

A systematic survey of deep learning in breast cancer

Xiang Yu  | Qinghua Zhou  | Shuihua Wang  | Yu-Dong Zhang 

School of Computing and Mathematical Sciences, University of Leicester, Leicester, Leicestershire, UK

Correspondence

Shuihua Wang and Yu-Dong Zhang, School of Computing and Mathematical Sciences, University of Leicester, Leicester, Leicestershire, LE1 7RH, UK. Email: shuihuawang@ieee.org and yudongzhang@ieee.org

Funding information

Royal Society International Exchanges Cost Share Award, Grant/Award Number: RP202G0230; Hope Foundation for Cancer Research, UK, Grant/Award Number: RM60G0680; Medical Research Council Confidence in Concept Award, UK, Grant/Award Number: MC_PC_17171

Abstract

In recent years, we witnessed a speeding development of deep learning in computer vision fields like categorization, detection, and semantic segmentation. Within several years after the emergence of AlexNet, the performance of deep neural networks has already surpassed human being experts in certain areas and showed great potential in applications such as medical image analysis. The development of automated breast cancer detection systems that integrate deep learning has received wide attention from the community. Breast cancer, a major killer of females that results in millions of deaths, can be controlled even be cured given that it is detected at an early stage with sophisticated systems. In this paper, we reviewed breast cancer diagnosis, detection, and segmentation computer-aided (CAD) systems based on state-of-the-art deep convolutional neural networks. The available data sets also indirectly determine CAD systems' performance, so we introduced and discussed the details of public data sets. The challenges remaining in CAD systems for breast cancer are discussed at the end of this paper. The highlights of this survey mainly come from three following aspects. First, we covered a wide range of the basics of breast cancer from imaging modalities to popular databases in the community; Second, we presented the key elements in deep learning to form the compactness for methods

mentioned in reviewed papers; Third and lastly, the summative details in each reviewed paper are provided so that interested readers can have a refined version of these works without referring to original papers. Therefore, this systematic survey suits readers with varied backgrounds and will be beneficial to them.

KEYWORDS

breast cancer, CAD systems, deep learning, systematic review

1 | INTRODUCTION

Breast cancer, one of the common cancers diagnosed among women worldwide, has become the second leading cause of cancer death narrowly after lung cancer.^{1–3} The situation is even worse for females in the UK, where new breast cancer incident cases registered in 2017 reached,^{4,5} accounting for 15.1% of all cancer cases.⁶ Mammography has been proven to be one of the most effective techniques to detect breast carcinoma in the early stage.⁷ With mammography screening, radiologists are considerably assisted in finding tumors in mammogram images with smaller sizes and more randomness in locations. It has been shown that mammography can detect main early symptoms such as microcalcifications and masses, reducing the death rate of breast carcinoma by around 15%.⁸ To facilitate the diagnosis of breast cancer, experts in the computer science community have developed useful computer-aided design (CAD) systems during the past decades. Those CAD systems can be mainly categorized into three classes including segmentation, detection, classification regarding their application tasks. However, many CAD systems, which can be referred to as traditional CAD systems, heavily rely on manually crafted features and thus significantly impairs the overall performance. Also, the robustness of these systems, which means the performance of these systems on new data, remains to be improved.

The situation was mitigated by the advent of deep learning. The concept of deep learning, or deep convolutional neural networks (CNNs), arose to the public when AlexNet⁹ achieved exceptional performance on image recognition challenges.¹⁰ Deep-learning methods are based on representation learning, where representations are from multiple levels. At each level, nonlinear but simple modules are combined to transform the representation from the lower into higher levels, where the lower-level features are more intuitive while the higher-level features are abstractive. Feature examples of ResNet18 can be seen in Figure 1.¹¹

As can be seen from Figure 1, early features are more intuitive as they are edges-like while the features in the second convolutional block showed more complex patterns. Similar patterns can be found in features produced by deeper layers. Complex functions, therefore, can be learned with the composition of the transformations. Compared to traditional machine-learning methods, one characteristic of deep learning is that less human intervention is required for similar pattern-recognition systems. Given the above advantages, deep learning turns out to be capable of solving problems with complicated structures and high dimensions from the domains of business, science, and government. In traditional fields such as image recognition,^{12–14} speech recognition,¹⁵ deep learning has predominantly outperformed traditional machine-learning methods. While in

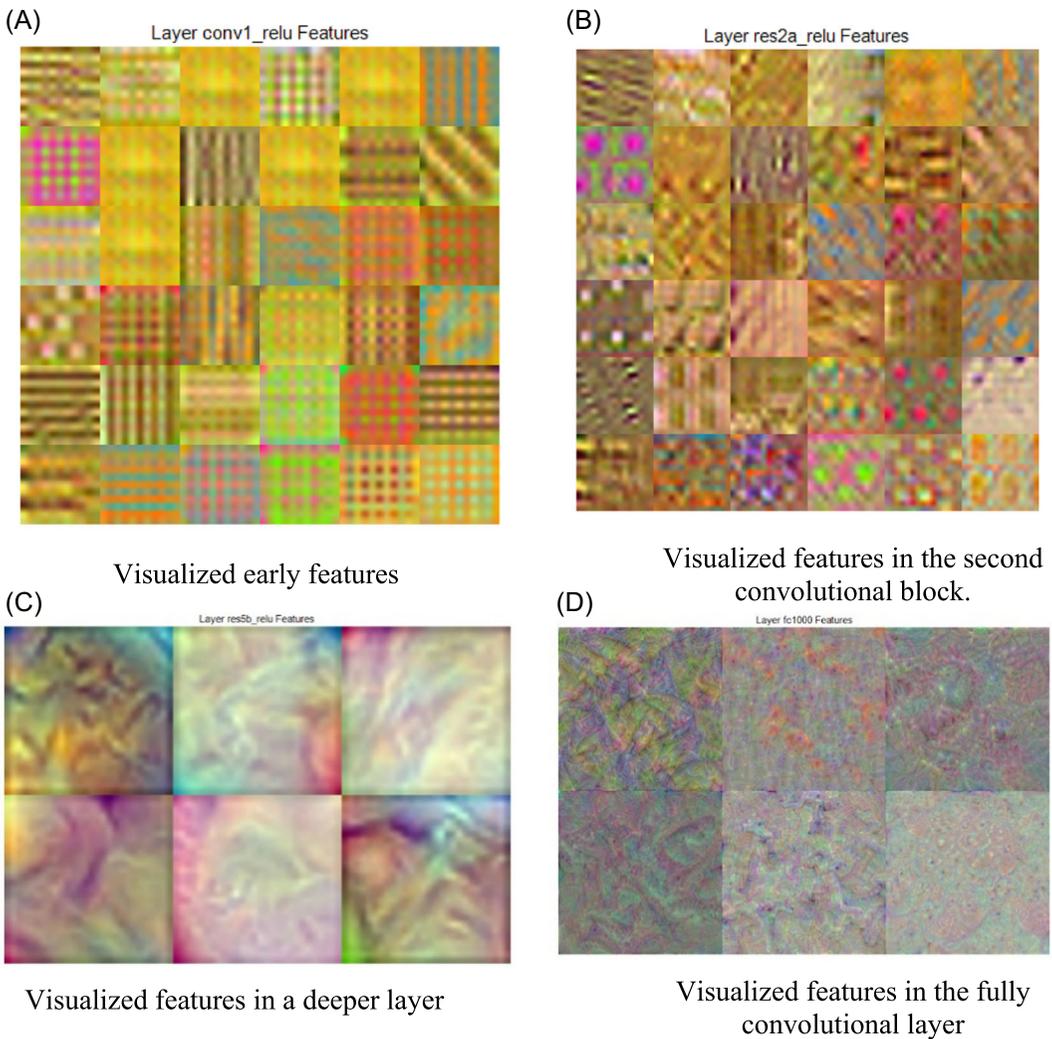


FIGURE 1 Visualized features in different depths of ResNet18 [Color figure can be viewed at wileyonlinelibrary.com]

other areas where traditional machine-learning methods have never been applied before, deep learning remains the surprising performance in fields such as natural language understanding,^{16,17} and language translation.^{18,19} Benefited from the fast development of deep learning, the deep-learning-based CAD systems outnumbered the traditional CAD systems by a large margin. The number of the publications in recent five years also witnessed the speeding development of the area, which can be seen in Figure 2.

While it is a speeding developing area, systematic reviews on the developed methods are lagged. In the survey work,²⁰ Debelee et al. reviewed the deep-learning-based works on image level and histopathology level. The reviewed topics included segmentation, classification, feature extraction, prediction, and detection. However, the related deep learning background is missing. In another work,²¹ Abdelrahman et al. provided a balanced review work that included some basics of deep learning and corresponding applications of deep learning. Nevertheless, the edge-cutting techniques of deep learning are missing while the application can be further extended. To mitigate the situation,

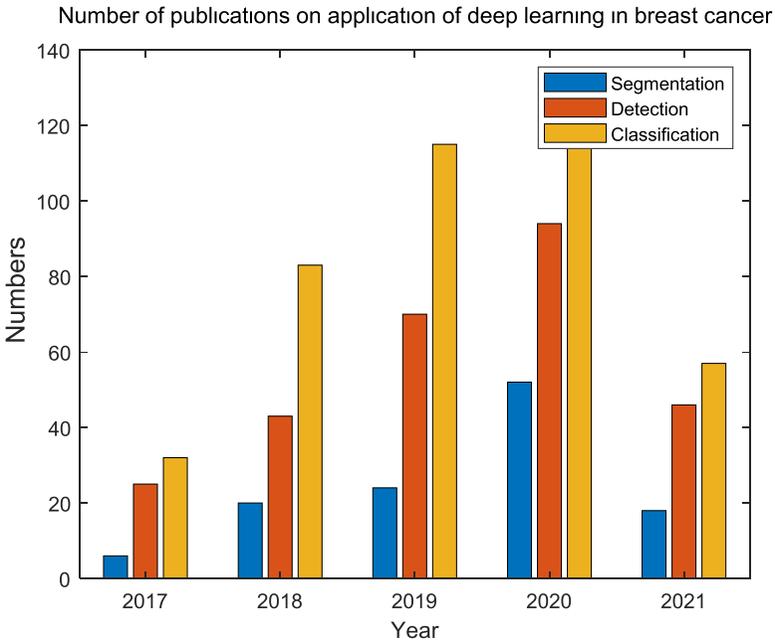


FIGURE 2 Publications on the application of deep learning in breast cancer in lastest 5 years [Color figure can be viewed at wileyonlinelibrary.com]

we aimed at providing a comprehensive review that covers the latest development in deep learning and the applications of the newly developed techniques in the diagnosis of breast cancer. The remainder of this paper is arranged as follows. In Section 2, we will emphasize the most popular breast imaging modalities and mammographic breast databases because imaging modalities directly determine the image resolutions and qualities. As a result, image modalities indirectly determined the potential performance of CAD systems. The breast databases provided a fair platform for algorithm evaluation. The size and rations between different categories in the databases may interfere with the final performance of algorithms. To catch up with the latest development of deep learning, we will have a brief introduction to basic and novel components of deep learning in Section 3. As abnormalities are the main symptoms of breast cancer, therefore we will take detection of abnormalities as the entry point for breast cancer detection. In Section 4, we will briefly review the abnormality classification system before the prevalence of deep learning. Section 5 and Section 6 will introduce the detection and segmentation work implemented by deep learnings, respectively. Later on, we will discuss the remaining challenges and future trends in breast cancer in Section 7. And this paper ends up with a conclusion in Section 8.

2 | IMAGING MODALITIES AND DATABASES

Imaging modalities are of great significance for image-based CAD systems as the quality of images provided by these modalities directly determine the speed, accuracy, and therefore, the overall performance of CAD systems. In this section, we will review and summarize the popular imaging modalities for breasts. The obtained images by different types of equipment in different modalities contribute to numerous data sets, which are usually used for training and

evaluation of the developed deep learning models. Considering this, some introduction to go-to data sets is amended here for compactness.

2.1 | Imaging modalities

Imaging modalities are the starting point of the detection of breast abnormalities. The quality of imaging by different techniques may have a profound impact on the performance of detection. In breast screening, there are the following commonly used imaging technologies. Notably, metrics including sensitivity, specificity, recall rates, positive predicted value, F-score, accuracy, and AUC are used to measure the performance of these imaging technologies.²⁰

Screen-film mammography (SFM) is a standard imaging modality for the detection of suspicious lesions at an early stage. SFM showed high sensitivity (100%) of abnormality detection in breasts mainly have fatty tissues. But for breasts with dense glandular tissue the sensitivity decrease dramatically. As a result, breast cancers that cannot be visualized take up from 10% to 20%. The film acts as the medium of image acquisition as well as displaying and storage. One drawback of this imaging technique is that no image improvements can be carried out once the film is produced. Some produced images inevitably suffer from lesser contrast, which requires the patients have to go through another imaging procedure and thus expose themselves to more radiation dose. Besides, the main problem with SFM is that it cannot be digitalized.

Digital mammography is also another important imaging modality that is important and effective for early-stage breast cancer screening.²² However, low specificity is featured as one of the limitations. Consequently, a large number of unnecessary biopsies are carried out, which leads to a waste of healthcare resources and stress on patients. Like SFM, patients may have to undergo multiple times digital mammograms (DM) and undertake the risk of ionizing radiation that could endanger patients' health. However, DM has the advantage over SFM in that it can be digitalized so that CAD systems can be utilized to improve the radiologist's sensitivity.

As the second choice to DM, ultrasound (US) is another widely used imaging modality, though operator-dependent, for breast lesion detection and differentiation. It is shown in Shin et al.²³ that ultrasound achieved promising performance on detection and discrimination of benign and malignant masses. Unwanted biopsies, therefore, can be reduced by US imaging modality. Compared to other imaging modalities such as DM and digital breast tomosynthesis, the US is an alternative imaging modality that is safe, low-cost, accurate, and highly universal.²⁴ However, the interpretation of US images is not straightforward, therefore requires in-depth knowledge of image features. Considering the US's benefit, US has been recommended to be a supplement to DM due to its safety and low cost.

Magnetic resonance imaging (MRI) produces a strong magnetic field that forces protons in the target organ to align with the field so that sensors can imaging the tissues and organs. It is used for high patient risk and clinical monitoring and diagnosis of breast cancer. An enhanced form of MRI named dynamic contrast-enhanced (DCE)-MRI can extract valuable information by providing higher volumetric resolution, which presents better lesion localization. Studies have shown that DCE-MRI is a useful tool for breast prognosis, diagnosis, and correlation with genomics.²⁵ In contrast to other imaging modalities, MRI has higher sensitivity of breast cancer diagnosis while DCE-MRI has a high sensitivity for cancer detection, even in dense breasts.^{26,27} However, MRI is not the best choice among all imaging modalities due to its expensiveness and lower specificity.

With the development of imaging technology, three-dimensional (3D) imaging technique has become a demanding technique. As a typical representative of the 3D imaging modality, digital breast tomosynthesis (DBT) can produce 3D images of the breasts at different angles but uses low dose X-rays²⁰ that the X-rays dose is similar to that of a regular mammogram. In the process of image acquisition of DBT, breasts are placed in the same position, and images are compressed in the same way as a mammogram while the difference lies in that the X-ray tube moves in a circular arc way around the breasts.^{28,29} However, compared to traditional mammography, DBT requires less time for imaging while better detail of dense tissues in the breast can be presented.^{30,31} Therefore, DBT has been a newly emerged breast cancer imaging modality that improves the accuracy and sensitivity of detection of breast cancer.^{28,32} However, DBT has its limitations. One is that malignant microcalcification could be missed if they were not on the DBT slice plane.³³ The recall rates for architectural distortion also increase.³⁴ As images are presented in 3D, it takes longer to interpret DBT slices compared to DM. The summary of the imaging modalities mentioned above is given in Table 1.

Although there are numerous breast cancer screening methods, DM is recognized as the most effective method for early detection of breast cancer.³⁵ For this purpose, we focus on the mammographic databases used in the community.

2.2 | Objectives of databases

To build a mammogram database, some basic information about mammograms should be provided, such as multiview and ground truth (GT). Generally, each breast has two recordings from a craniocaudal (CC) view, a top to bottom view, and a mediolateral oblique (MLO) view in mammographic imaging. Besides, breasts can be imaged from both the right-side and left-side, which therefore contributes to four mammograms from different sides and views regarding one breast. Examples are given in Figure 3.

TABLE 1 Breast imaging modalities

Modalities	Abbreviation	Advantages	Disadvantages
Screen-film mammography	SFM	High sensitivity for breasts with fatty tissues;	Low sensitivity for breasts with dense glandular tissue; Does not allow image improvements;
Digital mammography	DM	Can be digitalized;	Low specificity; Repetitive examination;
Ultrasound	US	Safe, low-cost, accurate, and highly universalize	Specific knowledge required for understanding
Magnetic resonance imaging	MRI	Higher sensitivity	Expensiveness and low specificity
Digital breast tomosynthesis	DBT	Better details; reduced imaging time	Increase of recall rate; Longer interpretation time.

Abbreviations: DBT, digital breast tomosynthesis; DM, digital mammograms; MRI, magnetic resonance imaging; SFM, Screen-film mammography; US, ultrasound.

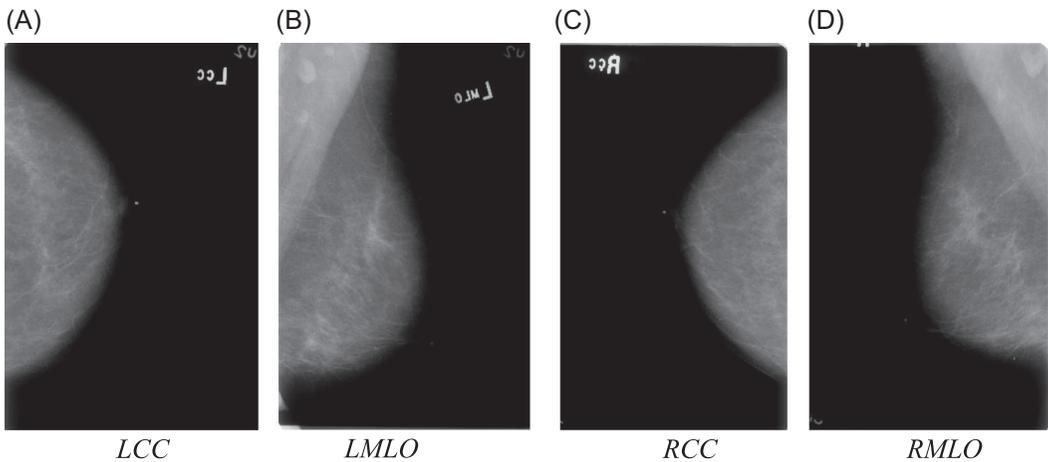


FIGURE 3 Mammograms in different views

Images can be printed into X-ray film or stored as digital mammography (Figure 3). After examining by radiologists, abnormal findings in mammograms were categorized into one of six categories according to the breast imaging report and data system (BI-RADS) scale, a standardized mammographic report system developed by the American College of Radiology (ACR). The six categories are Category 1, negative (no findings); Category 2, benign findings; Category 3, probable benign findings; Category 4, suspicious anomaly; Category 5, highly suggestive of malignancy; Category 6, proved malignancy by biopsy. Exams that are not conclusive are classified into the incomplete category that will be generally exempted from clinical usage. Breast composition tissue, another important characteristic referred to by ACR, has four categories ranging from low density (Category 1, fatty tissue) to high density (Category 4, dense tissue). Therefore, the GT, which was proved to be the correct category of the lesion with locations and boundaries, should be accessible. Additional information such as age, previous biopsies can also be useful for studying breast cancer in that more nonradiographic information would contribute to a higher performance CAD system.³⁶

For easy storage and utilization, digital mammograms are usually saved in digital imaging and communications in medicine (DICOM) format. One good feature is that images and metadata can be acquired from one single file. Another popular format is eXtensible Markup Language (XML), designed for convenient storage and data transportation.

2.3 | Public databases

In the breast cancer area, there are several popular databases. Though some of them may fail to meet all of the requirements, they do profoundly contribute to the advancement of breast cancer research.

The Mammographic Image Analysis Society Digital Mammogram Database (MIAS),³⁷ the most classical database, is still extensively used. Compared to contemporary databases, the resolution is reduced to 1024 by 1024 pixels, and the database itself is no longer maintained. In total, there are 161 cases, leading to 322 digitalized *MLO* images stored in the format of portable grey map file format (.pgm). The findings include normal, benign, and malignant. The normal

type accounts for the majority of the database. Breast density is provided but not classified regarding ACR standards. Only when its usage for CAD classification increased, it was then determined to classify the set of mammograms according to the standard. A high percentage of spiculated masses is provided. However, the percentage of benign findings is elevated according to some research.³⁸ The annotation of *MIAS* is given by the center and radius of a circle around the center. Various studies have pointed out several drawbacks of this kind of annotation. The annotation itself is insufficient that manual segmentation has to be conducted.³⁹ Besides, the digitalized mammogram's resolution hinders the detection of microcalcification (MCCs).⁴⁰ As *GT* contains more healthy tissue than lesion tissues, calcifications are not considered in some studies.⁴¹

The BancoWeb LAPIMO Database,⁴² a more recent database, contains 320 cases providing 1473 images from multiple views, including *CC*, *MLO*, and magnification. Just like *MIAS*, three types of sample images that are normal, benign, and malignant are given. Registered users can access the database at <http://lapimo.sel.eesc.usp.br/bancoweb/>. BI-RADS annotations, together with patient information, are accessible. Textual descriptions of all of the findings are available but annotations, as shown in the form of regions of interest (ROIs), only exist in some of the images.

Magic-5,⁴³ an Italian database set up in 2002, contributes 3369 images to the community. Those images, which were digitized with a resolution of 12 bits and saved in DICOM format, are coming from multi-view including *MLO*, *CC*, and lateral. Similar to *MIAS*, *GT* is given by the centers of masses and MCCs and the circles around the regions of interest. Patient age is approachable as additional information but no BI-RADS categorization while density classification is not provided in the form corresponding to ACR standard. The limitation of *Magic-5* is the heterogeneity because images are collected in different environments.

INbreast,³⁶ which has 410 images in total, contains six categories of findings: normal, calcification, masses, asymmetries, multiple finding, and architectural distortions. This database provides a substantial number of mammograms with calcifications and multiple findings. Another prominent characteristic of *INbreast* is the careful annotation provided by specialists. Unlike most of the databases that provide circles around the ROI, pixel-level contours surrounding lesions are allocated. When a cluster of MCCs appears, an ellipse enclosing the cluster is given. The XML format is used to store information of ROIs and related information of patients such as family history, ACR breast density, and BI-RADS classification distribution.

