Edinburgh Research Explorer

Dual Convolutional Neural Networks for Breast Mass Segmentation and Diagnosis in Mammography

Dual Convolutional Neural Networks for Breast Mass Segmentation and Diagnosis in Mammography

Heyi Li, Dongdong Chen, William H. Nailon, Mike E. Davies *Fellow, IEEE*, and David Laurenson

arXiv:2008.02957v2 [eess.IV] 11 Aug 2020

Abstract—Deep convolutional neural networks (CNNs) have emerged as a new paradigm for Mammogram diagnosis. Contemporary CNN-based computer-aided-diagnosis (CAD) for breast cancer directly extract latent features from input mammogram image and ignore the importance of morphological features. In this paper, we introduce a novel deep learning framework for mammogram image processing, which computes mass segmentation and simultaneously predict diagnosis results. Specifically, our method is constructed in a dual-path architecture that solves the mapping in a dual-problem manner, with an additional consideration of important shape and boundary knowledge. One path called the Locality Preserving Learner (LPL), is devoted to hierarchically extracting and exploiting intrinsic features of the input. Whereas the other path, called the Conditional Graph Learner (CGL) focuses on generating geometrical features via modeling pixel-wise image to mask correlations. By integrating the two learners, both the semantics and structure are well preserved and the component learning paths in return complement each other, contributing an improvement to the mass segmentation and cancer classification problem at the same time. We evaluated our method on two most used public mammography datasets, DDSM and INbreast. Experimental results show that DUALCORENET achieves the best mammography segmentation (in both high and low resolution) and classification simultaneously, outperforming recent state-of-the-art models.

Index Terms—Mammography Diagnosis, Dual-Path Network, Deep Learning

I. INTRODUCTION

BREAST cancer, according to the International Agency for Research on Cancer [1], is the most frequently diagnosed cancer. Screening mammography is widely employed and has shown its significance especially for invasive breast tumours when they are too small to be palpable or cause symptoms. The manual inspection of a mammogram typically requires the lesion's identification as either benign or malignant, and sometimes the according delineation. However, the manual inspection is tedious, subjective, and prone to errors [2]–[4]. Striving for the optimal health care, mammographical computer aided diagnosis (CAD) systems are designed as an alternative to a double reader, aiming to achieve similar inspection results to that of human experts (Fig. 1).

Many conventional machine learning algorithms have been proposed to tackle this problem, which typically comprises

E-mail: {heyi.li, d.chen}@ed.ac.uk. H. Li, D. Chen, M. Davies and D. Laurenson are with the School of Engineering, University of Edinburgh, UK. W. Nailon is with Oncology Physics Department, Edinburgh Cancer Centre, Western General Hospital, Edinburgh, UK. D. Chen and M. Davies are funded by the ERC Advanced grant 694888, C-SENSE.
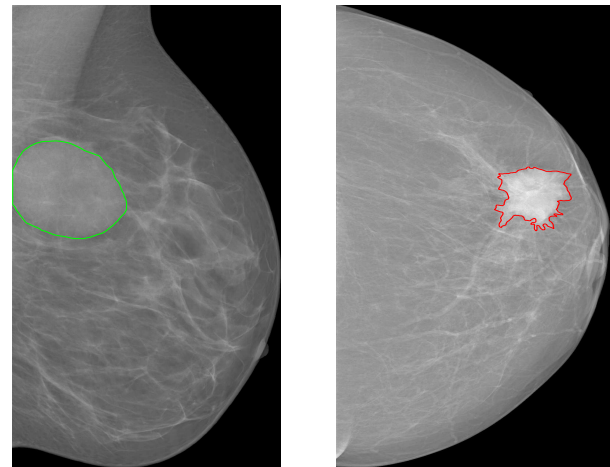
(a) Annotation of a benign mass   (b) Annotation of a malignancy

Fig. 1. Manual inspection examples for breast masses in full field digital mammography (FFDM) INBreast dataset [5]. (a) contains a benign mass delineated with green lines, which is of oval shape and circumscribe boundaries in a CC view. (b) shows a malignant mass in red lines captured in a MLO view, which is of irregular shape and spiculated boundaries.

various image processing operations (such as image segmentation, feature extraction, feature selection, and classification). The performance of a conventional CAD often relies heavily on cumbersome hand-engineered features [6], which are subsequently introduced into various classifiers. Oliver et al. [2] has demonstrated in their review paper that an accurate segmentation is the foundation of subsequent cancerous diagnosis, since the likelihood of malignancy depends on the shape and margin of lesions [7]. This statement has been empirically verified by a number of works [4, 6, 8, 9], which all claim that the most accurate breast mass diagnosis was obtained by the shape-related descriptors when compared with other conventional hand-crafted features. In fact, traditional machine learning algorithms are still popular in recent commercial CADs. However, there is significant room for improvement, especially for breast mass diagnosis.

In recent years, leveraging the insights from the success of deep neural networks (deep learning) [10] in computer vision tasks [11]–[18], a noticeable shift to deep learning based CADs has been seen. Some works have proposed the use of extracting segmentation-related features by CNNs with radiologists' pixel-level annotations, in order to further improve automatic diagnosis performance [19, 20]. However,
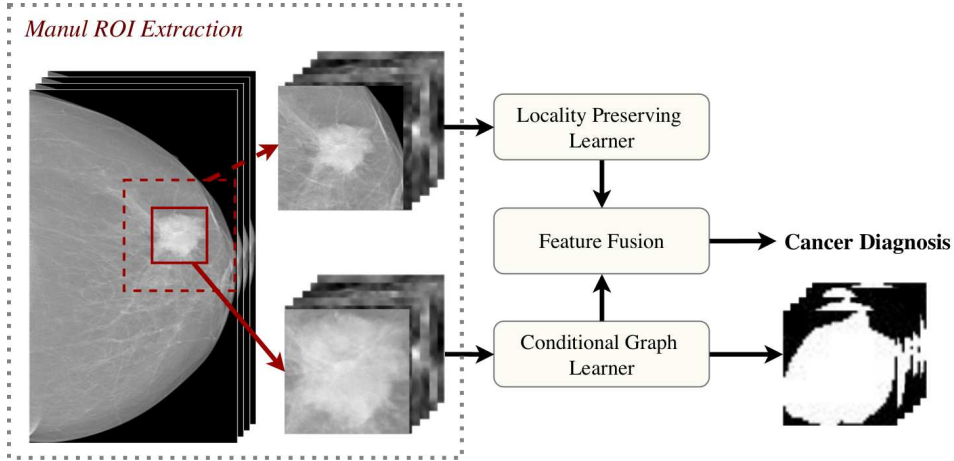
Fig. 2. The flow diagram of proposed DUALCORENET. With the extracted multi-scale ROIs as the input of LPL and CGL path separately, the DUALCORENET outputs both segmentation mask and diagnosis label.

this method requires large volume of accurate pixel-wise annotations, which are very difficult to obtain in practice. In order to enhance the network without using binary mask labeling, some authors explored the performance with an automatic segmentation algorithm [20]. Yet this automatic setup has caused a considerable performance drop. The poor performance is likely caused by the multi-stage process training. Based on these observations we are motivated to construct a CNN architecture trained in an end-to-end fashion, in order to *jointly* solve the breast cancer diagnosis (benign vs malignant) *and* the segmentation problem in mammography.

