2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)

DCE-MRI Breast Lesions Segmentation with a 3TP U-Net Deep Convolutional Neural Network

1st Gabriele Piantadosi
DIETI
University of Naples Federico II
Naples, Italy
gabriele.piantadosi@unina.it

2nd Stefano Marrone
DIETI
University of Naples Federico II
Naples, Italy
stefano.marrone@unina.it

3rd Antonio Galli
DIETI
University of Naples Federico II
Naples, Italy
anto.galli@studenti.unina.it

4th Mario Sansone
DIETI
University of Naples Federico II
Naples, Italy
mario.sansone@unina.it

5th Carlo Sansone
DIETI
University of Naples Federico II
Naples, Italy
carlo.sansone@unina.it

***Abstract*—Nowadays, Dynamic Contrast Enhanced-Magnetic Resonance Imaging (DCE-MRI) is increasingly succeeding as a complementary methodology for breast cancer, with Computer Aided Detection/Diagnosis (CAD) systems becoming essential technological tools to provide early detection and diagnosis of tumours. Several CADs make use of machine learning, resulting in a constant design of hand-crafted features aimed at better assisting the physician. In recent years, Deep learning (DL) approaches raised in popularity in many pattern recognition tasks thanks to their ability to learn compact hierarchical features that well fit the specific task to solve. If, on one hand, this characteristic suggests to explore DL suitability for biomedical image processing, on the other, it is important to take into account the physiological inheritance of the images under analysis. With this goal in mind, in this work we propose "3TP U-Net", an U-Shaped Deep Convolutional Neural Network that exploits the well-known Three Time Points approach for the lesion segmentation task. Results show that our proposal is able to outperform not only the classical (non-deep) approaches but also some very recent deep proposal, achieving a median Dice Similarity Coefficient of 61.24%.**

I. Introduction

Breast cancer is the tumour second cause of death in the USA [1], with early detection resulting to be the most important factor for a positive prognosis. World Health Organization suggests mammography as the main breast cancer screening methodology for its fast processing and high diagnostic value [2], but, unfortunately, this methodology is not suitable for under forty women (showing hyperdense glandular tissues). Over the time, researchers have been focusing on Dynamic Contrast Enhanced-Magnetic Resonance Imaging (DCE-MRI) as a complementary tool for early detection of breast cancer, demonstrating its potential both for staging newly diagnosed patients and in assessing therapy effects [3].

DCE-MRI consists in the temporal acquisition of 3D volumes before (pre) and after (post) the intravenous injection of a paramagnetic contrast agent, such as Gadolinium-based, resulting in 4-dimensional data (Fig. 1).

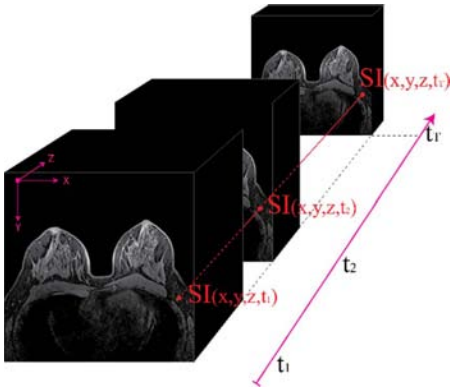


Fig. 1: A representation of the four dimensions (3 spatial + 1 temporal) of a typical breast DCE-MRI. In red, en exemplification showing a voxel (of coordinates x, y, z) acquisition over different time points (t_1 to t_T).

For each voxel, a Time Intensity Curve (TIC) is obtained. The TIC shows the absorption and the release of the contrast agent (Fig. 2) over time following the vascularisation characteristics of the tissue under analysis [4].

To take advantage of the huge amount of data produced by DCE-MRI, radiologists usually make use of Computer-Aided Detection and Diagnosis (CAD) systems, namely tools designed i) to assist in the detection and in the diagnosis of tumour lesions and ii) to reduce the inter- and intra-observer variability [5]. CAD usually consists of several modules to face different tasks, such as lesion detection, segmentation, diagnoses, etc. Among all, performing an accurate lesion segmentation is very important in order to provide the lesion classification module with input as free as possible of tissues not belonging to the lesion. This task is usually not trivial, time-consuming and error-prone and, therefore, over the years several papers proposed to face it by making use of machine