Digital Database for Screening Mammography (DDSM),⁴⁴ a relatively more significant database of 2620 cases, provides the community with a powerful resource. Given the large scale of the database, DDSM has been widely used since it was released in 1997. According to BI-RADS, information such as patient age, mass shape, mass margin, calcification distribution, and breast density is offered. Metadata is given in the form of .ics file that includes the above information. *GT* validation and ROI annotations further boost the popularity of the database. However, images of the database are saved in a nonstandard compression way while decompression code has not been updated for modern devices. The positions of abnormalities provided are too general to form a precise segmentation, making the evaluation of CAD algorithms more difficult. Also, researchers pointed out that some ROI annotations are questionable,⁴⁵ where suspicious regions are not shown in the image. With consideration to the problems mentioned above, Rebecca et al. cautiously selected a subset of *DDSM* named *Curated Breast Imaging Subset of DDSM (CBIS-DDSM)*.⁴⁶ This database selects 753 calcification cases and 891 mass cases from the original *DDSM* data set after removing questionable images and refining the ROI annotations. For better evaluation of algorithms, images with masses and calcifications in

the database are split into the training set and testing set, in which 20% of the cases are partitioned, respectively.

The OPTIMAM Medical image Database (OMI-DB) is a large repository containing over 2 million digital mammography images.⁴⁷ To construct the repository, images are collected from numerous sites while normal, benign, and malignant images are selectively collected from three sites. In total, there are 2,889,312 images from 173,319 women, among whom 154,832, 6909, 9690, and 1888 are found with normal breasts, benign findings, screen-detected cancers, and interval cancers, respectively. For screen-detected cancers, experienced radiologists annotated 7143 lesions by indicating attributes such as the location and area of lesions. Additionally, relevant tags such as screening history, biopsy results are also extracted for a searchable index. Besides the databases mentioned above, some databases are not widely used.^{4,42,48} Detailed characteristics of different databases can be seen in Table 2.

3 | DEEP LEARNING

The development of the neural network can be chronologically divided into four periods: enlightenment (1890–1969), trough (1969–1982), resurgence (1982–1986), and the current new era (1986 to now), as can be seen in Figure 4.

The initial effort towards artificial intelligence can be dated back to the 20th century. In 1943, McCulloch and Pitts⁴⁹ first came up with the idea that propositional logic can explain neural events and the relationships between them. This study was recognized as the seminal work for artificial neural networks (ANNs). The perceptron,⁵⁰ which was considered the first ANN in a strict sense and triggered broad interest in the community. However, the upsurge retreated because the perceptron's limitation, as pointed out by Minsky, is being unable to solve nonlinear problems.⁵¹ After this assertion, no significant breakthroughs were made during the following years (trough period), which would be the winter for ANNs. What rejuvenated the research interests of ANNs is the proposal of a recursive ANN termed Hopfield neural network in 1982⁵² and then the concept of deep learning comes to the public.

Regarding the learning pattern, deep learning can be subdivided into supervised models and unsupervised models. Supervised models mainly include traditional ANNs, CNNs and recurrent neural networks (RNNs), where CNNs are generally used to cope with array-like data and RNNs are more efficient in dealing with sequential data. Compared to ANNs, the components of CNNs are not simply the counterpart in ANNs but also pooling, flattening that shrink the dimensionality of features extracted by CNN blocks. Unlike ANNs and CNNs, RNNs rely on a special unit named long-term memory (LSTM) to extract temporal information. The typical architectures of the three types of models can be seen in Figure 5.

Supervised models are trained with cautiously annotated data. However, available annotated data sets tend to be quite limited under some situations while not all models have to be trained in a supervised pattern. Therefore, unsupervised models that correspond to unsupervised learning pattern also plays an important role in deep learning. Unsupervised models can be further subdivided into self-organizing maps (SOMs), Boltzmann machines, and auto-encoder. SOMs are usually deployed for feature dimension reduction as the output dimension is always reduced to 2. Compared to Boltzmann machines, restricted Boltzmann machines (RBMs), which is the basic component of deep belief networks (DBNs), are more popularly used for feature reduction. Also, autoencoder is another popular deep learning model for feature reduction. However, as CNNs are predominantly used in the area, we will shift our

TABLE 2 Characteristic of databases

	MIAS	BancoWeb	Magic-5	INbreast	DDSM	CBIS-DDSM	OPTIMAM
Year	1994	2010	1999	2012	1999	2017	2010-now
Number of cases	161	320	967	115	2620	1,644	
Views	<i>MLO</i>	<i>MLO, CC</i> and et al	–	<i>MLO, CC</i>	<i>MLO, CC</i>	<i>MLO, CC</i>	<i>MLO, CC</i>
Number of Images	322	1400	3369	410	10,480	3468	2,889,312
Image type	PGM	TIFF	DICOM	DICOM	LJPEG	DICOM	DICOM
Resolution	8 bits/pixel	12 bits/pixel	16 bits/pixel	14 bits/pixel	8 or 16 bits/pixel	16 bits/pixel	–
Image acquisition Mode	Screen film	Screen film	Screen film	Full-field	Screen film	Screen film	–
Lesion type	All kinds	All kinds	All kinds	All kinds	All kinds	MCCs and masses	All kinds
BI-RADS	No	No	Yes	Yes	Yes	Yes	
<i>GT</i>	Center and circle around	ROI partially available	Radiologist-drawn circle	Pixel-level contours	Pixel-level boundary	Pixel-level boundary	Pixel-level boundary
Clinical history	No	Yes	Yes	Yes	Age	Age	Yes
ACR breast density	No	No	Yes	Yes	Yes	Yes	–
Accessibility	Yes	Yes	No	Yes	Yes	Yes	Yes
Support	No	Yes	Yes	Yes	No	Yes	Yes

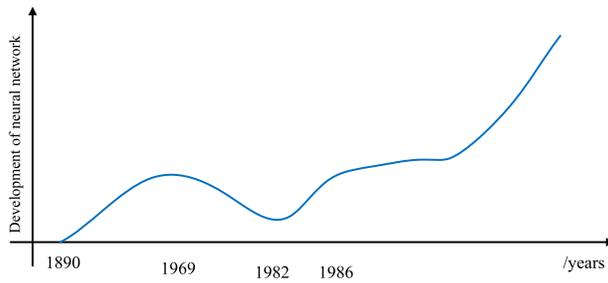


FIGURE 4 Development of neural network [Color figure can be viewed at wileyonlinelibrary.com]

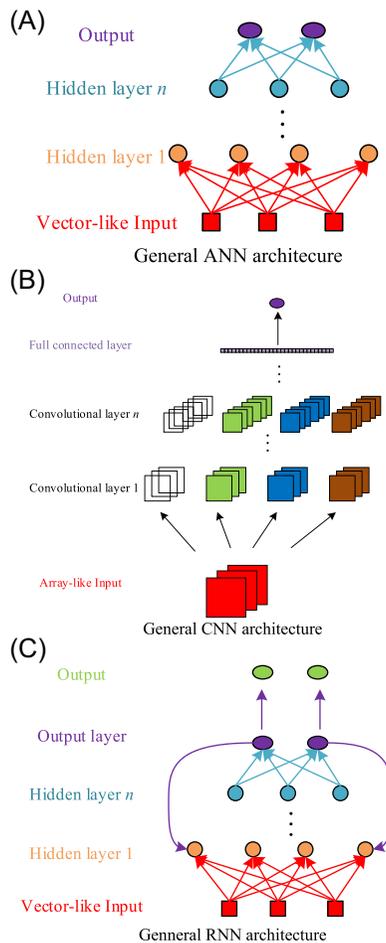


FIGURE 5 Typical schemes of ANN, CNN, and RNN models. (A) General ANN architecture, (A) general CNN architecture, and (C) general RNN architecture [Color figure can be viewed at wileyonlinelibrary.com]

focus to CNNs of deep learning. Interested readers are suggested to follow work.⁵³ Between the supervised learning models and unsupervised models, there are also semisupervised models including generative networks and attention mechanisms. The generative networks, such as variational autoencoders (VAEs) and generative adversarial networks (GANs) are actively

deployed for data augmentation in mammogram data sets. The attention mechanism, as a novel technique introduced into classifiers, pushes the classifiers towards only necessary features in the images.⁵⁴

In this section, we will introduce the supervised deep learning models including CNNs, fully convolutional neural networks (FCNNs). Usually, CNNs are deployed as classifiers for classification tasks while FCNNs are employed as semantic segmentation models for segmentation tasks. We aimed at a systematic survey paper here, so we believed there is a necessity to introduce the details of these two kinds of networks. Besides, both of them play important roles in breast disease CAD systems, a detailed illustration would be helpful for readers with different backgrounds to understand the paper. We end up this section by introducing some state-of-the-art applications based on deep learning as we would like to show a broader side of deep learning instead of restraining it on the application of breast diseases only.

3.1 | Convolutional neural network

As the most commonly used models, convolutional neural networks are the cornerstone of deep learning. Generally, CNNs consist of layers for convolution, activation, pooling, and fully connection, among which convolutional layers are of most significance. Convolutional layers are responsible for extracting features from images or feature maps produced by the previous layers to form new feature maps. A set of weighted windows, named filters or kernels, are sliding over feature maps to implement convolution in convolutional layers. When the filters convolve through local regions in feature maps, the weighted sum is calculated. An example of convolution can be seen in Figure 6. It is worth noting that all elements in the feature maps share the same filters in each channel while the receptive field refers to the size of the input perception range when calculating the convolutions. For traditional convolution, the receptive field equals the size of kernels.

where H , W , C respectively correspond to the height, width, and the number of input channels in the previous layer while H' , W' , C' are the counterparts of the output. Stride S is defined to specify the step of sliding windows when sliding through feature maps. Given the

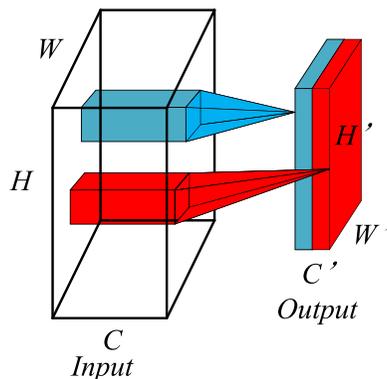


FIGURE 6 Convolution [Color figure can be viewed at wileyonlinelibrary.com]

size of filters D_F , D_F , F_C , $F_{C'}$ corresponding to the height, width, channels, and the number of groups, respectively, the following mathematical relationships can be inferred:

$$H' = \frac{H - D_F}{S} + 1 \quad (1)$$

$$W' = \frac{W - D_F}{S} + 1 \quad (2)$$

$$C' = F_{C'} \quad (3)$$

$$C = F_C \quad (4)$$

The weighted sum's feature maps are then passed to activation layers, which are units of nonlinearity activation functions such as ReLU.⁵⁵ The role of pooling layers in CNNs is to merge similar features into one and reduce the computational cost.⁵⁶ Pooling layers provide reliable motifs by coarse-graining the positions of features when relative positions of the features vary. Typical pooling units include Maximum pooling and Average pooling. Maximum pooling computes the maximum of the local patches of feature maps while Average pooling computes the average. Another important role of pooling layers is to reduce the dimensions of feature representation by neighboring pooling, where a unit takes the input from patches with columns and rows regularly left out. Convolutional layers, activation layers, and pooling layers are generally stacked together as a block which can be repetitive components to extract more high level but more abstract features for fully connected layers. More details about these layers will be given in the following sections. The general architecture of CNN is shown in Figure 7.

3.1.1 | Convolutions in CNNs

As one of the popular convolution techniques, group convolution was first introduced in AlexNet to solve the memory crash problem during network training. In group convolution, input feature maps are divided into different groups where they are convolved by different convolutional filters. The output feature maps are acquired by stacking the output after convolution. As it was initially proposed to solve issues of memory limitations, group convolution has been removed from most of the state-of-the-art networks thanks to the advancement of hardware. As pointed out in ShuffleNet,⁵⁷ multiple stacking of group convolutions weakens networks' representations (Figure 8), as features are only generated from fractions of previous

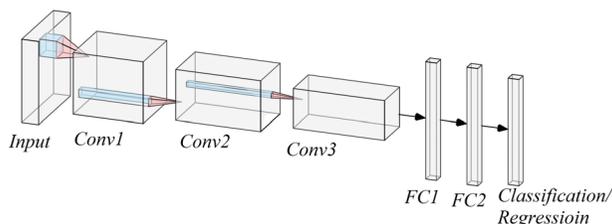


FIGURE 7 General architecture of CNNs [Color figure can be viewed at wileyonlinelibrary.com]

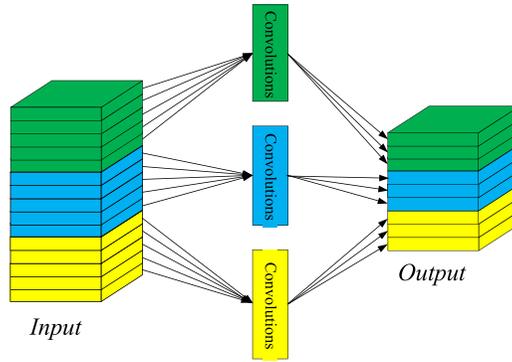


FIGURE 8 Group convolution [Color figure can be viewed at wileyonlinelibrary.com]

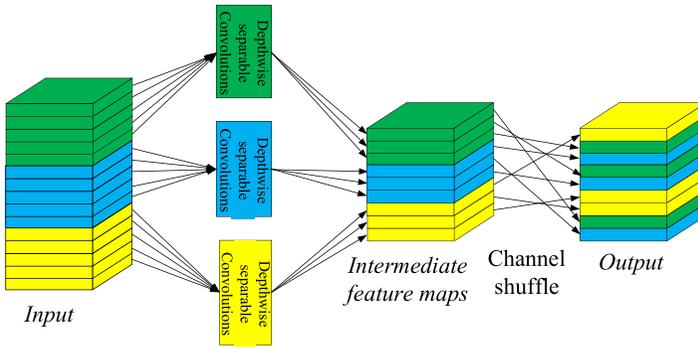


FIGURE 9 ShuffleNet convolution unit [Color figure can be viewed at wileyonlinelibrary.com]

channels. To resolve this in light-weighted networks, ShuffleNet proposed a novel two-stage convolution block. In the first stage, depthwise separable convolutions,⁵⁸ which will be introduced in detail later, are used in the group convolution to reduce the parameters. After convolution, the intermediate feature maps are shuffle to result in final output feature maps. Figure 9 shows the procedures involved in the ShuffleNet unit.

In traditional convolutions, the receptive field is restrained by the size of kernels. Larger receptive field usually means more learnable parameters as well as more complicated models. Effectively enlarging the receptive field of filters in convolutional operations without introducing extra parameters, dilated convolution, or Atrous convolution, has been widely used in the CNNs.⁵⁹ In a 2D case, given the input $X \in \mathbb{R}^{m \times n}$, convolutional kernel $W \in \mathbb{R}^{a \times a}$ (a is generally an odd number while $a \ll m, n$, “ \ll ” means much smaller), and the output $Y \in \mathbb{R}^{m' \times n'}$ then:

$$Y[i + a/2][j + a/2] = \sum_{c=1}^a \sum_{d=1}^a X[i + r \cdot c][j + r \cdot d]W[c][d] \tag{5}$$

r is the stride of convolution. Therefore, the size of the receptive field is determined by the stride r and the size of kernels. The size receptive field S has the following relationship with r and a :

$$S = a + (a - 1)(r - 1) \tag{6}$$

Compared to traditional 3 by 3 convolutional kernels that give the same size of the receptive field, a 3 by 3 dilated convolutional kernel gives a 5 by 5 receptive field when the stride r is 2 (Figure 10). However, the number of parameters is only 9 while it is 25 for a standard convolutional kernel to give the same size receptive field.

Along with the development, the design of the light-weighted for mobiles has also become a new focus. Depthwise separable convolutions are introduced to build lightweight models to embed deep learning models into portable devices like mobile phones.⁵⁸ Compared to traditional convolution, the proposed depthwise separable convolution has a lower computation cost. Figure 11 shows the convolutional operations in the depthwise separable convolution.

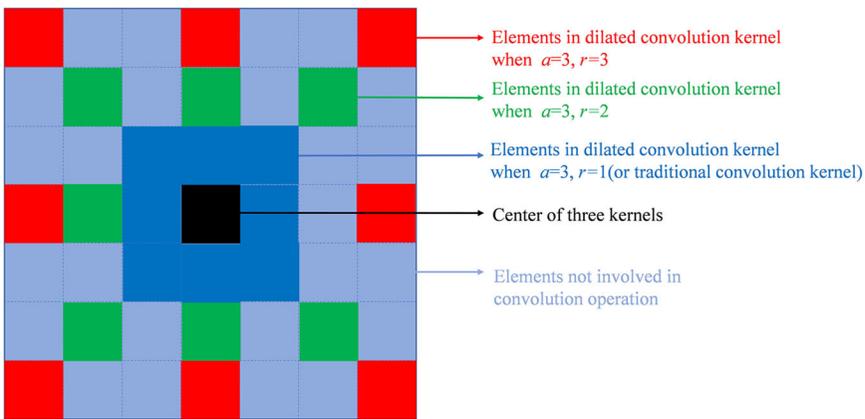


FIGURE 10 3 by 3 dilated convolution with different strides. Each kernel consists of one central element and eight surrounding elements while the distance of surrounding elements to the central element varies when stride r changes. When the stride r is 1, the dilated convolution kernel is as same as the traditional 3 by 3 convolution kernel [Color figure can be viewed at wileyonlinelibrary.com]

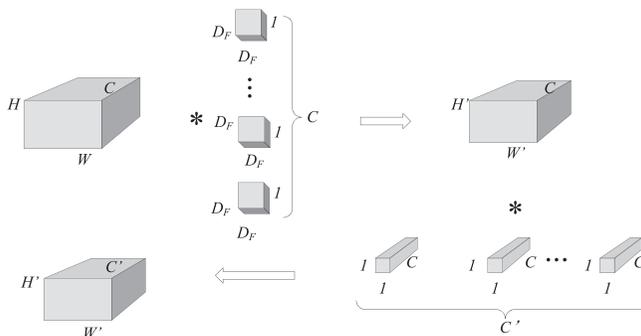


FIGURE 11 Depthwise separable convolution. Features maps of size $H \times W \times C$ are first convolved by the filters with size $D_F \times D_F \times 1 \times C$, which gives the resulting feature maps $H' \times W' \times C$. The second group of filters with size $1 \times 1 \times C$ are used to maintain the height and width of the feature maps while changing the number of channels from C to C'

For traditional convolution in Figure 6, the total computational cost is:

$$D_F \times D_F \times C \times H' \times W' \times C' \quad (7)$$

However, the total computational cost for depthwise separable convolution is:

$$D_F \times D_F \times C \times H' \times W' + H' \times W' \times C \times C' \quad (8)$$

A reduction rate of computation can be denoted as:

$$R_{reduction} = 1 - \frac{D_F \times D_F \times C \times H' \times W' + H' \times W' \times C \times C'}{D_F \times D_F \times C \times H' \times W' \times C'} \quad (9)$$

That is:

$$R_{reduction} = 1 - \frac{1}{C'} - \frac{1}{D_F^2} \quad (10)$$

The reduction rate is determined by the number of output feature maps and the size of filters, which implies meaningful reduction can be achieved when proposed C' and D_F is chosen. By introducing a fire module in Iandola et al.,⁶⁰ the proposed SqueezeNet achieved AlexNet-level accuracy while the number of parameters is more than 50 times fewer than that of AlexNet. The details about the fire module are shown in Figure 12. The variants of convolutions are shown in Table 3.

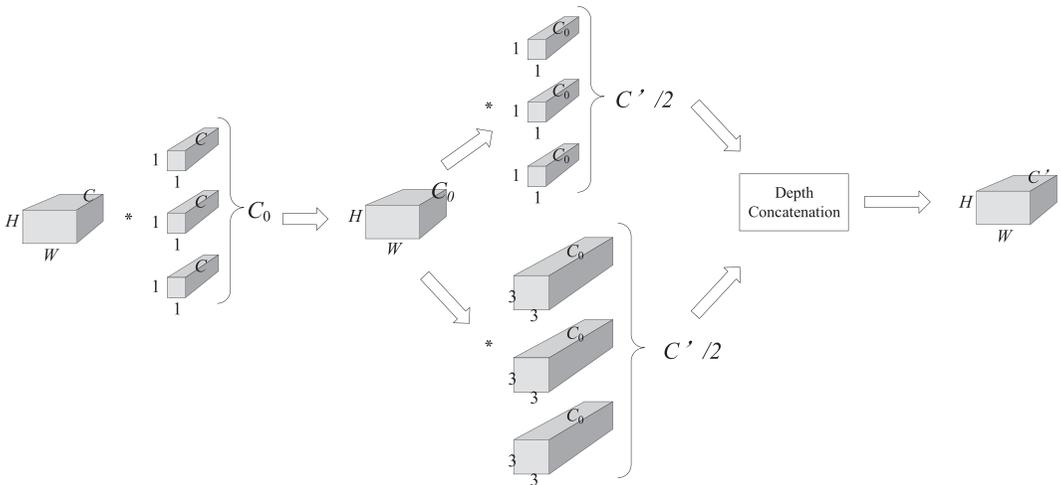


FIGURE 12 Fire module in SqueezeNet. The size of feature maps is firstly squeezed by $1 \times 1 \times C \times C_0$ ($C_0 < C$) filters. Then two filters with different sizes are convolved with resulting feature maps in the first step to extend the channels of feature maps to the expected C' . Depth concatenation is applied to concatenate two groups of feature maps obtained through the two different filters. The height and width of the input feature maps and the output feature maps are the same while there are more channels in the output feature maps. However, Maxpooling is used in SqueezeNet to shrink the height and width of the output feature maps

TABLE 3 Variants of convolutions

Name	Stride	Groups	Dilation rate	Features
Normal convolution	1	1	1	Basis
Strided convolution	≥ 2	1	1	Reduce computational costs
Grouped convolution	Flexible	≥ 2	1	Parameters reduction
Dilated convolution	Flexible	1	≥ 2	Larger receptive field
Depthwise separable convolution	Flexible	≥ 2	1	Parameters reduction

3.1.2 | Pooling

Pooling is commonly used in CNNs for feature reduction and prevention of overfitting. As the location of the input is precisely recorded by stacking convolutional layers, limited representation can be learned. Also, abundant features can be extracted by convolutional layers with multiple kernels. However, information abundance slows the computation and increases redundancy. Therefore, pooling is introduced to retain the most important features in the feature maps when down-sampling and compressing the extracted feature maps. In pooling layers, unlike the parameters in convolutional layers that are learnable, the parameters are usually fixed instead. Besides feature reduction, pooling is also well known for its spatial invariance that translation, rotation, and scale will not significantly change feature maps.