In this paper, we presented a multi-scale dual-path CNNs as shown in Fig. 2, to solve the image to diagnosis label mapping in a dual-problem manner. In particular, the dual-problem here especially refers to the segmentation and classification problem. A preliminary version of this work appeared in [21]. This paper extends [21] by providing a more detailed description of the work and more comprehensive experimental evaluations. Based on the accurate breast mass segmentation algorithm presented in [22] and the related breast mass classifiers [17, 19, 23]–[25], a **Dual**-path **Co**nditional **Re**sidual **Net**work (DUALCORENET) for mammography analysis is introduced. Firstly, a mass and its context texture learner called the *Locality Preserving Learner (LPL)* is built with stacks of convolutional blocks, achieving a mapping from relative large scale ROIs to class labels. Secondly, an integrated graphical and CNN model, called the *Conditional Graph Learner (CGL)* is employed to learn the relative small scale ROI to mask correlation, and the extracted segmentation features will be further used to improve the final mass classification performance. Additionally, we train the model with multi-scale ROIs, since the surrounding tissues of breast masses play a pivotal role for the accurate cancer diagnosis, whereas the contextual tissues are less irrelevant for the segmentation task. For a certain breast mass, a larger scale ROI is used as the input of LPL for richer features extraction, and a smaller scale ROI is employed by the CGL path. DUALCORENET achieves the best mammography segmentation and classification simultaneously, outperforming recent state-of-the-art models. The main contributions of this paper are the following:

1) To our knowledge, DUALCORENET is the first fully automatic dual-path CNN-based mammogram analysis model that takes advantage of an automatic segmented mask for mass classification;
2) Our method has achieved the best performance for breast mass segmentation in both low and high resolutions;
3) DUALCORENET has achieved comparable results with mass classification tasks on publicly available mammography datasets.

Organization. The rest of the paper is arranged as follows: Section II presents the related preliminary techniques, Section III introduces the proposed DUALCORENET methodology, and Section IV shows the experimental results. We conclude this paper in Section V.

II. PRELIMINARY

In this section, we will introduce three machine learning methods, which are related to our proposed DUALCORENET. Firstly, we will discuss the residual learning and the inception modules separately, which have improved the deep model generalization in CNNs. Then the conditional random Field (CRF), a type of graphical model, will be discussed for the medical image segmentation.

A. Residual Learning

Residual learning [26] is proofed a efficient way to accelerate the neural network training and avoid the gradient vanishing/exploding problems, which had been extensively applied method for myriad computer vision tasks. The main idea of of residual learning is the use of residual connections between the input and output to the neural network internal layers or blocks or even entire network. In particular, by letting the desired input to output mapping in a residual module as $\mathcal{H}(\mathbf{x})$, the residual function obtained in each module is defined with:

$$\mathcal{H}(\mathbf{x}) := \mathbf{x} + \mathcal{F}(\mathbf{x}). \tag{1}$$
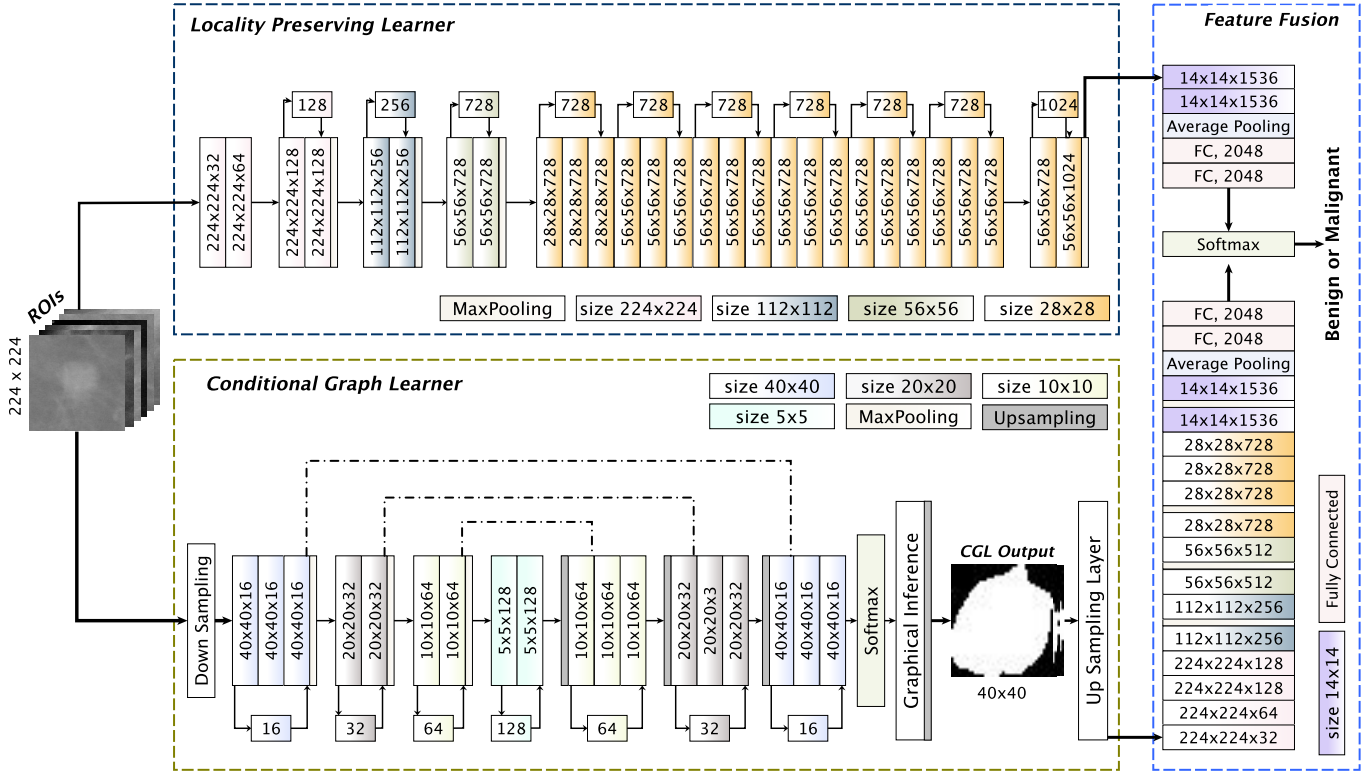
Fig. 3. Overall architecture of our proposed DUALCORENET architecture.

The gradients of a residual module in each layer are therefore pre-conditioned to be close to the identity function, solving the gradients vanishing problem. In this way, CNNs can be constructed with many more layers with efficacious training. Specifically, residual connections enable the deep neural network to learn the residual between the input and output of a module, instead of directly learning the mapping.

B. Conditional Random Fields

Conditional Random Fields (CRF) [27], as a variant of Markov Random Fields (MRFs), incorporates the label consistency with similar pixels and provides sharp boundary and fine-grained segmentation. Typically a CRF formulates the label assignment problem as a probabilistic inference problem, in which pixel labels are modeled as MRF random variables and conditioned upon the observations [27]. Given an observed deep latent feature $h(\mathbf{x})$, the Gibbs distribution of a fully connected CRF with $\mathcal{V}$ nodes and $\mathcal{E}$ edges is defined as:

$$p(\mathbf{y}|h(\mathbf{x})) = \frac{1}{Z(h(\mathbf{x}))} \exp\big(-\sum_{c \in \mathcal{C}} \phi_c(h(\mathbf{x}), y_c)\big), \quad (2)$$

where $\mathbf{y}$ is pixel-level label state, c is the one item in the clique set $\mathcal{C}$, y_c is the clique joint label, ϕ_c is the feature vector, and Z is the partition function.