2372-9198/19/$31.00 ©2019 IEEE
DOI 10.1109/CBMS.2019.00130

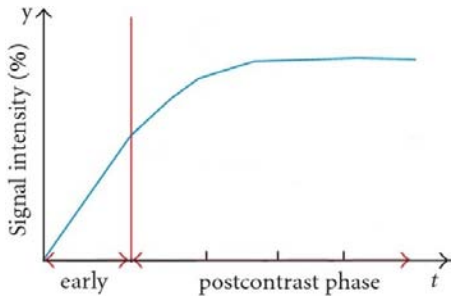


Fig. 2: Illustration of a Time Intensity Curve for a voxel: the t axes represents the different acquisitions along time, highlighting the pre-contrast (early) and postcontrast injection phases; the y axes reports the acquired signal (and thus the voxel luminance) variation.

learning. Although this allowed us to develop more accurate automated procedures to assist the physician in the lesion segmentation task, proposed works still were not able to identify the definitive set of features and segmentation model.

In recent years, Deep learning (DL) approaches raised in popularity in many pattern recognition tasks thanks to their ability to learn compact hierarchical features that well fit the specific task to solve. If, on one hand, this characteristic suggests exploring DL suitability for biomedical image processing, on the other, it is important to take into account the physiological inheritance of the images under analysis.

The aim of this work is to exploit Deep Learning for the automatic breast Lesion Segmentation task in DCE-MRI while taking advantage of past learned experience [6] by training the proposed architecture on images acquired at very specific time points. In particular, we propose **3TP U-Net**, an U-Shaped Deep Convolutional Neural Network [7] that exploits the well-known Three Time Points approach [8] to improve lesion segmentation performances.
The paper's sections are as follows: Section II outlines some related works; Section III describes the proposal, also introducing the dataset and the evaluation modality used; Section IV reports our results compared with those obtained by using some literature proposals; finally, Section V discusses the obtained results and provides some conclusions.

II. Related Works

The lesion segmentation task has been addressed in several ways, with some authors proposing fully automated approachs and other focusing only on semi-supervised ones [9]. Over the years several papers showed that the most effective way to perform the lesion segmentation by means of machine learning was by using Dynamic Feature, a set of characteristics modelling the shape of the TIC. More recently, after a Deep Convolution Neural Network (D-CNN or simply CNN) [10] won the 2012 Large Scale Visual Recognition Challenge [11] (a popular 1000 classes image classification contest), an increasing interest has starting to be paid by researches on the study and on the application of CNN in biomedical image processing [12], [13].

Focusing on breast lesion segmentation task, recently a work [14] faced the problem with deep learning, proposing a stacking of three parallel ConvLSTM [15] networks (to extract temporal and 3D features) over a 4-layer U-Net (to perform the segmentation). To the best of our knowledge, this is the only work using CNN for breast lesion segmentation proposed so far. However, since the task can be considered as a *semantic segmentation* in which the input image has to be divided into Region of Interests (ROIs) each referring to a lesion, every CNN segmentation model could be potentially used.

One of the first work that addressed semantic segmentation with CNN was SegNet [16], a deep convolutional encoder-decoder architecture, followed by a pixel-wise classifier. The role of the encoder network is to learn a compact representation of the input data, while the role of the decoder network is to map the encoded features to a segmentation mask. Similarly, the U-Net [7] model exploits an encoder-decoder architecture, enhanced by the presence of skipping connections between the two sides with the aim of i) exploiting encoding information to improve the decoding stage and ii) to reduce the gradient vanishing problem. Later in time, combining the ideas of MobileNets [17] Depthwise Separable Convolutions with U-Net, in 2017 MobileU-Net[1] was developed with the aim to build a high speed, low parameter semantic segmentation model. More recently, DeepLab [18] proposes to improve semantic segmentation by exploiting dilated (atrous) convolution to control the resolution at which feature responses are computed (DeepLabV1), Atrous Spatial Pyramid Pooling (DeepLab2), batch normalization (DeepLabv3) and depthwise separable convolutions (DeepLabv3+).

III. Material and Methods

The problem we want to face consists in the segmentation of 3D lesions in breast DCE-MRI. As aforementioned, over the years it has been shown the importance of intensity variations due to contrast agent flowing, with authors designing new dynamic features or exploiting deep networks specifically designed to learn temporal relations.