Max pooling and average pooling are two representative pooling operations. For max pooling, the maximum value from a local region of each feature map is selected as the most representative feature preserving texture features. Average pooling averages elements in a local area of each feature map and passes the averaged value as it is believed that overall information can be better preserved by average pooling. To produce feature maps with smaller sizes, the stride for pooling is usually greater than 1. However, overlapping pooling, where the stride is smaller than the kernel size, is also commonly used to produce feature maps with more information. Global pooling is another widely used pooling technique. Unlike other pooling layers that downsample patches of the input feature maps, global pooling down samples each feature map into a single value instead. Global pooling can summarize the presence of features in the images and sometimes can be an alternative to fully connected layers to transform feature maps to output predictions for the classification models. Depends on the rules of calculation, global pooling can be subdivided into global average pooling and global max pooling that are quite similar to average pooling and max pooling but with only single values as the output. Stochastic pooling,⁶¹ which chooses activation based on the probability distribution in local areas, is another alternative to commonly used max pooling and average pooling. It was claimed that stochastic pooling could be used to avoid over-fitting and reduce test errors.

A rank-based pooling was proposed by Shi and others,⁶² in which activations in pooling regions are ranked before performing pooling operation. Based on the ranked activations, three derived pooling methods including rank-based weighted pooling (RWP), rank-based average pooling (RAP), and rank-based stochastic pooling (RSP) can be implemented when corresponding weighting strategies applied. The rank-based pooling proposition was based on the observations that the ranking list would remain unchanged, though activation

values would change in a pooling region, contributing to more robust pooling performance and thus better performance of CNN models. Ordinal pooling, a novel pooling scheme that arranges a pooling region's activations in a sequence, was proposed recently in Kumar.⁶³ Different weights are assigned to activations regarding the order of activations in the sequence while these weights are learned through a standard gradient-based training. The pooling operation is then performed by summed activations that are multiplied by the weights. The proposed pooling scheme performs a hybrid behavior between the average pooling and max pooling in a differentiable manner. As indicated by the experiments in Kumar,⁶³ networks equipped with different pooling operations within pooling layers would be more advantageous. The proposed ordinal pooling facilitates the training process and eases the issue of choice between max pooling and average pooling. The summary of different pooling methods is given in Table 4.

3.1.3 | Activations and fully connected layers

The activation function is introduced into CNN for nonlinearity. In practice, data is not usually linearly classifiable. For CNN-based classifiers, the performance would significantly decrease if no nonlinear activation units were introduced into the classifiers. Sigmoid function and Tanh function are the two most popular activation functions in early neural networks. The sigmoid function could be expressed as follows:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (11)$$

where e stands for the natural number. The sigmoid function is introduced as the activation function because it is smooth and differentiable (Figure 13). However, one main shortcoming

TABLE 4 Pooling methods

Name	Stride	Strategies	Features
Max pooling	Equals to kernel size	Select maximum activations	Preserve texture feature
Average pooling	Equals to kernel size	Average activations	Preserve background information
Overlapping pooling	Less than kernel size	Usually, select maximum activations	Better representation
Global pooling	No		Summarize features into single values
Stochastic pooling	Equals to kernel size	Select activations based on the probability distribution	Prevention of overfitting
Rank-based pooling	Equals to kernel size	Select activations based on the rank of activations	More robustness
Ordinal pooling	Equals to kernel size	Select activations based on the order of activations	Facilitate training

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

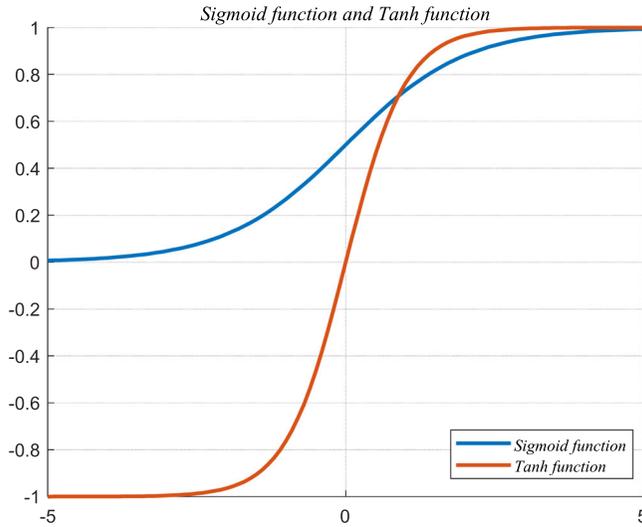


FIGURE 13 Sigmoid and Tanh function [Color figure can be viewed at wileyonlinelibrary.com]

brought by it is the gradient vanishing as networks going deeper. Besides, the sigmoid function's output is no longer zero-centered and therefore changes the distribution of input. Tanh function, which mitigates gradient vanishing to some extent and centers the output to be zero-centered, can be written as follows:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (12)$$

However, the limitations of the Tanh function like high computation cost and irresolvable gradient vanishing restrain the usage of Tanh function. The Rectified Linear Unit (ReLU) function then becomes the most commonly used⁵⁵ activation function as it converges quickly without introducing a gradient vanishing problem. The definition of ReLU can be given in the form of:

$$\text{ReLU}(x) = \begin{cases} x, & \text{when } x \geq 0 \\ 0, & \text{when } x < 0 \end{cases} \quad (13)$$

As the gradient vanishing no longer exists when ReLU is used as the activation function, deeper networks can be effectively trained. The ReLU activation function is shown in Figure 14.

However, as can be seen from Figure 14, some features that might be useful are lost as the output of the ReLU function is set to be zeros when activations are negative. To reduce feature loss due to activation functions, numerous ReLU variants are being introduced in the area. Leaky ReLU (LReLU) remains the negative of activations by multiplying the negatives with a

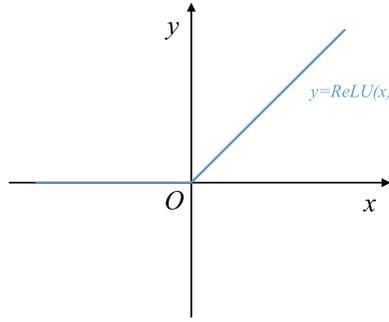


FIGURE 14 ReLU activation function [Color figure can be viewed at wileyonlinelibrary.com]

small positive constant of 0.01 while the values of positive activations remain the same as that of ReLU. Mathematically, LReLU can be expressed as follows:

$$LReLU(x) = \begin{cases} x, & \text{when } x \geq 0 \\ 0.01x, & \text{when } x < 0 \end{cases} \quad (14)$$

However, the fixed slope in LReLU can be inappropriate on some occasions. To the end of setting flexible slopes for negative activations, PReLU was proposed. PReLU can be written in the form as follows:

$$PReLU(x) = \begin{cases} x, & \text{when } x \geq 0 \\ \alpha x, & \text{when } x < 0 \end{cases} \quad (15)$$

where α is a small positive variable that is determined according to neural networks. In another strategy called randomized ReLU (RReLU), the value of the slope for negative activations is randomly chosen from a predefined range. Instead of determining the slope before training in PReLU, the slope in RReLU is selected from a given range in each training epoch. RReLU function can be extended as follows:

$$RReLU(x) = \begin{cases} x, & \text{when } x \geq 0 \\ \alpha' x, & \text{when } x < 0 \end{cases} \quad (16)$$

where α' follows a uniform distribution from c to d , where $c, d \in [0, 1]$. Besides the series of above-mentioned ReLU, there are also activation functions that introduce exponential operation for better approximations of negative activations. Exponential linear unit (ELU), a typical function that allows negative activations to be close to zero, has also been widely used as the activation function. ELU can be written as follows:

$$ELU(x) = \begin{cases} x, & \text{when } x \geq 0 \\ \beta(e^x - 1), & \text{when } x < 0 \end{cases} \quad (17)$$

where β determines the scale of closeness between the output of activations to zero. The larger β , the bigger distance. Two ELU functions with different values of β are shown in Figure 15.

ELU function can be easily extended to scaled exponential linear units (SELU) by introducing a scale factor γ . SELU, therefore, can be written as follows:

$$SELU(x) = \begin{cases} \gamma x, & \text{when } x \geq 0 \\ \gamma\beta(e^x - 1), & \text{when } x < 0 \end{cases} \quad (18)$$

In Table 5, we briefly summarized the activation functions for reference.

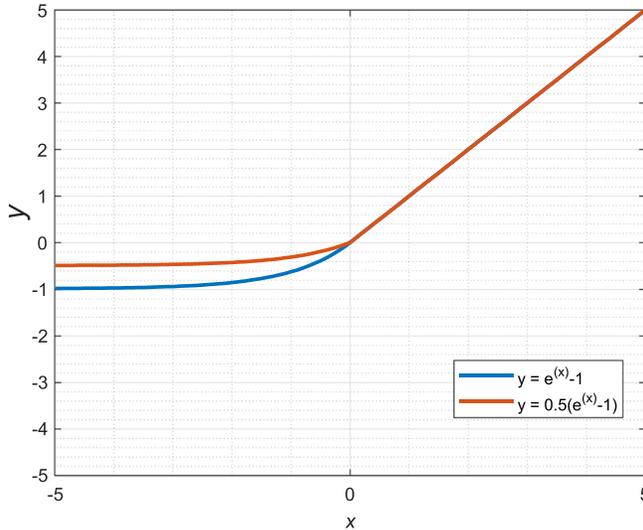


FIGURE 15 ELU functions when $\beta = 0.5$ and $\beta = 1$ [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Activation functions

Name	Symmetry about origin	Saturable	Convergence speed	Output range	Features
Sigmoid	No	Yes	Low	$(0, 1)$	Gradient vanishing
Tanh	Yes	Yes	Low	$(-1, 1)$	Zero-centered, faster than Sigmoid but still suffers from gradient vanishing
ReLU	No	No	Fast	$[0, +\infty)$	No gradient vanishing but may have dead neurons
Leaky ReLU	No	No	Restricted fast	$(-\infty, +\infty)$	Ease dead neurons issue
PReLU	No	No	Relatively fast	$(-\infty, +\infty)$	Faster than leaky ReLU
RReLU	No	No	Restricted Fast	$(-\infty, +\infty)$	Flexibility
ELU	No	No	Fast	$(-1, +\infty)$	Faster than ReLU
SELU	No	No	Restricted Fast	$(-1, +\infty)$	Flexibility

The fully connected layer works as the “classifier” in the whole CNN model. All previous layers including convolutional layers, pooling layers, and activation layers focus on extracting and mapping useful into lower-dimensional representations; the fully connected layer then maps those representations into the targeted space and then completes classification. Feature maps from the last activation layers are first vectorized as the input for the fully connected layers. By stacking multiple layers of the fully connected layers, features are weighted by different weights, and the dimensionality of the features is further reduced towards the targeted one. Compared to the number of parameters in convolutional layers, the number of parameters in the fully connected layers is usually larger. Also, the models tend to overfit when a large sum of parameters is introduced into the models. To resolve these issues, global pooling has become the substitute for the fully connected layers. After replacing the fully connected layers with global pooling layers, the number of parameters in models could be significantly reduced while the vectorization of features turns out to be much simpler.

3.1.4 | Novel convolutional architectures

Since the proposal of AlexNet,⁹ numerous high-performance CNN models have been developed in the past few years with novel architectures. AlexNet, not only the biggest winner in ImageNet Large-Scale Visual Recognition Challenge 2010 (ILSVRC-2010)¹⁰ contest and ILSVRC-2012 competition⁶⁴ with only 8-layer in architecture but also played the role as the cornerstone for the development of deep CNN in the following years. It has been noted that the number of learnable parameters in CNN is largely determined by the number of neurons in the fully connected layers. Therefore, the removal and replacement of fully connected layers raised great concern. In 2013, Min et al. proposed to replace fully connected layers with global average pooling layers and therefore, reduce the size of prior-arts greatly.⁶⁵ Increasing the depth and width of the network is an intuitive way to improve network performance. In 2015, the authors of VGG net proposed to increase the performance of CNN by increasing the depth of networks.⁶⁶ Only 3×3 kernels are used throughout the proposed models, which showed remarkable performance improvement in the ImageNet challenge. GoogLeNet,¹² the champion of ILSVRC-2014, introduced novel architecture named Inception to increase the width and depth of the network while kept the computational costs expedient. Three convolution kernels with different sizes are paralleled in three columns to extract rich features while the parallel pooling path, according to the authors, was added for possible benefit. All of the outputs from each path are concatenated to form the final output feature maps. However, one problem brings in by concatenation of feature maps is the unavoidable increase of the output feature maps. Therefore, 1×1 convolutions were introduced to reduce the dimensions during convolution operations. The details of Inception module is given in Figure 16.

However, the deeper the network, the more difficult it is in training. To mitigate the difficulty in training sessions of deep CNNs, researchers developed residual learning, a certain kind of shortcut connection technique.¹¹ An illustration of residual learning is shown in Figure 17.

Given the desired output $H(I)$ as the underlying mapping of input I after stacked nonlinear layers, the difference between $H(I)$ and I can be denoted as $F(I)$, which is

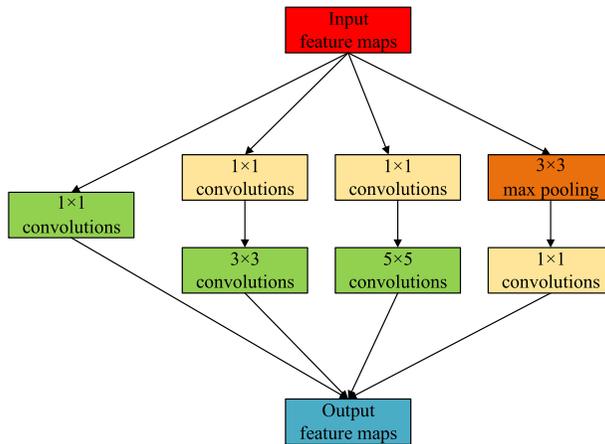


FIGURE 16 Inception module in GoogLeNet [Color figure can be viewed at wileyonlinelibrary.com]

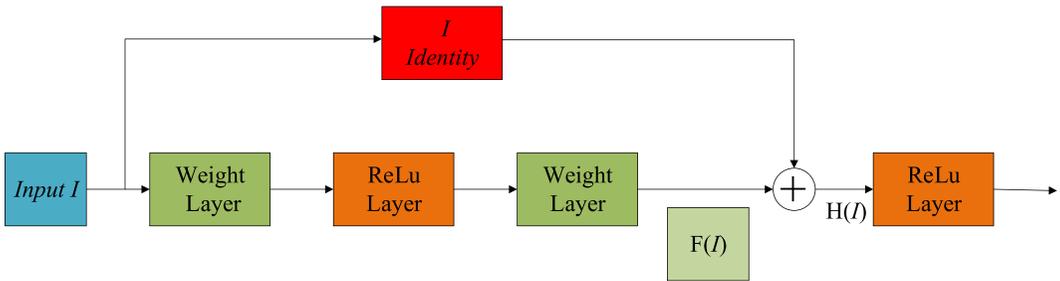


FIGURE 17 Residual learning block [Color figure can be viewed at wileyonlinelibrary.com]

$$F(I) = H(I) - I \quad (19)$$

Therefore, $H(I)$ can be represented by $F(I)$ and I by:

$$H(I) = F(I) + I \quad (20)$$

It was believed that $F(I)$ is easier to be optimized than $H(I)$ because $F(I)$ is mapped from I by the nonlinear units while $H(I)$ is unreferenced to I . There are also numerous shortcut connection techniques found to be useful to improve the performance of CNNs. DenseNet, which incorporates the novel convolution blocks codenamed dense blocks, showed great improvement on the popular image recognition tasks, including ImageNet and CIFAR-10.^{67,68} In dense block, several convolution layers are stacked for feature extraction where feature maps generated in the previous convolution layers are stacked as input for following convolution layers in the block. Details can be seen in Figure 18. Given the number of channels of the input feature maps is k_0 , each convolution layer produces k channels feature maps, then the number of channels of feature maps of i th layer can be expressed as follows:

$$k_i = k_0 + (i - 1) * k \quad (21)$$

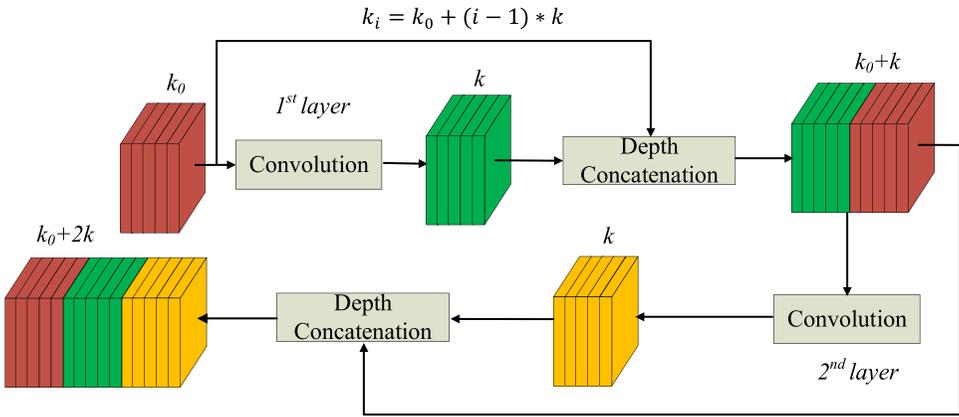


FIGURE 18 Dense block [Color figure can be viewed at wileyonlinelibrary.com]

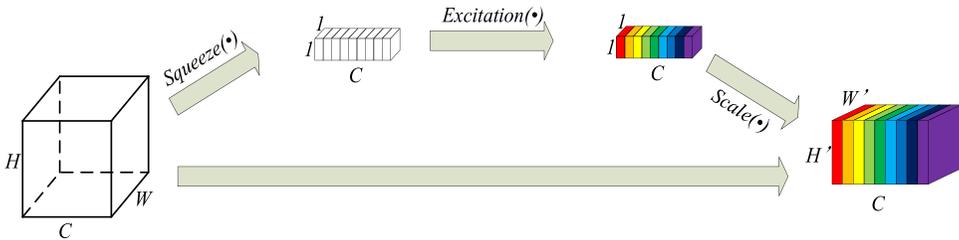


FIGURE 19 SE block [Color figure can be viewed at wileyonlinelibrary.com]

To improve the CNNs representational power by channel relationship, a new architectural unit named SE block (Figure 19) was developed by Hu and others.⁶⁹ SE block is a two-stage convolution block that consists of the *squeeze* phase and *excitation* phase. During the *squeeze* phase, a channel descriptor of size $1 \times 1 \times C$ (C corresponds to the channels of feature maps) depicts global information in each channel and therefore enables useful information to be enhanced channel-wise. In the *excitation* phase, specific activations learned based on the relationships between channels are taken as the output. The output has the same size as the input feature activation. One advantage of the SE block is that it can be inserted into any depth of the state-of-the-art network with little computational costs introduced.

Given the input feature map $X \in \mathbb{R}^{H \times W \times C}$, then *squeeze* operation generates a $1 \times 1 \times C$ feature map $R \in \mathbb{R}^{1 \times 1 \times C}$ by global average pooling, which can be denoted as follows:

$$R_i = \frac{1}{H * W} \sum_{m=1}^H \sum_{n=1}^W X_i(m, n) \tag{22}$$

where R_i and X_i stand for i th feature map in R and X , respectively. For *excitation* operation, two fully connected layers with parameters $W_1 \in \mathbb{R}^{r \times C}$ and $W_2 \in \mathbb{R}^{C \times r}$ are stacked to extract channel-wise relationships. The feature map R' after *excitation* operation can then be denoted as follows:

$$R' = \text{sigmoid}(W_2 \delta(W_1 R)) \quad (23)$$

where δ is the ReLU function.⁵⁵ Then in *Scale* operation, R' pointwise multiplies with the input feature map X and leads to channel enhanced feature maps. By iteratively updating the parameters W_1 and W_2 , more representative channels are enhanced while the less representative channels are depressed instead. With the development of CNNs, researchers have endeavored to propose new models with computational efficiency and comparable even higher accuracy on ImageNet. The summary of the novel convolutional architectures can be seen in Table 6.

3.2 | Fully convolutional neural network

CNNs are widely used as solutions for image-level tasks such as image classification and detection. However, there are also demanding needs on pixel-level tasks such as segmentation. Therefore, FCNNs are developed. FCNNs can be adapted from CNNs, where fully connected layers are replaced with deconvolutional layers and upsampling layers,⁷⁰ the backward versions of convolutional layers and subsampling layers (Figure 20).

Given the size of input of deconvolution i , the size of kernel k , the padding size p , the stride s , then the output size o can be denoted as follows:

TABLE 6 Novel convolutional architectures

Name	Convolutions	Size of kernels	Shortcut connection	Features
Inception module	Padded	$1 \times 1, 3 \times 3, 5 \times 5, 7 \times 7$	No	Higher efficiency on feature utilization
Residual learning block	Normal	$1 \times 1, 3 \times 3, 5 \times 5, 7 \times 7$	Residual learning	Mitigate training difficulty for deep networks
Dense block	Normal	$1 \times 1, 3 \times 3, 5 \times 5, 7 \times 7$	Dense connection	Allow features to be reused in consequent layers
SE block	Normal	$1 \times 1, 3 \times 3, 5 \times 5, 7 \times 7$	Squeeze and Excitation	Improve CNN representation through channel relationship

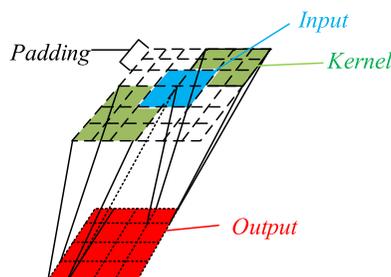


FIGURE 20 Deconvolution. The blue grid is the input while the green grids are kernels for deconvolution. The output is denoted in red grids where input is firstly padded [Color figure can be viewed at wileyonlinelibrary.com]

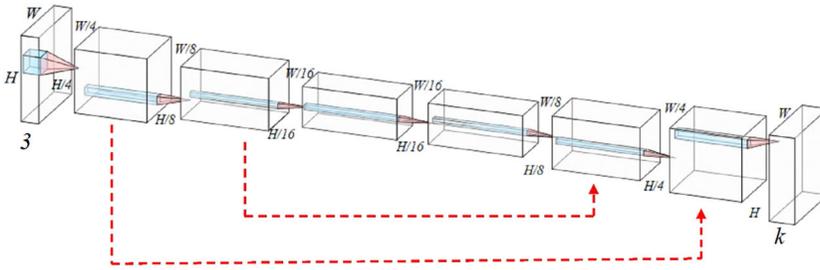


FIGURE 21 Architecture of FCNNs. Dashed red arrows denote the shortcut connections between feature maps [Color figure can be viewed at wileyonlinelibrary.com]

$$o = s(i - 1) + 2p - k + 2 \quad (24)$$

Unlike CNNs, which generate probability scores of different classes, FCNNs generate score maps of the same size as the input images. Each pixel in the input images is classified into one of k targeted categories, and segmentation results come from the probability map of pixels. Feature maps in earlier layers are also combined in FCNNs through upsampling layers with deconvolutional results to improve prediction accuracy. An example of FCN is shown in Figure 21.

