Spatio-Temporal Crossmodality Learning from Sparse Sensing Data: A Case Study on Microwave Ablation

Zhihua Wang*, Zihang Lai*, Stefano Rosa, Allen Chi-Shing Yu, Linhai Xie, Hanbin Ma, Andrew Markham

Abstract—**Many physical phenomena, such as temperature, pressure, magnetic potential etc., exist as spatially and temporally varying fields. These can be modelled by non-linear partial differential equations and numerically solved using techniques such as Finite Element Modelling (FEM). However, a key challenge in physics modelling is to be able to invert the problem and infer the field across space and time given only spatially sparse readings. This challenge is compounded by imprecise knowledge about the exact properties of the physical world and complex, time-varying cross-modal interactions of physics coupling e.g. the coupling of electric fields and thermal diffusion. In this work, we tackle this problem by embedding the physical simulator as the forward model with a recurrent, generative network that is conditioned on observed and predicted sensor readings. We further deploy this method in modelling and predicting the destruction of a tumour using microwave ablation. Using sparse temperature measurements, we demonstrate that not only can we accurately model the instantaneous tissue damage in the face of time-varying material properties, we can also predict the future damage spread.**

Index Terms—**Deep learning, generative adversarial networks, crossmodal perception, spatio-temporal prediction, bioheat transfer**

I. INTRODUCTION

A Large number of properties of the physical world, such as temperature or pressure, can be represented as a function of spatial location and time e.g. $f(x, y, z, t)$, which characterises how the scalar field varies corresponding to the physical properties. These can be accurately modelled in the form of partial differential equations (PDEs) which can be solved, subject to knowledge of initial conditions and material properties. Due to the complexity of deriving an analytical solution, in practice, this is typically done numerically, through the use of solvers such as Finite Element Modelling (FEM) simulation. These are captured by multiphysics simulator, which combine different physical properties and models e.g. temperature and Joule heating. FEM can be regarded as a *forward* model which can predict the future temporal evolution of the field(s) of interest, as shown in Figure 1(a), where the
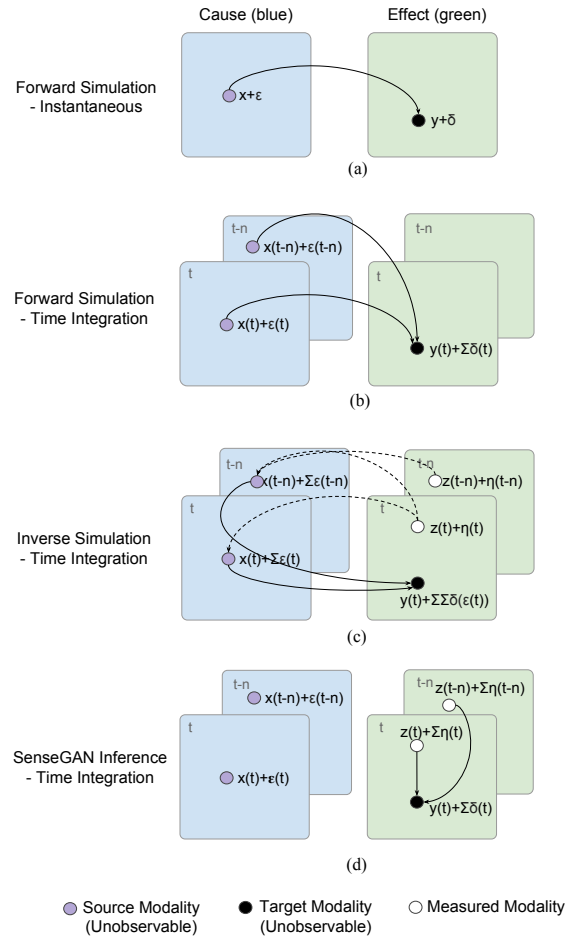
Z. Wang, Z. Lai, S. Rosa, L. Xie and A. Markham are with the Department of Computer Science, University of Oxford, Oxford, United Kingdom, e-mail: zhihua.wang@cs.ox.ac.uk.

A. Yu is with the School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China

H. Ma is with the CAS Key Laboratory of Bio-Medical Diagnostics, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, China

*These two authors contributed equally

Fig. 1. Solving a multiphysics coupling task with simulation and with the proposed SenseGAN, in order to infer an unobservable variable y. Left side (blue) represents the causes in a physical process, whereas right side (green) represents effects. (a) forward simulation, (b) time dependent forward simulation, (c) time dependent inverse simulation, (d) time depentdent SenseGAN as proposed method

cause represents initial conditions and *effect* represents the desired fields results. The error ϵ in the source x leads to an error δ in the target y. In time dependent applications shown in Figure 1(b), error $\Sigma\delta(t)$ accumulates over time. However, the inverse problem, that of predicting the coupling fields across space and into the future, given a few sensor readings,
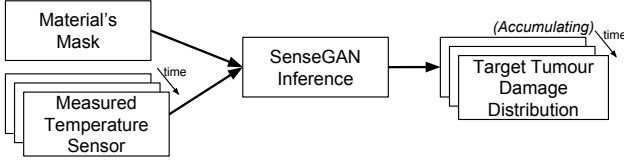
Fig. 2. SenseGAN for tumor damage estimation based on temperature sensing data in the process of microwave ablation.
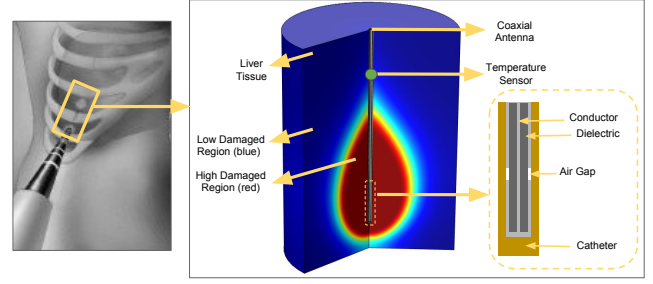

Fig. 3. Microwave ablation of a tumour. A thin catheter is percutaneously inserted into the tumour and an electrical current injected, inducing localized heating which diffuses across the tumour, causing cell death. The key is to kill the tumour cells, without damaging healthy tissues, which requires accurate prediction of the ablated region. The false colours show damage likelihood: red areas are damaged, blue areas are not affected, areas in between are partially damaged.

is an open challenge, especially in cross-modal applications, as shown in Figure 1(c). When source x is unobservable, an observation of z is used to determine its causal factor x thence target y can then be inferred. However, this process is often too sophisticated to solve in real world, and also leading to error accumulation due to double integration. We propose a machine learning method to solve this sophisticated inverse physics problem namely SenseGAN. As shown in Figure 1(d), a neural network learns the correlation between measured modality and target modality, eliminating the accumulation of errors.

In a wide range of applications, the ability to infer the spatio-temporal distribution of a scalar field, given a small sub-set of measurements is of great importance. A technique which is widely used in statistics for example is that of Kriging Interpolation, which is effectively a form of Gaussian Process Regression, providing an best unbiased linear predictor of the interpolated field, given a set of measurements. However, it makes assumptions about the statistical moments of the random field e.g. that the first moment (mean) is homogeneous or there is strong, pre-defined correlation between second order moments (variance).