In this work we take into account this DCE-MRI fundamental characteristic by properly feeding the segmentation CNN with images acquired at very specific time points: the idea is to perform the segmentation slice-by-slice, considering the different acquisitions (along time) of the same slice as channels within the image. In particular, since DCE-MRI voxels are usually highly anisotropic, we choose to extract slices along the projection with the higher resolution.

A. 3TP U-Net

The proposed approach consists of three main steps:

- Removing all the foreign tissues (bones, muscles, etc) and air background, by using a proper breastmask [19]
- Extraction of *temporal slices*

[1] https://github.com/GeorgeSeif/Semantic-Segmentation-Suite

Fig. 3: Proposed segmentation schema.

Fig. 4: 3TP Slice creation: slices extracted from the pre-contrast (t_0), from 2 (t_1) and 6 minutes (t_2) after the CA injection volumes are fused in a 3 channels image.

- Slice-by-slice segmentation with a U-Net CNN

The core of the 3TP U-Net are the last two stages (Fig. 3). In fact, while the use of a breastmask is mainly needed to reduce noise in the data, allowing the net to focus only on the breast tissues, the choice of combining some well defined temporal acquisitions as channels of the same image is proposed to i) make the approach more general while ii) feeding the net with data that can effectively synthesize the whole contrast agent course. This last assertion is sustained by the studies conducted by Degani [8] that showed how breast lesion analysis can be improved by focusing on just three well-defined temporal acquisitions: t_0 (pre-contrast), t_1 (2 minutes after CA injection) and t_2 (6 minutes after CA injection). Therefore, in order to exploit this finding, for each slice we create the equivalent three channels image by combining the slices picked at the temporal instants proposed by the Three Time Points (3TP) method, as first, second and third channel of the same image (Fig. 4).

The so obtained images are fed to a U-Shaped CNN (Fig. 5), an encoder-decoder architecture originally designed for biomedical electron microscopy (EM) images multi-class pixel-wise semantic segmentation [7]. The original U-Net has been modified by setting output feature-map to one channel allowing a faster convergence during the training step and by applying zero-padding with a size-preserving strategy in order to maintain the output shapes. Moreover, batch normalization after each convolution is introduced with the aim of making training step more stable.

Each side of the net consists in the repeated application of some layers. In particular, the "descending" super-layer structure consists of:

- 2D Convolution (3x3 kernel, zero-padding, stride 1)
- Rectified Linear Unit (ReLU) activation
- Batch Normalization
- 2D Convolution (3x3 kernel, zero-padding, stride 1)
- Rectified Linear Unit (ReLU) activation
- Batch Normalization
- Pooling (max-pool, stride 2)

with each super-layer performing a down-sampling step halving the size of the image (using the max-pool operation) and doubling the size of the feature channels. Similarly, the right side performs the expansive path, with the "ascending" super-layer structure consisting of:

- 2D Up-Convolution (2x2 kernel, zero-padding, stride 1)
- Concatenation with feature-map from the same level
- 2D Convolution (3x3 kernel, zero-padding, stride 1x1)
- Rectified Linear Unit (ReLU) activation
- Batch Normalization
- 2D Convolution (3x3 kernel, zero-padding, stride 1)
- Rectified Linear Unit (ReLU) activation
- Batch Normalization

The Up-Convolution is used to halve the feature channels by mean of a trainable kernel. Then, the 2D Convolution (1x1 kernel, zero-padding, stride 1) maps each of the 64 output features to the network output. Finally, a probabilistic output is obtained by using the Sigmoid activation function.

B. Dataset

The proposal has been evaluated by using a private dataset composed of 35 women breast DCE-MRI 4D data (average age 40 years, in range 16-69) with benign or malignant lesions histopathologically proven.

All patients underwent imaging with a 1.5T scanner (Magnetom Symphony, Siemens Medical System, Erlangen, Germany) equipped with breast coil. DCE T1-weighted FLASH 3D coronal images were acquired (TR: 9.8ms, TE: 4.76 ms; FA: 25 degrees; FoV 370x185 mm^2; Image: 256x128; Thickness: 2 mm; Gap: 0; Acquisition time: 56s; 80 slices spanning entire breast volume). One series (t_0) was acquired before intravenous injection and 9 series (t_1-t_9) after. In particular, the intravenous injection consists of 0.1 mmol/kg of a positive paramagnetic contrast agent (gadolinium-diethylene-triamine penta-acetic acid, Gd-DOTA, Dotarem, Guerbet, Roissy CdG Cedex, France). In order to perform the injection, an automatic system was used (Spectris Solaris EP MR, MEDRAD, Inc.,Indianola, PA) and the injection flow rate was set to 2 ml/s followed by a flush of 10 ml saline solution at the same rate. The lesions ground-truth was manually carried out by.