Recently, due to its flexibility and efficiency, CRF has been applied on medical images for various segmentation problems [28, 29] in which the CRF was mainly used as a post processing step which usually lead to an inherent shrinking bias problem [30]. In contrast, in this paper, the CRF is used

as a specific neural network layer that comprised of the unary potentials and the pair-wise potentials connecting all other pixels with the aim of modeling long-range connections in arbitrarily large neighbourhoods and simultaneously preserve the advantageous fast inference [27, 29], in order to provide a balanced partitioning.

III. METHODOLOGY

In this section, we will first define the notations used throughout the paper and formulate the problem to solve. After that, the two paths are separately discussed, which is followed by an introduction of two different feature aggregation methods.

A. Notations and Problem Formulation

Given N mammograms and the biopsy-confirmed annotations from human experts, the dataset can be then noted as $\{X, Y, \mathcal{Z}\} = \{\{\mathbf{x}^{(n)}\}, \{\mathbf{y}^{(n)}\}, \{z^{(n)}\}\}_{n=1}^{N}$, where $\mathbf{x}^{(n)} \in \mathbb{R}^{H \times W}$ represents the nth mass-contained ROI with spatial dimension $H \times W$, $\mathbf{y}^{(n)} \in \{0 : \text{Normal pixel}, 1 : \text{Mass pixel}\}^{H \times W}$ is the pixel-level annotation corresponding to the nth ROI, and $z^{(n)} \in \{0, 1\}$ is a scalar indicating the diagnosis class label, and "0" and "1" represents the benign and the malignant class, respectively. Note that the cropped ROIs contain only one mass in our scenario. The main targets solved by DUALCORENET can be formulated as follows: (1) given a mass contained mammogram ROI, DUALCORENET is desired to map the original images to binary masks, so that mass pixels are segmented: $\mathbf{x}^{(n)} \rightarrow \mathbf{y}^{(n)}$. (2) With the original

(a) Module A

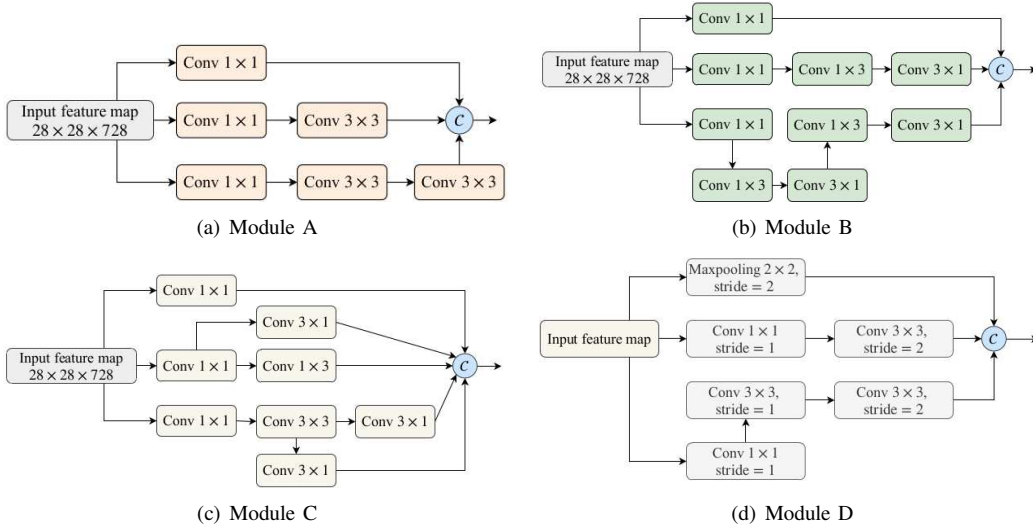(b) Module B

(c) Module C

(d) Module D

Fig. 4. Three convolution modules applied in DUALCORENET, where the original input feature map are all of the dimension $28 \times 28 \times 728$ and the expanded features are concatenated in each module. (a) *Module A* replaces a 5×5 convolution with two 3×3 convolutions. (c)*Module B* replaces a 7×7 with smaller kernel convolutions. (c) *Module C* expands a filter bank of 8×8 grids. (a) *Module D* is a dimension reduced convolution module, which halves the module input's spatial dimension.

mammogram images and the obtained pixel-level labels, learn a nonlinear mapping to the diagnosis label: $\mathbf{x}^{(n)}, \mathbf{y}^{(n)} \to \mathbf{z}^{(n)}$, where .

B. Motivation

Practically, radiologists make decisions about breast masses with their shape and boundary features. The more irregular the shape is, the more likely the mass is malignant [31], i.e., the classification results generally heavily rely on the segmentation results [2]. In analogy, decoupling a complicated learning task into several sub-tasks that are easier to solve is has also been proved an efficient learning paradigm in machine learning, there many methods aim to use multi-paths neural networks to solving image classification or other image restoration problems [16, 32]–[40]. However, the decoupling of breast mass diagnosis problem is seldom studied and the multi-path architecture has never been exploited with the dual segmentation and classification problem. Based on these, we aim to close this gap by solving these two problems in one, thus further improving the mammography analysis.

C. The proposed DUALCORENET

In this paper, we propose the DUALCORENET architecture, which decouples the differentiation of benign and malignant classes into dual problems: segmentation and classification. In the classification task, each input ROI sample (with surrounding tissues) will be classified into cancer category or not; whereas in the segmentation task, each pixel is labeled as either 0 or 1 so that mass pixels can be accurately identified within the tight bounding box ROI.

The DUALCORENET takes a batch of multi-scaled mammogram ROIs as the input and outputs the mass segmentation masks and the diagnosis labels simultaneously (Fig. 2). The mass segmentation computes the mapping from smaller scale ROI to binary masks, i.e. $X \mapsto Y$. The mass classification

solves the mapping of $X \mapsto Z$, by which the larger scale ROIs are mapped into diagnosis labels. Based on this idea, the DUALCORENET is constructed, which is comprised of the *Locality Preserving Learner* (LPL) and the *Conditional graph learner* (CGL) paths.

1) Locality Preserving Learner: The LPL path is constructed to learn the hierarchical and local intrinsic features from large scale ROIs. Large scale ROIs includes both textural and contextual information which are pivotal for mass classification [3, 6]. Inpisred by the well-known CNN backbone architecture, *e.g.,* VGG [42], the ResNet [43, 44], and the DenseNet [45], in this paper we propose an effective architecture, especially for mammography diagnosis, as illustrated in Fig. 3, and the several separable convolution modules Fig. 4 along residual connections are employed.

