# Breast Imaging by Cascaded CNN from Joint Microwave and Ultrasonic Data

Valentin Noël
*Systèmes et Applications des Technologies de l'Information et de l'Energie*
*ENS Paris-Saclay*
*Laboratoire des Signaux et Systèmes*
*Univ. Paris-Saclay, CNRS, CentraleSupélec*
Gif-sur-Yvette, France
valentin.noel1@ens-paris-saclay.fr

Yingying Qin
*Department of Physics and Technology*
*UiT - The Arctic University of Norway*
Tromsø, Norway
yingying.qin@uit.no

Thomas Rodet
*Systemes et Applications des Technologies de l'Information et de l'Energie*
*ENS Paris-Saclay*
Gif-sur-Yvette, France
thomas.rodet@ens-paris-saclay.fr

Dominique Lesselier
*Laboratoire des Signaux et Systèmes*
*Univ. Paris-Saclay, CNRS, CentraleSupélec*
Gif-sur-Yvette, France
dominique.lesselier@centralesupelec.fr

***Abstract*—In the context of early breast tumor characterization, combining electromagnetic (EM) and ultrasound (US) modalities is of interest, since both non-ionizing and low-cost, and harboring complementary resolution features. Here, a new Convolutional Neural Network (CNN) structure is proposed, denoted as Structurally-Aware Complex Cascaded Neural Network (SACC-CNN). It consists of two parts, the Structurally-Aware Reconstruction Net (SARNet) and the Structurally-Aware Classification Net (SACNet). SACNet outputs the tissue type map which is then fed to the SARNet, which reconstructs the EM and US parameters. These two parts can be seen as two independent modules. A physics-guided loss function is implemented in the SARNet network to enhance structural similarity. Main features of the approaches, illustrated by simulation, are described.**

***Index Terms*—breast imaging, electromagnetics, ultrasound, data fusion, cascaded convolutional neural networks**

## I. Introduction

Breast tumors are most common and their detection is critical at early stage [1]. Hence, developing technologies to image a small tumor at low cost and low risk is an important issue. Currently, X-ray mammography is the gold standard for this detection. Despite of high-resolution of the result, it has a number of limitations including low sensitivity, ionizing radiation, discomfort from breast compression, and detection quite affected by breast density. Fusion of complementary multimodality data has been investigated by three of the authors of this contribution [2], and it can be achieved by successive imaging with one modality used as prior. Usually structure information, like tissue interface or region, can be extracted from a high-resolution image. Such information can be used as regularization term to guide inversion with another modality, e.g., [3]–[5]. Joint inversion is another way to fuse data where those are inverted simultaneously. Structural similarity can be employed also to combine the parameters of different modalities in the inversion procedure [6].

Convolutional neural networks (CNN) may be a good way forward. As an example, among many, in [7], a CNN structure is designed to use the measured fields as the input and predict the scatterer map directly. State-of-the-art networks for tumor detection and/or breast image reconstruction include the U-Net and some of its extensions as developed in [8] and also Artifical Neural Networks (ANN) as shown in [9]. Such a type of methods can provide reconstruction results in real time yet gains little from the physical knowledge available. To build up neural networks so as they benefit from the physics of the problem at hand is increasingly investigated. In this context, several works taking place in a different physical framework have been on physics-inspired neural networks (PINN). Efforts have mainly focused on the design of a physical framework that can be translated in different ways. In breast imaging, one has to reconstruct the image of the breast and first and foremost classify its components as tissues with/without tumor.

Previous works have illustrated different approaches for tackling such a problem, some with deep learning. To our knowledge, none of the networks have been inspired/guided by the physics of this same problem though clearly beneficial to the quality of reconstruction and/or classification. In the present contribution one is particularly interested in the accuracy of the breast image and especially in structural similarity between image of the breast and ground truth.