In Figure 21, W , H , 3 are the width, height, and the number of channels of the input image. k is the number of classes that pixels belong to. Intermediate feature maps are symmetrically downsampled and upsampled due to convolution and deconvolution. The depth, which is the level of downsampling, could be even deeper if necessary. The red dashed lines denote the skip connection. The output feature maps of the second and third convolutional layers are concatenated to the second last and last deconvolutional layers by upsampling.

Novel FCNNs have achieved promising results in the applications of many fields. DeepLab is an FCN that merges the methods from Deep CNNs (DCNNs) and probabilistic graphical models into one model to address the pixel-level classification problem.⁷¹ Because of the insufficiency of responses at the final DCNN layer to provide accurate object segmentation, the responses are then combined with a fully connected conditional random field (CRF). On the PASCAL VOC-2012 semantic image segmentation task, the proposed framework achieved 71.6% IOU accuracy on the test set, which was the best to date. In 2017, A new version of DeepLab surprisingly achieved 79.7 percent mIOU in the test set of PASCAL VOC-2012 semantic image segmentation task.⁵⁹ In this version of DeepLab, the authors proposed to combine atrous convolution with the spatial pyramid pooling (SPP) technique and thus formed so-called atrous spatial pyramid pooling (ASPP). Atrous convolution, which was initially designed for wavelet decomposition in Holschneider et al.,⁷² enables enlarged receptive fields of filters in convolutional layers without bringing in the extra number of parameters. Thanks to SPP and atrous convolution, convolutional layers in the ASPP framework can capture image context and objects in images at multiple scales.

U-net, an end-to-end FCN, won the ISBI cell tracking challenge 2015 by a large margin with fast speed.⁷³ Two paths, including a contracting path for context capture and an expanding path that provides a symmetric extension of feature maps, enable precise localization. The training strategy proposed turns out to use the available annotated samples and data augmentation more efficiently. It was proved on the ISBI challenge for segmentation that the proposed

network can be trained end-to-end by using few images with exceeding performance than that of the prior best method. A simplified architecture of U-Net is shown in Figure 22.

Another semantic pixel-wise segmentation architecture is SegNet, where the core lies in the trainable segmentation engine comprising of AE (Figure 23). The encoder structure is the same as topological the VGG16 network⁶⁶ while the decoder is responsible for mapping low-resolution features maps from the encoder to feature maps that have the same size of input images.

The pooling indices, which were computed in the max-pooling layer of the corresponding encoder, were used in the decoder. By doing such, learning to upsampling is removed. Dense feature maps are acquired through the sparse upsampled maps that were convolved with trainable filters. Compared to the other state-of-the-art architectures such as DeepLab-LargeFov,⁷¹ DeconvNet,⁷⁴ SegNet is more efficient both on memory and computational time. When comparing the architectures in terms of the number of trainable parameters, it is also much smaller. In the latest version of DeepLab, which was known as DeepLabv3+, methods from SPP, encoder-decoder, separable convolution are included to forge a faster and stronger encoder-decoder network. The accuracy on the test set of PASCAL VOC 2012 is then further improved to 89%.

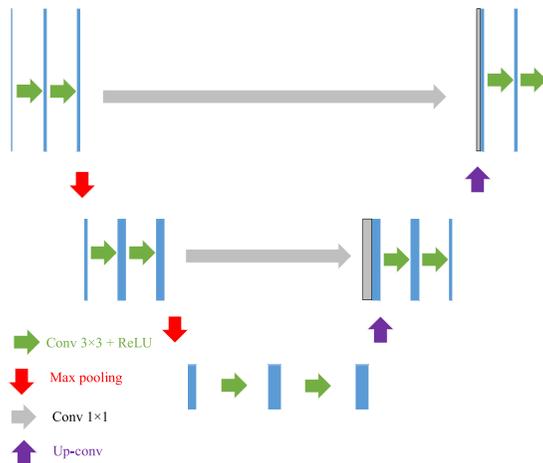


FIGURE 22 Simplified architecture of U-net [Color figure can be viewed at wileyonlinelibrary.com]

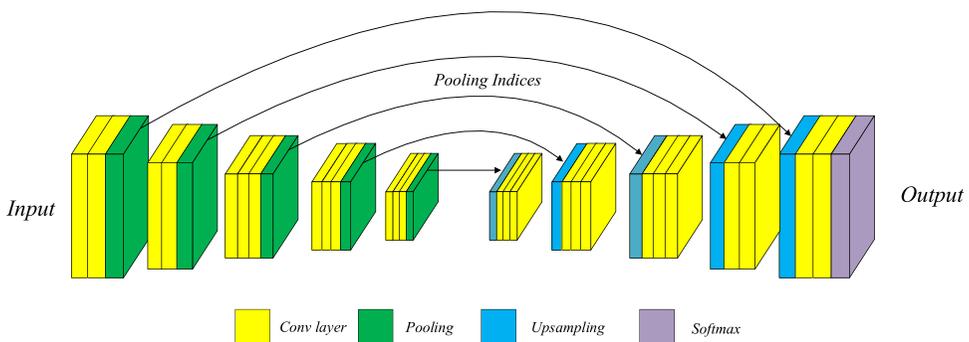


FIGURE 23 Architecture of SegNet [Color figure can be viewed at wileyonlinelibrary.com]

3.3 | Optimization techniques

Usually, optimization is also an important element in the successful application of deep learning. Along with the development of novel architectures of deep learning models, novel optimization algorithms have been extensively explored as well. In this section, we will briefly review the useful optimization techniques including regularization technique and weight initialization technique.

Regularization technique can be subdivided into three categories including regularization through data, network architecture, and regularization term respectively. In data-based regularization techniques, batch normalization and dropout are two representative techniques, where the dropout technique is especially effective for overfitting mitigation. Similar to dropout, there is another drop connection technique called DropConnect that randomly sets the subset of weights to be zero.⁷⁵ Other variants of dropout can be seen in References [76–78]. Novel architectures can also serve as a regularization technique. Weight sharing, which allows part of the networks to share the weights between them, reduces the number of training parameters and therefore improves the generality of deep learning models. Besides the weight sharing scheme, proper activation functions also help to improve the performance of deep learning models. The introduction of rectified linear unit (ReLU) successfully solved the gradient vanishing problem brought by the sigmoid function while ReLU tends to be more expressive than the sigmoid function. Additionally, transfer learning and meta-learning are the other two useful methods that can be treated as regularization methods. It turns out to be more intuitive to implement regularization through regularization terms or so-called regularizers by adding them to the loss functions. The regularization term is independent of the target but introduces desired properties into the models. Some examples of regularizer-based regularization methods can be seen in Lasserre et al.⁷⁹ and Goodfellow.⁸⁰

Weight initialization technique is also part of optimization as the purpose of weight initialization is to improve the performance of deep learning models. The most intuitive weight initialization way is to set the weight to be zeros or constant. However, the issue is that the distribution of the weights in each learnable layer remains the same and therefore impairs the learning capability of the deep learning models. Random selection of weights from a normal distribution can substitute for constant or zero initialization. However, gradient vanishing and explosion can happen due to randomness. A further consequence is that the deep learning model may never converge. The improved version of the random selection of weights is to select weights from a bounded uniform distribution. Other initialization techniques can be found in Glorot and Bengio⁸¹ and He et al.⁸²

3.4 | Applications of deep learning

Deep CNNs are widely used in computer vision tasks for image classification, object detection, and image caption. Besides the mentioned networks with novel convolutional units and novel architectures, numerous state-of-the-art networks show high performance on the image classification task.^{13,14,83} Traditionally, image classification is two-stage, which comprises feature extraction and classification. In feature extraction, handcrafted features are firstly extracted for the following trainable classifiers.⁸⁴ However, a major shortcoming is that the systems' performance heavily relies on the design of feature extraction. The advent of deep CNNs, however, has successfully set a new milestone for image classification. In usual life, deep CNN-based

image classification methods have been merged into various scenarios such as Gaming, automotive, and manufacturing.^{85,86} In autonomous driving, image recognition, which is image classification problems, in essence, plays a key role in identifying moving and static objects, including pedestrians, pathways, traffic lights, and road signs. In manufacturing, image classification can be used to detect defects within the manufacturing process and improve the quality of industrial products.

Besides image classification for the initial proposal of CNNs, object detection is also an area where CNNs have been extensively applied. Object detection is composed of two subtasks: object localization and object classification. Localization determines the positions of objects in the given images, while classification follows to classify the found objects into a specific category. For a long, object detection remains to be an attractive yet challenging task given the complexity of images with little gain in performance. The sliding window strategy is popularly used in the community to generate bounding boxes (BB), regions where objects are likely to show up. However, the process seems to be inefficient and even inaccurate. Besides, combinations of low-level descriptors and shallow models can not bridge the semantic gap. When CNNs were brought into the area, object detection has been greatly advanced because of its capability to learn more complex features and the flexibility to learn more informative representations. Ideal object detection systems are expected to accurately detect and recognize objects with high efficiency, measured by computational time, memory, and storage.⁸⁷

Deep-learning-based object detection frameworks can be classified into two types. One followed the traditional pipeline of detection by generating the region proposals first and then have the proposals classified into different categories.^{88–91} One typical region-based network is RCNN.⁸⁹ Though it showed high performance on detect, the framework suffers from several flaws. One is the optimization difficulty due to multistage training. In RCNN, four stages including region proposals computation, CNN model training and finetuning, class-specific SVM classifiers training, and bounding box regressor training. However, each stage has to be individually trained, which leads to a time-consuming training process and makes it even harder to be optimized. Also, when it comes to the testing phase, testing is slow because the features of each region proposed are extracted by CNNs. To speed up the detection process, SPPNet, which solved the problem with fixed-size input, was proposed to generate the fixed length of features for the arbitrarily sized candidate regions in the test image.⁹⁰ An improved version of RCNN named Fast RCNN enabled the simultaneous training of a softmax classifier for classification and a class-specific bounding box regression. Due to this improvement, Fast RCNN was typically three times faster than RCNN/SPPNet during the training session.⁹² Based on a novel region proposal block named region proposal network (RPN), faster RCNN can break the speed bottleneck in Fast RCNN.⁹³ In RPN, a certain number of anchors of different aspect ratios and scales, or the reference boxes, are generated at feature maps produced by convolution. Each anchor was then mapped to a vector with lower dimensionality for two siblings FC layers, which are responsible for object classification and box regression. Contrary to reliance on external region proposals of fast RCNN, RPN in faster RCNN uses internal resources within the network and therefore contributes to further speedup of detection. In R-FCN,⁹⁴ the authors proposed to construct position-sensitive score maps by using specialized convolution layers as the output of a fully convolutional network. R-FCN can achieve comparable accuracy to faster RCNN with backbone ResNet101 with shorter running times. Mask RCNN, an extended version of faster RCNN, focuses on pixel-wise object instance segmentation. In mask RCNN, a new FCN branch was added to output binary masks for each RoI when another

branch simultaneously predicts the class and box offset. The performance of mask RCNN is startling that it achieved top results on COCO object instance segmentation and bounding box object detection.

The aforementioned two-stage region-based showed powerful performance on general object detection. However, these approaches are too computationally expensive for mobile devices due to the limited computational capability and storage. Hence, another strategy that treats the detection as a regression or classification problem by employing a unified framework has been in the focus of research as well. The typical representative of unified frameworks is *DetectorNet*, *OverFeat*, *YOLO*, *SSD*, and *CornetNet*.^{88,95–98} As the first attempt at deploying CNN for object detection, *DetectorNet* used AlexNet with the final softmax classifier layer replaced by a regression layer. Foreground pixels of the object are detected by one network while the object's top, bottom, left, and right halves have to be detected by four additional networks. During the training session, many crops of the images have to be taken as the input of multiple networks, which slows the detection process. As the champion of ILSVRC2013 localization and detection competition, *OverFeat* remains one of the most influential networks in object detection. *OverFeat* used the sliding window to generate object candidates while fully connected layers, which restrain input size, are removed to form a fully convolutional network. The input image was scaled up to six different scales for multiple-scaled features to improve the overall performance. For each scaled input, a grid of predictions was generated. The offset max-pooling was applied for more views and therefore increases the robustness. When objects were identified, bounding box regressors were applied to generate bounding box predictions while a greedy merge strategy was deployed to combine each prediction. YOLO is another expedient framework that runs at 45 FPS in real-time.⁹⁶ Unlike two-staged frameworks that predict detections through features from local regions, YOLO uses global features from the entire image. Specifically, an input image is divided into $S \times S$ grid, within which each subgrid C -class probabilities, confidence scores, and B bounding box locations are predicted. However, more localization errors were reported of YOLO when compared to Fast RCNN due to the improper setting of scale, aspect ratio. Single shoot detector (SSD), a framework that is even faster than YOLO without sacrificing too much detection performance, was well known for multiscale detection. Early SSD layers were maintained as popular networks while extra convolution layers are stacked to the top to provide multi-scale feature maps, based on which category scores and box offsets are predicted. The proposal of *CornetNet* comes from the questioning about the predominant role of anchors. Authors of *CornetNet* claimed that anchor boxes would cause issues including a huge imbalance between positive and negative samples, slowing down the training, and so on. To resolve this, the author formulated the bounding box detection problem as paired keypoints, which are top-left and bottom-right, detection problems. The backbone of *CornetNet* is two stacked Hourglass networks while a simple corner pooling approach was attached for better corner localization.⁹⁹ The advantage of *CornetNet* is that it outperformed all previous one-stage detectors. However, the inference time is slower than the rest of one stage detector. Nevertheless, an improved version of *CornetNet* named CenterNet further raised MS COCO AP to 47.0% though it was slower than *CornetNet*.¹⁰⁰

One of the areas that benefit most from the advancement of CNNs is medical image analysis. Thanks to the massive growth in the volume of medical data and computing power of the hardware, deep learning has been successfully applied for risk predictions and automation analysis.¹⁰¹ In Nasr-Esfahani et al.,¹⁰² a simple CNN with only two convolutional layers was developed for the detection of Melanoma. One hundred and seventy images, including 70 melanoma and 100 nevi, are analyzed in the research. Data augmentation was applied to

generate a total of 6120 images, while 80% of them are partitioned into the training set. The developed model showed 81% sensitivity and 80% specificity on the test set. Glaucoma, a great threat to aged people who are over 60, could lead to blindness. To effectively detect glaucoma, researchers evaluated the capability of three deep CNNs, which are VGG19, ResNet152, and DenseNet201, on the detection task on a data set that has 3312 images.¹⁰³ The experiment showed that all networks involved for a binary classification task have an area under the curves (AUCs) of 0.9 or more, which indicates the plausibility of applying deep CNNs for glaucoma detection. Pneumothorax is a thoracic disease that can induce a life-threatening emergency.¹⁰⁴ Timely reviewing radiographs is demanding for the diagnosis but becomes challenging due to the huge volumes of images. The state-of-the-art networks, including Xception and ResNet, have been transferred for a binary task where images with large- or moderate-sized pneumothorax are considered positive. The experimental results on an internal data set with 13,292 frontal chest X-rays showed promising results that the highest specificity and sensitivity are over 0.80. However, the models showed a performance decline on the external data set. Nevertheless, the application of deep learning to pneumothorax detection seems to be promising. There are also prominent works on other disease detection by utilizing deep learning techniques.^{105–107}

4 | BREAST ABNORMALITIES ANALYSIS BEFORE DEEP LEARNING

Traditional CAD systems for breast abnormalities analysis followed strict pipelines during the process of development, which makes them better explainable. Also, there are some similarities between traditional CAD systems and deep learning-based ones, such as they share similar measurements for evaluations and some deep-learning-based works are inspired by the traditional ones. So before we formally move to applications of deep learning in breast abnormalities analysis, we briefly introduce the pipelines before deep learning and then continue with measurements for evaluation.

4.1 | Pipelines before deep learning

In the early stage of breast cancer, breast mass and microcalcifications are the two most palpable signals. Therefore, the community has been focused on developing novel detection models of breast mass and microcalcifications for a long time. Traditional CAD systems mainly consist of four components including preprocessing, segmentation, feature extraction and selection, and final analysis as can be seen in Figure 24. In traditional CAD systems, preprocessing procedure is a crucial component that affects the overall performance of CAD systems. Preprocessing focuses on improving the quality of images for better understanding or visualization by utilizing image processing techniques such as histogram equalization and image denoising.^{108,109} Inspired by the work presented in Zuiderveld,⁵ M. Sundaram proposed an improved contrast enhancement method termed HM-CLAHE to improve the contrast of mammograms. Based on wavelet processing, an algorithm that denoises and enhances mammograms was proposed.¹¹⁰ The acquisition of ROIs from raw images, or detection, is the most important step that directly affects the performance of the following components. We will have a detailed introduction of detection in a later section.

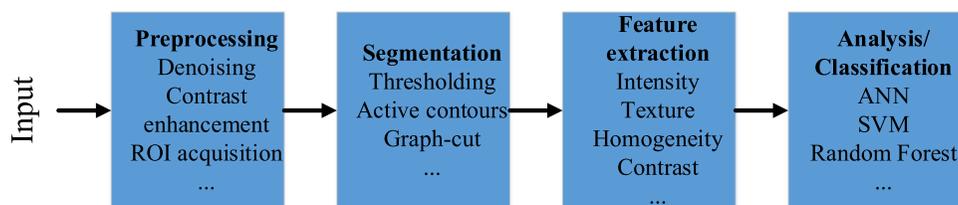


FIGURE 24 Data flows of traditional CAD systems [Color figure can be viewed at wileyonlinelibrary.com]

Preprocessing is usually required to standardize the input images for analysis. The raw images could be of low resolution and contaminated by noises. Especially, mammograms are usually taken from different views so that standardizing images in different views also requires preprocessing. Also, the images may suffer from low contrast. Contrast limit adaptive histogram enhancement (CLAHE) is a widely used image contrast enhancement method. In Yu et al.,¹¹¹ CLAHE is used to enhance the image contrast while it was repurposed as the data augmentation method. In another work,¹¹² Borges et al. explored the effect of denoising on the localization of breast microcalcification. In another research work,¹¹³ a framework for mammogram image denoising was proposed.

Segmentation, especially for mammograms that contain the small size of microcalcification, is also another critical component that plays an indispensable role in CADx systems. The popular algorithms include thresholding, active contours, and graph-cut, and so forth. An algorithm termed the self-guided region-growing method realized the task of microcalcification by using the intuitive region-growing method on mammograms being enhanced by histogram expansion.¹¹⁴ A fuzzy logic-based segmentation method that applied fuzzy c-means clustering on mammograms enhanced by morphological tophat algorithm was proposed by Bhattacharya and Das.¹¹⁵ Compared to the task of calcification segmentation, mass segmentation turns out to be less challenging given that the size of masses is generally much larger than the size of calcification deposits. An edge-based segmentation method was proposed to determine the boundary for a suspicious mass region by Zhang and others.¹¹⁶ Contrast stretching was first applied to improve the contrast of ROIs acquired from mammograms, followed by a denoising method to remove possible noises. The energy texture image is then computed correspondingly, which is severed to detect the edges in images. Finally, the boundary of mass is determined by forming enclosed edges. Active contour,^{117–119} as a powerful segmentation tool, has been widely used in mass segmentation.^{120,121} Hao and others¹²⁰ proposed a hybrid method that combines a random walk algorithm and Chan-Vese (CV) active contour. Noises in the ROIs are firstly depressed while a set of seeds for random walks is set. Based on the result of random walks, the modified CV model improves the result by updating the probability matrices, which can be converted to foreground and background regarding mass.

Feature extraction aims at extracting discriminative features for classification from segmented ROIs containing masses and calcifications. After segmentation, the features are mainly morphological and texture, which are important indicators of malignancy. The extracted features, usually in the form of vectors, are then forwarded to the classifier, which analyses the features and classifies ROIs into the most probable categories. The classical classifiers that are used to classify objects are also applicable to the classification scenarios. Support vector machines (SVM), as one of the powerful tools in supervised learning, reduces the error of learning

machines by deploying a structural risk minimization mechanism. SVM has been widely used for breast cancer detection.^{122,123} Artificial neural network (ANN), which was developed based on human perception and can model complex nonlinear functions, becomes another useful tool for the classification of breast cancer.^{124–126} However, traditional CADx systems are also suffering from problems such as lack of robustness and substantial intervention of humans though they have been widely accepted.

4.2 | Evaluation criteria

In the situation of mass and microcalcification classification, the categories are generally binary. The metrics to evaluate the performance of classifiers are *Accuracy*, *Sensitivity*, and *Specificity*, *Precision*, and *F1* score. *TP* is the number of positive ROIs that have been correctly classified as positive. Similarly, true-negative (*TN*) is the number of negative ROIs that are classified as negative. False-positive (*FP*) and false-negative (*FN*) stand for misclassified negative ROIs and positive ROIs. The *Accuracy*, also called *Recall*, can then be represented by *TP*, *TN*, *FP*, and *FN* by

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (25)$$

The *Sensitivity* measures the ability of the classifiers on detecting positive ROIs by

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (26)$$

As the counterpart of *Sensitivity*, *Specificity* is used to evaluate the performance of the classifiers on detecting negative ROI, which can be denoted as follows:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (27)$$

Precision, which shows the percentage of *TP* ROIs detected out from all ROIs that are classified as positive, can be expressed by *TP* and *FP* by

$$\text{Precision} = \frac{TP}{TP + FP} \quad (28)$$

F1 demonstrates the classification ability of the classifiers:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Accuracy}}{\text{Precision} + \text{Accuracy}} \quad (29)$$

For free-response systems, free-response receiver operating characteristic (FROC) curves are used to characterize the performance of these systems when all decision thresholds are applied.¹²⁷

To efficiently evaluate the regression models for repetitive thresholding with different thresholds, area under the ROC curve (AUC) has been widely introduced for its free from parameters and simplicity. AUC measures the entire area of the ROC curve from (0,0) to (1,1) in the two-dimensional space while the bigger area, the more desirable the classifier is. An example of AUC is given in Figure 25.