In this research, we tackle the challenging biomedical task of microwave ablation (MWA) of tumours in various organs, where an accurate inference of destructed target tumour fraction based on a cross-modal sensor field is achieved, as shown in Figure 2. As a minimally invasive technique, percutaneous MWA involves the insertion of a thin catheter into a tumour. The catheter has an electrically conductive tip, which is coupled to a powerful signal generator. High-frequency radio fields are injected into the catheter, where the injected current causes Joule heating in the immediate vicinity of the tip, and conductive heating diffuses through the tumour and surrounding tissue [1], detailed setups are illustrated in Figure 3. The challenge is to necrotise the tumour through hyperthermal coagulation, without damaging the surrounding tissue. In our experiments, we use machine learning method to determine the fraction of tumour destruction intra-operatively in real time, as well as prediction of future treatment outcome, as doctor's aiding tools during treatment. Compared with conventional method like physics modelling based on Arrhenius heat equation [2], we instead establish a novel model namely SenseGAN, which takes sensor reading to predict destruction. More accurate results are obtained compared with physics modelling due to its ability to eliminate errors, including antenna mismatch interference and time-varying material properties.

We use FEM simulation as a tool to validate our proposed method in this research. A pre-operative simulation of the ablation process is performed with the aid of FEM, using *ex-vivo* characteristics of the tissue material properties (both electrical and thermal) [3] and a 3-D model of the liver tumour obtained using CT or MRI scans as an input to the FEM model. With the aid of this model, clinicians can create a schedule for the timing, intensity and duration of the electrical current that needs to be applied. However, there are a number of limitations of this forward simulation method. One factor is that the actual physical properties of the media are not known precisely enough in advance. For example, cirrhotic livers have an electrical impedance that is almost twice as high as a healthy liver [4], yet this cannot be determined pre-operatively in a non-invasive way. Moreover, complex multi-physics effects occur during treatment, altering the thermal and electrical properties of the tissues themselves. As a result, the tissue around the catheter tip becomes dehydrated, greatly increasing its electrical impedance [5]. Although FEM simulations provide a valuable baseline for treatment, the inability to assess the actual damage in real-time is a key limitation for MWA [6].

In order to address this issue, we propose to exploit the strengths of FEM as a forward simulator and combine it with machine learning method of a variational conditional generative adversarial network (GAN), which improved the accuracy compared with conventional FEM simulation. This is achieved by the sparse observation of sensor measurements and cross-modality inference. Several thermistors along the shaft of the catheter provide sparse spatial measurements about the distribution of thermal energy, as shown in Figure 3. The proposed cross-modal modelling framework, SenseGAN, is able to reconstruct the unobservable damage, by supplying thermal measurements. In this example, we demonstrate how sampling from eight 1-D thermistors linearly spread along the catheter can provide diagnostic information about the 2D radius of damage. In particular, we combine a variational conditional GAN with a recurrent network to predict not only across space but also across time. Despite this research

is applied to microwave ablation, this generic framework SenseGAN can also be used in other similar tasks of cross-modal spatio-temporal prediction.

II. RELATED WORK

A. Cross-modal Learning

Learning a correlation between sensor modalities presents considerable challenges, because of the different distributions and representations of the data. This is often known as the heterogeneity gap. Cross-modal learning methods aim to learn a common representation for correlating the heterogeneous data. This is especially the case when some modalities are easily observable/measurable, and others are difficult to directly observe. Traditional methods learn linear projections of the features of different modalities into a common space. For instance, *Canonical Correlation Analysis* (CCA) and similar methods aim to optimise the statistical data distribution by performing statistical correlation analysis [7]. *Cross-modal Factor Analysis* (CFA) minimises the Frobenius norm between data from different modalities. Other methods use graph regularisation for correlating the data from different modalities in a common space [8]. Deep neural networks have been applied to the cross-modal learning problem [9]–[13]. However, most applications focused on learning cross-modal representations in the context of computer vision and in particular image-retrieval applications, rather than cross-modal of physics fields.

B. Deep Neural Networks

Long Short-Term Memory networks (LSTM) were introduced [14] to model temporal behaviour. As a modified version of recurrent neural networks (RNN), LSTM addresses the vanishing point problem through the inclusion of gating cells which allow the network to selectively store and forget memories. *Gated Recurrent Units* (GRU) [15] are a variant of LSTM that have been proved to achieve comparable performances with LSTM with fewer parameters.

Deep generative models such as *Generative Adversarial Networks* (GAN) [16] and *Variational Autoencoders* (VAE) [17] have recently shown outstanding results in learning latent encodings and generating high-dimensional representations. In VAE the latent representation of the input is reparameterized to follow a mixed Gaussian distribution. This constraint makes the encoder more efficient, creating information-rich latent variables. In the original GAN formulation the discriminator network is trained to classify real and fake examples. However, the loss function can be difficult to converge and training is often unstable. WGAN [18] proposed to use Wasserstein distance with weight clipping for stabilizing training. Recently, Gulrajani et.al. [19] proposed to penalize the norm of the discriminator gradient with respect to its input, further improving the training stability.

In *Conditional GAN* (cGAN) the generated output is conditioned on external conditional information. cGAN allows to tackle problems where the input-to-output mapping is of the type one-to-many. *Invertible conditional GAN* (IcGAN) [20] combine an autoencoder with a cGAN and have been proved to be able to learn a good latent representation of the inputs.

Prompted by recent successes of applying deep learning to large, diverse, and high-dimensional datasets, which is also gaining traction in many biomedical tasks, such as cancer detection [21], cancer classification [22], anti-cancer drug prescription [23], and patient outcome prediction [24]. For the task of detecting skin cancer, Esteva et al. developed a deep convolutional neural network that rivals the accuracy of a well-trained dermatologist [21]. More recently, Way et al. deployed a VAE using more than 10,000 cancer datasets for cancer stratification [25]. However, these researches seldom take into the account of both spatial and temporal features of the data. By considering both space and time in the design of the mathematical model, Wang et al. showed a 3-fold increase in cancer therapy response [26]. We envision the incorporation of spatio-temporal information in our deep learning model could represent the complex biological outcomes more faithfully.

C. Thermal ablation

Thermal ablation refers to the surgical removal of tissue, most commonly malignant tumours, via extreme temperature change. The removal is commonly performed percutaneously, where a needle-shaped applicator delivers focal heat energy around the tumour. The heat causes tissues to coagulate, followed by tissue necrosis. Compared to traditional open surgery, thermal ablation allows real-time imaging guidance for more precise control, and it is a clinically more effective due to its lower rate of mortality [27]. Some primary cancers, such as hepatocellular carcinoma (HCC), have limited long-term survival benefit when chemotherapy is used. The presence of tumour heterogeneity in HCC poses another challenge for targeted pharmacologic treatment [28]. In fact, major cancer-driving genes in HCC, such as *TP53* and *CTNNB1*, are yet to be druggable [29]. Therefore, surgical removal of tumours via thermal ablation remains as an important treatment for cancer.

Microwave ablation (MWA) is a recently developed technique of thermal ablation. As previously shown in Figure 3, a thin microwave antenna is inserted into the tumour. The antenna consists of a conductor, a dielectric, a catheter and an air gap. Power is injected into the conductor, electric current flows into the surrounding dielectric and eventually is emitted as microwaves through the air gap. The microwaves heat up the tumour, leading to a coagulated region where the cancer cells are destroyed. MWA can create an ablation region faster than its predecessor, namely radiofrequency ablation (RFA) [30]. Besides, the precision of the ablation zone is improved, due to its relative independence of contact distance between tissue and applicator [31]. To date, MWA has been successfully applied to the removal of hepatocellular carcinoma [27], kidney tumours [32], lung tumours [32], and osteomas [32].