Indeed, improvement of the structural similarity between breast image and ground truth should indicate an improvement of the quality of the image with respect to the existing inner structures of the breast as well as an improvement concerning the reliability of its physics. So, the neural network must be set up in order to favor a reconstruction maximizing the struc-

tural similarity of the image obtained. Starting from a multi-task neural network developed previously, two approaches to enhance the aforementioned physical meaning are proposed. The network thus developed is a cascaded multi-task physics-influenced neural network composed of a main neural network divided into two sub-CNN.

As focused onto fusion of electromagnetic (EM) and ultrasound (US) data, a double-stream complex cascaded convolutional neural network framework with structural awareness is proposed herein. In the numerical experiments, a pending breast —in contrast with the demanding compressed breast of X-ray mammography— immersed within an unbounded coupling medium is assumed.

The contribution is organized as follows. In Section II the modeling is presented. In Section III, the structure of the network is sketched. In Section IV numerical experiments on breast phantoms are proposed. The conclusion is Section V. Notice that recent works on breast imaging and data fusion by the co-workers as mentioned before [10] have been illustrated on phantoms and datasets like those of now, yet going to tailored CNN stands in contrast to these earlier investigations.

## II. Electromagnetic and ultrasound modeling

A two-dimensional geometry is considered, fluid-like materials supporting only compressional waves in the US case, while the EM case is concerned only with Transverse-Magnetic polarization. Time-harmonic waves are assumed with time dependence $\exp(-i\omega t)$ for both US and EM. The breast is located within a domain of interest (DoI) $\mathcal{D}$. The known background medium is of complex relative permittivity $\epsilon_b$, permeability $\mu_b$, and, correspondingly, wavenumber $k_b^{\text{em}} = \omega\sqrt{\epsilon_0\epsilon_b\mu_b}$ in EM, speed of sound $c_b$, attenuation $\alpha_b$, and wavenumber $k_b^{\text{ac}} = \omega/c_b + i\alpha_b$ in US. Permeability and density are constant everywhere. The spatial distributions of complex relative permittivity $\epsilon_r(\mathbf{r}) = \epsilon_r'(\mathbf{r}) + i\epsilon_r''(\mathbf{r})$, sound speed $c(\mathbf{r})$, and attenuation $\alpha(\mathbf{r})$ are the unknowns. For each modality, $N_i$ probes are located at $\mathbf{r}_v'$, $v = 1, 2, \ldots, N_i$ on a circle $\mathcal{S}$, as sketched in Fig. 1. Each one illuminates the DoI and scattered fields are collected by all.

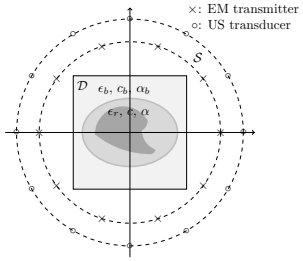


Fig. 1: Sketch of configuration.

To work with the model numerically, $\mathcal{D}$ is discretized into $M = H \times W$ small squared pixels centered at $\mathbf{r}_m$, $m = 1, 2, \ldots, M$, the size of which is to differ betwen the direct problem (acquiring data) and the inverse one (probing the breast). US and EM contrasts are defined as

$$\chi^{\text{ac}}(\mathbf{r}) = \frac{(k^{\text{ac}}(\mathbf{r}))^2 - (k_b^{\text{ac}})^2}{(k_b^{\text{ac}})^2}, \tag{1}$$

$$\chi^{\text{em}}(\mathbf{r}) = \frac{(k^{\text{em}}(\mathbf{r}))^2 - (k_b^{\text{em}})^2}{(k_b^{\text{em}})^2}, \tag{2}$$

and the contrast source for the $v$th incidence as

$$J_v(\mathbf{r}) = \chi(\mathbf{r}) F_v^t(\mathbf{r}), \tag{3}$$

$F$ pressure field $P$ or electric field $E$. Superscript "$t$" denotes total field. Discretized source-type integral equations (i.e., observation and state equations) involving the free-space Green function can be derived using a standard pulse-basis point-matching Method of Moments, those detailed in [11].