For segmentation, the commonly used evaluation criteria are pixel accuracy (PA), mean pixel accuracy (MPA), mean intersection over union ($MIoU$), and frequency weighted intersection over union ($FWIoU$).¹²⁸ Given a k -class segmentation task (background is included as one class) and the segmentation results \mathbf{C} , the corrected segmented pixel that belongs to class i ($i \in [1, k]$) can be denoted as \mathbf{C}_{ii} while the pixel in \mathbf{C} is denoted as \mathbf{C}_{ij} when it is mis-segmented into class j ($j \in [1, k]$). Then PA , as the simplest metric that only calculates the ratio between the number of correctly classified pixels and the total number of the pixels in GT , can be written by

$$PA = \frac{\sum_{i=1}^k \mathbf{C}_{ii}}{\sum_{i=1}^k \sum_{j=1}^k \mathbf{C}_{ij}} \quad (30)$$

MPA is slightly improved that correct pixels are computed within each class first and then be averaged by the total number of classes.

$$MPA = \frac{1}{k} \sum_{i=1}^k \frac{\mathbf{C}_{ii}}{\sum_{j=1}^k \mathbf{C}_{ij}} \quad (31)$$

$MIoU$, as a standard metric for segmentation, calculates the ratio between the intersection and the union of two sets, which are GT and the predicted segmentation. According to the segmentation results, the predicted pixels can be categorized into three classes: TP, FN, and FP, as shown in Figure 26. For the calculation of $MIoU$, IoU is firstly calculated within each class and then be averaged.

$$MIoU = \frac{1}{k} \sum_{i=1}^k \frac{\mathbf{C}_{ii}}{\sum_{j=1}^k \mathbf{C}_{ij} + \sum_{j=1}^k \mathbf{C}_{ji} - \mathbf{C}_{ii}} \quad (32)$$

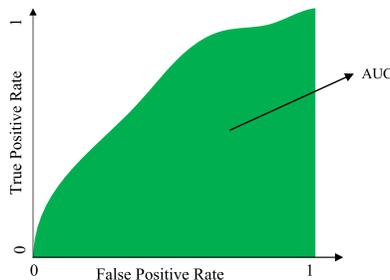


FIGURE 25 Area under the ROC curve (AUC) [Color figure can be viewed at wileyonlinelibrary.com]

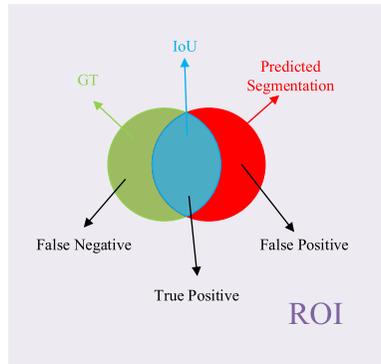


FIGURE 26 Intersection of Union (IoU) [Color figure can be viewed at wileyonlinelibrary.com]

$FWIoU$ is an improved $MIoU$ that considers the class importance concerning the appearance frequency of each class.

$$FWIoU = \frac{1}{\sum_{i=1}^k \sum_{j=1}^k C_{ij}} \sum_{i=1}^k \frac{\sum_{j=1}^k C_{ij} C_{ii}}{\sum_{j=1}^k C_{ij} + \sum_{j=1}^k C_{ji} - C_{ii}} \quad (33)$$

The dice coefficient (*Dice*), also known as the *F1 score*, is another common segmentation performance metric.

$$Dice = \frac{2 * \sum_{i=1}^k C_{ii}}{\sum_{j=1}^k C_{ji} + \sum_{i=1}^k C_{ii}} \quad (34)$$

Among all of the metrics, $MIoU$ turns out to be the most popular one thanks to its simplicity and typicalness.

5 | ABNORMALITY DETECTION BY CNNs

In this section, we will introduce the application of deep learning in the detection of abnormalities. As breast mass and microcalcification are the two main breast cancer symptoms, we mainly focused on works that detect these two kinds of abnormalities. Multiview detection is commonly used in the clinical diagnosis of breast abnormality. Usually, if a suspicious region in left *CC* (*LCC*) view is found, the corresponding region in the left *MLO* (*LMLO*) view and right *CC* (*RCC*) view will be checked. This region's likelihood to be abnormal increases when the region in *LMLO* is also suspicious while turns out to be normal in *RCC*.¹²⁹ Generally, one view of mammograms is sufficient to make reliable diagnosis decisions. However, it was shown in Bekker and others¹³⁰ that multiview could be more beneficial to the detection of breast abnormalities.

5.1 | Mass detection

The formation of mass is one of the most common symptoms of breast cancer. Extensive research on mass detection has been done, and multiple studies have shown high performance.^{131–135} The detection of mass can generally divide into two steps: localization of mass and classification. With the help of deep CNNs, these two steps can be realized easier but more efficiently. Masses in mammograms can be grossly classified into being benign and being malignant. Usually, the mass region in a mammogram is quite different from surrounding tissues, as can be seen from the examples shown in Figure 27. However, mass detection could be hampered by the high-density breast, which makes mass less palpable. As automatic mass detection remains a challenge, some research focused on developing semiautomated or manual intervention involved detection systems that classify manually detected mass into benign and malignant categories.¹³⁶ Using the high-level features extracted by deep CNNs, those CADx systems outperformed the CADx systems built on traditional machine learning methods.

Arevalo and others¹³⁶ developed a CNN-based mass classification system. In this study, a data set named BCDR-F03, which was no longer maintained was, analyzed. In the data set, 344 patients contributed to 736 film images in total, of which 310 of them were malignant mass lesions while the rest were benign mass lesions. The data set provides manually segmented lesions; therefore, ROIs were extracted according to the given bounding boxes and then were rescaled to a fixed size. Data augmentation was applied to flip and rotate the ROIs to generate seven new artificial ROIs from one ROI. The ROIs were then enhanced by global and local contrast normalization before they were fed to a deep CNN model for classification. The CNN models with different depths were deployed to produce deep learning representations for a following linear SVM, the classifier of the proposed model. To measure the efficacy of deep learning representation, manual descriptors such as histogram of oriented gradients (HOG) and Histogram of gradient divergence (HGD) were compared. Experimental results showed that CNN models with three convolutional layers performed best, giving the highest accuracy at 0.860.

A hybrid method for the detection of breast mass, which was similar to the work by Dhungel and others,¹³⁷ was presented in Dhungel and others.¹³⁸ The detection of ROI consists of three stages. In the first stage, the deep belief network (m-DBN) and a Gaussian mixture model (GMM) were introduced to generate mass candidates. The following two stages are aimed at refining the candidates from coarse to fine. In the first stage, the m-DBN model was applied to input images of coarse resolution by utilizing a grid-based method. Each grid point was classified into positive and negative regarding classification results on the fixed size patch that takes the grid point as the center point. All positive grid points were then further classified in a finer resolution stage. The coarse-to-fine refinement was repeated three times while GMM works on the finest resolution image for pixel-wise classification. When training the m-DBN model, positive patches were defined when central points fell into the annotated mass. Negative patches, where the central points did not belong to the annotated mass, were randomly sampled in the first stage. Final estimated positive labels and negative labels were obtained through thresholding posterior probabilities produced by the GMM model. The pixel-wise classification was combined through union operations that connected components analysis determines the mass candidate. In the second stage, two-stage R-CNNs were stacked to reduce the false-positive rate. The training samples for R-CNNs were considered positive when the overlap between the estimated bounding box and the annotated box was greater than 0.2. Otherwise, they were considered negative. Instead of using the softmax layer for classification, SVM, which

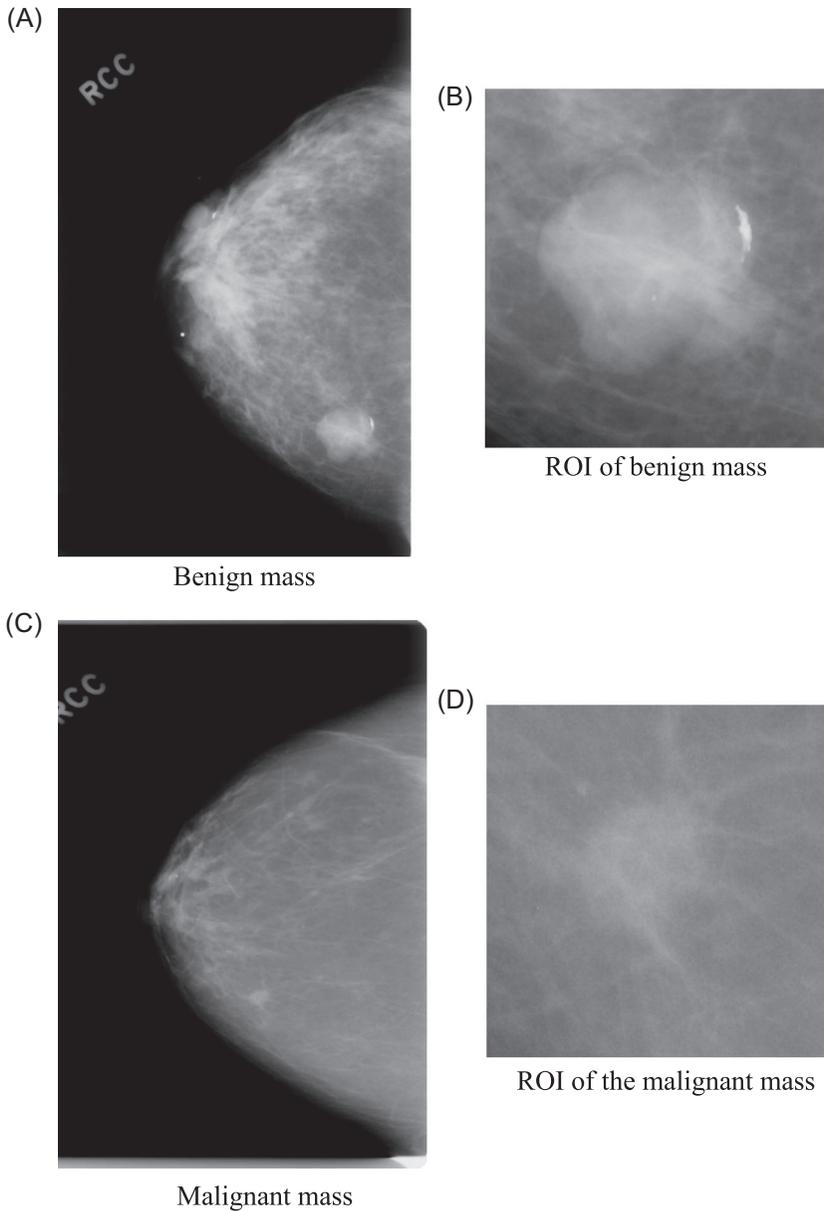


FIGURE 27 Mammograms of mass and corresponding ROIs

takes the features from the last fully connected layer as the input, was reported to produce better classification results. The candidates who survived the first cascade of the R-CNN were then passed to the second cascade of R-CNN for further false-positive reduction. As reported, the false-positive rate was still high even after R-CNNs stage; therefore, a third subsequent stage that used random forest (RF) as a final classifier to remove false-positive candidates was attached. A huge number of hand-crafted features, such as area, perimeter-to-area ratio, circularity, and rectangularity, were extracted from the second stage's survived candidates. Finally, 781 hand-crafted features were used as the input of RF. Besides these three-staged

detection procedures, one more hypothesis refinement step was followed to adjust the bounding boxes. The final mass detection method gives a TPR of 0.95 ± 0.02 at a FPI = 5 for the testing set while TPR of 0.95 ± 0.02 at a FPI = 3 when an Intersection over Union (IoU) ≥ 0.2 . The training set was comprised of 60% of images from the INbreast data set, while 20% of images were partitioned into the testing set. The false-positive regions have to be manually excluded for the following modules such as segmentation.

More works on developing automated detection systems of breast mass have been done.¹³¹ In Agarwal and others,¹³¹ a sliding window detection algorithm is the most classical detection method that can be implemented simply by scanning a window of a specific size on images where objects to be detected. By incorporating the sliding window technique, a patched-based CNN model aimed at automatic mass detection in full-field digital mammograms (FFDM) was proposed. Patches with fixed 224×224 pixels are obtained from the original mammograms when the sliding window slides over the images at the stride of 56×56 pixels. Whether a patch to be labeled as negative or positive depends on whether the center pixels in the patches were located in the regions of masses or not. Three state-of-the-art networks including VGG16,⁶⁶ ResNet50,¹³⁹ and InceptionV3¹³ were utilized to produce the mass probability of the patches. By aggregating each patch's mass possibility, a mass probability map (MAM) can be formed for the determination of bounding boxes. The MAM was then thresholded by a predefined value to generate the final bounding box. The data sets involved in the work are CBIS-DDSM and INbreast while CBIS-DDSM was used for pretrained and INbreast was used for validation. To validate the performance of networks when they were pretrained on data sets from a different domain, two scenarios have been explored: one is direct transferring networks pretrained on ImageNet to the targeted INbreast, however, in another scenario, CBIS-DDSM was used to finetune networks trained on ImageNet, and the finetuned networks were then trained on INbreast. When evaluating the developed models on INbreast using five cross-validations, a mass was considered to be detected when there was at least 0.2 overlapping between the bounding box and the annotated ground truth. Finally, it was reported that the best model, based on InceptionV3, achieved a TPR of 0.98 ± 0.02 at 1.67 false-positives per image (FPI).

A simultaneous detection and classification system was proposed by Al-antari and others.¹⁴⁰ In their framework, YOLO, which was known for its fast speed and high accuracy, was utilized to search for the possible mass region from a full mammogram. An augmented mass data set from INbreast data set was analyzed in the study. There are in total of 107 cases in both reviews of mammograms in the original INbreast data set. However, few cases have more than one mass in mammograms leading to 112 masses. Before the detection of mass, all 112 images are augmented by rotating them with different angles eight times, where the difference between two adjoining angles is 45° . As a consequence, the augmented data set was comprised of 896 mass images in total. Four cross-validation method was introduced to evaluate the proposed detection method, where 75%, 6.25%, and 18.75% of the 896 mass images were partitioned into the training set, validation set, and testing set respectively. The reported results on the testing set showed an averaged detection accuracy at 98.96% in a fast way that each image was processed within only 3 s.

In another research, Kooi and others¹³² did comparative research to evaluate the performance of a CNN-based lesion detection system and the traditional CAD system. A two-stage procedure for the detection of the lesions has been proposed where candidates were generated in the first stage and then be classified in the second stage. An algorithm proposed in Karssemeijer and Brake¹⁴¹ was introduced to produce five-pixel features for each pixel in the mammogram images. As hypothesized by the authors, every pixel could be the center of a

lesion. Therefore, positive samples for training candidate detection classifiers are pixels inside each annotated malignant lesion, which are taken as the center pixels of the patches. To form an unbiased data set for training, 1 in 300 pixels is randomly sampled from normal images. Furthermore, the training set, validation set, and test set were split on a patient level to prohibit any bias. All positive samples are augmented with scale and translation transformations. After data augmentation, the numbers of positive patches and negative patches were respectively 334,752 and 853,800 in the training set. Afterward, a random forest classifier was trained to generate the likelihood image, where local optima were taken as seed points for the extraction of the potential masses. The extracted candidates are then analyzed by CNN and reference system, where the reference system relies on manual features, such as contrast features and texture features, extracted from the segmented patches. The CNN architecture, which consists of five convolutional layers and two fully connected layers, is relatively simple compared to the advanced networks. Though the AUC of the reference system reached 0.91, the AUC of CNN was even higher at 0.929. Therefore, the conclusion that CNN outperformed the state-of-the-art system has been supported by the experiment. To specify the CNN with optimal depth, Arevalo et al. tested the performance of CNNs with various depths. To validate the best model, they compared the model with a traditional CAD system method that uses 17 hand-crafted features and methods that use HOG and HGD descriptors.^{136,142} The conclusion was that the performance improved when learned features and hand-crafted features were combined.

While X-ray images, digital breast tomosynthesis (DBT) has emerged as a novel tomographic technique to eliminate limitations in conventional breast screening methods. By performing a series of low-dose radiographic exposures, breast tissue is imaged in a 3D manner, which allows characterized findings including normal structures to be imaged.¹⁴³ The summation of overlapping breast tissue, which could be misinterpreted as breast cancer, can be reduced by this technique. Given its exclusive advantages in localizing lesions, DBT has been extensively introduced in the detection method based on deep learning.^{144,145}

In another study, Samala and others¹⁴⁴ developed an automatic DBT-based mass detection system. The authors combined 2282 digital mammograms and 324 DBT volumes to form a merged data set. An experienced breast radiologist manually marked the mass of interest for later-on reference. The developed system consists of two stages within which the first stage prescreens the false-positive ROIs while the second stage recognizes TP regions from all regions that survived from the first stage. In the prescreening stage, first-order features are introduced to generate coarse candidate regions, while second-order features are extracted from the candidate regions to reduce false-positive ROIs. The data augmentation method was used to generate 45,072 mammographic ROIs and 37,450 DBT ROIs. A deep CNN with only four convolutional layers and three fully connected layers is used as the classifier. The transfer learning technique was applied for a better performance of the classifier on the DBT data set. Initially, the CNN was trained on the mammographic ROIs data set. The last convolutional layer and all fully connected layers are then fine-tuned on the DBT ROIs. Experimental results showed that the trained CNN by mammographic ROIs data set achieved an AUC of 0.99. However, the trained CNN only showed 0.81 AUC on the training set of DBT data sets without further training. After fine-tuning the trained CNN on the training set of the DBT data set, AUC rise to 0.90, which showed the effectiveness of transfer learning. On breast level detection, the CNN-based system showed priority to another feature-based system on sensitivity at 91%–83% while the false-positive rate is only 1 per image.

In another work, Yousefi and others¹⁴⁵ designed three breast mass detection models. The detection system takes 2D slices from 3D DBT as the input. From 87 DBT volumes, a total of

5040 2D slices were obtained. For all three modules, preprocessing procedures, including image denoising and pectoral muscle removal, are designed before the detection of mass. For the hand-crafted feature-based module, ROIs are firstly selected for the following hand-crafted feature extraction component. The extracted features are then reduced in dimensionality by dimension reduction component. Finally, multiple-instance learning was used as the final classifier for classification. In the second module, the deep cardinality-restricted Boltzmann machine (CaRBM)¹⁴⁶ was used as the feature extractor. In the third module, deep CNN was used as the feature extractor instead. Experimental results showed that the model based on deep CNN has the best performance that it achieved the accuracy at 86.81%, sensitivity at 86.6%, specificity 87.5%, and 0.87 AUC, respectively.

Faster-RCNN has been introduced in the work¹⁴⁷ by Fan et al. for breast mass detection on a DBT data set which includes 89 patients with 105 masses. Before detection, CC and MLO views are fused for the identification of nipples in the algorithm proposed in Zhou et al.¹⁴⁸ To reduce the unnecessary computational cost when detecting candidate regions by the RCNN-based system, background and skin were excluded by a dynamic multiple thresholding algorithm. The backbone of faster-RCNN in the work was AlexNet, the champion of image classification tasks in 2012. The training of faster-RCNN was comprised of four steps. The first step is aimed at RPN training in the detection network to generate proposal bounding boxes as the candidate masses. In Step 2, another separated classifier was trained using the bounding boxes generated in the first step. In Step 3, the weights in particular layers of RPN trained in the first step were updated while the convolutional layers, or the shared layers, were initialized by the trained classifier in Step 2. In Step 4, the proposal bounding boxes acquired in Step 3 were used to train the classifier network to update unique layers. Consecutive slices in the form of 2D were fed to the trained detection network. The 3D detection results were formed based on the detection results on each 2D slice. Bounding boxes in different slices were taken as the same mass when the overlap-ratio was greater than 0.5. The bounding box in a slice was ignored when there were no overlapping bounding boxes found in the neighboring slices. By merging all of the consecutive founding boxes, the final mass likelihood score was defined. To make a comparison with the model developed by Samala and others,¹⁴⁴ the authors re-implemented the CAD system by replacing the four convolutional layers with a deeper network that has 22 layers. The reimplemented system was referred to as the DCNN-based system. Also, no transferring from mammography data was introduced as it was reported that direct training on DBT ROI images would result in a better performance of DCNN models.¹⁴⁹ In terms of AUC, the developed RCNN-based system achieved 0.96 while the DCNN-based system was 0.92. The sensitivity of the RCNN-based system was as same as that of the DCNN-based one on the lesion-based level. However, the FPs per volume of the RCNN-based system was 1.54 while the DCNN-based one had the FPs per volume at 2.81. For the breast-based level, 0.76 FP was reported for the developed system while it was 2.25 for the DCNN-based system. The summary of the stated systems is given in Table 7.

5.2 | Microcalcification detection

In mammograms, microcalcifications show up in the form of opacities that have different appearances to other breast tissues. The main features of microcalcification are varied shape, size, distributions, and morphology. The challenges for accurate classification of microcalcifications come from two aspects. One is that the small sizes of microcalcifications are easy

TABLE 7 The state-of-the-art mass detection systems

Authors	Method	Automated	Data set for training	Data set for evaluation	Validation method	Number of images for evaluation	Performance
Agarwalet al. ¹³¹	Sliding window + DCNNs	Yes	CBIS-DDSM	INbreast	5 Cross-validation	~23	TPR of 0.98±0.02 at 1.67 FPI
Kooi et al. ¹³²							
Arevalo et al. ¹³⁶	DCNNs	No	BCDR-03	BCDR-03	Hold out Validation	~294	0.860
Dhungel et al. ¹³⁸	Deep belief network + Gaussian mixture model	Yes	INbreast	INbreast	5 Cross-validation	82	TPR of 0.95 ± 0.02 at 5 FPI
Al-antari et al. ¹⁴⁰	YOLO	Yes	INbreast	INbreast	4 Cross-validation	168	0.9896
Samala et al. ¹⁴⁴	Gradient filed convergence map + DCNN	Yes	Merged data set (2282 digital mammogram + 324 DBT volumes)	Private data set	Hold out	94 DBT views	0.91 sensitivity at 1 FPI
Yousefi et al. ¹⁴⁵	Hand-crafted ROI selection + DCNN	Yes	Private data set (87 DBT volumes)	Private data set	Hold out	-	Accuracy at 0.8681, Sensitivity at 0.8750, AUC at 0.8700
Ming et al. ¹⁴⁷	Faster-RCNN	Yes	-	-	-	-	0.90 sensitivity at 0.76 FPI

to be overlooked by radiologists or CAD systems, especially when surrounding tissues cover small calcifications. For CAD systems, the number of microcalcification pixels, called positive class, in mammogram images is very small compared to the number of non-microcalcification, called negative class. The imbalanced rate between positive and negative classes impedes most classification systems.¹⁵⁰ The other is that the dense background of the breast makes the microcalcifications less recognizable. Nevertheless, microcalcifications are of high spatial resolution, which can be useful information for methods that can draw on high spatial frequencies such as wavelet transform.^{151,152} With the advancement of deep learning, more efforts have been spent on designing higher-efficiency microcalcification detection systems. Mammograms with microcalcifications are shown in Figure 28.