While some clinicians are using FEM to model the parameters of MWA [3], others are simply judging the power settings based on personal experience and the manufacturer's recommendations [33]. In a study of the long-term outcomes of microwave ablation, clinicians used anywhere from 45W to 140W during the ablation of 416 tumours [33]. As discussed earlier, the dynamic physical properties of tissues can significantly affect the electrical impedance [4], [5]. This calls

for a need to develop accurate yet systematic techniques that optimise the power settings for MWA.

D. Modelling of thermal ablation zones

Early methods for the simulation of thermal ablation zones were based on greatly simplified geometries and physical properties of a tumour. For instance, some researchers assumed that the tumour tissue is homogeneous, and the ablation zone is a symmetrical ellipsoid [34]–[36]. Heuristics were often used to account for the dynamic changes in the material during ablation, such as the heat-sink effect of surrounding blood vessels. Since the computational power was limited, these approximations were necessary to enable fast calculation of ablation zones.

To model the ablation zone more accurately, several numerical methods were introduced, such as the FEM method [37] and the Finite Difference Time Domain (FDTD) method [38]. These methods allow the modelling of complex geometries and the non-linear combination of physical properties, which could better reflect the actual surgical conditions. FEM is more widely used than FDTD in the field of MWA for two primary reasons [39]. First, efficient commercial implementations of FEM such as COMSOL (Burlington, MA, USA) is available. Second, FEM methods, in particular COMSOL, were validated against analytical solutions, where less than 5% of differences were observed [40]. More comprehensive reviews of thermal ablation modelling methods are available from Schumann et al. [41] and Jason et al. [39].

III. MICROWAVE ABLATION MODELLING

Microwave ablation couples the physics of electromagnetic fields and bioheat transfer. We solve the non-linear, time dynamic coupling with a multi-physics simulator COMSOL. Tissue destruction is modelled with the Arrhenius rate equation, with parameters supplied from the biomedical literature [42].

$$\frac{\delta \alpha}{\delta t} = A e^{-\frac{\delta E}{RT}}. \qquad (1)$$

where A is the frequencies factor (s^{-1}), δE is the activation energy for irreversible damage reaction (J/mol), R is the gas constant and T is the temperature (K).

$$\theta = 1 - e^{-\alpha}. \qquad (2)$$

where θ is the fraction of necrotic tissue.

It should be noted that the equation is sensitive to parameter values, motivating the need for a sensor-driven approach, due to patient-variability. Ideally, the simulator can provide accurate prediction in the ablated (necrotised) region. However, in practice, the impedance of the antenna increases during the therapy due to the dehydrating tissue. This in turn leads to a mismatch of microwave power transfer and ultimately inaccurate prediction in tissue destruction.

In our experimental setup, eight temperature sensors are linearly placed along the catheter (Figure 3). In addition, binary masks representing the materials around the catheter are available (being part of the setup). We change the position of the air gap inside the antenna from top to bottom. The change

in air gap position will also lead to different microwave radiation patterns and variations in tissue destruction distribution. The ultimate goal is to learn to predict the damage of the ablated region over a period of 10 minutes, based only on the measurements from eight 1-D temperature sensors.

IV. THE PROPOSED APPROACH

The approach is composed of two modules: a *Temporal-Prediction Network* (TPN), which predicts a future sequence of sparse sensor measurements given a previous sequence of sparse sensor measurements, and a *Spatio-Prediction Network* (SPN), which takes in a series of images representing occupancy masks for different materials around the sensor, and predicts the future value of the required scalar field. The SPN is conditioned on a sequence of sparse sensor measurements.

Based on these two modules, we propose two architectures: SenseGAN, in which the SPN is conditioned on current temperature measurements in order to predict the fraction of ablated tumor at the current time instant. And SenseGAN+, in which the SPN is conditioned on the output of the TPN on order to predict future tumour ablation.

We will first describe in detail the SPN and TPN modules, then we describe the SenseGAN and SenseGAN+ architectures.

A. Spatio Prediction Network

The Spatio-Prediction Network (SPN) for predicting damaged tissue is implemented as a conditional variational autoencoder with adversarial training (cVAEGAN), composed of an encoder and a generator network in Figure 4 (bottom). The autoencoder is trained in an adversarial fashion with the addition of a disciminator network at training time. At prediction time, only the generator is used to generate the predictions.

The normal encoding of a VAE offers a sensible way to encode the smooth distribution of real data, as opposed to a traditional autoencoder which has no implicit underlying distribution of the latent space. In our application scenario, we assume that the volumetric region of ablation is axially symmetrical, reducing the 3-D problem to a 2-D problem.

The generator takes as input an image of dimension $256{\times}256{\times}4$ as a conditioning input. Each channel is a binary image mask and represents one layer of a different material mask around the area to be modelled (e.g., catheter, liver tissue, air-gap etc). Note that for different problems, the number of material masks can be easily altered. The encoder is composed of eight convolutional layers, with $4{\times}4$ filters, each followed by a $2{\times}2$ max pooling layer and a LReLU nonlinearity. The last layer is flattened into a fully connected layer. The fully connected layer is followed by two layers μ and σ, that encode a normal latent representation of the input. Random values are sampled from the mixture model.

The sampled latent vector is concatenated with an encoded sensor measurement vector c. The c vector encodes a sequence of temperature measurements. The decoder follows the inverse structure of the encoder, with one fully connected layer followed by eight deconvolutional layers. Corresponding layers in
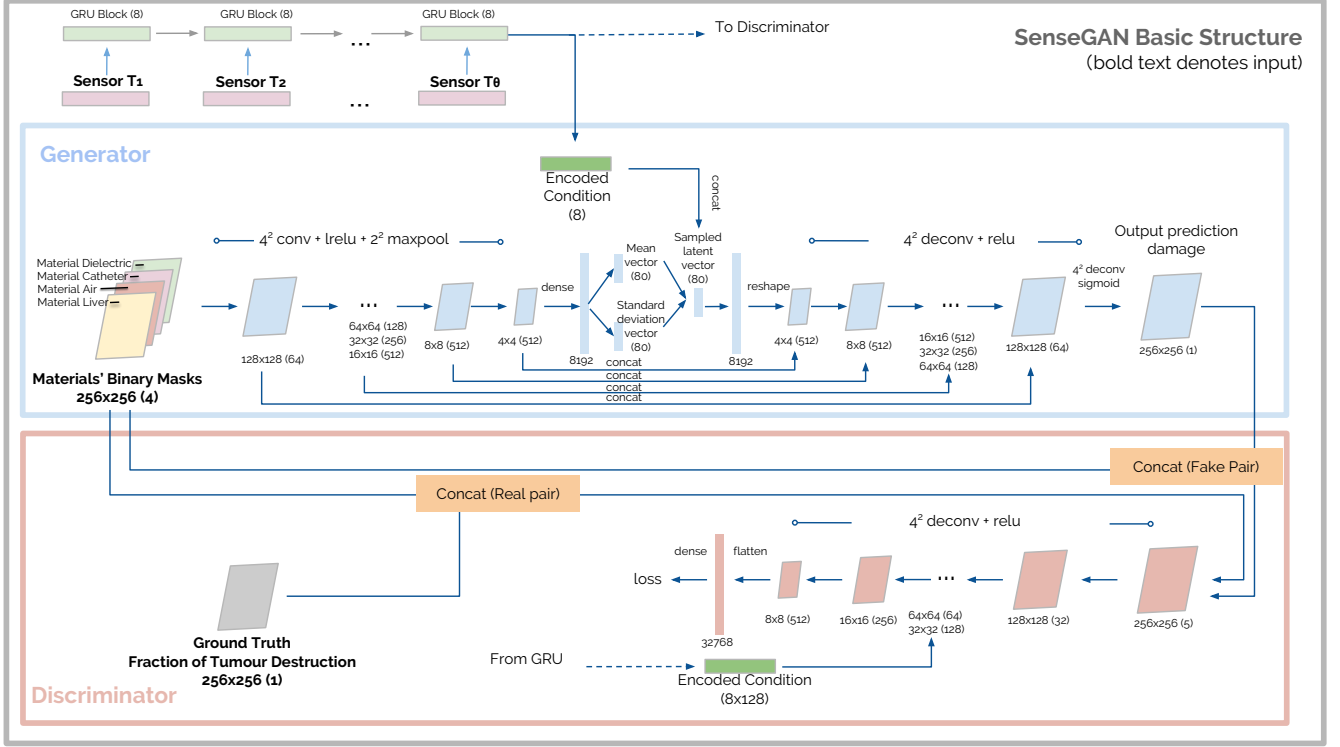
Fig. 4. The SenseGAN architecture for predicting the fraction of tumour destruction at current time.