## III. CNN scheme

The main task is to reinforce the structural similarity of the image, so it is of primary interest to compare first a specialized non-cascaded CNN structure with multi-stream input and a multi-task learning strategy (CNN-MM) with its cascaded version (SACC-CNN) which takes more into account the morphological structure of the object based on segmentation, then to define a physics-guided cost function network (PG-SACC-CNN) guiding to a better structural similarity (and thus to more realistic physics).

### A. CNN structure

As indicated, the CNN-MM structure is inspired by a two-stream CNN proposed in [12]. In the present work, EM data at a single frequency and US data at two frequencies are used. To introduce physical knowledge, the input of the proposed network is chosen as contrast-source and field quantities $\omega$ and $f$ inside the DoI obtained from backpropagation as

$$\overline{\omega}_v = \overline{\overline{G}}_s^H \overline{F}_v^s, \; \overline{f}_v^s = \overline{\overline{G}}_d \overline{\omega}_v, \tag{4}$$

where $\overline{\overline{G}}_s^H$ is the conjugate transpose of $\overline{\overline{G}}_s$, $\overline{a}$ denoting a vector and $\overline{\overline{a}}$ a matrix. Notice that in back-propagation methods, a complex parameter is used in (4) to minimize the quadratic error in the scattered field, here this normalised EM-specific retro-projection parameter is simply set to 1.

For each modality at given frequency, $\omega_v$ and $f_v$ are reshaped as $H \times W$ images and will allow to limit the long correlation at the data level and also allow the network to be efficient with less layers or with smaller kernel sizes. $\Re\{\omega_v\}$, $\Im\{\omega_v\}$, $\Re\{f_v\}$ and $\Im\{f_v\}$ obtained from all incidences are separated in imaginary part input and real part input with size of $H \times W \times 4N_i$. To achieve the reconstruction, the SARNet network has an output of size $H \times W \times 4$ to predict EM parameters $\epsilon_r'$ and $\epsilon_r''$, and US parameters $c$ and $\alpha$ simultaneously.

Apart from the regression task, an auxiliary classifier (i.e., SACNet) is introduced to provide a segmentation image. A softmax classifier yields the probability distribution over predicted output classes (tissue types and background) for each

pixel and the said pixel is classified into the class with highest score. This segmentation task is quite relevant to the regression task. Such multi-task learning can help training the network and improve the generalization capability.

Another interest adds to, since, still with the objective of feeding the network with physical knowledge, the classification output could inform the reconstruction and thus potentially increase its quality. Such an architecture is the so-called cascaded architecture sketched in Alg. 1, Alg. 2 and Fig.2. The resulting Cascaded CNN is composed of two main blocks, one being the first CNN in charge of the classification whose output is directly input to the second one, in charge of the reconstruction. Their respective architectures are directly inspired from [2].

**Algorithm 1** *Sub-network SACNet.*

1) **for** $m = 1$ to $nb\_iters$ **do**
   a) 2 input channels: Real and Imaginary parts input.
      **for** $i = 1$ to $nb\_input\_channels$ **do**
         $\mathbf{x_i} \leftarrow \text{Conv2D}(\mathbf{x_i}, y, kernel\_size = (1,1))$
         $\mathbf{x_i} \leftarrow \text{BatchNormalization}()(\mathbf{x_i})$
         $\mathbf{x_i} \leftarrow \text{LeakyReLU}(\alpha = 0.1)(\mathbf{x_i})$
      **end for**
   b) Repeat step a).
   c) $\mathbf{x} \leftarrow \text{Concatenate}([\mathbf{x_1}, \mathbf{x_2}])$
      $\mathbf{x} \leftarrow \text{Conv2D}(\mathbf{x}, y, kernel\_size = (3,3))$
   d) Creation of $J = 4$ shortcuts $s_j$, $j = 1, 2, ..., J$
      **for** $j = 1$ to $J$ **do**
         $s_j = \mathbf{x}$
         $\mathbf{x} \leftarrow \text{BatchNormalization}()(\mathbf{x})$
         $\mathbf{x} \leftarrow \text{LeakyReLU}(\alpha = 0.1)(\mathbf{x})$
         $\mathbf{x} \leftarrow \text{Conv2D}(\mathbf{x}, y, kernel\_size = (3,3))$
         Repeating once these 3 steps.
         $\mathbf{x} \leftarrow \text{Add}()([s_j, \mathbf{x}])$
      **end for**
   e) $\mathbf{x} \leftarrow \text{Conv2D}(\mathbf{x}, y, kernel\_size = (3,3))$
      $\mathbf{x} \leftarrow \text{LeakyReLU}(\alpha = 0.1)(\mathbf{x})$
   **end for**
2) **Seg. output** $\leftarrow \text{Conv2D}(\mathbf{x}, y, kernel\_size = (3,3),\ softmax)$.

