

Deformable Gabor Feature Networks for Biomedical Image Classification

Xuan Gong^{1†}, Xin Xia^{2†}, Wentao Zhu³, Baochang Zhang², David Doermann¹, Li'an Zhuo²

¹University at Buffalo ²Beihang University ³Kwai Inc.

{xuangong, doermann}@buffalo.edu {xiaxin, bczhang, lianzhuo}@buaa.edu.cn
wentaozhu91@gmail.com

Abstract

In recent years, deep learning has dominated progress in the field of medical image analysis. We find however, that the ability of current deep learning approaches to represent the complex geometric structures of many medical images is insufficient. One limitation is that deep learning models require a tremendous amount of data, and it is very difficult to obtain a sufficient amount with the necessary detail. A second limitation is that there are underlying features of these medical images that are well established, but the black-box nature of existing convolutional neural networks (CNNs) do not allow us to exploit them. In this paper, we revisit Gabor filters and introduce a deformable Gabor convolution (DGConv) to expand deep networks interpretability and enable complex spatial variations. The features are learned at deformable sampling locations with adaptive Gabor convolutions to improve representativeness and robustness to complex objects. The DGConv replaces standard convolutional layers and is easily trained end-to-end, resulting in deformable Gabor feature network (DGFN) with few additional parameters and minimal additional training cost. We introduce DGFN for addressing deep multi-instance multi-label classification on the INbreast dataset for mammograms and on the ChestX-ray14 dataset for pulmonary x-ray images.

1. Introduction

Automated medical imaging techniques for cancer screening are widely used for lesion analysis [8], but the traditional pipeline for computer aided diagnosis is typically built based on hand-crafted features [25]. These features are not flexible and have poor generalization on unseen data. Deep features, however, are data-driven and are becoming the approach of choice in medical image analysis. Deep learning has achieved great success on skin cancer diagnostics [6], organs at risk delineation for radiotherapy [32] and

pneumonia detection from chest x-ray images [21] for example.

One challenge for deep learning is that it is data hungry and often requires expensive and detailed annotation [10, 24]. For cancer screening training and validation data in medical images, image-level description of the clinical diagnosis may not be sufficient to train for clinical diagnosis [34]. Another challenge arises from CNN itself. CNNs are widely considered black boxes and difficult to interpret. This becomes a greater challenge for weakly supervised learning in biomedical image analysis, whose performance depends highly on powerful representations to handle complicated spatial variations, such as lesion sizes, shapes and viewpoints.

Gabor wavelets [7] are widely considered the state-of-the-art hand-crafted feature extraction method, enhancing the robustness of the representation to scale and orientation changes in images. The advantage of Gabor transforms for specific frequency analysis makes them suitable to interpret and resist to dense spatial variations widely existing in biomedical images. Recently, Gabor convolutional networks (GCNs) [17] have used Gabor filters to modulate convolutional filters and enhance representation ability of CNNs. [17] only consider rigid transformations of kernels, however, and not deformable transformations on features that are required for medical image analysis. Thus the robustness of Gabor filters to spatial variations has not been fully investigated to facilitate feature extraction in CNNs.

On the other hand, deformable convolutional networks (DCNs) [3] augment spatial sampling locations and provide generalized transformations such as anisotropic aspect ratios, demonstrating effectiveness on sophisticated vision tasks such as object detection. We will show that the tailored combination of Gabor filters and deformable convolutions in a dedicated architecture can better characterize the spatial variations and enhance feature representations to facilitate medical image analysis.

In this paper, we investigate deeply into Gabor wavelets with deformable transforms to enhance the networks interpretability and robustness to complex data variations.

[†]Equal contribution.

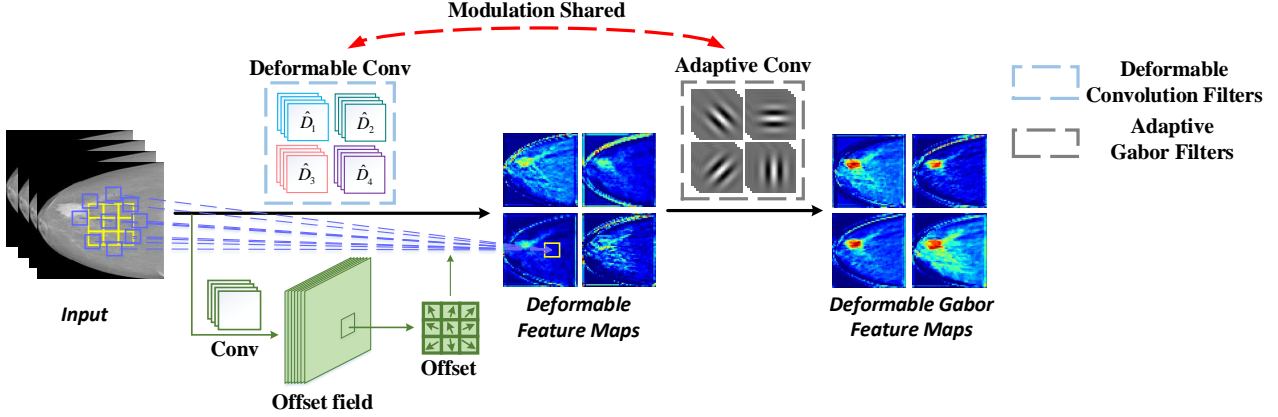


Figure 1. The framework of our deformable Gabor Convolution (DGConv).