the encoder are concatenated into decoders following the shape of U-Net [43]. The deconvolutional layers are followed by a ReLU non-linearity, save for the last one which is followed by a sigmoid function.

The predicted output is an image of dimensions $256 \times 256 \times 1$ which represents the cumulative tissue damage. The reconstruction loss $\mathcal{L}_{ae}$ for the autoencoder is defined as the L1 distance between the true and the predicted output:

$$\mathcal{L}_{ae} = \mathbf{E}||y - G(x, z, c)||_1. \qquad (3)$$

where G is the generator network, x and y are the corresponding input and output samples respectively, z is the latent representation and c is the conditioning vector.
A second loss is given by the Kullback-Leibler divergence D_{KL} of the latent representation from the normal distribution:

$$\mathcal{L}_{KL} = D_{KL}(N(\mu, \sigma)||N(0, I)). \qquad (4)$$

where $\mathcal{L}_{KL}$ enforces a normal distribution over the values of the latent representation.
The total loss $\mathcal{L}_{VAE}$ for the autoencoder is:

$$\mathcal{L}_{VAE} = \mathcal{L}_{ae} + \mathcal{L}_{KL}. \qquad (5)$$

The discriminator network takes as input a "real" pair, given by the concatenation of the input material mask x and the real output y, and a "fake" pair, composed by the concatenation of the material mask x and the predicted output $\hat{y}$. The encoded sequence of temperature measurements is concatenated to the third layer.

The structure follows that of the encoder, with eight convolutional layers with filter size 4×4, 2×2 max pooling,

followed by LReLU non-linearities. The convolutional layers are followed by a fully connected layer.

We use L2 loss for the discriminator loss $\mathcal{L}_{GAN}$. The discriminator learn a high-level representation of the outputs and forces the generator to produce sharper and more realistic compared to a standard VAE architecture.
The total SPN loss is thus:

$$\mathcal{L}_{SPN} = \alpha \mathcal{L}_{VAE} + (1 - \alpha)\mathcal{L}_{GAN}. \qquad (6)$$

where α is a constant that balances the VAE and GAN losses. Rather than outputting a single scalar output, as in traditional GAN architectures, the discriminator maps the original 256×256 image into 70×70 overlapping patches, and outputs a scalar value indicating the "real" or "fake" value for each patch.

B. Temporal Prediction Network

In order to capture the temporal evolution of the sensor inputs and to be able to predict future measurements, we design a TPN based on a sequence-to-sequence architecture.
The network is implemented as a recurrent autoencoder [44], and the details of the architecture are shown in Figure 5 (top). The encoder is composed by an LSTM layer with hidden state of dimension 256, and produces a latent encoding of the input sequence with dimension 256. The latent vector is fed into the decoder, implemented as another LSTM network with hidden state dimension of 256. The decoder generates a predicted sequence of future temperature measurements from the latent encoding.
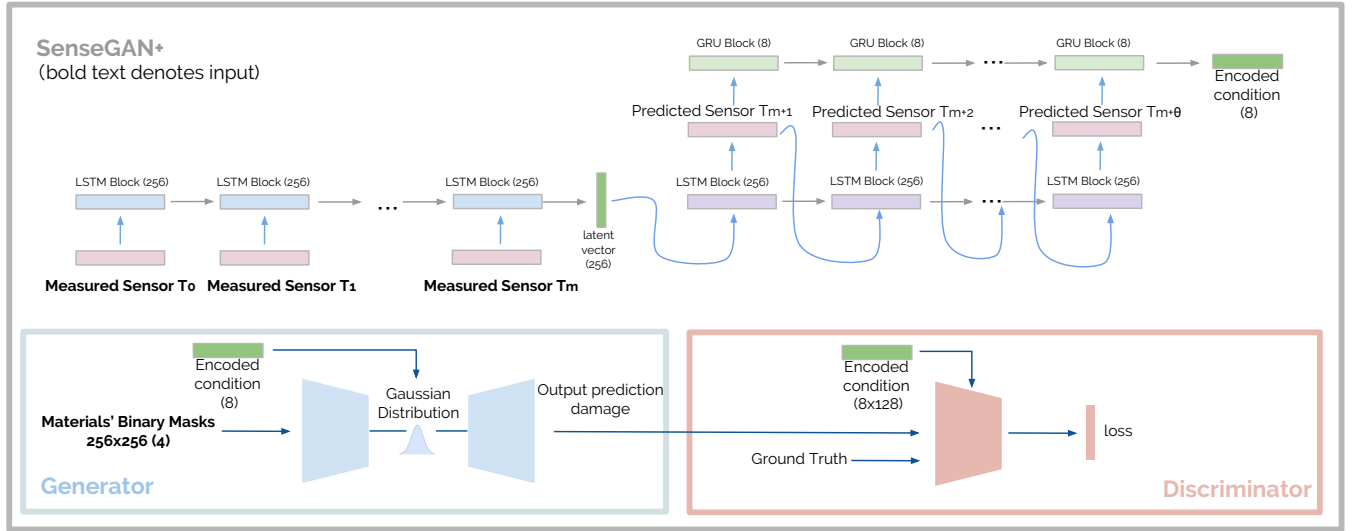
Fig. 5. The SenseGAN+ architecture for predicting the fraction of tumour destruction at future time steps given a sequence of past sparse temperature measurements.

The predicted sequence is then fed to a *Gated Recurrent Unit* (GRU) recurrent network with an hidden state of dimension 8, in order to extract a latent encoding c. The hidden state of the GRU blocks is initialised with random uniform values at the start. The latent encoding vector is then concatenated to the sampled latent representation and to the discriminator of the SPN network.