**Algorithm 2** *Sub-network SARNet.*

1) **for** $m = 1$ to $nb\_iters$ **do**
   a) 3 input channels: Real and Imaginary part inputs + Seg. output.
      **for** $i = 1$ to $nb\_input\_channels$ **do**
         $\mathbf{x_i} \leftarrow \text{Conv2D}(\mathbf{x_i}, y, kernel\_size = (1,1))$
         $\mathbf{x_i} \leftarrow \text{BatchNormalization}()(\mathbf{x_i})$
         $\mathbf{x_i} \leftarrow \text{LeakyReLU}(\alpha = 0.1)(\mathbf{x_i})$
      **end for**
   b) Repeat step a).
   c) $\mathbf{x} \leftarrow \text{Concatenate}([\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}])$
      $\mathbf{x} \leftarrow \text{Conv2D}(\mathbf{x}, y, kernel_s ize = (3,3))$
   d) Creation of $J = 4$ shortcuts $s_j$, $j = 1, 2, ..., J$
      **for** $j = 1$ to $J$ **do**
         $s_j = \mathbf{x}$
         $\mathbf{x} \leftarrow \text{BatchNormalization}()(\mathbf{x})$
         $\mathbf{x} \leftarrow \text{LeakyReLU}(\alpha = 0.1)(\mathbf{x})$
         $\mathbf{x} \leftarrow \text{Conv2D}(\mathbf{x}, y, kernel\_size = (3,3))$
         Repeating once these 3 steps.
         $\mathbf{x} \leftarrow \text{Add}()([s_j, \mathbf{x}])$
      **end for**
   e) $\mathbf{x} \leftarrow \text{Conv2D}(\mathbf{x}, y, kernel\_size = (3,3))$
      $\mathbf{x} \leftarrow \text{LeakyReLU}(\alpha = 0.1)(\mathbf{x})$
   f) Creation of $J = 1$ shortcut $s_1$
   **end for**
2) **Reg. output** $\leftarrow \text{Conv2D}(\mathbf{x}, y, kernel\_size = (3,3),\ sigmoid)$.

### *B. Loss function*

The networks considered, i.e., CNN-MM and (Physics-Guided) SACC-CNN, are using different loss functions. The

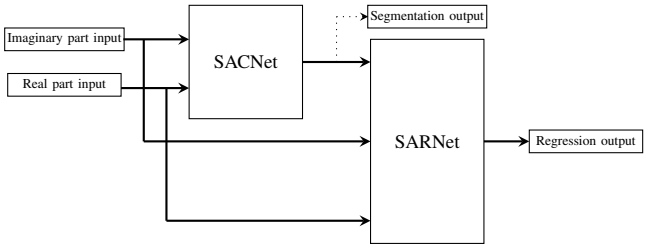


Fig. 2: Sketch of architecture of network SACC-CNN.