In this paper, the LPL path is constructed with 11 convolutional layers. In particular, the first four consecutive layers of the LPL employ naive convolutional layers along with the dimension reduction block (Fig. 4(d)), allowing the network to learn features at every chosen spatial scale. The number of feature maps consistently increase in the first four convolutional layers, from 16 (input) to 728 (the 4th layer output) feature maps, whereas the spatial dimension reduces from 224×224 to 56×56. Regarding the spatial downsampling, the maxpooling layer is employed in the 2nd layer, whereas the separable convolution in Figure 4(d) is utilized in the 3rd and 4th layers. Instead of using maxpooling followed by the convolutions, this dimension reduction method concatenates different scaled features generated by one or two convolutional operators directly. After that, two blocks of each *Block-A* (5th and 6th layers), *Block-B* (7th and 8th layers), and *Block-C* (9th and 10th layers) are separately constructed. These depth-wise separable convolutional layers produce the same number of feature maps and the identical spatial dimension, i.e. $28 \times 28 \times 728$. By utilizing 1×1 convolutions in the depth-wise separable blocks, the cross-channel correlations are learned first, resulting in

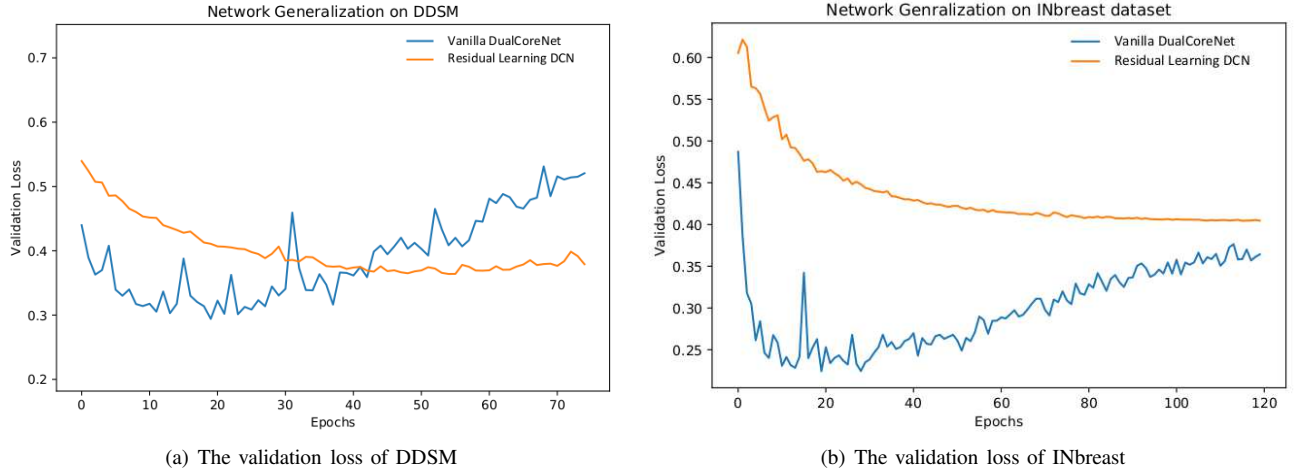(a) The validation loss of DDSM   (b) The validation loss of INbreast

Fig. 5. The residual learning and vanilla DUALCORENET validation loss on DDSM and INbreast datasets.

TABLE I

HIGH-RESOLUTIONAL BREAST CANCER LOCAL ANALYSIS WITH BOTH SEGMENTATION AND CLASSIFICATION PERFORMANCE OF STATE-OF-ART ALGORITHMS. THE SEGMENTATION PERFORMANCE IS ASSESSED BY THE DI (%) MATRIX AND THE CLASSIFICATION PERFORMANCE (MALIGNANT VS BENIGN) IS EVALUATED WITH THE AUC SCORE. THE "PROCESSED" COLUMN REPRESENTS THE PRE-PROCESSING AND POST-PROCESSING OF THE LIST ALGORITHMS RESPECTIVELY.

Methodology	Dataset	Segmentation	Classification	End-to-end
Arevalo *et al.* [23] (2016)	private	-	0.82	✗
Dhungel *et al.* [19] (2016)	INbreast	-	0.91	✗
Kooi *et al.* [24] (2017)	private	-	0.80	✗
Dhungel *et al.* [20] (2017)	INbreast	85.0	0.76	✗
Al-antari *et al.* [41] (2020)	INbreast	92.36	**0.95**	✗
DUALCORENET	INbreast	**93.69**	0.93	✓
	DDSM	**92.17**	**0.85**	✓

a much smaller feature space. Thereby, the LPL is enabled to learn richer features with much fewer parameters, hence alleviating the overfitting problem markedly with the same amount of training data.

Lastly, the generated deep features in the 11$^{\text{th}}$ layer are activated by the softmax non-linearity. The loss associated to the LPL layer is defined with categorical cross-entropy as:

$$\ell_{\text{LPL}} = -\sum_{n=1}^{N} \log p(z^{(n)} \mid \boldsymbol{x}^{(n)}; \boldsymbol{\theta}_1) \quad (3)$$

where z is the class indicator and $\boldsymbol{\theta}_1$ is the corresponding parameter set in LPL.

2) Conditional Graph Learner: The CGL path aims to extract segmentation-related or geometrical features from the resulted binary mask produced by an image to pixel-level label mapping. However, adapting CNNs to pixel-level labelling tasks is a significant challenge, since convolutional filters produce coarse outputs and max-pooling layers further reduce the sharpness of segmented boundaries. Although many methods have been utilized for this problem [22, 46, 47], unfortunately the balanced partitioning with high pixel resolution is still a challenge to solve [30].

Thereby, we propose a novel breast mass segmentation CNN architecture for the CGL path, as shown in Fig. 3, which

TABLE II

BREAST CANCER DIAGNOSIS PERFORMANCE (MALIGNANT VS BENIGN) OF THE LPL PATH IN DUALCORENET.

Dataset	Pre-training	Augmentation	AUC score
DDSM	none	none	0.73
DDSM	none	flips	0.73
DDSM	none	flips, random crops	0.74
DDSM	ImageNet	none	0.79
DDSM	ImageNet	flips	0.79
DDSM	ImageNet	flips, random crops	**0.85**
INbreast	none	none	0.80
INbreast	none	flips	0.85
INbreast	none	flips, random crops	0.84
INbreast	DDSM	none	0.86
INbreast	DDSM	flips	0.89
INbreast	DDSM	flips, random crops	**0.93**