Intuitively, the GRU cells are able to capture the time-varying evolution of the sparse temperature inputs. This was originally inspired by taking a deep insight into the Arrhenius rate equation (Eq. 1), where the fraction of necrotic tissue can be described at a high level as the integral over time of a function that is dependent on the transient temperature. At last, the sequence-to-sequence architecture is used to predict future sequences of measurements.

C. SenseGAN Framework

The SenseGAN architecture as shown in Fig. 4 is composed of an SPN module conditioned on an array of temperature measurements at current time t. The SPN is able to leverage large amounts of simulated data in order to learn a direct mapping from sensed temperatures to fraction of necrotic tissue by implicitly modelling the wide range of impedance parameters of liver tissue, by being conditioned on the surrounding material masks.

D. SenseGAN+ Framework

We then propose an extension of the SenseGAN architecture, SenseGAN+, for the task of extrapolating future tissue damage from a sequence of past temperature measurements. In this architecture the STN is conditioned on the output of a TPN module. Similar to SenseGAN, the encoded output of the TPN is concatenated to the sampled latent vector of the generator, as well as to the discriminator.

The rationale behind the proposed architecture is that, while the SPN is able to learn the spatial mapping between the different input and output data modalities, the TPN focuses on learning to approximate the time-varying dependence between the fraction of necrotic tissue and the transient temperature.

V. EXPERIMENTS

A. Experimental Settings

Synthetic ablated region and sensor measurement dataset are generated using above modelling method. The sequential sensor data is composed of 4000 temperature data points, with 100 randomly selected mismatched begin-end impedance pairs over 40 time steps (i.e. 15 s to 600 s with an interval of 15 s). Each data point is a vector of 8 temperature readings. We use the Adam optimizer for training the TPN network, with a learning rate of 1e-5. The objective is L1 loss. The training time is around two hours on an Nvidia Titan X GPU.

The SPN is trained on a dataset of 2000 images, with 100 randomly selected mismatched impedance pairs over 20 possible time steps (30s to 600s with a time step of 30s). We train the SPN using the Adam optimizer, with a learning rate of 2e-4, using L1 loss. Training time is around 12 hours on an Nvidia Titan X GPU. Testing time is 125ms per prediction.

In order to evaluate our results, we employ two accuracy metrics: the L1 distance l and the accuracy rate η based on an error threshold of τ_{th}, respectively defined as

$$l = \mathbf{E}||\hat{y} - y||_1.$$

and

$$\eta = \frac{\sum_{ij} I\left(\left|\hat{y}_{ij} - y_{ij}\right| < \tau_{th}\right)}{r * c}.$$

where I is an indicator function, $\hat{y}_{ij}$ is the predicted value for the fraction of necrotic tissue at position (i, j), y_{ij} is the true value for the fraction of necrotic tissue at position (i, j), τ_{th} is the error threshold, r and c are the number of rows i and numbers columns j respectively.
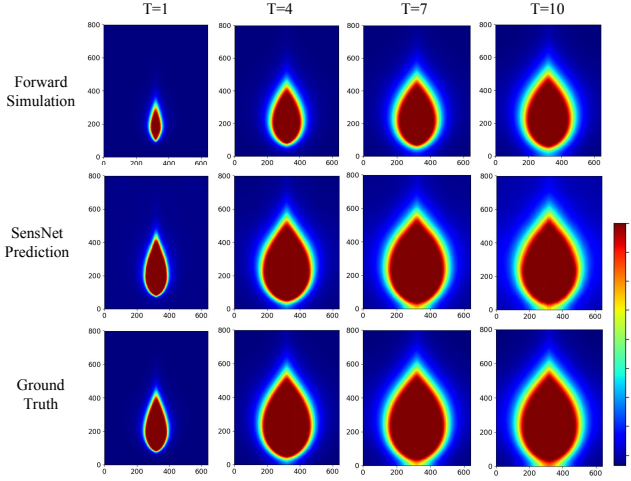
Fig. 6. Fraction of necrotic tissue over time steps, with the source suffering from a randomly mismatched impedance. Top row: forward simulation based on power consumption; Middle row: inverse prediction using sparse temperature sensing data; Bottom row: ground truth
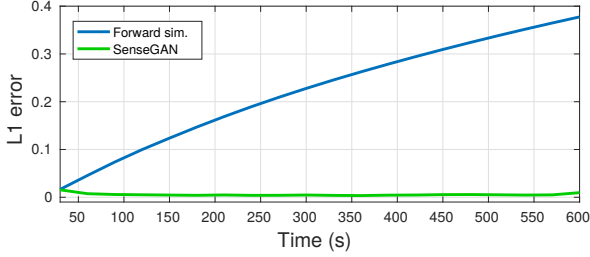

Fig. 7. The accuracy l over 600 seconds, showing how SenseGAN can handle impedance uncertainty.
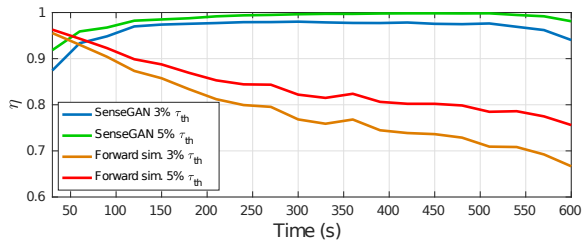

Fig. 8. The accuracy η at $\tau = 3\%, 5\%$ over 600 seconds, showing how SenseGAN can handle impedance uncertainty.

B. Implicit inference of unknown impedance

Liver tissue can have a wide range of impedance (electrical conductivity) parameters, depending on its level of disease and perfusion [4]. It is very complicated to determine these parameters remotely e.g., with diagnostic imaging, yet they play a significant role in the level of tissue damage for a given input power. In this experiment we demonstrate how, given a random impedance mismatch, SenseGAN can correctly predict the area of necrotic tissue, as informed by the temperature observations made along the length of the catheter. This is illustrated in Fig 6, where it can be seen how the radius of damage for the baseline forward prediction, under the assumption of matched impedance, is significantly smaller
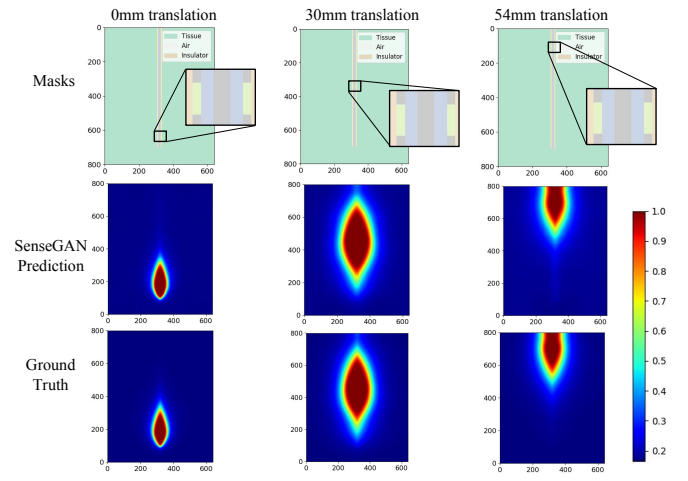

Fig. 9. Fraction of necrotic tissue with various mask positions. Top row: various mask positions; Middle row: inverse prediction using sparse temperature sensing data; Bottom row: ground truth

TABLE I
EVALUATION OF PREDICTION ACCURACY FOR VARIOUS MASK POSITIONS
FOR η WITH $\tau = 3\%, 5\%$, AND l.