C. Experimental Setup

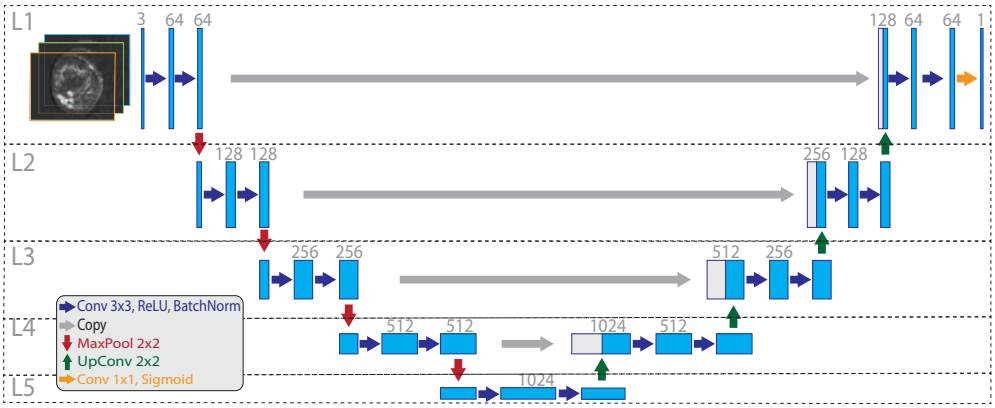
The proposed CNNs have been evaluated using the high-level neural networks API Keras (Python 3.6) with TensorFlow 1.6 as back-end. Python scripts have been executed on a physical server hosted in our university HPC center[2] equipped with 2 x Intel(R) Xeon(R) Intel(R) 2.13GHz CPUs (4 cores), 32GB RAM and a Nvidia Titan Xp GPU (Pascal family) with 12GB GRAM. In order to train the proposed U-Net model for Lesion segmentation, a task-specific loss has been minimized:

$$\text{loss} = 1 - \text{DSC}(y_{\text{network}}, y_{\text{gold}_{\text{s}}\text{tandard}}) \quad (1)$$

$$\text{DSC} = (2 \cdot n(GS \cap SEG))/(n(GS) + n(SEG)) \quad (2)$$

[2]http://www.scope.unina.it

Fig. 5: The 3TP U-Net model for semantic segmentation of a breast slice in DCE-MRI. The left side implements the contracting path, where the spatial-sizes (represented by the filters receptive field and by the output sizes) decrease and the feature-size increases. The right side implements the expansive path, with the aim of increasing the image sizes.



where *DSC* is the Dice Similarity Coefficient, calculated considering the number of voxels per each volume $n(\cdot)$. The network kernel weights have been initialized from a random standard distribution $\mathcal{N}(0, \sqrt{2/(\mathrm{fan_i} + \mathrm{fan_o})})$ [20], where fan_i and fan_o are respectively the input and output size of the convolution layer, while the bias weights have been initialized to a constant value of 0.1. ADAM optimizer [21] was used, with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\mathrm{lr} = 0.001$ using an inverse time decay strategy. All the techniques were compared in terms of per-patient DSC value, since it is a simple and useful summary measure of spatial overlap widely used in biomedical image segmentation [22].

IV. Results

This section reports the results of the described approach compared with some deep and non-deep literature proposals. To evaluate the effectiveness of the proposed approach, our results were compared with those obtained by applying the algorithms described in Section II, by using a 10-fold Cross Validation (CV) ensuring that, to obtain a reliable and fair evaluation, the slices from the same subject are always separated across the CV folds. Given the characteristics of the used dataset, slices were extracted along the coronal projection. For the sake of completeness, we also considered three non-deep approaches based on dynamic features: the one from Fusco et al. [23] that proposed to use a multilayer perceptron (MLP); the approach proposed by Torricelli et al. [24] that suggests to use a simple threshold on the signal enhancement; an our previous work [25] in which we proposed to use a Support Vector Machine trained on features selected by an experienced radiologist. All the algorithms have been reimplemented and evaluated to the best of our interpretation of the authors' papers to guarantee a fair comparison. Finally, before comparing with other approaches, we analyzed the impact that the 3TP input has on the segmentation model.