is expected to not only precisely segment high-resolutional breast mass but also to control the model complexity. To do that, a CRF inference layer is applied in the low resolutional latent space and a concatenation is added to connect the high-resolutional feature maps, so that exhaustive textural features are interlaced and fully used.
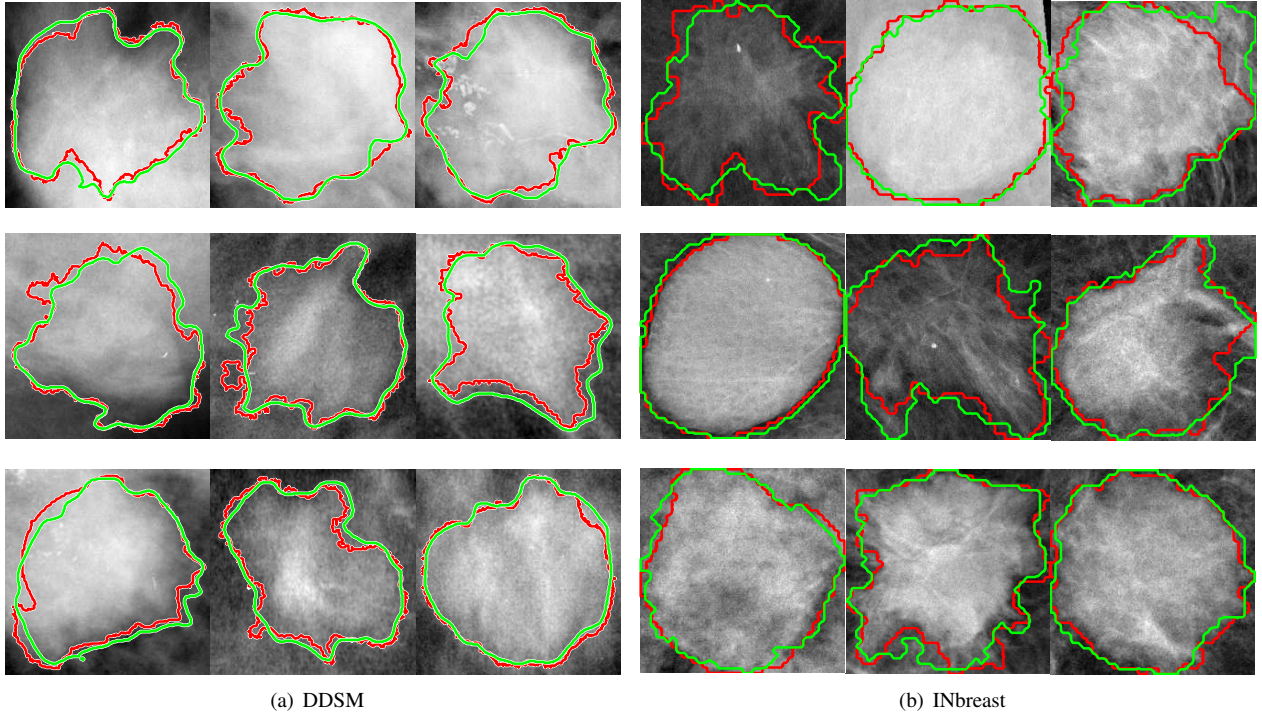
(a) DDSM   (b) INbreast

Fig. 6. Exampled high-resolutional breast mass segmentation results on DDSM and INbreast datasets, with an visualized comparison between radiologists' annotation (red lines) and DUALCORENET segmentation results (green lines).

TABLE III

QUANTITATIVE BREAST MASS SEGMENTATION PERFORMANCE (DICE COEFFICIENT, %) OF DUALCORENET AND SEVERAL STATE-OF-THE-ART METHODS ON TEST SETS. THE "PROCESSED" COLUMN REPRESENTS FOR THE PRE-PROCESSING AND POST-PROCESSING OF THE LIST ALGORITHMS RESPECTIVELY.

Methodology	INbreast	DDSM	Spatial Dimension	Pre/Post-processed
Beller et al. [48] (2005)	-	70	-	- / -
Cardoso et al. [49] (2015)	88	-	40×40	- / -
Dhungel et al. [50] (2015)	88	87	40×40	✓/ ✓
Dhungel et al. [31] (2015)	90	90	40×40	✓/ ✓
Zhu et al. [28] (2018)	89.36 ± 0.37	90.62 ± 0.16	40×40	✓/ ✗
Al-antari et al. [51] (2018)	92.69	-	40×40	✓/ ✓
U-Net [46] (2015)	88.54 ± 1.17	83.85 ± 2.13	40×40	✗/ ✗
Li et al. [22] (2018)	$\mathbf{93.66} \pm 0.10$	$\mathbf{92.23} \pm 0.26$	40×40	✗/ ✗
U-Net [46] (2015)	89.79 ± 0.33	90.42 ± 0.37	224×224	✗/ ✗
Dhungel et al. [20] (2017)	85	-	original	✓/ ✓
Wang et al. [52] (2019)	91.10	91.69	256×256	✓/ ✓
Singh et al. [53] (2020)	92.11	-	256×256	✓/ ✓
DUALCORENET (2020)	$\mathbf{93.69} \pm 0.17$	$\mathbf{92.17} \pm 0.03$	224×224	✗/ ✗

In particular, the fully connected CRF can be defined as follows:

$$p(\mathbf{y}|h(\mathbf{x})) = \frac{1}{Z(h(\mathbf{x}))} \exp \left(\sum_{i \in V} \phi_u(h(\mathbf{x}_i)) + \sum_{i,j \in E} \phi_p(\mathbf{y}_i, \mathbf{y}_j \mid h(\mathbf{x})) \right) \quad (4)$$

where Z is the partition function, and $h(\mathbf{x})$ is the deep latent feature of input $\mathbf{x}$ calculated by the softmax layer in CGL path. ϕ_u is the unary potential function, which is initialized as $h(\mathbf{x})$.

ϕ_p is the pair-wise potential function which is formulated as

$$\phi_p(\mathbf{y}_i, \mathbf{y}_j \mid h(\boldsymbol{x})) = \mu(\mathbf{y}_i, \mathbf{y}_j) \sum_m \mathbf{w}^{(m)} k_G^{(m)}(\mathbf{x}_i, \mathbf{x}_j), \quad (5)$$

where $\mathbf{y}_i$ and $\mathbf{y}_j$ are the predicted labels of connected nodes for position i and j respectively. $\mu(\mathbf{y}_i^{(n)}, \mathbf{y}_j^{(n)})$ is the label compatibility defined by the Pott's Model [54]. $\mathbf{w}^{(m)}$ is the learned weight and $k_G^{(m)}$ is the pre-defined weighed Gaussians over feature vectors at position i and j [22, 27].

In order to improve the segmentation performance for the unbalanced mass contained ROIs, we proposed to minimize the below two dice losses g and f for the CGL path to output a high-resolutional binary label mask.
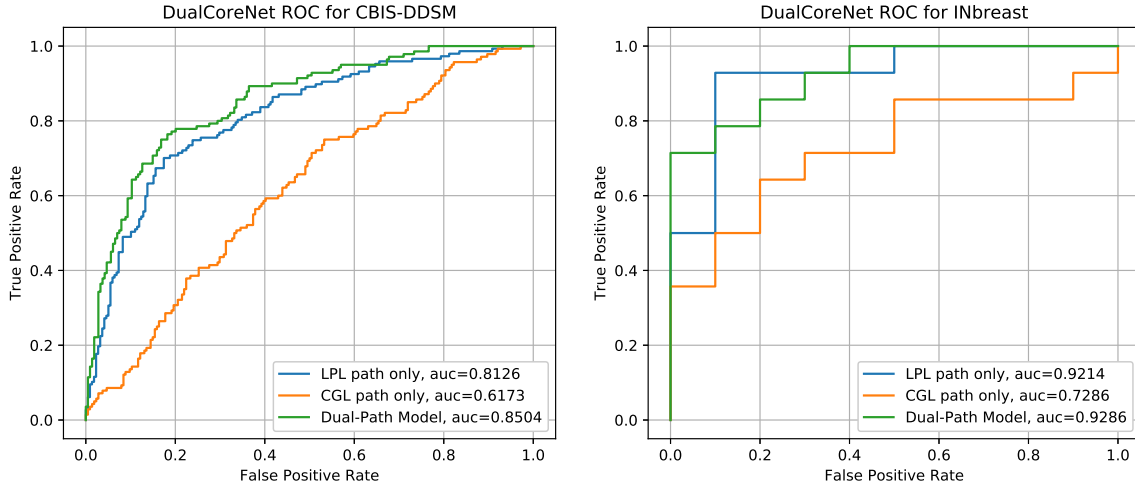
Fig. 7. Mass classification ROC curves of the DUALCORENET experimented on INbreast and CBIS-DDSM for identifying the cancerous masses from the union of Benign and Malignant ROIs.

$$\ell_{\text{CGL}} = 1 - \underbrace{\frac{2\sum \mathbf{y} \cdot p(\mathbf{x}, \boldsymbol{\theta}_2)}{\sum \mathbf{y} + \sum p(\mathbf{x}, \boldsymbol{\theta}_2)}}_{\text{CNN}} + \gamma \cdot 1 - \underbrace{\frac{2\sum \mathbf{y} \cdot p(\mathbf{y}|h(\mathbf{x}))}{\sum \mathbf{y} + \sum p(\mathbf{y}|h(\mathbf{x}))}}_{\text{CRF}},$$

(6)

where $p(\mathbf{x}, \boldsymbol{\theta}_2)$ is the output of the final softmax layer in the CGL path, $\boldsymbol{\theta}_2$ are the parameters, γ is the trade-off factor in the CGL path.