Mask shift / mm	0	6	12	18	24
η @ 3% τ_{th}	0.946	0.971	0.972	0.990	0.970
η @ 5% τ_{th}	0.976	0.989	0.981	0.998	0.991
l / 10^{-2}	0.752	0.618	0.699	0.479	0.739
Mask shift / mm	30	36	42	48	54
η @ 3% τ_{th}	0.953	0.967	0.966	0.964	0.976
η @ 5% τ_{th}	0.972	0.985	0.986	0.991	0.992
l / 10^{-2}	0.945	0.570	0.685	0.605	0.535

than reality. In this case, by just using the offline calculated treatment plan, the damage would not be sufficient to cover overall tumour region.

The cumulative impact of impedance mismatch on the L1 error is shown in Fig. 7 which shows how the SenseGAN is able to accurately track the actual level of damage over time, whereas the forward prediction gets cumulatively worse over time. The error results for 3% and 5% are also shown in Fig. 8, showing how SenseGAN can estimate the area of ablation correctly, whereas the forward model slowly diverges from reality.

In summary, this experiment demonstrates that SenseGAN is able to, in the face of high levels of uncertainty about tissue impedance, accurately infer the true region of ablation just by using a few 1-D sensor measurements. This demonstrates that the network has intuitively learned a complex spatio-temporal relationship between multiple modalities.

C. Generalisation to mask variations

In this experiment, we consider how well the model can generalise to variations in the material masks. In particular, we examine the most significant effect, the location of the antenna port in the catheter. Under the assumption of a constant operating frequency, changing the location of the antenna port alters the radiated power pattern, which is a complex function of wavelength and the surrounding dielectric
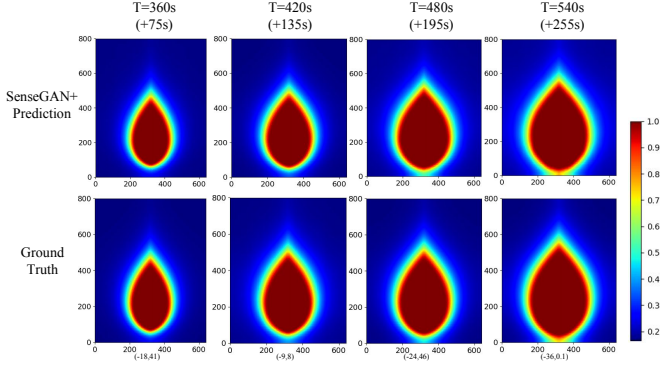
Fig. 10. Predicted fraction of necrotic tissue in future time steps. Top row: inverse prediction based on predicted future temperature data; Bottom row: ground truth

Fig. 11. Quantitative results for tissue damage predication on four other materials: bone, prostate, kidney, lung.

TABLE II

THE PERCENTAGE OF CORRECT TEMPERATURE PREDICTIONS FOR THE NEXT 300 SECONDS WHOSE ERROR IS UNDER CERTAIN THRESHOLDS T_{th} (IN KELVIN). TOP ROW: PREDICTIONS WITH 30 SECONDS INTERVAL; BOTTOM ROW: PREDICTIONS WITH 15 SECONDS INTERVAL.

T_{th} / Kelvin	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\Delta 5$
Accurate rate, 30s	0.909	0.981	0.995	0.999	1
Accurate rate, 15s	0.911	0.980	0.995	0.999	1

TABLE IV

EVALUATION OF PREDICTION ACCURACY FOR VARIOUS MATERIALS FOR η WITH $\tau = 3\%, 5\%$.

Material	Bone	Prostate	Kidney	Lung	mean
η @ 3% τ_{th}	0.936	0.937	0.962	0.961	0.949
η @ 5% τ_{th}	0.978	0.973	0.984	0.986	0.980

TABLE III

EVALUATION IN ABLATION PREDICTION ACCURACY UP TO NEXT 285 SECONDS FOR η WITH $\tau = 3\%, 5\%$, AND l.

Time / s	+15	+45	+75	+105	+135
η @ 3% τ_{th}	0.906	0.902	0.910	0.915	0.905
η @ 5% τ_{th}	0.945	0.947	0.951	0.957	0.953
l / 10^{-2}	1.148	1.109	1.006	0.964	1.000
Time / s	+165	+195	+225	+255	+285
η @ 3% τ_{th}	0.915	0.906	0.914	0.924	0.913
η @ 5% τ_{th}	0.955	0.955	0.965	0.967	0.964
l / 10^{-2}	0.944	0.983	0.938	0.908	0.939

properties. As the separation distance increases, the radiated lobe becomes broader as it approaches a quarter wavelength, before becoming smaller again for larger separations.

The results for three different mask locations are shown in Fig. 9. Note that in these predictions, only few pixels have moved which represents the location of the port, yet the model is able to capture this small variation and infer the large changes in the lobe pattern. This is also shown in more detail in Table I.

This experiment demonstrates that SenseGAN is able to generalize to variations not only in the properties of the surrounding media, but also their geometry. This further shows that SenseGAN is robust to model uncertainty, making it suitable for real-world prediction problems.

D. Prediction of future damage

Lastly, in this experiment, we examine how closely we can extrapolate future tissue damage, given a time-series of observed temperature data. We first examine how well the
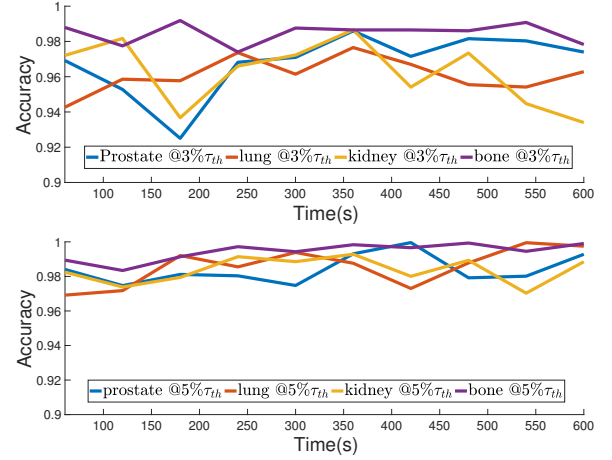
recurrent network can predict future temperature readings, as shown in Table II, with different forecasting intervals. This shows that the RNN is able to forecast the true temperature 98% of the time over the entire 300 s window below an error of 2 Kelvin, and 99.5% below an error of 3 Kelvin.

Next, we examine the overall accuracy of damage forecasting, which is a critical requirement for accurate feedback driven control of thermal ablation. Fig. 10 shows the extremely close match between the true and predicted zones of ablation over a 255 s window. The numerical accuracy for different time steps are also shown in Table III.

This result demonstrates the power of coupling the TPN and SPN to predict tissue damage into the future, with high accuracy.

E. Generalisation to other materials

In this experiment, we consider four other materials to show that our proposed method can be transferred to other medical tasks. Specifically, we use the same experiment set up as previous ones, but instead of changing port position, we change the material from liver to four other different materials: bone, lung, prostate, and kidney, which can also be operated with microwave ablation treatment. We observe that our model can also outperform conventional forward simulation on all of these materials. Table IV shows the average IoU with thresholds of 3% and 5% on these new materials. They all achieve over 93% accuracy, with little variance over different materials. This suggests that our model have good generalisation ability, and is affected very little by the change of material properties. Figure 11 reinforces this conclusion with high accuracy over different time steps,

Fig. 13. Prediction error in time step 4, 7 and 10. Top row: Square root of per pixel MAE of our prediction (We square root the error for clearer presentation). Middle row: an example prediction at each time step. Bottom row: corresponding ground truth of the example prediction.

surrounding damage regions have more errors compared with the rest regions. This error is due to the spreading of dynamic destruction, which brings more uncertainty.