loss function for the regression task in CNN-MM is chosen as the mean of the absolute error and of the squared error as

$$L_{reg} = \frac{1}{2NHW} \sum_n \sum_{h,w} \left( |\hat{\tilde{\zeta}}^n_{h,w} - \tilde{\zeta}^n_{h,w}| + |\hat{\tilde{\zeta}}^n_{h,w} - \tilde{\zeta}^n_{h,w}|^2 \right) \tag{5}$$

for parameter $\zeta$, where $\hat{\tilde{\zeta}}^n_{h,w}$ is the predicted value of the $n$th sample located at $(h, w)$ and $\tilde{\zeta}^n_{h,w}$ is the ground truth. $N$ is the number of samples in a training batch. For the SARNet, the same loss function was first used. However, as considered in [13], a loss function can be tuned and used as a physics-guideline for the learning task of a network. In this precise context, it would be of particular interest to train the SARNet such that it focuses on structural similarity between the training set and their ground truths in order to improve the definition of the reconstructed (internal and external) contours.

To that extent, a SSIM (Structural Similarity Index Metric) metric is implemented in the proposed physics-guided loss function to compare the structural similarity of the full image and ground truth (respectively denoted as $x$ and $y$), with

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{6}$$

in which $\mu_x$ and $\mu_y$ are local means of the images, $\sigma_x$ and $\sigma_y$ are local standard deviations, $\sigma_{xy}$ is local cross-covariance, and $C_1$ and $C_2$ are regularization constants. Other implementations, by example a smaller-sized SSIM sliding window, could hypothetically give a better granularity to the structural similarity. Accordingly, this knowledge leads to the new loss function

$$L_{regnew} = L_{reg} + \beta(1 - SSIM(x,y)), \tag{7}$$

$\beta$ being a regularization term tuned at each epoch through a bandit-based Hyperband algorithm as introduced in [14].This loss function will hence prioritize more the structural similarity than the quadratic error.

In the segmentation task, each pixel is classified into a tissue type or background medium. Softmax classifier is used for this multi-class classification problem. The loss function is cross-entropy as

$$L_{seg} = -\frac{1}{NHW} \sum_n \sum_{h,w} \sum_k t^n_{h,w;k} log(s^n_{h,w;k}), \tag{8}$$

letting $t^n_{h,w;k}$ and $s^n_{h,w;k}$ the ground truth and predicted probability of the pixel for Class $k$.

## IV. Numerical simulations

### A. Measurement setup and dataset

The networks are trained and tested with realistic breast phantoms available in the on-line UWCEM repository [15]. Breast composition varies from person to person and the phantoms in the repository are classified into four classes according to radiographic density: almost entirely fatty, scattered fibroglandular tissue, heterogeneously dense breast, and extremely dense tissue.

Since dense breasts face a higher risk of cancers [16], breast phantoms categorized into the last two classes, denoted as Classes 3 and 4, are used to generate the dataset. Those are tumor-free, so an artificial tumor is added into the glandular part with random location. Its size is set randomly with a radius between $6\,\mathrm{mm}$ and $8\,\mathrm{mm}$ to add variability.

In the simulations, for the EM case, 20 antennas are evenly set on a circle of radius $0.1\,\mathrm{m}$, working as ideal sources and receivers simultaneously at a single $1\,\mathrm{GHz}$ frequency. For the US case, 64 transducers operate at 100 and $200\,\mathrm{kHz}$ on a $0.12$ m-radius circle. The DoI for simulation has a resolution of $0.5\,\mathrm{mm}$, as given in the repository, involving $300 \times 300$ pixels. Scattered fields are obtained by solving the forward problem. Corresponding calculations are run in Matlab. An example of input of the network, $\Re\{\omega\}$, $\Im\{\omega\}$, $\Re\{f\}$, $\Im\{f\}$ as back-propagated is in Fig. 3, emphasizing the absence of direct interpretation of these images in terms of contrasts in particular.