3) Fusion Module: So far, the textual features and shape features have been extracted by the LPL and CGL path, it is natural to integrate these two separate features in the feature fusion block, in order to further improve the diagnosis performance. To do that, we propose a fusion module as shown in Fig. 3. In particular, we first use two transformation blocks, each of which consists of several convolution layers followed by an average pooling and two fully connected layers, to transfer the output feature maps of LPL and CGL paths, respectively, such transformed feature maps from two paths will feed into a softmax layer to output the final classification result. The overall categorical cross-entropy based loss for classification task is defined as:

$$\ell_{\text{Fusion}} = -\sum_{n=1}^{N} \log p\big(z^{(n)} \mid \boldsymbol{x}^{(n)}; \boldsymbol{\theta}\big),$$

(7)

where z is the diagnosis class indicator and $\boldsymbol{\theta}$ is the entire network parameter vector.

Finally, by integrating the losses for LPL path *e.g.* (3), CGL path *e.g.* (6) and Fusion modules *e.g.* eq:fusion, the entire DUALCORENET loss is thereby defined as:

$$\ell_{\text{DUALCORENET}} = \ell_{\text{Fusion}} + \alpha \cdot \ell_{\text{LPL}} + \beta \cdot \ell_{\text{CGL}},$$

(8)

where α and β are two trade-off factors to control the importance of LPL and CGL paths, which are empirically set as $\alpha = \beta = 1$ in our experiments.

D. Implementation

We use the Adam to optimize our DUALCORENET . In order to alleviate overfitting and improve generalization, we use several training techniques for the DUALCORENET model. (1) Regarding the initialization of the LPL path, the DDSM dataset is first trained in the LPL path and the parameters are then fine-tuned in the INbreast dataset. (2) The dropout layers are employed with 50% random parameters dropping.

IV. EXPERIMENTS

A. Materials

In this paper, we validate the proposed DUALCORENET with two public mammography datasets: CBIS-DDSM [55] and INbreast [5]. The CBIS-DDSM is a modernized subset of Digital Database for Screening Mammography (DDSM) [56], in which 2478 digitized mammograms are formatted in DICOM format. On top of that, the CBIS-DDSM has already been partitioned into the training (1318 masses) and test set (378 masses). In this paper, we adopt the same data division method as the website did. In the matter of class balance, either training or test set involves the equivalent amount of two classes of lesions. Particularly, the malignant and benign ratio for the training and test sets are both roughly $1 : 1.07$. The INbreast dataset [5] is a Full Field Digital Mammography (FFDM) dataset, which was acquired at a Breast Centre in Hospital de So Joo, Porto, Portugal. There are a total of 115 cases in the INbreast dataset, which contains 410 mammogram images. Analogously, there are also two mammographic views for each breast and the images were annotated by human experts in a pixel-level labeling fashion. Regarding the data division for the evaluation of DUALCORENET, the INbreast data set is divided by patients into a training set and a test set as 80%: 20%.

In terms of the ROIs selection, masses are center cropped by two scales. One scale is the rectangular tight bounding box padded with 5 pixels on each boarder, which is utilized
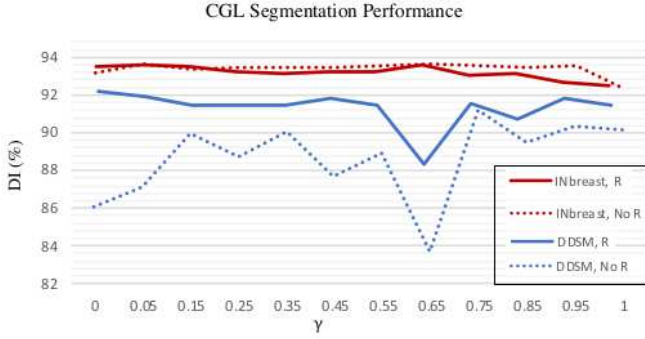
Fig. 8. CGL segmentation performance for DDSM and INbreast datasets in DUALCORENET. Dice coefficients (%) are averaged over 10 experiments with different γ in loss (6) and residual configurations ("R" represents residual and "No R" means without residual skips) on INbreast and DDSM datasets.

Fig. 9. LPL diagnosis performance (Malignant vs Benign) by various backbone networks

by the CGL to explore the segmentation related features. The other scale is to crop the contextual rectangular region with proportional padding, so that mass-centered ROI includes regions 2 times the size of the bounding box. These contextual ROIs are utilized by the LPL path to extract latent and hierarchical features from masses and their surrounding tissues. The selected ROIs are then all resized into identical dimension 224×224 by bicubic interpolation. Accordingly, the ground truth binary masks are cropped and resized but with the nearest neighbor interpolation. To avoid overfitting and provide better generalization, the selected ROIs are augmented with horizontal and vertical flips and random crop (with augmentation probability of 50% for each instance) after data division.

B. Results and Analysis

1) Comparison with State-of-the-art: We first compared the propose DUALCORENET with five related state-of-the-art breast mass diagnosis methods [19, 20, 23, 24, 41]. The results are listed in Table I where the performances of these compared methods are obtained from the results presented by their papers. In particular, [20] and [41] have solved both breast mass segmentation and classification problem. Compared with these methods, DUALCORENET is the only algorithm which has experimented on the DDSM dataset and DUALCORENET has achieved leading diagnosis performance (second-best) on INbreast dataset. DUALCORENET produces a 0.93 and 0.85 AUC score for mass diagnosis on INbreast and DDSM dataset, respectively. When compared with [41], there is only a 0.001 AUC difference, which is mainly because [41] randomly divided the training and test set after data augmentation. In our paper, however, we first divide the original data into training and test set, which are followed by augmentation. Furthermore, DUALCORENET has obtained the best segmentation performance when compared with all other algorithms, yielding 93.69% and 92.17% for INbreast and DDSM dataset.