VI. CONCLUSION

We presented a new method for estimating spatially and temporally varying fields governed by multi-modal, coupled partial differential equations, and in particular inferring such fields given only spatially sparse readings. We proposed SenseGAN, a deep architecture based on a recurrent, generative network that is conditioned on observed and predicted sensor readings. SenseGAN can optimise the power settings of microwave ablation for more effective treatment. Although we only considered the problem of microwave ablation in this paper, this general framework of learning how to invert a physical model is of general and widespread use to other applications that involve inverse physics and multi-physics coupling.
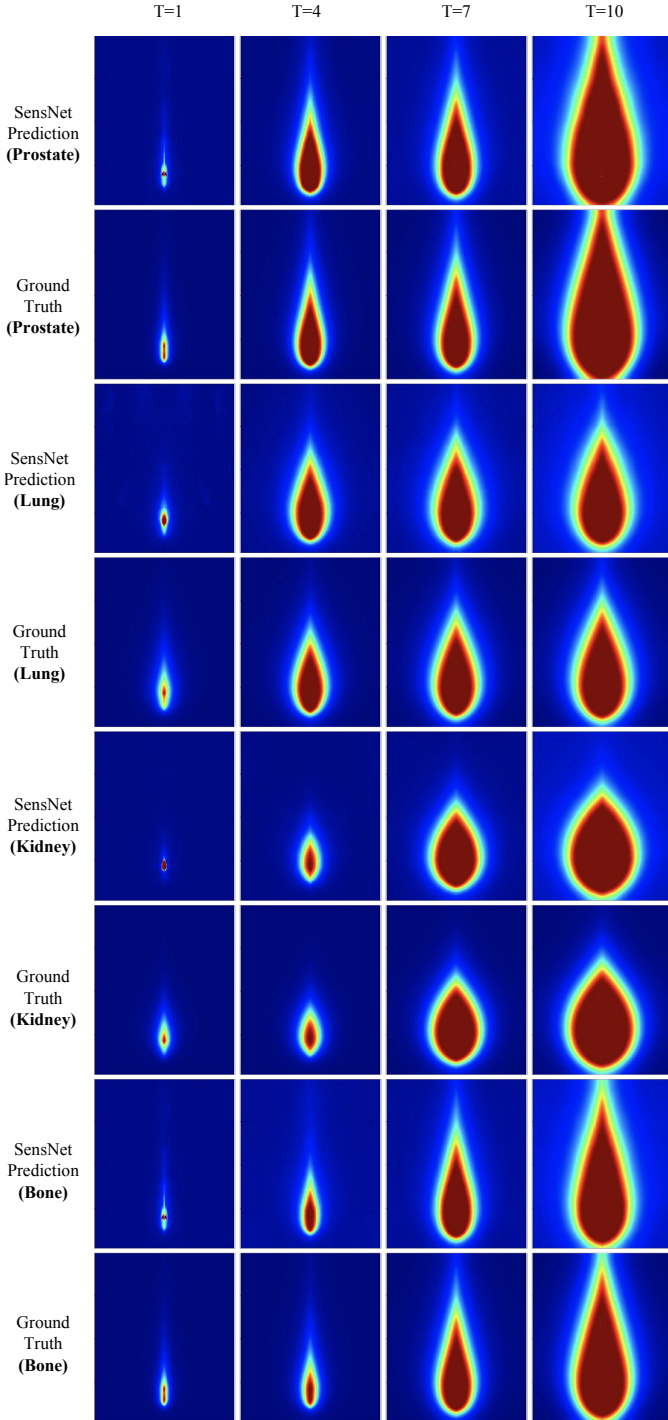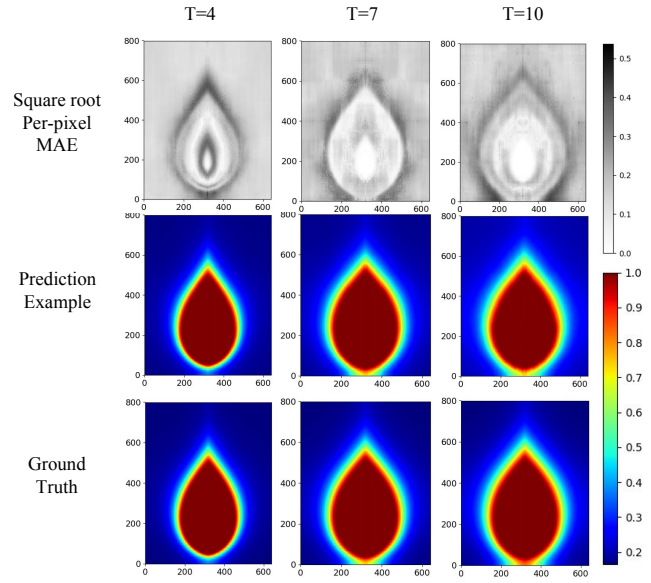

Fig. 12. Qualitative results for tissue damage predication on 4 other materials: bone, prostate, kidney, lung.

similar to our result of prediction of liver damage, where few qualitative results can be found in Figure 12. We also show some randomly sampled qualitative results of tissue damage prediction over these materials. Notice indeed errors are minor, and general shapes, which differ in these materials, are well predicted across all new materials. We also analyse the prediction error of our method. The error of our prediction is shown in Fig 13, where the first row is the plot of mean absolute error (MAE) per pixel. We can see that the boundaries

REFERENCES

[1] K. F. Chu and D. E. Dupuy, "Thermal ablation of tumours: biological mechanisms and advances in therapy," *Nature Reviews Cancer*, vol. 14, no. 3, p. 199, 2014.

[2] J. M. Hill and T. R. Marchant, "Modelling microwave heating," *Applied mathematical modelling*, vol. 20, no. 1, pp. 3–15, 1996.

[3] C. Rieder, T. Kroeger, C. Schumann, and H. K. Hahn, "Gpu-based real-time approximation of the ablation zone for radiofrequency ablation," *IEEE transactions on visualization and computer graphics*, vol. 17, no. 12, pp. 1812–1821, 2011.

[4] S. Laufer, A. Ivorra, V. E. Reuter, B. Rubinsky, and S. B. Solomon, "Electrical impedance characterization of normal and cancerous human hepatic tissue," *Physiological measurement*, vol. 31, no. 7, p. 995, 2010.

[5] C. L. Brace, P. F. Laeseke, D. W. van der Weide, and F. T. Lee, "Microwave ablation with a triaxial antenna: results in ex vivo bovine liver," *IEEE transactions on microwave theory and techniques*, vol. 53, no. 1, pp. 215–220, 2005.

[6] X. Chen and G. M. Saidel, "Mathematical modeling of thermal ablation in tissue surrounding a large vessel," *Journal of biomechanical engineering*, vol. 131, no. 1, p. 011001, 2009.

[7] N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G. R. Lanckriet, R. Levy, and N. Vasconcelos, "A new approach to cross-modal multimedia retrieval," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 251–260.