Mordang and others¹⁵⁰ developed an automatic microcalcification detection system in multivendor mammography. In this study, a CNN-based system was used to analyze 1606 mammograms, amongst which 490 are collected with Holgic digital mammography systems, 1044 acquired with GE Senographe systems, and 72 with Siemens Mammomat Inspiration systems. Medio-lateral oblique and cranial-caudal views of all left and right breasts are available. To detect microcalcifications, each pixel in a mammogram needs to be classified as microcalcification, the positive class, or non-microcalcification, the negative class. Patches around the positive pixels are taken as positive patches for training while negative patches are obtained when the center pixel is not a microcalcification pixel. By doing so, there are in total 11,711 positive patches and 27,946,799 negative patches in the training set while the positive patches and negative patches in the testing set are 5298 and 18,320,976, respectively. Each patch is of size 13 by 13 pixels while the individual microcalcification is centered. A hard negative mining strategy was proposed to address the class imbalance between the classes given that non-microcalcification outnumbered microcalcification significantly. In the proposed hard negative mining strategy, the CNN is firstly trained on a small data set. After training, the CNN classifies the whole data set to remove easy samples. A second CNN is trained on a larger data set that includes the hard negative samples to distinguish hard negative samples. This study's CNN architecture is two repetitions of two consecutive convolution layers with a Max-pooling layer, and three layers of fully connected layers. The same data was used to train a cascade of classifiers to compare with the proposed model. The cascade classifier consists of nodes that each node is a classifier for the binary classification task. In each node, patches are filtered out when the scores are below a predetermined threshold. Therefore, patches that survive in the last node receive the final score of the last patches. The ROC analysis was applied for the comparison of the two systems. The result showed that the proposed CNN method outperformed the cascaded classifiers by about 6% mean sensitivities.

In Wang and Yang,¹⁵¹ Wang et al. proposed a context-sensitive deep neural network that takes both local features of a microcalcification as well as surrounding tissue background into consideration. The authors hypothesized that the localized nature of individual microcalcifications is desirable for the classifier to obtain the image features within a small region around the microcalcification. However, the surrounding background of microcalcification is believed to be beneficial for the suppression of potential FPs. The authors claimed that a direct application of an image window containing a large area surround the microcalcification would be problematic given that individual microcalcifications could remain very close to each other. Therefore, the authors proposed using two image windows as the input to the classifier simultaneously, within which one window is for microcalcification feature characterization while the other is for surrounding image background property description. Two networks, global subnetwork, and local subnetwork are deployed for feature extraction of surrounding

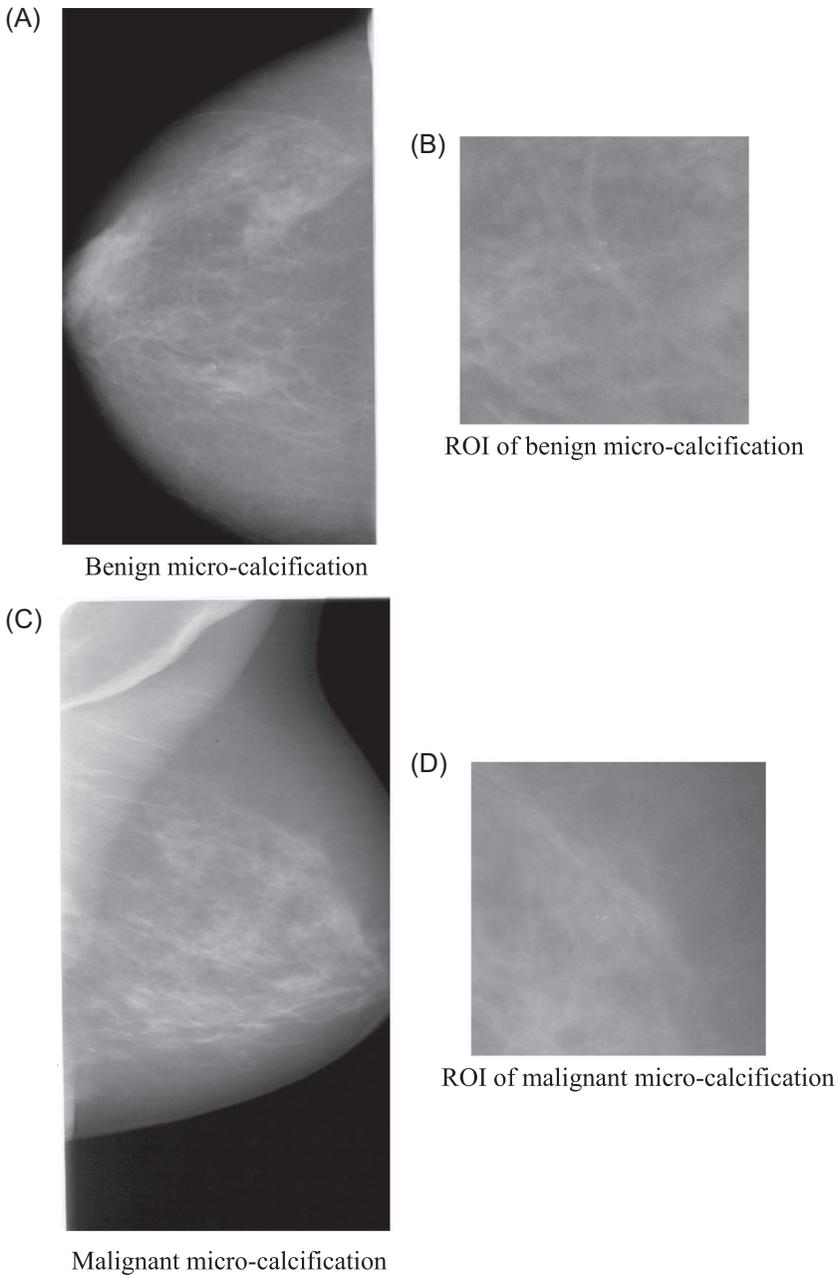


FIGURE 28 Mammograms of microcalcifications and corresponding ROIs

background image and microcalcification image. To optimize the global subnetwork structure, the authors varied the number of convolutional layers from three layers to eight layers gradually. A batch normalization layer and a nonlinearity layer are concatenated for each convolutional layer to refine the feature maps. The number of convolutional layers in the local subnetwork varies from one to four when optimizing the architecture. As a result, 24 variants of the combined networks are evaluated in the work while the best-performing variant is chosen

for testing. A merged data set, which includes 521 screen-film mammograms (SFM) images from 297 cases and 188 full-filed digital mammogram images from 95 cases, is used for the research.

By combining image context with deep CNNs, the FP rate was proved to be reduced by the experiment on the set of 292 mammograms. Among all of the images, 167 cases (300 images) are randomly portioned into the training set while 67 cases (117 images) and 158 cases (292 images) are used for validation and testing, respectively. To mitigate the inhomogeneity in the tissue background, a background subtraction step was applied. In the background subtraction method, each location's background is estimated by calculating the average intensity of a circular region that takes 7 pixels as the diameter while the location is taken as the center. The resultant images are then normalized to yield mean zero and unit standard deviation images. Positive samples are cropped out from mammogram images at each marked microcalcification location by two windows while 20 negative samples are randomly cropped at the non-microcalcification background. The proposed method is then validated on patch-level for individual microcalcifications and image-level for microcalcifications clusters detection in mammograms. In the experiment of the classification of microcalcification and non-microcalcification samples, the best architecture of the context-sensitive DNN classifier is determined when the numbers of convolutional layers in the global subnetwork and the local subnetwork are 7 and 4, respectively. For comparison, three models including a unified SVM detector, a CNN cluster detector, and a local DNN classifier, are introduced. The experimental results of individual microcalcification detection showed that the FROC of the proposed context-sensitive DNN is higher than the compared methods. For the context-sensitive DNN, a sensitivity of 79.7% is achieved while the FP rate of 1.03 FPs/cm². In the second experiment on detecting microcalcification clusters in mammograms, the same conclusion is drawn that the model with the best performance proposed method achieved TP as 87.40% while the FP rate of 0.5 clusters/image, which is better compared to 84.6% of the Unified SVM classifier.

Rehman and others¹⁵³ developed a so-called intelligent system for the detection of microcalcification by integrating DNN. The developed system was designed for two objectives. The first one is to classify breast cancer into being benign and being malignant based on mammogram images. The second objective of the proposed system was to automatically select ROIs, which exempts the system's reliance on manual selection of ROIs. The developed system can be divided into three major phases; within each phase are different subtasks. In the first stage, noise in the images was reduced, and features were selected for the second stage, which built the classification model based on three common layers including the convolutional layer, pooling layer, and fully connected layer for the binary classification task. In the third stage, the performance of the model and the result were visualized for better understanding. In the work, 1273 mammogram images were analyzed, and the classification accuracy reached 95.6%. However, details about the feature selection and the architecture of CNNs were missing in the work.

Wang et al. evaluated deep learning-based models on large data sets for the detection of microcalcification in Wang et al.¹⁵⁴ The proposed method was trained on the training set contains 677 benign images and 323 malignant images, which led to 1000 lesions in total, where each image contributes a lesion. For evaluation, an independent test set including 97 benign and 107 malignant images was used. Three scenarios, including isolated microcalcifications, isolated masses, and a combination of microcalcifications and masses, are considered. A deep learning-based semi-automated segmentation method was developed for the three different scenarios to yield 15, 26, and 41 features, respectively. When microcalcification was considered

alone, the proposed deep learning-based discrimination classifier showed an accuracy of 87.3%, which was slightly higher than the accuracy of 85.8% given by SVM. However, both methods showed poor performance on the analysis of masses that the accuracy of the two methods was only 61.3%. Nevertheless, the accuracy of the proposed model and SVM improved to 89.7% and 85.8%, respectively, when the analysis was carried out on the combination of microcalcifications and masses. Therefore, the experimental results supported the conclusion that deep learning methods are more superior for the detection of microcalcifications.

Akselrod-Ballin and others¹⁵⁵ proposed a mass and calcifications detection system based on Faster-RCNN. The detection framework comprises three stages, which aim at initial detection, classification, and classification refinement. The input mammogram images are firstly divided into an overlapping grid for training the deep CNN. In the first stage, a deep FCNN was trained to detect windows likely to have objects of interest included. In this stage, sliding windows at the stride of 32 pixels slide over the input images to generate grids. For each grid location, nine seedboxes in different sizes and aspect ratios are being classified. According to the IoU between ground truth object boxes, these boxes are labeled as object anchors when IoU is greater than 0.5. Finally, only the boxes with top prediction scores are kept. In the second stage, a faster R-CNN detection network is trained to classify potential object windows which survived from the first stage into the mass, benign calcifications, and malignant calcifications. Here, calcifications include microcalcifications and macrocalcification, which refers to bigger bits of calcium. VGGNet is used as the backbone of two networks. The number of windows in the training phase from the RPN is 2000 while it reduces to 500 during the testing phase. In the third stage, the TP and FP candidate boxes in the training set are selected to train VGG-16, the final classifier to recognize the TP and the hard negatives. The experiment was carried out on the INbreast data set and an Internal data set with 3500 images, amongst which 750, 360, and 2400 images are with mass, malignant calcifications, and benign calcification, respectively. On INbreast, a TPR at 0.4 with 1 FPI was achieved for all of the calcifications. When small calcifications were excluded, the TPR rose to 0.85, so did FPI to 1.5. A comparison with the other two methods indicated that the proposed method had shown better performance. A summary of the microcalcification detection systems is given in Table 8.

5.3 | Other detection scenarios in breast cancer

While mass and microcalcifications are two dominating symptoms for the detection of breast cancer, there are also other valuable works aiming at detecting breast cancer depending on other minor symptoms.

Architectural distortion (AD) is the third common yet the most easily to be missed abnormality that follows mass and microcalcification. However, it is reported that detection of AD will be helpful for breast cancer detection.¹⁵⁶ To this end, Costa et al. transferred the VGG-16 network for detection on clinical mammogram data sets with 280 images.¹⁵⁷ Positive patches for training are cropped out from original images where the center of AD is shown in the patches while negative patches are randomly sampled within the breast area. By doing so, a patch data set containing 44,224 ROIs with the same number for positive and negative patches is formed. Data augmentation including rotation, horizontal and vertical flip is carried out leading to 123,844 ROIs in total. To specify the best training approaches, several VGG-16 instances are trained in different depths. The best AUC achieved was 0.89 by a network that was fine-tuned at a certain level. In another work, Oyelade et al. customized a deep CNN for a

TABLE 8 The state-of-the-art microcalcification detection systems

Authors	Method	Purpose	Data set for training	Data set for evaluation	Validation method	Number of images (or patches) for evaluation	Performance
Mordang et al. ¹⁵⁰	DCNN	Microcalcification detection	Hologic, GE, and Siemens data	Hologic, GE, and Siemens data	Hold-out	18,326,274 patches	Mean sensitivity at 0.6914 ± 0.0041
Wang et al. ¹⁵¹	Context-sensitive DCNN	Microcalcification detection	Private merged data set	Private merged data set	Hold-out	292 images	sensitivity of 79.7% with FP rate of 1.03 FPs/cm ²
Rehman et al. ¹⁵³	DCNN	Microcalcification severity prediction	Private merged data set	Private merged data set	-	1273	accuracy of 95.6%
Wang et al. ¹⁵⁴	DCNN	Microcalcification and mass severity prediction	Private data set	Private data set	Hold-out	204	accuracy of 87.3%
Akselrod et al. ¹⁵⁵	Faster-RCNN	Mass and calcifications detection	INbreast + private data set	INbreast + private data set	Hold-out	310	TPR at 0.85 FPI at 1.5

three-class classification task that classifies obtained patches from MIAS, DDSM, INbreast into Normal, Benign, and Malignant categories.¹⁵⁸ Data augmentation is applied as well to enhance the performance of the developed network. Finally, the model was reported with an accuracy of 93.75%. A detailed survey about the detection of AD can be found in Oyelade and Ezugwu.¹⁵⁶

Axillary lymph node (ALN) is another common abnormality in scenarios of breast cancer. The status of ALN is quite helpful for prognosis and therapy decisions. To avoid unnecessary surgeries for removal of negative ALN, Zheng et. al explored the performance of different deep learning models on distinguishing N0 status nodes and $N_+(\geq 1)$ status nodes,¹⁵⁹ where N0 means the disease-free axilla while $N_+(\geq 1)$ indicates any axillary metastasis. The involved deep learning models are ResNet50, ResNet101, InceptionV3, and VGG19 while ResNet50 performed best among all of them. To further improve the performance of the proposed model, additional clinical information was added to the second last fully connected layer of the model on basis of ResNet50, which was termed as ResNet50+C. By doing so, the best performance given by ResNet50+C achieved an accuracy of 74.6 on an independent test cohort containing 118 images.

Also, mass and microcalcification detection-oriented systems require a huge size of annotations, which turn out to be expensive and time-consuming. Therefore, breast cancer detection systems in weak or unsupervised manners have received wide exploration by researchers in the field. Lotter et al.,¹⁶⁰ proposed an annotation-efficient deep learning-based method for the classification of benign and malignant mammograms. The developed algorithm is trained in a three-stage manner, where strong supervision that requires patch-level annotations is introduced in the first two stages of training. In the substage of the third stage, called stage 3A, the proposed algorithm outputs scores and bounding boxes for 2D mammograms after being trained in a weakly supervised manner. In stage 3B, however, the algorithm is extended for 3D mammograms breast cancer detection that it outputs score, bounding boxes, and slice numbers in 3D volumes. Finally, the proposed method was reported to beat five breast-imaging specialists by an average increase in sensitivity of 14%. Another image-level breast cancer detection system can be found in Eskreis-Winkler et al.,¹⁶¹ where Eskreis et al. deployed CNN to distinguish cancer-containing slices from MRI images. On an independent test set containing 706 images, the developed system was reported an accuracy of 92.8% on the binary classification task while the sensitivity and specificity reached 89.5% and 94.3%, respectively. Other detection and classification works that are related to breast cancer are listed in Table 9.

6 | ABNORMALITY SEGMENTATION BY CNNs

In this section, our main focus will be CNN-based works for breast mass and microcalcifications to maintain the consistency of our survey. However, other segmentation scenarios that may be beneficial to the detection of breast cancer will be amended as well. Segmentation of breast mass and microcalcifications is also another important branch for the design of CAD systems. Effective segmentation of masses and microcalcifications can lead to overall accuracy improvement while reducing the FP and FN. However, segmentation remains the big challenge due to the low contrast, ambiguous boundaries, and irregularities in size, location, and shape.^{165,166} In breast abnormality segmentation, numerous works deployed classical methods, but the exploration of utilizing deep learning models is quite limited.¹⁶⁷ Compared to traditional segmentation methods, CNN-based segmentation methods, mainly FCNNs, turn out to

TABLE 9 Other application scenarios of deep learning in breast cancer

Authors	Method	Purpose	Data set	Number of images (or patches) for evaluation	Performance
Costa et al. ¹⁵⁷	DCNN (VGG-16)	Architectural distortion	Private data set	28	AUC of 0.89
Oyelade et al. ¹⁵⁸	DCNN (Custom)	Architectural distortion	Merged data set (MIAS, INbreast, DDSM)	–	Accuracy at 93.75%
Zheng et al. ¹⁵⁹	DCNN (ResNet50, ResNet101, InceptionV3, VGG19)	Axillary lymph node classification	Private data set	118	Accuracy of 74.6%
Lotter et al. ¹⁶⁰	DCNN	Beingness and malignancy classification	Merged data set (DDSM, OMI-DB)	285	Outperform breast-imaging experts by an average 14% sensitivity.
Eskreis et al. ¹⁶¹	DCNN	Healthy and cancerous mammogram classification	Private data set	706	Accuracy of 92.8% (648/706)
Albayrak et al. ¹⁶²	DCNN	Mitosis detection	MITOSATYPIA-14	–	Precision of 1.00, Recall of 0.94, F-measure 0.97
Chen et al. ¹⁶³	FCNN + DCNN	Mitosis detection	MITOSATYPIA-14	496	Precision of 0.46, Recall of 0.51, F1 score of 0.48.
Swiderski et al. ¹⁶⁴	DCNN	Lesion recognition	DDSM	1122	An AUC of 0.919 for abnormal vs. normal; An AUC of 0.909 for malignant vs. nonmalignant

be more powerful.⁷⁰ In this section, we will mainly review segmentation works on mass and microcalcification by deep learning models.

6.1 | Mass segmentation by FCNNs

Al-antari and others¹⁶⁸ proposed a full resolution convolutional network (FrCN) to implement segmentation. In the proposed system, masses were firstly detected by the proposed mass detection module and were fed to the proposed FrCN for segmentation. To enhance the contrast of the detected masses, the authors performed the contrast-limited adaptive histogram equalization (CLAHE) method, a widely-used classical image contrast enhancement algorithm. Compared to FCN, SegNet, and U-net that used multiple max-pooling and subsampling layers, the proposed FrCN removed the max-pooling and subsampling layers to maintain the full spatial resolution of the original input. When max-pooling and subsampling layers in encoder networks were introduced, the spatial resolution of the feature maps was reduced. Decoder networks can be used to recover the resolution but will introduce more parameters into the networks. As a result, segmentation models based on deep learning suffer from the loss of details. The key modification of max-pooling and subsampling removal enables accurate pixel-to-pixel mass segmentation by keeping as many details as possible. The FrCN proposed was based on VGG-16 with the last three fully connected layers being replaced with three full convolutional layers, while max-pooling and subsampling layers were removed. As reported by the authors, an overall accuracy of 92.97% on detection of mass was produced by the segmentation module while the Dice (F1-score) of 92.69% and Jaccard similarity coefficient metrics of 86.37% are given, respectively.

The level set method,^{169,170} a traditional image segmentation method, was applied to refine the segmentation result given by deep structured learning methods in Dhungel et al.¹³⁸ Dhungel and others¹³⁸ designed a deep structured network to segment a low-resolution input patch. The network consists of two convolutional layers that produce 6 feature maps and 12 feature maps, respectively. For each convolutional layer, a subsampling layer is followed to reduce the size of feature maps. The output of the input patch, which can be reshaped to maintain the same size as the input patch, is the possibility map produced by the proposed CNN model. A further refinement step was carried out by integrating the Chan-Vese active contour model¹⁷¹ on the segmented image. The proposed segmentation method was evaluated on a subset from INbreast data set. 41 benign masses and 75 malignant masses are included in the subset, while 60% of them are used for training while the rest are divided into the validation set and the test set, respectively. The experimental results showed that the proposed method has a Dice index on the training data at 0.85 ± 0.01 , similar to the result at 0.85 ± 0.02 on the test set.

As one of the most representative deep learning segmentation models, U-net has been widely used for breast mass segmentation.^{5,73-174} Li and others¹⁷² proposed to use an improved version of U-Net codenamed Conditional Residual U-Net (CRU-Net) that integrated CRF and residual learning for segmentation of mass. The residual learning serves to improve the feature extraction while CRF helps to improve the pixel-level segmentation results. A total of 290 ROIs from DDSM-BCRP and INbreast are used in the research. One-hundred and seventy-four of the ROIs are from DDSM-BCRP while the rest are from INbreast. All of the ROIs in both data sets are evenly partitioned into the training set and the testing set. Before segmentation, each ROI is resized to 40 by 40 pixels by bicubic interpolation. The best

performance on ROIs from DDSM-BCRP is 91.43 ± 0.02 while it is 93.66 ± 0.10 on the ROIs from INbreast. Another U-Net-based mass segmentation that combines densely connected networks with attention gates (AG) is proposed by Li and others.¹⁷³ In the developed method, the network is comprised of an encoder and a decoder. Dense blocks, which were borrowed from DenseNet, formed the encoder. The decoder of the proposed network is the decoder of U-Net integrated with AGs. The proposed method is tested on the public data set DDSM while the evaluation metrics are F1-score, mean IoU, sensitivity, specificity and overall accuracy. As a result, the proposed method showed an overall accuracy at $78.38 \pm 0.04\%$, which is the highest one compare to that of the other state-of-the-art methods.

Annotation for breast cancer is a process that not only requires specialty from experienced radiologists but also consumes plenty of time. In situations where only limited annotations are available, shape priors can improve the segmentation results. Maicas and others¹⁷⁵ proposed a novel fully automated breast mass segmentation on an MRI data set. In the proposed method, a deep learning model was introduced to generate shape priors for the following segmentation module based on globally optimal inference in a continuous space (GOCS). Mass detection was firstly derived from the method proposed by Dhungel et al.¹⁷⁶ The difference is that the number of cascade stages is reduced while no hand-crafted features are used. Finally, the detection method gives a TPR of 0.85 at 3.66 false-positive regions per patient. The deep learning model has three convolutional layers within which 10, 20, and 20 filters are used for gross segmentation, which is later used as the shape prior for GOCS. The data set comes from 117 patients, where 58 patients who contributed 72 lesions (23 benign and 49 malignant) are partitioned into the training set. Sixty-nine lesions (23 benign and 46 malignant) from the remaining 59 patients are used to evaluate the proposed segmentation method. The mean and median Dice coefficient is used to assess the accuracy of segmentation on both the training and testing sets. The proposed segmentation algorithm then showed the mean Dice on the testing set at 0.77 ± 0.14 while the median Dice is 0.82. More mass segmentation methods have been listed in Table 10.