Additionally, we evaluated the effectiveness of residual learning employed in DUALCORENET, we compare the test loss on two datasets between the configuration with or without residual skips in the best performing architecture (Fig. 3). Gen-
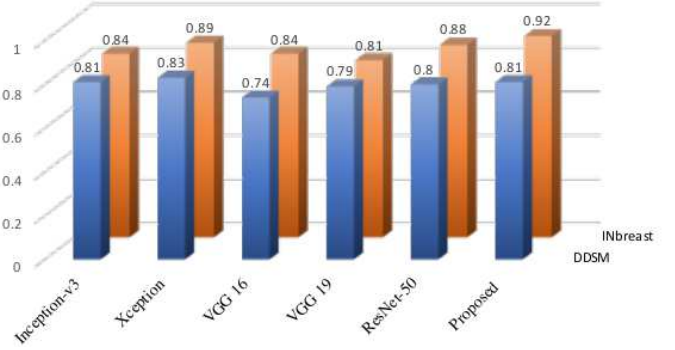
erally, a desired validation loss is expected to be stable after some training epochs, after consistently decreasing with the increasing training epoch. As shown in Fig. 5, the validation loss of the vanilla DUALCORENET either did not generalize or even slightly increased on both datasets. On the contrary, the residual learning DUALCORENET has shown a good ability of generalization, in which the validation loss was decreasing and saturated with the increasing number of training epochs. It is noted that the validation loss of the residual learning is much higher than that without residual skips (Figure 5). This is caused by the weight decay regularization term. Since the number of parameters in the residual learning is larger than that in the no residual connection DUALCORENET .

In addition, as shown in Table II, the cancer diagnosis performance of DUALCORENET (Malignant vs Benign) with different regularization configurations (augmentation or pre-training) have been listed. It can be noticed that the pre-training has markedly improved the model performance and augmentation method further competently increased the generalization. When DUALCORENET is trained with pre-training and data augmentation, the diagnosis AUC score for DDSM and INbreast has achieved 0.85 and 0.93, respectively.

2) The importance of Dual-paths: We are interested in studying the importance of dual-paths. We evaluate the CGL and the LPL path for the segmentation and classification performance, each of which is individually trained by ℓ_{CGL} and ℓ_{LPL} only.

Regarding the segmentation performance of CGL, the dice coefficients comparison between related works have been listed in Table III. For all the listed algorithms, only mass contained ROIs are provided as the input. In terms of low-resolutional mass segmentation with output dimension 40×40, [22] has achieved the best performance with 93.66% and 92.23% on INbreast and CBIS-DDSM, respectively. With respect to the high-resolutional segmentation, the DUAL-CORENET is so far the best performing algorithm for both datasets, with a 93.69% DI score in INbreast and 92.17% in DDSM dataset. Fig. 8 has shown the segmentation performance when CGL is trained with various γ values. It can be noticed that the over all segmentation performance on the INbreast dataset is significantly better than that on the

DDSM dataset, which is mainly attributed to the higher quality data of INbreast. Specifically for the INbreast dataset, the no residual connections and residual configuration generally performs equivalently. However, the best performance was obtained by no residual connection configuration when γ is 0.65. Note that the worst performance on INbreast (with either residual skips or not) was at $\gamma = 1$, where the CGL segmentation loss is contributed to the CNN and graphical model as ratio $1:1$. In terms of the DDSM dataset, the overall better performance was produced by the residual learning configuration. The best performance was obtained at $\gamma = 0.42$, where the segmentation DI achieves 92.17%.

The visualized segmentation results of CGL and radiologists' delineations can be seen in Fig. 6 for both datasets. It can be noticed that, the proposed segmentation method performs very well with higher resolution mammograms, in which fine boundary details and irregular shape contours are both well depicted. There are no resulting spurious regions in DUALCORENET, which is mainly due to the structural consistency restriction by the graphical inference layer. Although we implement the graphical inference in a small spatial size before converting to high resolutions, the performance is not affected. By doing so, a more efficient inference can be obtained with much less parameters and computing time.

As for the classification performance of the LPL path, we compared the LPL classification performance with various backbone networks, such as the Inception-v3, Xception, VGG16, VGG19, and ResNet-50 (Fig. 9). It can be seen that all networks perform better on the INbreast dataset. The best performances for INbreast and DDSM are obtained by the proposed LPL architecture (0.92 AUC score) and Xception network, respectively. However, the performance difference margin between our LPL (0.81 AUC score) and the Xception (0.83 AUC score) on DDSM is very small, with only a 0.02 AUC score. Generally speaking, the proposed LPL architecture achieved the leading performance for breast cancer diagnosis on both INbreast and DDSM datasets.

3) Ablation study on the training loss: Finally, We compared the DUALCORENET with with different loss function configurations.:[1] $\ell_{\text{Fusion}} + \ell_{\text{CGL}}$, $\ell_{\text{Fusion}} + \ell_{\text{LPL}}$, and $\ell_{\text{Fusion}} + \ell_{\text{CGL}} + \ell_{\text{LPL}}$. The diagnosis ROC curves of DUALCORENET with the Fusion loss combined with the LPL path, CGL path and dual-paths loss are shown in Fig. 7. It can be noticed that DUALCORENET performs best when activating both paths. And the second best performing training loss configuration is obtained by the Fusion and LPL path loss. This indicates that the features integrated from bot paths can learn richer information from the data.

V. CONCLUSIONS

In this paper, we propose an innovative dual-path CNN architecture called DUALCORENET for segmentation and classification problem. DUALCORENET first embeds the original mammography data into two heterogeneous data domains (i.e. the original image and the binary mask domain), where deep

features are jointly learned. By integrating the conditional graph learner path and the locality preserving learner path, our DUALCORENET works in a simple but effective way to jointly learn segmentation and classification. The integrated intrinsic localized textural features and semantic information extracted from binary masks contribute to an interpretable and more discriminative representation, which can maximize the similarity margins between benign and malignant instances in the deep latent space. Extensive experiments have shown that our method outperforms the state-of-the-arts on both breast mass segmentation and classification tasks in mammography. In addition DUALCORENET performs better on higher quality dataset (the INbreast dataset).

REFERENCES

[1] P. Boyle, B. Levin *et al.*, *World cancer report 2008.* IARC Press, International Agency for Research on Cancer, 2008.

[2] A. Oliver, J. Freixenet, J. Marti, E. Perez, J. Pont, E. R. Denton, and R. Zwiggelaar, "A review of automatic mass detection and segmentation in mammographic images," *Medical image analysis*, vol. 14, no. 2, pp. 87–110, 2010.

[3] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[4] B. Swiderski, S. Osowski, J. Kurek, M. Kruk, I. Lugowska, P. Rutkowski, and W. Barhoumi, "Novel methods of image description and ensemble of classifiers in application to mammogram analysis," *Expert Systems with Applications*, vol. 81, pp. 67–78, 2017.

[5] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: toward a full-field digital mammographic database," *Academic radiology*, vol. 19, no. 2, pp. 236–248, 2012.

[6] C. Varela, S. Timp, and N. Karssemeijer, "Use of border information in the classification of mammographic masses," *Physics in medicine & biology*, vol. 51, no. 2, p. 425, 2006.

[7] A. Jalalian, S. B. Mashohor, H. R. Mahmud, M. I. B. Saripan, A. R. B. Ramli, and B. Karasfi, "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review," *Clinical imaging*, vol. 37, no. 3, pp. 420–426, 2013.

[8] I. Domingues, E. Sales, J. Cardoso, and W. Pereira, "Inbreast-database masses characterization," *XXIII CBEB*, 2012.

[9] S. Dhahbi, W. Barhoumi, and E. Zagrouba, "Breast cancer diagnosis in digitized mammograms using curvelet moments," *Computers in biology and medicine*, vol. 64, pp. 79–90, 2015.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[11] D. Chen, J. Lv, and Z. Yi, "Graph regularized restricted boltzmann machine," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 6, pp. 2651–2659, 2018.

[12] W. Zhu, Q. Lou, Y. S. Vang, and X. Xie, "Deep multi-instance networks with sparse label assignment for whole mammogram classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 603–611.