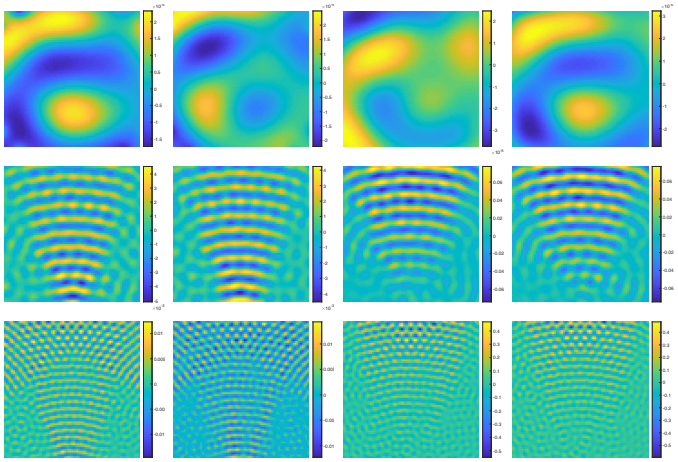

Fig. 3: Input images (produced by the first antenna/transducer) of a sample. Top to bottom: EM case, US case at 100 kHz, US case at 200 kHz. Left to right: $\Re\{\omega\}$, $\Im\{\omega\}$, $\Re\{f\}$, $\Im\{f\}$.

To enlarge the dataset, data augmentation is employed. Different levels of noise are added to the scattered fields from $15\,\mathrm{dB}$ to $30\,\mathrm{dB}$ above the average SNR. The DoI is discretized into $75 \times 75$ pixels to compute the input data $\omega$ and $f$ with equations (4). Tissue type image and distribution of EM and US parameters with same discretization are used as output. Rotation and flipping are operated on both input and output of a sample to generate new data. The whole data set contains 2920 samples, 1536 samples are from Class 3 and 1384 from Class 4, subsequently divided into a 2336 samples training set and a 584 samples test set.

The ranges to normalize the parameters are $[1, 70]$, $[0, 30]$, $[1450, 1650]$ and $[0, 5]$ for $\epsilon_r'$, $\epsilon_r''$, $c$ and $\alpha$, respectively. Here, attenuation at $200\,\mathrm{kHz}$ is predicted. In the segmentation task, pixels are classified into six categories according the tissue type given in the repository, namely background medium, skin, fat, transitional, glandular and tumorous tissue.

### B. Implementation details

All networks are implemented in Keras with Tensorflow backend, run on a laptop with NVIDIA CPU Quadro P620. The networks are trained with ADAM solver with an initial learning rate of $5\ 10^{-4}$ which decays with a factor of $0.15$ each 10 epochs. Batch size is 10, and 60 epochs are run total.

### C. Quantitative assessment

The network performance is assessed by two metrics for different tasks. For segmentation, Intersection-over-Union (IoU) for Class $i$ is calculated as

$$IoU_i = \frac{1}{N} \sum_n \frac{p_{ii}^n + 1}{\sum_{k=1}^K p_{ik}^n + \sum_{k=1}^K p_{ki}^n - p_{ii}^n + 1}, \qquad (9)$$

where $p_{ik}^n$ is the number of pixels in test sample $n$ labeled as Class $i$ but predicted into Class $k$, and $K$ is the number of classes. Note that, in the calculation, a smoothness number of 1 is used. For regression, the relative error is used for evaluation. Given the output of the network $\hat{\tilde{\zeta}}$ and the ranges for normalization, the predicted parameter value $\hat{\zeta}$ can be calculated. The relative error between $\hat{\zeta}$ and the ground truth $\zeta$ is defined as

$$Err = \frac{1}{N} \sum_n \sqrt{\frac{\sum_{w,h} |\hat{\zeta}_{w,h}^n - \zeta_{w,h}^n|^2}{\sum_{w,h} |\zeta_{w,h}^n|^2}}. \qquad (10)$$

Quantitative assessment, IoU, Err and SSIM on the test set are given in Table I, II and III, respectively. To provide better comparisons, a CNN inspired from [17] called Res-Net 50, based on a U-Net model using skip-connections, has also been implemented, since it has been yielding good results for image reconstruction.