6.2 | Microcalcification segmentation by FCNNs

Individual microcalcifications are challenging to recognize even for experienced radiologists due to the heterogeneity of breast tissue and the variability. Classical methods, such as thresholding or morphological filtering, generally use low-level features, resulting in unsatisfactory outcomes that even totally fail to work. Therefore, more works focus on the detection of clusters of microcalcification instead. Segmentation of microcalcifications, however, does not necessarily have to be embedded in the detection system. As a consequence, limited deep learning-based works are especially aimed at microcalcification segmentation tasks.

Valvano et al.¹⁷⁷ developed a segmentation system for microcalcification based on CNN, which was reported to achieve 99.99% accuracy of detection while the false-positive rate of 0.005% in 283 mammograms. In their work, a CNN detector was designed to classify patches obtained from mammograms to be positive samples or negative samples before segmentation. Using the sliding window of $N \times N$ pixels with the stride of $N/2$, the ROIs were classified into positive samples if microcalcification was presented, or negative samples otherwise. The authors successfully converted the detection problem into a binary classification problem, which turned out to be more computational expedient. The CNN architecture of detection and

TABLE 10 The state-of-the-art mass segmentation methods

Authors	Method	Features	Number of images (or patches) in the testing set	Performance
Al-antari et al. ¹⁶⁸	FrCN	No pooling used	168	Accuracy of 92.97%, F1-score of 92.69%, Jaccard index 86.37%
Dhungel et al. ¹³⁸	DCNN	Pixel classification + level set method refinement	~23	Dice index at 0.90 ± 0.02
Li et al. ¹⁷²	CRU-Net	CRF + Residual learning	145	91.43 ± 0.02 on DDSM-BCRP and 93.66 ± 0.10 on INbreast
Li et al. ¹⁷³	Attention dense-u-net	Dense encoder + Attention integrated decoder	–	78.38 ± 0.04%
Byra et al. ¹⁷⁴	Attention gated U-Net	Entropy maps + Attention gated	81	average Dice score of 0.60
Maicas et al. ¹⁷⁵	GOCS +DCNN	Two-staged segmentation based on shape prior produced by DCNN	69	mean Dice 0.77 ± 0.14

segmentation were the same that comprises six convolutional layers with the size of kernels 3×3 . Each pixel in ROIs was then classified into the foreground that belongs to the microcalcification, and background accordingly while the masks provided by radiologists were taken as ground truth. When training the proposed model, 231 out of 283 mammograms were randomly partitioned into the training set while 25 in the remaining mammograms were built as a validation set. Finally, the method was examined on the test set that consists of 27 mammograms. While the accuracy of microcalcification segmentation was pretty high, the proposed segmentation method remains to be tested on larger data sets.

Similar work was presented in Valvano's other work.¹⁷⁸ Each pixel in the mammogram was picked as the center point of a sliding window, which was later passed to the constructed deep CNN for binary classification. Therefore, the segmentation problem is converted to a classification problem. According to the classification result, pixels were then partitioned into microcalcification and non-microcalcification. The proposed network has six convolutional layers and three fully connected layers. The size of the input is 99 by 99. Two-hundred and thirty-eight mammograms were used to train and validate the proposed network while another independent 52 images are used for testing. In contrast to accuracy at 58% by the classical approach, the proposed method's accuracy reached 83.7% as reported by the author.

Hossain¹⁷⁹ proposed a modified U-Net in an automatic microcalcification detection system as the segmentation module. In the detection system, the Laplacian filter is used to improve the contrast of mammograms as microcalcifications are brighter than surrounding pixels. The pectoral region is segmented by K-means pixel-wise clustering for breast region acquisition. A fuzzy C-means clustering algorithm is used to detect the suspicious regions, which are divided into negative and positive patches. The modified U-Net has four convolutional layers in the encoder part, while the input size is 32×32 . The five cross-validations method is utilized to evaluate the developed system. When dividing the data set, 60% of positive patches from the patch selection procedure are used to train the U-Net while the remaining 40% of them are evenly divided into the validation set and the testing set, respectively. Finally, a mean accuracy at 98.2% with the sensitivity at 98.4%. The remaining measurement metrics are precision

TABLE 11 The state-of-the-art microcalcification segmentation methods

Authors	Method	Features	Number of images (or patches) for evaluation	Performance
Gabirele et al. ¹⁷⁷	DCNN	High accuracy	27	Accuracy of 99.99% with a false-positive rate of 0.005%
Gabirele et al. ¹⁷⁸	DCNN	Conversion from segmentation to pixel classification	52	Accuracy of 83.7%
Hossain et al. ¹⁷⁹	Preprocess + Modified U-Net	High accuracy	–	a mean accuracy at 98.2%, sensitivity at 98.4%.
Hou et al. ¹⁸⁰	U-Net	Multitask	–	a dice coefficient of 0.49

at 94.7%, F-measure at 98.5%, Dice index at 97.8%, and Jaccard Index at 97.4%. However, the details about the size of the data set are missing the work, which is important to an objective evaluation of the developed system. In Table 11, we summarized the state-of-the-art microcalcification segmentation methods.

6.3 | Other segmentation scenarios in breast cancer

Pectoral muscle segmentation is crucial to advance the performance of CAD systems.¹⁸¹ For breast cancer detection, the detection accuracy, and speed can be improved once the breast region can be restricted to smaller and accurate areas. Therefore, there are also numerous works aiming at designing novel pectoral muscle segmentation algorithms.¹⁸¹ In,¹⁸¹ Ali et al. proposed a novel U-Net codenamed Deep Res U-Net for pectoral muscle segmentation by adding skip connections into the network. Also, the activation function has been updated from ReLU to exponential linear unit (ELU) to avoid dying neurons because of ReLU. Before segmentation, image enhancement is carried out to remove noises by applying different filters. After trained with images from MIAS, INbreast, and DDSM, the developed network was reported a mean intersection over union (IoU) of 97%, dice similarity coefficient (DSC) of 96%, and accuracy of 98% on 1974 images using 10-fold cross-validation. To compare the performance of the deep learning method and the traditional method, called texture-field orientation (TFO) method, on segmentation of pectoral muscle, Ma et al. developed three DCNN models based on U-Net.¹⁸² The performance of the proposed models finally gave a mean percent overlap area (POA) of $93.7\% \pm 6.9\%$ while the POA of TFO was only $86.9\% \pm 16.0\%$. Other similar works can be found in Wang et al.¹⁸³ and Liu et al.¹⁸⁴ More pectoral muscle segmentation works can be found in Moghbel et al.¹⁸⁵

The subjective assessment of breast density can serve as a reliable predictor of breast cancer than other automated or semiautomated methods.¹⁸⁶ Dense breast tissues make breast cancer more difficult being identified and could increase the risk of breast cancer.¹⁸⁷ Therefore, segmentation of breast tissue for subjective assessment is also a valuable procedure. To this end, Saffari et al. developed a so-called effective conditional generative adversarial network for the segmentation of dense tissues in mammograms.¹⁸⁷ After segmentation, a classification network was followed to classify the segmented mammograms into one of four classes. Finally, the proposed deep learning-based classification demonstrated precision, sensitivity, and specificity on a test set containing 410 images of 97.85%, 97.85%, and 99.28%, respectively. Compared to supervised learning, unsupervised learning models are more preferable. Given this, Kallenberg et al. proposed a novel unsupervised deep learning method that addresses breast density segmentation and mammographic risk scoring.¹⁸⁸ In the proposed deep learning model, stacked convolutional layers are used as a sparse autoencoder to extract features in an unsupervised manner. A two-layer neural network was then deployed as the classifier, which was trained in a supervised manner. Percentage mammographic density, a golden standard in mammographic density scoring is used as the measurement in this study. The proposed model, which was named convolutional sparse autoencoder, yielded an AUC of 0.59 on the Dutch Breast Cancer Screening data set containing 394 cancer cases and 1182 healthy controls. More segmentation works that are related to breast cancer are listed in Table 12.

TABLE 12 Other segmentation works related to breast cancer

Authors	Method	Purpose	Features	Number of images (or patches) for evaluation	Performance
Ali et al. ¹⁸¹	FCNN (U-Net)	Pectoral muscle segmentation	Novel architecture	198	A mean Intersection over Union (IoU) of 97%, dice similarity coefficient (DSC) of 96%, and accuracy of 98%.
Ma et al. ¹⁸²	FCNN (U-Net)	Pectoral muscle segmentation	Novel architecture	203	Mean POA of 93.7%±6.9%
Soleimani et al. ¹⁸⁹	DCNN (VGG16)	Pectoral muscle segmentation	A combined method based deep learning and graph-based image processing	–	Dice similarity coefficient (DSC) and accuracy (ACC) on MIAS, CBIS-DDSM and INbreast are 97.22 ± 1.96% and 99.64 ± 0.27%, respectively.
Kim et al. ¹⁹⁰	FCNN (U-Net)	Pectoral muscle segmentation	–	65	A mean sensitivity of 95.55%, mean specificity of 99.88%, mean accuracy of 99.67%, and mean Dice similarity coefficient (DSC) of 95.88%
Saffari et al. ¹⁸⁷	FCNN (U-Net) and DCNN	Breast tissue segmentation and classification	High accuracy;	410	Precision, sensitivity, and specificity of 97.85%, 97.85%, and 99.28%.
Kallenberg et al. ¹⁸⁸	DCNN (Sparse autoencoder)	Breast density segmentation	Unsupervised	–	An AUC of 0.59
Dalmış et al. ¹⁹¹	FCNN (U-Net)	Breast and fibroglandular tissue segmentation	Multiple stage segmentation	22	The highest DSC of 0.944
Zhang et al. ¹⁹²	FCNN (U-Net)	Breast and fibroglandular tissue segmentation	Improved U-Net with residual learning	224	The mean DSC was 0.86 ± 0.05 for breast, 0.83 ± 0.06 for FGT; and the mean accuracy was 0.94 ± 0.03 for breast and 0.93 ± 0.04 for FGT

7 | REMAINING CHALLENGES AND FUTURE TRENDS

The remaining challenges in designing breast cancer CAD systems come from aspects of data sets and the intrinsic issues with deep learning models.

The available large-scale annotated data sets are still challenging. The number of large-scale data sets that provide combinational ground truth and correct annotations is still limited. The reason behind this could be the expensive costs during image acquisition and maintenance. Also, it was pointed out in 197 that the current corpus is leaning toward mass detection. As a result, the related works are skewed as well. Therefore, it would be more beneficial to the development of breast cancer CAD systems if there were much more balanced and large-scaled data sets. To train CNNs to achieve high performance, the number of samples should be large enough though there are semi- or unsupervised systems aiming at mitigating the situation. However, the performance of semi or unsupervised systems is still not comparable to systems trained in a supervised manner. Besides, the ground truth should be provided in systematic ways while the annotations should not be subjective. Corresponding accurate annotations for different tasks are of great importance, given the difference between different tasks such as segmentation and classification. As shown in some databases,^{37,43} annotated ROIs, which are denoted by circles around pixels of interest, are not appropriate to be the ground truth for segmentation. To avoid subjectivity, different radiologists should be involved to contribute to relatively objective annotations. While some data sets are well annotated, only a limited number of data sets are publicly available. Given the above difficulties, few databases meet all of the listed requirements. To cope with insufficient data sets for training and evaluating deep learning models, techniques such as transfer learning and data augmentation are developed.^{193,194} For transfer learning, there are numerous works that transfer models pretrained on nonbreast or even nonmedical images while limited research investigated knowledge transferring from one medical image modality to breast cancer or from one breast cancer data set to another. This could be one of the future directions for transfer learning. Similarly, there are also works exploring new data augmentation methods.¹¹¹ To summary, available large-scale annotated data sets could be the most straightforward yet challenging solution to the challenge.

Another challenge is the intrinsic problems with deep learning models that prevent the development of these models. While most systems based on deep learning outperformed systems based on traditional methods, modern CAD systems' overall performance is still far from satisfactory. For mass detection, how to effectively detect masses that are surrounded by dense tissues needs to be addressed by more intelligent deep learning algorithms. Moreover, for deep learning algorithms, a combination of domain knowledge with deep learning would provide a more interpretable understanding and better reasoning behind the direct convolutional operations. Also, determining which deep learning architectures can achieve high accuracy while remaining computationally reasonable, is another aspect of interest to improve CAD systems' performance. Therefore, future breast cancer CAD systems will benefit a lot from the upgraded deep learning models such as attention-based models.¹⁹⁵

Interpretability has long been an obvious flaw to denounce deep learning models as these models are more like black boxes rather than meaningful hand-crafted features in traditional CAD systems. To address this, deep learning works for visualization have been developed.^{196,197} So, future works for breast cancer CAD systems are more likely to embed these frameworks for better interpretation of the models.

8 | CONCLUSION

This paper simply reviewed the works on CAD systems for breast cancer. Concerning the hot topics, systems that focused on detection, segmentation, and classification of mass and microcalcification are covered in this paper. The popular mammographic databases were well-introduced at the beginning of this paper because they were an indispensable part of the evaluations of CAD systems. Some basic concepts about deep learning were recapped for a better understanding of CNNs. Traditional CAD systems have been introduced to make comparisons between conventional CAD systems and deep learning-based systems. As can be concluded, transfer learning is a popular technique when it comes to breast mass and calcification classification while the state-of-the-art detection frameworks are still taken as the leading method for mass and calcification detection. While CNNs are widely used in solving classification and detection problems, FCNNs are generally deployed for segmentation tasks. However, we found that segmentation works regarding mass significantly outnumbered the works on calcification as mass detection and classification tend to more interesting in the area. In conclusion, deep-learning-based methods have become a predominating choice in the implementation of breast cancer CAD systems. There are also some limitations of this survey paper. One is that we only focused on CAD systems for breast mass and calcification analysis but missed other topics such as breast density and breast asymmetry, which relatively lack attention in the area. Another is the coverage of this survey is still limited. While we aimed at presenting details of CAD systems, there are still considerable meaningful works that are out of our reach.

ACKNOWLEDGMENTS

This paper is partially supported by the Royal Society International Exchanges Cost Share Award, UK (RP202G0230); Hope Foundation for Cancer Research, UK (RM60G0680); Medical Research Council Confidence in Concept Award, UK (MC_PC_17171); British Heart Foundation Accelerator Award, UK; Sino-UK Industrial Fund, UK (RP202G0289); Global Challenges Research Fund (GCRF), UK (P202PF11).

ORCID

Xiang Yu  <https://orcid.org/0000-0002-4495-3749>

Qinghua Zhou  <https://orcid.org/0000-0002-3327-0440>

Shuihua Wang  <https://orcid.org/0000-0003-4713-2791>

Yu-Dong Zhang  <https://orcid.org/0000-0002-4870-1493>

REFERENCES

1. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal AJ. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA Cancer J Clin.* 2017;67(6):439-448.
2. WHO. Cancer; 2019. Cancer data. <https://www.who.int/news-room/fact-sheets/detail/cancer>
3. Fund WCR. Breast cancer statistics; 2018. <https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics>
4. Oliveira JE, Gueld MO, Araújo AdA, Ott B, Deserno TM. Towards a standard reference database for computer-aided mammography. In: *Proceedings of SPIE, the International Society for Optical Engineering*; 2008: 69151Y. 69151-66915Y. 69159.
5. Zuiderveld K. Contrast limited adaptive histogram equalization. In: *Proceedings of Graphics gems IV*; 1994:474-485.

6. Sarah Caul JB. Cancer registration statistics, England: 2017. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2017>
7. Welch HG, Prorok PC, O'Malley AJ, Kramer BSJ. Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. *N Engl J Med.* 2016;375(15):1438-1447.
8. Løberg M, Lousdal ML, Bretthauer M, Kalager MJBCR. Benefits and harms of mammography screening. *Breast cancer research: BCR.* 2015;17(1):63.
9. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Proceedings of Advances in neural information processing systems*; 2012:1097-1105.
10. A. Berg, JD, and, Fei-Fei, L. Large scale visual recognition challenge; 2010. <http://www.image-net.org/challenges/LSVRC/2010/>
11. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, Nevada, 2016:770-778.
12. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: *Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition*; 2015:1-9.
13. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016:2818-2826.
14. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
15. Sainath TN, Mohamed A-r, Kingsbury B, Ramabhadran B. Deep convolutional neural networks for LVCSR. In: *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*; 2013:8614-8618.
16. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing. *IEEE Computat intell Magazine.* 2018;13(3):55-75.
17. Kumar A, Irsoy O, Ondruska P, et al. Ask me anything: dynamic memory networks for natural language processing. In: *Proceedings of International Conference on Machine Learning*; 2016:1378-1387.
18. Osterhout RF. Language translation with head-worn computing. Google Patents; 2017.
19. Dong D, Wu H, He W, Yu D, Wang H. Multi-task learning for multiple language translation. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*; 2015:1723-1732.
20. Debelee TG, Schwenker F, Ibenenthal A, Yohannes D. Survey of deep learning in breast cancer image analysis. *Evolving Syst.* 2020;11(1):143-163.
21. Abdelrahman L, Al Ghamdi M, Collado-Mesa F, Abdel-Mottaleb M. Convolutional neural networks for breast cancer detection in mammography: a survey. *Comput Biol Med.* 2021;131:104248.
22. Jalalian A, Mashohor SB, Mahmud HR, Saripan MIB, Ramli ARB, Karasfi B. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clin Imaging.* 2013;37(3):420-426.
23. Shin SY, Lee S, Yun ID, Kim SM, Lee KM. Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images. *IEEE Trans Med Imaging.* 2018;38(3):762-774.
24. Byra M, Sznajder T, Korzinek D, et al. Impact of ultrasound image reconstruction method on breast lesion classification with neural transfer learning. *Pattern Recogn Image Anal.* 2019:41-52.
25. Zhang J, Cain EH, Saha A, Zhu Z, Mazurowski MA. Breast mass detection in mammography and tomosynthesis via fully convolutional network-based heatmap regression. In: *Proceedings of Medical Imaging Computer-Aided Diagnosis*, Vol. 10575; 2018.
26. Antropova NO, Abe H, Giger ML. Use of clinical MRI maximum intensity projections for improved breast lesion classification with deep convolutional neural networks. *J Med Imaging.* 2018;5(1):014503.
27. Group MS. Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). *The Lancet.* 2005; 365(9473):1769-1778.
28. Gur D, Abrams GS, Chough DM, et al. Digital breast tomosynthesis: observer performance study. *Am J Roentgenol.* 2009;193(2):586-591.
29. Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution X-ray imaging observer study. *Radiology.* 2012;262(3): 788-796.

30. Fotin SV, Yin Y, Haldankar H, Hoffmeister JW, Periaswamy S. Detection of soft tissue densities from digital breast tomosynthesis: comparison of conventional and deep learning approaches. In: *Proceedings of Medical Imaging Computer-Aided Diagnosis*, Vol. 97850X; 2016.
31. Poplack SP, Tosteson TD, Kogel CA, Nagy HM. Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. *Am J Roentgenol*. 2007;189(3):616-623.
32. Gennaro G, Toledano A, Di Maggio C, et al. Digital breast tomosynthesis versus digital mammography: a clinical performance study. *Eur Radiol*. 2010;20(7):1545-1553.
33. Hooley RJ, Durand MA, Philpotts LE. Advances in digital breast tomosynthesis. *Am J Roentgenol*. 2017; 208(2):256-266.
34. Lourenco AP, Barry-Brooks M, Baird GL, Tuttle A, Mainiero MB. Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. *Radiology*. 2015;274(2): 337-342.
35. Vahabi M. Breast cancer screening methods: a review of the evidence. *Health Care Women Int*. 2003;24(9): 773-793.
36. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast: toward a full-field digital mammographic database. *Acad Radiol*. 2012;19(2):236-248.
37. Gale AG. The mammographic image analysis society digital mammogram database. In: *Proceedings of the 2nd International Workshop on Digital Mammography* 1994; 1069:375-378.
38. Rangayyan RM, Mudigonda NR, Desautels JL. Boundary modelling and shape analysis methods for classification of mammographic masses. *Med Biol Eng Comput*. 2000;38(5):487-496.
39. Oliver A, Freixenet J, Marti J, et al. A review of automatic mass detection and segmentation in mam-mographic images. *Med Image Anal*. 2010;14(2):87-110.
40. Dominguez AR, Nandi A. Detection of masses in mammograms using enhanced multilevel-thresholding segmentation and region selection based on rank. In: *Proceedings of IASTED International Conference on Biomedical Engineering*, Austria: Innsbruck; 2007:370-375.
41. Llobet R, Paredes R, Pérez-Cortés JC. Comparison of feature extraction methods for breast cancer de-tection. In: *Proceedings of Iberian Conference on Pattern Recognition and Image Analysis*; 2005:495-502.
42. Matheus BRN, Schiabel H. Online mammographic images database for development and comparison of CAD schemes. *J Digit Imaging*. 2011;24(3):500-506.
43. Tangaro S, Bellotti R, De Carlo F, et al. MAGIC-5: an Italian mammographic database of digitised images for research. *Radiol Med*. 2008;113(4):477-485.
44. M Heath KB, Kopans D, R Moore PK Jr. The digital database for screening mammography. In: *Proceedings of the 5th international workshop on digital mammography*; 2000:212-218.
45. Song E, Jiang L, Jin R, Zhang L, Yuan Y, Li Q. Breast mass segmentation in mammography using plane fitting and dynamic programming. *Academic Radiol*. 2009;16(7):826-835.
46. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci Data*. 2017;4:4.
47. Halling-Brown MD, Warren LM, Ward D, et al. OPTIMAM mammography image database: a large scale resource of mammography images and clinical data. *Radiol: Artif Intell*. 2020;3(1):e200103.
48. Antoniou ZC, Giannakopoulou GP, Andreadis II, Nikita KS, Ligomenides PA, Spyrou GM. A web-accessible mammographic image database dedicated to combined training and evaluation of radiologists and machines. In: *Proceedings of 2009 9th International Conference on Information Technology and Applications in Biomedicine*; 2009:1-4.
49. McCulloch WS, WJTbomb P. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biol*. 1943;5(4):115-133.
50. Rosenblatt F. The perceptron, a perceiving and recognizing automaton Project Para: Cornell Aeronautical Laboratory; 1957.
51. Minsky M, Papert SA. *Perceptron: An Introduction on Computational Geometry*; 1996: MIT Press.
52. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA*. 1982;79(8):2654-2558.
53. Usama M, Qadir J, Raza A, et al. Unsupervised machine learning for networking: Techniques, applica-tions and research challenges. *IEEE Access*. 2019;7:65579-65615.