[13] S. Shams, R. Platania, J. Zhang, J. Kim, and S.-J. Park, "Deep generative breast cancer screening and diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 859–867.

[14] D. Chen, J. Lv, and Z. Yi, "Unsupervised multi-manifold clustering by learning deep representation," in *Workshops at the 31th AAAI conference on artificial intelligence (AAAI)*, 2017, pp. 385–391.

[15] M. Golbabaee, D. Chen, P. A. Gómez, M. I. Menzel, and M. E. Davies, "Geometry of deep learning for magnetic resonance fingerprinting," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 7825–7829.

[16] D. Chen and M. E. Davies, "Deep decomposition learning for inverse imaging problems," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[17] G. Carneiro, J. Nascimento, and A. P. Bradley, "Automated analysis of unregistered multi-view mammograms with deep learning," *IEEE transactions on medical imaging*, vol. 36, no. 11, pp. 2355–2365, 2017.

---

[1]Note we observed in our experiments that if we trained the model with only use ℓ_{Fusion} would lead to a not converged model.

[18] D. Chen, M. E. Davies, and M. Golbabaee, "Compressive mr fingerprinting reconstruction with neural proximal gradient iterations," in *International Conference on Medical image computing and computer-assisted intervention (MICCAI)*, 2020.

[19] N. Dhungel, G. Carneiro, and A. P. Bradley, "The automated learning of deep features for breast mass classification from mammograms," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 106–114.

[20] ——, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention," *Med. Image Anal.*, vol. 37, pp. 114–128, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.media.2017.01.009

[21] H. Li, D. Chen, W. H. Nailon, M. E. Davies, and D. Laurenson, "A deep dual-path network for improved mammogram image processing," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1224–1228.

[22] ——, "Improved breast mass segmentation in mammograms with conditional residual u-net," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Springer, 2018, pp. 81–89.

[23] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," *Computer methods and programs in biomedicine*, vol. 127, pp. 248–257, 2016.

[24] T. Kooi, B. van Ginneken, N. Karssemeijer, and A. den Heeten, "Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network," *Medical physics*, vol. 44, no. 3, pp. 1017–1027, 2017.

[25] H. Li, D. Chen, W. H. Nailon, M. E. Davies, and D. I. Laurenson, "Signed laplacian deep learning with adversarial augmentation for improved mammography diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2019, pp. 486–494.

[26] 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*, 2016, pp. 770–778.

[27] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in neural information processing systems*, 2011, pp. 109–117.

[28] W. Zhu, X. Xiang, T. D. Tran, G. D. Hager, and X. Xie, "Adversarial deep structured nets for mass segmentation from mammograms," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 2018, pp. 847–850.

[29] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[30] F. Guo, M. Ng, M. Goubran, S. E. Petersen, S. K. Piechnik, S. Neubauer, and G. Wright, "Improving cardiac mri convolutional neural network segmentation on small training datasets and dataset shift: A continuous kernel cut approach," *Medical Image Analysis*, vol. 61, p. 101636, 2020.

[31] N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep learning and structured prediction for the segmentation of mass in mammograms," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 605–612.

[32] D. Liu, Z. He, D. Chen, and J. Lv, "An improved dual-channel network to eliminate catastrophic forgetting," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020.

[33] Z. Wang, X. Liu, H. Li, L. Sheng, J. Yan, X. Wang, and J. Shao, "Camp: Cross-modal adaptive message passing for text-image retrieval," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 5764–5773.

[34] Z. Zheng, L. Zheng, M. Garrett, Y. Yang, M. Xu, and Y.-D. Shen, "Dual-path convolutional image-text embeddings with instance loss," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 16, no. 2, pp. 1–23, 2020.

[35] J. Fan, X. Cao, P.-T. Yap, and D. Shen, "Birnet: Brain image registration using dual-supervised fully convolutional networks," *Medical image analysis*, vol. 54, pp. 193–206, 2019.

[36] B. C. Kim, J. S. Yoon, J. S. Choi, and H. I. Suk, "Multi-scale gradual integration CNN for false positive reduction in pulmonary nodule detection," *Neural Networks*, 2019.

[37] X. W. Gao, R. Hui, and Z. Tian, "Classification of ct brain images based on deep learning networks," *Computer methods and programs in biomedicine*, vol. 138, pp. 49–56, 2017.

[38] X. Liu, F. Hou, H. Qin, and A. Hao, "Multi-view multi-scale cnns for lung nodule type classification from ct images," *Pattern Recognition*, vol. 77, pp. 262–275, 2018.

[39] F. Ciompi, K. Chung, S. J. Van Riel, A. A. A. Setio, P. K. Gerke, C. Jacobs, E. T. Scholten, C. Schaefer-Prokop, M. M. Wille, A. Marchiano *et al.*, "Towards automatic pulmonary nodule management in lung cancer screening with deep learning," *Scientific reports*, vol. 7, p. 46479, 2017.

[40] D. Liang, L. Lin, X. Chen, H. Hu, Q. Zhang, Q. Chen, Y. Iwamoto, X. Han, Y. Chen, R. Tong, and J. Wu, "Multi-stream scale-insensitive convolutional and recurrent neural networks for liver tumor detection in dynamic ct images," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 794–798.

[41] M. A. Al-antari, M. A. Al-masni, and T. S. Kim, "Deep Learning Computer-Aided Diagnosis for Breast Lesion in Digital Mammogram," *Adv. Exp. Med. Biol.*, vol. 1213, pp. 59–72, 2020.

[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[43] 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*, 2016, pp. 770–778.

[44] ——, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.

[45] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[46] 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.

[47] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1529–1537.

[48] M. Beller, R. Stotzka, T. O. Müller, and H. Gemmeke, "An example-based system to support the segmentation of stellate lesions," in *Bildverarbeitung für die Medizin 2005*. Springer, 2005, pp. 475–479.

[49] J. S. Cardoso, I. Domingues, and H. P. Oliveira, "Closed shortest path in the original coordinates with an application to breast cancer," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 29, no. 01, p. 1555002, 2015.

[50] N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep structured learning for mass segmentation from mammograms," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 2950–2954.

[51] M. A. Al-antari, M. A. Al-masni, M.-T. Choi, S.-M. Han, and T.-S. Kim, "A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification," *International journal of medical informatics*, vol. 117, pp. 44–54, 2018.

[52] R. Wang, Y. Ma, W. Sun, Y. Guo, W. Wang, Y. Qi, and X. Gong, "Multi-level nested pyramid network for mass segmentation in mammograms," *Neurocomputing*, vol. 363, pp. 313–320, 2019.

[53] V. K. Singh, H. A. Rashwan, S. Romani, F. Akram, N. Pandey, M. M. K. Sarker, A. Saleh, M. Arenas, M. Arquez, D. Puig *et al.*, "Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network," *Expert Systems with Applications*, vol. 139, p. 112855, 2020.

[54] R. B. Potts, "Some generalized order-disorder transformations," in *Mathematical proceedings of the cambridge philosophical society*, vol. 48, no. 1. Cambridge University Press, 1952, pp. 106–109.

[55] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data*, vol. 4, p. 170177, 2017.

[56] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The digital database for screening mammography," in *Proceedings of the 5th international workshop on digital mammography*. Medical Physics Publishing, 2000, pp. 212–218.