progress in Biophysics and Molecular Biology*, 93(1):3–83, 2007. Effects of ultrasound and infrasound relevant to human health.

[13] Dale R. Wagner, Masaru Teramoto, Trenton Judd, Joshua Gordon, Casey McPherson, and Adrianna Robison. Comparison of a-mode and b-mode ultrasound for measurement of subcutaneous fat. *Ultrasound in Medicine & Biology*, 46(4):944–951, 2020.

[14] Josep M. Llovet, Josep Fuster, and Jordi Bruix. The barcelona approach: Diagnosis, staging, and treatment of hepatocellular carcinoma. *Liver Transplantation*, 10(S2):S115–S120, 2004.

[15] Frank Preiswerk, Cheng-Chieh Cheng, Jie Luo, and Bruno Madore. *Synthesizing Dynamic MRI Using Long-Term Recurrent Convolutional Networks*, pages 89–97. Springer International Publishing, September 2018.

[16] Klaus Greff, Rupesh K. Srivastava, Jan Koutník, Bas R. Steunebrink, and Jürgen Schmidhuber. Lstm: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10):2222–2232, 2017.

[17] Fakultit Informatik, Yoshua Bengio, Paolo Frasconi, and Jfirgen Schmidhuber. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. *A Field Guide to Dynamical Recurrent Neural Networks*, March 2003.

[18] Optel. Opbox ver 2.1 mini ultrasonic box with integrated pulser and receiver, 1989. Accessed on Date April 9, 2024.

[19] François Durand, Jean Marc Regimbeau, Jacques Belghiti, Alain Sauvanet, Valérie Vilgrain, Benoît Terris, Vincent Moutardier, Olivier Farges, and Dominique Valla. Assessment of the benefits and risks of percutaneous biopsy before surgical resection of hepatocellular carcinoma. *Journal of Hepatology*, 35(2):254–258, 2001.