TABLE I: IoU of different networks on test set.

| Network | Medium | Skin | Fat | Transition | Gland | Tumor | Average |
|---|---|---|---|---|---|---|---|
| Res-Net | 0.997 | 0.836 | 0.863 | **0.422** | 0.678 | 0.709 | 0.753 |
| CNN-MM | 0.995 | 0.834 | 0.860 | 0.419 | 0.677 | 0.715 | 0.751 |
| SACC-CNN | 0.997 | 0.836 | 0.862 | 0.420 | 0.678 | 0.717 | 0.753 |
| PG-SACC-CNN | **0.999** | **0.838** | **0.866** | **0.422** | **0.680** | **0.718** | **0.755** |

From the quantitative assessment, it can be seen that SAC-Cascaded CNNs to reconstruct both tissue type and parameter value works best on the test set. Physics-Guided SACC-CNN also slightly outperforms SACC-CNN in both classification and reconstruction. One should note that the SSIM results shown here are mean-SSIM from the entire test set and include

TABLE II: Err of different networks on test set.

| Network | $\epsilon_r'$ | $\epsilon_r''$ | $c$ | $\alpha$ |
|---|---|---|---|---|
| Res-Net | 0.213 | 0.195 | 0.004 | 0.134 |
| CNN-MM | 0.260 | 0.265 | 0.005 | 0.135 |
| SACC-CNN | 0.253 | 0.257 | 0.004 | 0.133 |
| PG-SACC-CNN | **0.156** | **0.159** | **0.003** | **0.121** |

TABLE III: Structural Similarity of the different images obtained from networks on test set.

| Network | mean-SSIM | lowest SSIM | highest SSIM |
|---|---|---|---|
| Res-Net | 0.796 | 0.504 | 0.892 |
| CNN-MM | 0.723 | **0.512** | 0.875 |
| SACC-CNN | 0.771 | 0.510 | 0.882 |
| PG-SACC-CNN | **0.810** | 0.511 | **0.903** |

high variations with at least one outlier corresponding to the lowest SSIM reported in Table III. The SSIM difference can be spotted in Fig. 4, where one can see that the contours of the reconstructed parameter $\epsilon_r'$ are continuous for the PG-SACC-CNN whereas they are not for the other ones.

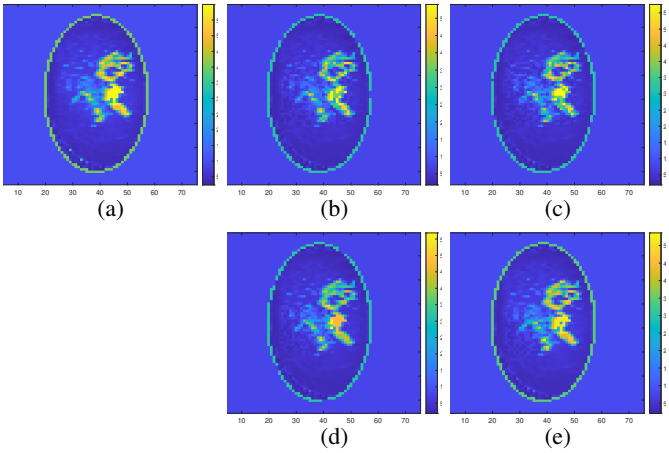


Fig. 4: Ground truth (a) and imaging with (b) CNN-MM, (c) ResNet 50, (d) SACC-CNN, (e) PG-SACC-CNN.

It can be concluded that a SACC-CNN performs only slightly better on this dataset than a CNN-MM, while a Physics-Guided SACC-CNN moderately outperforms both of them and performs better than the Res-Net for the reconstruction part, while leading to slightly better results, yet not so significantly, for the classification part.

Meanwhile, one can achieve real-time reconstruction. In average, with Fast Fourier Transform implemented, the input data of the network can be computed in 0.5681 seconds and it takes 0.2718 seconds to get the prediction results of one sample on Intel Core i7 10875H CPU (2.30 GHz).