[8] X. Zhai, Y. Peng, and J. Xiao, "Heterogeneous metric learning with joint graph regularization for cross-media retrieval." in *AAAI*, 2013.

[9] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 689–696.

[10] N. Srivastava and R. Salakhutdinov, "Learning representations for multimodal data with deep belief nets," in *International conference on machine learning workshop*, vol. 79, 2012.

[11] D. Wang, P. Cui, M. Ou, and W. Zhu, "Deep multimodal hashing with orthogonal regularization." in *IJCAI*, vol. 367, 2015, pp. 2291–2297.

[12] L. Castrejon, Y. Aytar, C. Vondrick, H. Pirsiavash, and A. Torralba, "Learning aligned cross-modal representations from weakly aligned data," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2940–2949.

[13] Z. Wang, S. Rosa, L. Xie, B. Yang, S. Wang, N. Trigoni, and A. Markham, "Defo-net: Learning body deformation using generative adversarial networks," in *IEEE Intl Conference on Robotics and Automation (ICRA)*, 2018.

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1162/neco.1997.9.8.1735

[15] K. Cho, B. van Merriënboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734. [Online]. Available: http://www.aclweb.org/anthology/D14-1179

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[17] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *ICLR*, 2014.

[18] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *CoRR*, vol. abs/1701.07875, 2017.

[19] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," 2017.

[20] G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Álvarez, "Invertible conditional gans for image editing," *arXiv preprint arXiv:1611.06355*, 2016.

[21] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.

[22] J. A. Golden, "Deep learning algorithms for detection of lymph node metastases from breast cancer: Helping artificial intelligence be seen," *Jama*, vol. 318, no. 22, pp. 2184–2186, 2017.

[23] K. Preuer, R. P. Lewis, S. Hochreiter, A. Bender, K. C. Bulusu, and G. Klambauer, "Deepsynergy: Predicting anti-cancer drug synergy with deep learning." *Bioinformatics*, vol. 1, p. 9, 2017.

[24] K. Chaudhary, O. B. Poirion, L. Lu, and L. X. Garmire, "Deep learning based multi-omics integration robustly predicts survival in liver cancer," *Clinical Cancer Research*, pp. clincanres–0853, 2017.

[25] G. P. Way and C. S. Greene, "Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders," *Pac Symp Biocomput*, vol. 23, pp. 80–91, 2018.

[26] Z. Wang, R. Kerketta, Y.-L. Chuang, P. Dogra, J. D. Butner, T. A. Brocato, A. Day, R. Xu, H. Shen, E. Simbawa *et al.*, "Theory and experimental validation of a spatio-temporal model of chemotherapy transport to enhance tumor cell kill," *PLoS computational biology*, vol. 12, no. 6, p. e1004969, 2016.

[27] E. A. F. T. S. O. T. Liver *et al.*, "Easl–eortc clinical practice guidelines: management of hepatocellular carcinoma," *Journal of hepatology*, vol. 56, no. 4, pp. 908–943, 2012.

[28] J.-C. Nault and A. Villanueva, "Intratumor molecular and phenotypic diversity in hepatocellular carcinoma," *Clinical Cancer Research*, 2015.

[29] J. M. Llovet and V. Hernandez-Gea, "Hepatocellular carcinoma: reasons for phase iii failure and novel perspectives on trial design," *Clinical cancer research*, 2014.

[30] L. S. Poulou, E. Botsa, I. Thanou, P. D. Ziakas, and L. Thanos, "Percutaneous microwave ablation vs radiofrequency ablation in the treatment of hepatocellular carcinoma," *World journal of hepatology*, vol. 7, no. 8, p. 1054, 2015.

[31] Z. Issa, J. M. Miller, and D. P. Zipes, *Clinical Arrhythmology and Electrophysiology: A Companion to Braunwald's Heart Disease E-Book: Expert Consult: Online and Print*. Elsevier Health Sciences, 2012.

[32] C. L. Brace, "Radiofrequency and microwave ablation of the liver, lung, kidney, and bone: what are the differences?" *Current problems in diagnostic radiology*, vol. 38, no. 3, pp. 135–143, 2009.

[33] U. Leung, D. Kuk, M. D'angelica, T. Kingham, P. Allen, R. DeMatteo, W. Jarnagin, and Y. Fong, "Long-term outcomes following microwave ablation for liver malignancies," *British Journal of Surgery*, vol. 102, no. 1, pp. 85–91, 2015.

[34] E. S. McCreedy, R. Cheng, P. F. Hemler, A. Viswanathan, B. J. Wood, and M. J. McAuliffe, "Radio frequency ablation registration, segmentation, and fusion tool." *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 10, no. 3, pp. 490–6, jul 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16871716 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2614074

[35] T. Butz, S. K. Warfield, K. Tuncali, S. G. Silverman, E. van Sonnenberg, F. A. Jolesz, and R. Kikinis, "Pre- and Intra-operative Planning and Simulation of Percutaneous Tumor Ablation." Springer, Berlin, Heidelberg, oct 2000, pp. 317–326. [Online]. Available: http://link.springer.com/10.1007/978-3-540-40899-4_32

[36] C. Villard, L. Soler, and A. Gangi, "Radiofrequency ablation of hepatic tumors: simulation, planning, and contribution of virtual reality and haptics," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 8, no. 4, pp. 215–227, aug 2005. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/10255840500289988

[37] P. Phasukkit, S. Tungjitkusolmun, and M. Sangworasil, "Finite-Element Analysis and <i>In Vitro</i> Experiments of Placement Configurations Using Triple Antennas in Microwave Hepatic Ablation," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 11, pp. 2564–2572, nov 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19628446 http://ieeexplore.ieee.org/document/5170082/

[38] A. Littmann, A. Schenk, B. Preim, G. Prause, K. Lehmann, A. Roggan, and H.-O. Peitgen, "Planning of anatomical resections and in situ ablations in oncologic liver surgery," *International Congress Series*, vol. 1256, pp. 684–689, jun 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0531513103003194

[39] J. Chiang, P. Wang, and C. L. Brace, "Computational modelling of microwave tumour ablations." *International journal of hyperthermia : the official journal of European Society for Hyperthermic Oncology, North American Hyperthermia Group*, vol. 29, no. 4, pp. 308–17, jun 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23738698 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3768158

[40] M. Rivera, J. Molina, M. Trujillo, and E. Berjano, "Theoretical modeling of RF ablation with internally cooled electrodes: Comparative study of different thermal boundary conditions at the electrode-tissue interface," *Mathematical Biosciences and Engineering*, vol. 6, no. 3, pp. 611–627, jun 2009. [Online]. Available: http://www.aimsciences.org/journals/displayArticles.jsp?paperID=4261

[41] C. Schumann, C. Rieder, J. Bieberstein, A. Weihusen, S. Zidowitz, J. H. Moltz, and T. Preusser, "State of the art in computer-assisted planning, intervention, and assessment of liver-tumor ablation." *Critical reviews in biomedical engineering*, vol. 38, no. 1, pp. 31–52, 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21175402

[42] V. Neagu, "A study of the effects of geometry on the efficiency of single slot microwave ablation antennas used in hepatic tumor hyperthermia," in *International Conference on Advancements of Medicine and Health Care through Technology; 12th-15th October 2016, Cluj-Napoca, Romania*. Springer, 2017, pp. 131–136.

[43] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[44] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.