Table I reports results obtained when all the 10 temporal acquisition was used (10 channels image were created and the U-Net was modified accordingly): results show that the proposed approach improves the lesion segmentation effectiveness by better enhancing the dynamic course of the contrast agent while reducing the noise impact.

Table II compares our best result with respect to the methods depicted in section III-C. Results show that deep-based approaches always outperform non-deep ones, with our 3TP U-Net performing better in terms of median DSC. Figure 7 directly compares the results from each approach to better highlight the benefit or our proposal with respect to the competitors. It is worth noticing that the ConvLSTM-based approach proposed in [15] is not reported here since, despite our best efforts, we were not able to make the proposed architecture converged on our dataset.

Finally, in order to allow a visual comparison between the results obtained by using the approaches reported in Table II, Figure 6 shows segmented ROIs for a slice of one of the patients. As a reference, also the ground truth has been reported.

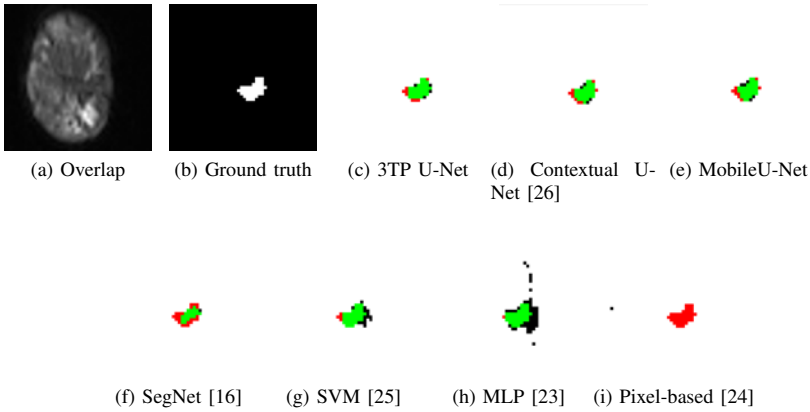
V. Discussions and Conclusions

The aim of this work was to *exploit Deep Learning for the automatic breast Lesion Segmentation* task in DCE-MRI, *while taking advantage of a past learned experience* by training the proposed architecture on images acquired at very specific time points. To this aim, we proposed **3TP U-Net**, an U-Shaped Deep Convolutional Neural Networks [7] with the well-known Three Time Points (3TP) approach [8] suitable exploited in order to improve lesion segmentation performances.

The choices made were not solely guided by the enthusiasm for deep learning, but the result of precise analysis of its ability to learn compact hierarchical features that well fit the specific task to solve, relieving the domain experts of continuously designing task-oriented hand-crafted features, particularly in domains lacking effective expert-designed ones [27]. Similarly, the 3TP approach usage allows us to take into account for the past discoveries about DCE-MRI automated analysis, in a

TABLE I: Proposed approach variants comparison. DSC median values and relative 95% confidence intervals are reported (obtained with a 10-folds CV).

Input Timepoints	DSC [%]	ACC [%]	SPE[%]	SEN[%]
10TP	57,69%±15,71%	**99,99%±0,01%**	100,00%	53,93%±19,44%
3TP	**61,24%±11,84%**	99,98%±0,02%	100,00%	**68,28%±9,73%**



(a) Overlap (b) Ground truth (c) 3TP U-Net (d) Contextual U-Net [26] (e) MobileU-Net

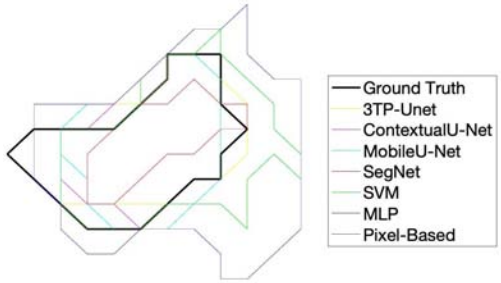
(f) SegNet [16] (g) SVM [25] (h) MLP [23] (i) Pixel-based [24]

Fig. 6: The final results of a zoomed pre-contrast slice from the input MRI volume (a). In particular (b) represents the ground truth, and (c)-(i) the results of the approaches in Table II. The colours represent the True Positive (Green), True Negative (White), False Negative (Red) and False Positive (Black) voxels.