## V. Conclusion

Aiming at real-time breast imaging, a Physics Guided SAC-Cascaded CNN structure has been proposed to combine EM and US data. Apart from the regression task to predict parameter values, an auxiliary classifier is used to classify each pixel to achieve model segmentation. The cascaded architecture improves the structural similarity of the image provided, while further implementing a physics-guided loss function reduces the error on the predicted parameter values and enhances the structural similarity of the image.

Many physics-guided loss functions for inverse scattering problems can however be imagined, and investigation of their complementarity in order to create a stronger physics-guided loss function is in need.

## References


[1] G. Baldassarre and B. Belletti, "Molecular biology of breast tumors and prognosis," *F1000Research*, vol. 5, 2016.

[2] Y. Qin, T. Rodet, M. Lambert, and D. Lesselier, "Microwave breast imaging with prior ultrasound information," *IEEE Open Journal of Antennas and Propagation*, vol. 1, pp. 472–482, 2020.

[3] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal, "Bayesian reconstruction of functional images using anatomical information as priors," *IEEE Transactions on Medical Imaging*, vol. 12, no. 4, pp. 670–680, 1993.

[4] L. M. Neira, B. D. Van Veen, and S. C. Hagness, "High-resolution microwave breast imaging using a 3-D inverse scattering algorithm with a variable-strength spatial prior constraint," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 11, pp. 6002–6014, 2017.

[5] A. Li, E. L. Miller, M. E. Kilmer, T. J. Brukilacchio, T. Chaves, J. Stott, Q. Zhang, T. Wu, M. Chorlton, R. H. Moore, *et al.*, "Tomographic optical breast imaging guided by three-dimensional mammography," *Applied Optics*, vol. 42, no. 25, pp. 5181–5190, 2003.

[6] L. A. Gallardo and M. A. Meju, "Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints," *Journal of Geophysical Research: Solid Earth*, vol. 109, no. B3, 2004.

[7] P. Ran, Y. Qin, D. Lesselier, and M. Serhir, "Subwavelength microstructure probing by binary-specialized methods: Contrast source and convolutional neural networks," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 2, pp. 1030–1039, 2021.

[8] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical image segmentation based on u-net: A review," *Journal of Imaging Science and Technology*, vol. 64, no. 2, pp. 20508–1, 2020.

[9] M. Ambrosanio, S. Franceschini, V. Pascazio, and F. Baselice, "Microwave breast imaging via neural networks for almost real-time applications," *arXiv preprint arXiv:2103.12522*, 2021.

[10] Y. Qin, T. Rodet, M. Lambert, and D. Lesselier, "Joint inversion of electromagnetic and acoustic data with edge-preserving regularization for breast imaging," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 349–360, 2021.

[11] Y. Qin, T. Rodet, M. Lambert, and D. Lesselier, "Microwave breast imaging with prior ultrasound information," *IEEE Open Journal of Antennas and Propagation*, vol. 1, pp. 472–482, 2020.

[12] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," *arXiv preprint arXiv:1406.2199*, 2014.

[13] Z. Liu, M. Roy, D. K. Prasad, and K. Agarwal, "Physics-guided loss functions improve deep learning performance in inverse scattering," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 236–245, 2022.

[14] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6765–6816, 2017.

[15] E. Zastrow, S. Davis, M. Lazebnik, F. Kelcz, B. Van Veen, and S. Hagness, "Database of 3D grid-based numerical breast phantoms for use in computational electromagnetics simulations," tech. rep., Department of Electrical and Computer Engineering University of Wisconsin-Madison, 2008.

[16] V. A. McCormack and I. dos Santos Silva, "Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis," *Cancer Epidemiology and Prevention Biomarkers*, vol. 15, no. 6, pp. 1159–1169, 2006.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Las Vegas, United States), pp. 770–778, 2016.