54. Li C, Xu J, Liu Q, et al. Multi-view mammographic density classification by dilated and attention-guided residual learning. *IEEE/ACM Trans Comput Biol Bioinf.* 2020;11:2339-2351.
55. Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. In: *Proceedings of ICML*; 2013.
56. LeCun Y, Bengio Y, Hinton GJ. Deep learning. 2015;521(7553):436.
57. Zhang X, Zhou X, Lin M, Sun J. Shufflenet: an extremely efficient convolutional neural network for mobile devices. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2018:6848-6856.
58. Howard AG, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv.* 2017;170404861.
59. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans Pattern Anal Mach Intell.* 2017;40(4):834-848.
60. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv preprint arXiv.* 2016;160207360.
61. Zeiler MD, Fergus R. Stochastic pooling for regularization of deep convolutional neural networks. *arXiv preprint arXiv.* 2013;13013557.
62. Shi Z, Ye Y, Wu Y. Rank-based pooling for deep convolutional neural networks. *Neural Net.* 2016;83:21-31.
63. Kumar A. Ordinal pooling networks: for preserving information over shrinking feature maps. *arXiv preprint arXiv.* 2018;180402702.
64. Deng J, Satheesh ABS, Su H, Khosla A, Fei-Fei L. ILSVRC-2012. <http://www.image-net.org/challenges/LSVRC/2012/>; 2012.
65. Lin M, Chen Q, Yan S. Network in network. *arXiv preprint arXiv.* 2013;13124400.
66. Simonyan Karen ZA. Very deep convolutional networks for large-scale image recognition. *arXiv.* 2014
67. Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images; 2009.
68. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2017:4700-4708.
69. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition*; 2018:7132-7141.
70. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 7-12, 2015; 2015:3431-3440.
71. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Semantic image segmentation with deep convolutional nets and fully connected crfs. *IEEE Trans Pattern Anal Mach Intell.* 2018;40(4):834-848.
72. Holschneider M, Kronland-Martinet R, Morlet J, Tchamitchian P. A real-time algorithm for signal analysis with the help of the wavelet transform. *Wavelets.* Springer; 1990:286-297.
73. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Proceedings of International Conference on Medical Image Computing And Computer-assisted Intervention*; 2015:234-241.
74. Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation. In: *Proceedings of the IEEE International Conference On Computer Vision*; 2015:1520-1528.
75. Srivastava Nitish HG, Alex KrizhevskyIlya Sutskever, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learning Res.* 2014;15(1):1929-1958.
76. Ba J, Frey B. Adaptive dropout for training deep neural networks. In: *Proceedings of Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada; 2013:3084-3092.
77. Konda K, Bouthillier X, Memisevic R, Vincent P. Dropout as data augmentation. *Statistics.* 2015;29:1050.
78. Gal Y, Ghahramani Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: *Proceedings of International Conference On Machine Learning*, New York City, NY; 2016:1050-1059.
79. Lasserre JA, Bishop CM, Minka TP. Principled hybrids of generative and discriminative models. In: *Proc 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* Las Vegas, NV; 2006:87-94.
80. Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep learning.* MIT press; 2016.

81. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence And Statistics*; 2010:249-256.
82. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE International Conference On Computer Vision*, Boston, MA; 2015:1026-1034.
83. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2017. pp. 1492-1500.
84. Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput.* 2017;29(9):2352-2449.
85. Luckow A, Cook M, Ashcraft N, Weill E, Djerekarov E, Vorster B. Deep learning in the automotive industry: Applications and tools. In: *Proceedings of 2016 IEEE International Conference on Big Data (Big Data)*; 2016:3759-3768.
86. Rana N, Chien C. Deep machine learning based Image classification in hard disk drive manufacturing. In: *Proceedings of Metrology, Inspection, and Process Control for Microlithography XXXII*; 2018:105850Y.
87. Zhao Z-Q, Zheng P, Xu S-t, Wu X. Object detection with deep learning: a review. *IEEE Transactions On Neural Networks And Learning Systems*; 2019.
88. Szegegy C, Toshev A, Erhan D. Deep neural networks for object detection. In: *Proceedings of Advances in Neural Information Processing Systems*; 2013:2553-2561.
89. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2014:580-587.
90. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell.* 2015;37(9):1904-1916.
91. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: *Proceedings of the IEEE international Conference on Computer Vision*; 2017:2961-2969.
92. Girshick R. Fast R-CNN. In: *Proceedings of the IEEE International Conference On Computer Vision*; 2015:1440-1448.
93. Ren Shaoqing HK, Girshick Ross, Jian S. Faster R-CNN: towards real-time object detection with region proposal networks. In: *Proceedings of Advances in Neural Information Processing Systems*; 2015:91-99.
94. Dai J, Li Y, He K, Sun J. R-FCN: object detection via region-based fully convolutional networks. In: *Proceedings of Advances in Neural Information Processing Systems*; 2016:379-387.
95. Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv.* 2013;13126229.
96. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016:779-788.
97. Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector. In: *Proceedings of European Conference on Computer Vision*; 2016:21-37.
98. Law H, Deng J. Cornernet: Detecting objects as paired keypoints. In: *Proceedings of the European Conference on Computer Vision (ECCV)*; 2018:734-750.
99. Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation. In: *Proceedings of European Conference on Computer Vision*; 2016:483-499.
100. Duan K, Bai S, Xie L, Qi H, Huang Q, Tian Q. Centernet: keypoint triplets for object detection. In: *Proceedings of the IEEE International Conference on Computer Vision*; 2019:6569-6578.
101. Fourcade A, Khonsari R. Deep learning in medical image analysis: A third eye for doctors. *Journal of stomatology, oral and maxillofacial surgery.* 2019;120(4):279-288.
102. Nasr-Esfahani E, Samavi S, Karimi N, et al. Melanoma detection by analysis of clinical images using convolutional neural network. In: *Proceedings of 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*; 2016:1373-1376.
103. Phan S, Satoh Si, Yoda Y, Kashiwagi K, Oshika T, Group JOIRR. Evaluation of deep convolutional neural networks for glaucoma detection. *Jpn J Ophthalmol.* 2019;63(3):276-283.
104. Taylor AG, Mielke C, Mongan J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: a retrospective study. *PLOS Med.* 2018;15(11):e1002697.

105. Arik SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. *J Med Imaging*. 2017;4(1):014501.
106. Uthoff RD, Song B, Sunny S, et al. Point-of-care, smartphone-based, dual-modality, dual-view, oral cancer screening device with neural network classification for low-resource communities. *PLoS One*. 2018;13(12):e0207493.
107. Alaskar H, Hussain A, Al-Aseem N, Liatsis P, Al-Jumeily D. Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images. *Sensors*. 2019;19(6):1265.
108. Gonzales RC, Woods RE. *Digital image processing*. Prentice Hall; 2002.
109. Dromain C, Balleyguier C. Contrast-enhanced digital mammography. *Digital mammography*. Springer; 2010:187-198.
110. Arianna Mencattini MS, Roberto L, Manuela F, Caselli F. Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing. *IEEE Trans Instrumentation Measurement*. 2008;57(7):1422-1430.
111. Yu X, Kang C, Guttery DS, Kadry S, Chen Y, Zhang Y-D. ResNet-SCDA-50 for breast abnormality classification. *IEEE/ACM Trans Computat Biol Bioinform*; 2020.
112. Borges LR, Caron RF, de Azevedo-Marques PM, Vieira MAC. Effect of denoising on the localization of microcalcification clusters in digital mammography. In: Bosmans H, Marshall N, VanOngeval C, eds. *15TH INTERNATIONAL WORKSHOP ON BREAST IMAGING (IWBI2020)*, 2020.
113. Hamed G, Marey M, Amin SE, Tolba MF. A proposed model for denoising breast mammogram images. In: International conference on Computer Engineering and Systems; 2018:652-657.
114. Shanmugavadivu P, Narayanan SGL. Segmentation of microcalcification regions in digital mammograms using self-guided region-growing. In: *Proceedings of 2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET)*, Dec 13-14, 2012; 2012:274-279.
115. Bhattacharya M, Das A. Fuzzy logic based segmentation of microcalcification in breast using digital mammograms considering multiresolution. In: *Proceedings of International Machine Vision and Image Processing Conference (IMVIP 2007)*, Sept. 5-7, 2007:98-105.
116. Zhang Y, Tomuro N, Furst J, Raicu DS. Image enhancement and edge-based mass segmentation in mammogram. In: *Proceedings of Medical Imaging 2010: Image Processing*; 2010:76234P.
117. Chan Tony FVLA. Active contours without edges. *IEEE Trans Image Process*. 2001;10(2):266-277.
118. Li C, Kao C-Y, Gore JC, Ding Z. Minimization of region-scalable fitting energy for image segmentation. *IEEE Trans Image Process*. 2008;17(10):1940-1949.
119. Li C, Xu C, Gui C, Fox MD. Level set evolution without re-initialization: a new variational formulation. In: *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*; 2005:430-436.
120. Hao X, Shen Y, Xia S-r. Automatic mass segmentation on mammograms combining random walks and active contour. *J Zhejiang Univ Sci*. 2012;13(9):635-648.
121. Liu F, Gong Z, Chen Y, Gu Y. Segmentation of mass in mammograms by a novel integrated active contour method. *Int J Computat Sci Eng*. 2015;11(2):207-215.
122. Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Syst Appl*. 2009;36(2):3240-3247.
123. El-Naqa I, Yang Y, Wernick MN, Galatsanos NP, Nishikawa RM. A support vector machine approach for detection of microcalcifications. *IEEE Trans Med Imaging*. 2002;21(12):1552-1563.
124. Abbass HA. An evolutionary artificial neural networks approach for breast cancer diagnosis. *Artif Intell Med*. 2002;25(3):265-281.
125. Lo S-CB, Lin J-SJ, Freedman MT, Mun SK. Application of artificial neural networks to medical image pattern recognition: detection of clustered microcalcifications on mammograms and lung cancer on chest radiographs. *J VLSI Signal Process Syst Signal Image Video Technol*. 1998;18(3):263-274.
126. Wang S, Chen P, Zhang Y, Liu A, Ling W. Abnormal breast detection in mammogram images by feed-forward neural network trained by Jaya algorithm. *Fundamental Informat*. 2017;151(1-4):191-211.
127. Bandos AI, Rockette HE, Song T, Gur D. Area under the free-response ROC curve (FROC) and a related summary index. *Biometrics*. 2009;65(1):247-256.
128. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Garcia-Rodriguez J. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv*. 2017;170406857.

129. Li Y, Chen H, Cao L, Ma J. A survey of computer-aided detection of breast cancer with mammography. *J Health Med Inf.* 2016;4(7).
130. Bekker AJ, Greenspan H, Goldberger J. A multi-view deep learning architecture for classification of breast microcalcifications. In: *Proceedings 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*; 2016:726-730.
131. Agarwal R, Diaz O, Lladó X, Yap MH, Martí R. Automatic mass detection in mammograms using deep convolutional neural networks. *J Med Imaging.* 2019;6(3):031409.
132. Kooi T, Litjens G, Van Ginneken B, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal.* 2017;35:303-312.
133. Menon RV, Raha P, Chakrabarti I. Classification of breast mass in ultrasound images using CAD: a survey. In: *Proceedings of 2016 International Conference on Systems in Medicine and Biology (ICSMB)*; 2016:31-35.
134. Varela C, Tahoces PG, Méndez AJ, Souto M, Vidal JJ. Computerized detection of breast masses in digitized mammograms. *Comput Biol Med.* 2007;37(2):214-226.
135. Akselrod-Ballin A, Karlinsky L, Alpert S, Hasoul S, Ben-Ari R, Barkan E. A region based convolutional network for tumor detection and classification in breast mammography. *Deep learning and data labeling for medical applications.* Springer; 2016:197-205.
136. Arevalo J, González FA, Ramos-Pollán R, Oliveira JL, Lopez MAG. Representation learning for mammography mass lesion classification with convolutional neural networks. *Comput Methods Programs Biomed.* 2016;127:248-257.
137. Dhungel N, Carneiro G, Bradley AP. Automated mass detection in mammograms using cascaded deep learning and random forests. In: *Proceedings of 2015 international conference on digital image computing: techniques and applications (DICTA)*; 2015:1-8.
138. Dhungel N, Carneiro G, Bradley AP. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Med Image Anal.* 2017;37:114-128.
139. He Kaiming ZX, Ren Shaoqing, Jian, S. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016:770-778.
140. Al-antari MA, Al-masni MA, Choi M-T, Han S-M, Kim T-S. A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *Int J Med Inform.* 2018;117:44-54.
141. Karssemeijer N, Brake GMt. Detection of stellate distortions in mammograms. *IEEE Trans Med Imaging.* 1996;15(5):611-619.
142. Arevalo J, González FA, Ramos-Pollán R, Oliveira JL, Lopez MAG. Convolutional neural networks for mammography mass lesion classification. In: *Proceedings of 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*; Aug. 25-29, 2015:797-800.
143. Peppard HR, Nicholson BE, Rochman CM, Merchant JK, Mayo III RC, Harvey JA. Digital breast tomosynthesis in the diagnostic setting: indications and clinical applications. *Radiographics.* 2015;35(4):975-990.
144. Samala RK, Chan HP, Hadjiiski L, Helvie MA, Wei J, Cha K. Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. *Med Phys.* 2016;43(12):6654-6666.
145. Yousefi M, Krzyżak A, Suen CY. Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning. *Comput Biol Med.* 2018;96:283-293.
146. Swersky K, Sutskever I, Tarlow D, Zemel RS, Salakhutdinov RR, Adams RP. Cardinality restricted boltzmann machines. In: *Proceedings of Advances in Neural Information Processing Systems*; 2012:3293-3301.
147. Fan M, Li Y, Zheng S, Peng W, Tang W, Li L. Computer-aided detection of mass in digital breast tomosynthesis using a faster region-based convolutional neural network. *Methods.* 2019;166:103-111.
148. Zhou C, Chan HP, Paramagul C, et al. Computerized nipple identification for multiple image analysis in computer-aided diagnosis: computerized nipple identification on mammograms. *Med Phys.* 2004;31(10):2871-2882.
149. Samala R, Chan H, Hadjiyski L, Wei J, Cha K, Helvie M. Comparison of mass detection for digital breast tomosynthesis (DBT) with and without transfer learning of deep-learning convolution neural network

- (DLCNN) from digitized screen-film mammography (SFM) and digital mammography (DM) RSNA Program Book, PH256-SD-WEB5; 2016:845.
150. Mordang J-J, Janssen T, Bria A, Kooi T, Gubern-Mérida A, Karssemeijer N. Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. In: *Cham*; 2016:35-42.
 151. Wang J, Yang Y. A context-sensitive deep learning approach for microcalcification detection in mammograms. *Pattern Recogn*. 2018;78:12-22.
 152. Zhang W, Yoshida H, Nishikawa RM, Doi K. Optimally weighted wavelet transform based on supervised training for detection of microcalcifications in digital mammograms. *Med Phys*. 1998;25(6):949-956.
 153. Rehman MA, Ahmed J, Waqas A, Sawand A. Intelligent system for detection of microcalcification in breast cancer. *Int J Adv Comput Sci Appl*. 2017;8(7):382-387.
 154. Wang J, Yang X, Cai H, Tan W, Jin C, Li L. Discrimination of breast cancer with microcalcifications on mammography by deep learning. *Sci Rep*. 2016;6:27327.
 155. Akselrod-Ballin A, Karlinsky L, Hazan A, et al. Deep learning for automatic detection of abnormal findings in breast mammography. In: *Cham*; 2017:321-329.
 156. Oyelade ON, Ezugwu AES. A state-of-the-art survey on deep learning methods for detection of architectural distortion from digital mammography. *IEEE Access*. 2020;8:148644-148676.
 157. Costa AC, Oliveira HC, Borges LR, Vieira MA. Transfer learning in deep convolutional neural networks for detection of architectural distortion in digital mammography. In: *Proceedings of 15th International Workshop on Breast Imaging (IWBI2020)*; 2020:115130N.
 158. Oyelade ON, Ezugwu AE. A deep learning model using data augmentation for detection of architectural distortion in whole and patches of images. *Biomed Signal Process Control*. 2021;65:102366.
 159. Zheng X, Yao Z, Huang Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun*. 2020;11(1):1-9.
 160. Lotter W, Diab AR, Haslam B, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nature Med*. 2021:1-6.
 161. Eskreis-Winkler S, Onishi N, Pinker K, et al. Using Deep Learning to Improve Nonsystematic Viewing of Breast Cancer on MRI. *J Breast Imaging* 2021.
 162. Albayrak A, Bilgin G. Mitosis detection using convolutional neural network based features. In: *Proceedings of 2016 IEEE 17th International Symposium on Computational Intelligence and Informatics (CINTI)*; 2016:000335-000340.
 163. Chen H, Dou Q, Wang X, Qin J, Heng P. Mitosis detection in breast cancer histology images via deep cascaded networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*; 2016.
 164. Swiderski B, Kurek J, Osowski S, Kruk M, Barhoumi W. Deep learning and non-negative matrix factorization in recognition of mammograms. In: *Proceedings of Eighth International Conference on Graphic and Image Processing (ICGIP 2016)*; 2017:102250B.
 165. Domínguez AR, Nandi AK. Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. *Pattern Recogn*. 2009;42(6):1138-1148.
 166. Rahmati P, Adler A, Hamarneh G. Mammography segmentation with maximum likelihood active contours. *Med Image Anal*. 2012;16(6):1167-1186.
 167. Xian M, Zhang Y, Cheng H-D, Xu F, Zhang B, Ding J. Automatic breast ultrasound image segmentation: a survey. *Pattern Recogn*. 2018;79:340-355.
 168. Al-antari MA, Al-masni MA, Choi M-T, Han S-M, Kim T-S. A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *Int J Med Inform*. 2018;117:44-54.
 169. Osher Stanley FR. *Level Set Methods and Dynamic Implicit Surfaces*. Springer Science & Business Media; 2006.
 170. Chunming L, Chenyang X, Changfeng G, Fox MD. Level set evolution without re-initialization: a new variational formulation. In: *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 431, June 20-25, 2005:430-436.
 171. Chan TF, Vese LA. Active contours without edges. *IEEE Trans Image Process*. 2001;10(2):266-277.
 172. Li H, Chen D, Nailon WH, Davies ME, Laurenson D. Improved breast mass segmentation in mammograms with conditional residual u-net. *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Springer; 2018:81-89.

173. Li S, Dong M, Du G, Mu X. Attention dense-u-net for automatic breast mass segmentation in digital mammogram. *IEEE Access*. 2019;7:59037-59047.
174. Byra M, Jarosik P, Dobruch-Sobczak K, et al. Breast mass segmentation based on ultrasonic entropy maps and attention gated U-Net. *arXiv preprint arXiv*. 2020;200110061.
175. Maicas G, Carneiro G, Bradley AP. Globally optimal breast mass segmentation from DCE-MRI using deep semantic segmentation as shape prior. In: *Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, April 18-21, 2017; 2017:305-309.
176. Dhungel N, Carneiro G, Bradley AP. Automated mass detection in mammograms using cascaded deep learning and random forests. In: *Proceedings of 2015 international conference on digital image computing: techniques and applications (DICTA)*; 2015:1-8.
177. Valvano G, Santini G, Martini N, et al. Convolutional neural networks for the segmentation of microcalcification in mammography imaging. *J Healthc Eng*. 2019;2019:9360941.
178. Valvano G, Della Latta D, Martini N, et al. *Evaluation of a Deep Convolutional Neural Network method for the segmentation of breast microcalcifications in Mammography Imaging*. Singapore; 2018:438-441.
179. Hossain MS. Microcalcification segmentation using modified U-net segmentation network from mammogram images. *J King Saud Univ-Comput Inform Sci*. 2019.
180. Hou R, Grimm L, Mazurowski M, et al. *A multitask deep learning method in simultaneously predicting occult invasive disease in ductal carcinoma in-situ and segmenting microcalcifications in mammography*. SPIE; 2020.
181. Ali MJ, Raza B, Shahid AR, et al. Enhancing breast pectoral muscle segmentation performance by using skip connections in fully convolutional network. *Int J Imaging Syst Technol*. 2020;30(4):1108-1118.
182. Ma X, Wei J, Zhou C, et al. Automated pectoral muscle identification on MLO-view mammograms: Comparison of deep neural network to conventional computer vision. *Med Phys*. 2019;46(5):2103-2114.
183. Wang K, Khan N, Chan A, Dunne J, Highnam R. Deep learning for breast region and pectoral muscle segmentation in digital mammography. In: *Proceedings of Pacific-Rim Symposium on Image and Video Technology*; 2019:78-91.
184. Liu W, Liu C, Wei Y. Utilizing deep learning technology to segment pectoral muscle in mediolateral oblique view mammograms. In: *Proceedings of 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)*; 2020:97-101.
185. Moghbel M, Ooi CY, Ismail N, Hau YW, Memari N. A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography. *Artif Intell Rev*. 2019:1-46.
186. Astley SM, Harkness EF, Sergeant JC, et al. A comparison of five methods of measuring mammographic density: a case-control study. *Breast Cancer Res*. 2018;20(1):1-13.
187. Saffari N, Rashwan HA, Abdel-Nasser M, et al. Fully automated breast density segmentation and classification using deep learning. *Diagnostics*. 2020;10(11):988.
188. Kallenberg M, Petersen K, Nielsen M, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE Trans Med Imaging*. 2016;35(5):1322-1331.
189. Soleimani H, Michailovich OV. On segmentation of pectoral muscle in digital mammograms by means of deep learning. *IEEE Access*. 2020;8:204173-204182.
190. Kim YJ, Yoo EY, Kim KG. Deep learning based pectoral muscle segmentation on MIAS mammograms. 2020.
191. Dalmiş MU, Litjens G, Holland K, et al. Using deep learning to segment breast and fibroglandular tissue in MRI volumes. *Med Phys*. 2017;44(2):533-546.
192. Zhang Y, Chen J-H, Chang K-T, et al. Automatic breast and fibroglandular tissue segmentation in breast MRI using deep learning by a fully-convolutional residual neural network U-net. *Academic Radiol*. 2019; 26(11):1526-1535.
193. Agarwal R, Diaz O, Yap MH, Lladó X, Martí R. Deep learning for mass detection in full field digital mammograms. *Comput Biol Med*. 2020;121:103774.
194. Castro E, Cardoso JS, Pereira JC. Elastic deformations for data augmentation in breast cancer mass detection. In: *Proceedings of 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*; 2018:230-234.

195. Woo S, Park J, Lee J-Y, Kweon IS. Cbam: convolutional block attention module. In: Proc Proceedings of the European conference on computer vision (ECCV); 2018:3-19.
196. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*; 2017:618-626.
197. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: *Proceedings of European Conference on Computer Vision*; 2014:818-833.

How to cite this article: Yu X, Zhou Q, Wang S, Zhang Y. A systematic survey of deep learning in breast cancer. *Int J Intell Syst.* 2022;37:152-216.

<https://doi.org/10.1002/int.22622>