TABLE II: Lesion segmentation results obtained with a 10-folds CV. Median values and relative 95% confidence intervals are reported (obtained with a 10-folds CV).

Authors	Method	DSC [%]
Our Proposal	3TP U-Net	**61,24%±11,84%**
Contestual U-Net [26]	Contestual U-Net	57,69%±12,30%
MobileU-Net	MobileU-Net	47,10%±10,90%
Badrinarayanan et al. [16]	SegNet	31,60%±15,27%
Marrone et al. [25]	SVM	19,07%±10,28%
Fusco et al. [23]	MLP	02,91%±01,51%
Torricelli et al. [24]	Pixel-based	02,80%±02,50%

Fig. 7: Segmentation results of the approaches in Table II overlapped and compared with the ground truth.



simple, but effective way. In literature, most authors focusing only on lesion detection, localization or classification [28], [29] and, to the best of our knowledge, there is only a work that also faced the lesion segmentation in breast DCE-MRI by using a deep approach [14]. Differently from that work, in which authors propose a stacking of three parallel ConvLSTM [15] networks (to extract temporal and 3D features) over a 4-layer U-Net (to perform the segmentation), our *3TP U-Net* not only is *simpler* (since there are no additional nets to train for learning temporal features), but allows a more *strict control* on what data will be used to perform the segmentation, by picking only some specific acquisitions from the original 4D volume. Moreover, this last point lays the foundations to define a protocol independent approach, since the proposed *3TP U-Net* just needs 3 volumes acquisition closer to very specific time intervals (defined in minutes) related to the contrast agent injection. Results show that the 3TP U-Net operating slice-by-slice over the 4D DCE-MRI volume can be effectively used to improve the lesion segmentation process. In particular, Table I not only shows that using "Three Time Points" allows us to strongly improve performance, but also the benefits of the breastmask to attenuate noise caused by extraneous voxels. Table II provides an insightful perspective on the lesion segmentation task, showing that non-deep approaches always perform worse with respect to deep ones. It is worth noticing that the wide confidence intervals reported in tables I and II, are mostly affected by the reduced number of patients considered in this study. Therefore, although results demonstrate the effectiveness of the proposed approach, in order to provide clearer insights with narrower dispersion indexes, it is very important to increase the number of involved subjects. Future works will evaluate to what extent the proposed 3TP U-Net is robust to multi-protocol and multi-tissue MRIs. Moreover, in order to improve the obtained results, we are planning to extend the proposed approach to other scanning procedures for detection and diagnosis of tumours (i.e. DWI, PET or CT) by

exploiting the ability of last generation scanner to acquire more than one kind of images at once (for example PET-TC or PET-MRI). On the other hand, it is worth noting that when using images from different tools, the best results can be achieved by relying on a very robust co-registration technique and a deep knowledge of the most descriptive characteristics of each study. Finally, we are considering to improve the obtained results by mixing the three orthogonal acquisition planes in a novel 3D approach and by considering a deeper network (more than 5 layers).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research, the availability of the Calculation Centre SCoPE of the University of Naples Federico II and his staff. The authors are also grateful to Dr. Antonella Petrillo, Head of Division of Radiology and PhD Roberta Fusco, Department of Diagnostic Imaging, Radiant and Metabolic Therapy, "Istituto Nazionale dei Tumori Fondazione G. Pascale" of Naples, Italy, for providing data. This work is part of the Synergy-net: Research and Digital Solutions against Cancer project (funded in the framework of the POR Campania FESR 2014-2020).

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA: a cancer journal for clinicians*, vol. 68, no. 1, pp. 7–30, 2018.

[2] R. A. Smith, D. Saslow, K. A. Sawyer, W. Burke, M. E. Costanza, W. P. Evans, R. S. Foster, E. Hendrick, H. J. Eyre, and S. Sener, "American cancer society guidelines for breast cancer screening: update 2003," *CA: a cancer journal for clinicians*, vol. 53, no. 3, pp. 141–169, 2003.

[3] J. Levman, T. Leung, P. Causer, D. Plewes, and A. L. Martel, "Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines," *IEEE Transactions on Medical Imaging*, vol. 27, pp. 688–696, 2008.

[4] S. Sourbron and D. L. Buckley, "Tracer kinetic modelling in mri: estimating perfusion and capillary permeability," *Physics in Medicine & Biology*, vol. 57, no. 2, p. R1, 2011.

[5] J. J. Fenton, S. H. Taplin, P. A. Carney, L. Abraham, E. A. Sickles, C. D'Orsi, E. A. Berns, G. Cutter, R. E. Hendrick, W. E. Barlow, and J. G. Elmore, "Influence of computer-aided detection on performance of screening mammography." *The New England journal of medicine*, vol. 356, no. 14, pp. 1399–409, apr 2007.

[6] D. Gragnaniello, G. Poggi, C. Sansone, , and L. Verdoliva, "Biometric spoofing detection by a domain-aware convolutional neural network," in *2016 12th International Conference on Signal-Image Technology and Internet-Based Systems*. IEEE, 2016, pp. 193–198.

[7] 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.

[8] H. Degani, V. Gusis, D. Weinstein, S. Fields, and S. Strano, "Mapping pathophysiological features of breast tumors by mri at high spatial resolution," *Nature medicine*, vol. 3, no. 7, pp. 780–782, 1997.

[9] G. Piantadosi, S. Marrone, R. Fusco, M. Sansone, and C. Sansone, "Comprehensive computer-aided diagnosis for breast t1-weighted dce-mri through quantitative dynamical features and spatio-temporal local binary patterns," *IET Computer Vision*, vol. 12, no. 7, pp. 1007–1017, 2018.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," 2012.

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[12] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[13] S. Marrone, G. Piantadosi, R. Fusco, A. Petrillo, M. Sansone, and C. Sansone, "An investigation of deep learning for lesions malignancy classification in breast dce-mri," in *International Conference on Image Analysis and Processing*. Springer, 2017, pp. 479–489.

[14] M. Chen, H. Zheng, C. Lu, E. Tu, J. Yang, and N. Kasabov, "A spatio-temporal fully convolutional network for breast lesion segmentation in dce-mri," in *International Conference on Neural Information Processing*. Springer, 2018, pp. 358–368.

[15] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in neural information processing systems*, 2015, pp. 802–810.

[16] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.

[17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[18] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *arXiv preprint arXiv:1802.02611*, 2018.

[19] G. Piantadosi, M. Sansone, and C. Sansone, "Breast segmentation in mri via u-net deep convolutional neural networks," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 3917–3922.

[20] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[21] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[22] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells III, F. A. Jolesz, and R. Kikinis, "Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports," *Academic radiology*, vol. 11, no. 2, pp. 178–189, 2004.

[23] R. Fusco, M. Sansone, A. Petrillo, and C. Sansone, "A multiple classifier system for classification of breast lesions using dynamic and morphological features in dce-mri," in *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer, 2012, pp. 684–692.

[24] P. Torricelli, A. Pecchi, G. Luppi, and R. Romagnoli, "Gadolinium-enhanced mri with dynamic evaluation in diagnosing the local recurrence of rectal cancer," *Abdominal imaging*, vol. 28, no. 1, pp. 0019–0027, 2003.

[25] S. Marrone, G. Piantadosi, R. Fusco, A. Petrillo, M. Sansone, and C. Sansone, "Automatic Lesion Detection in Breast DCE-MRI," in *Image Analysis and Processing (ICIAP)*. Springer Berlin Heidelberg, 2013, pp. 359–368.

[26] Z. Guo, L. Zhang, L. Lu, M. Bagheri, R. M. Summers, M. Sonka, and J. Yao, "Deep LOGISMOS: deep learning graph-based 3d segmentation of pancreatic tumors on CT scans," *CoRR*, 2018.

[27] Q. V. Le, W. Y. Zou, S. Y. Yeung, and A. Y. Ng, "Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3361–3368.

[28] I. A. Illan, J. Ramirez, J. Gorriz, M. A. Marino, D. Avendano, T. Helbich, P. Baltzer, K. Pinker, and A. Meyer-Baese, "Automated detection and segmentation of nonmass-enhancing breast tumors with dynamic contrast-enhanced magnetic resonance imaging," *Contrast Media & Molecular Imaging*, vol. 2018, 2018.

[29] A. Gubern-Mérida, R. Martí, J. Melendez, J. L. Hauth, R. M. Mann, N. Karssemeijer, and B. Platel, "Automated localization of breast cancer in dce-mri," *Medical image analysis*, vol. 20, no. 1, pp. 265–274, 2015.