Unlike previous hand-crafted filters, the newly designed module learns Gabor filters end-to-end, thus improving its adaptiveness to the input data. As illustrated in Figure 1, our deformable Gabor convolution (DGConv) includes deformable convolutions and adaptive Gabor convolutions that share the same modulation information. The deformable convolutions are endowed with local offset transforms to make the feature sampling locations learnable. The adaptive Gabor convolutions further facilitate the capture of visual properties such as spatial localization and orientation selectivity of the input objects, enhancing the generated deformable Gabor features with various dense transformations. To balance the performance and model complexity, we only employ deformable Gabor convolution (DGConv) to extract high level deep features. We integrate this new Gabor module into deep multi-instance multi-label networks, leading to deformable Gabor feature networks (DGFNs) to deal with large variations of objects in medical images. The contributions of this work are summarized as follows:

- Deformable Gabor feature network (DGFN) exploits deformable features and learnable Gabor features in one block to improve the interpretability of CNNs. The noise-resistant property inherited from Gabor features is successfully validated on CIFAR-10 with a 2% accuracy improvement over the baseline method.
- DGFN features both the adaptiveness to deformation and robustness to generalize spatial variations common in natural images. Their enhanced representative ability are shown to be beneficial for medical image analysis.
- The proposed Gabor module is generic and flexible, which can be easily applied to existing CNNs, such as ResNet and DenseNet.

2. Related Work

2.1. Deformable Convolutional Networks

CNNs have achieved great success for visual recognition but are inherently limited to spatial variations in object size, pose and viewpoint [16, 28]. One method that has been used to address this problem is data augmentation which adds training samples with extensive spatial variations using random transformations. Robust features can be learned from the data but at the cost of an increased number of model parameters and additional training resources. Another method is to extract spatial invariant features with learned transformations. Ilse *et al.* [14] first proposed spatial transformer networks to learn invariance to translation, scale, rotation and generic warping, giving neural networks the ability to actively and spatially transform feature maps. Deformable convolutional networks (DCNs) [3] introduced offset learning to sample the feature map in a local and efficient manner which can be trained end-to-end.

2.2. Gabor Convolution Networks

Gabor wavelets [7] exhibit strong characteristics of spatial locality, scale and orientation selectivity, and insensitivity to illumination change. The recent rise of deep learning has lead to the combination of Gabor filters and convolution neural networks. Previously Gabor wavelets were only used to initialize deep networks or used in the pre-processing [15, 31]. [22] replaced selected weight kernels of CNNs with Gabor filters to reduce training cost and time. Recent work has integrated Gabor filters into CNNs intrinsically to enhance the resistance of deep learned features to spatial changes [17]. However, the receptive field of the integrated Gabor filters is fixed and known, and such prior knowledge characterizes limited spatial transformations thus impeding the generalization of complicated spatial variations and new unknown tasks. In this work, we

go further by tailoring Gabor filters with learnable modulation masks and deformable transforms. The steerable property of Gabor filters is therefor inherited into the deformable convolutions and its representativeness to spatial variations is fully exploited.

2.3. Multi-Instance Learning for Weakly Supervised Image Analysis

There have been a number of previous attempts to utilize weakly supervised labels to train models for image analysis [23]. Papandreou *et al.* [20] proposed an iterative approach to predict pixel-wise labels in segmentation using image-level labels. Different pooling strategies were proposed for weakly supervised localization and segmentation respectively [27, 2]. Wu *et al.* [29] combined CNN with multi-instance learning (MIL) for image auto-annotation. Deep MIL with several efficient inference schemes was proposed for lesion localization and mammogram classification [33]. Attention based MIL further employed neural attention mechanisms as the inference [13]. Wan *et al.* [26] proposed a min-entropy latent model for weakly supervised object detection, which reduces the variance of positive instances and alleviates the ambiguity of the detectors. Unlike previous methods, our method uses a novel feature representation network to handle large variations of objects in medical images and improve overall image classification.

3. Deformable Gabor Convolution

Without loss of generality, the convolution operation described here is in 2D.

3.1. Deformable and Adaptive Gabor Convolution

To extract highly representative features, we combine the deformable convolution (DConv) with an adaptive Gabor convolution (GConv) by sharing modulation information. As illustrated in Figure 2, both the deformable convolution and Gabor transforms are adjusted with the learned masks.

Deformable Convolution: We are given U standard convolution filters of size $H \times H$, which after being modulated by V scale kernels of size $H \times H$, result in $U \times V$ modulated convolution filters of size $H \times H$. We define:

$$\hat{D}_{u,v} = C_u \circ S_v, \quad (1)$$

where $\hat{D}_{u,v}$ indicates the deformable convolution filter, \circ is element wise product operation, C_u is the u^{th} convolution filter, and S_v is the v^{th} kernel to modulate the convolution filter. In our implementation, the deformable transforms [3] augment $\hat{D}_{u,v}$ with translated offsets which are learned from the preceding feature maps through additional convolutions.

Consider a 3×3 kernel convolution, $\mathcal{R} = \{(-1, -1), \dots, (1, 0), (1, 1)\}$, with a dilation of 1,

for example. Given \mathbf{r}_0 as the 2D position of output feature and \mathbf{r}_n as the location of \mathcal{R} , the deformable convolution filter \hat{D} can be operated on as follows*:

$$F_y(\mathbf{r}_0) = \sum_{\mathbf{r}_n \in \mathcal{R}} \hat{D}(\mathbf{r}_n) \times F_x(\mathbf{r}_0 + \mathbf{r}_n + \Delta \mathbf{r}_n) \quad (2)$$

where F_x and F_y indicate the input and output feature respectively. The learned offset $\Delta \mathbf{r}_n$ updates the offset location to $\mathbf{r}_n + \Delta \mathbf{r}_n$ and adjusts the receptive field of input F_x on which \hat{D} is implemented.

Adaptive Gabor Convolution: Adaptive Gabor filters are generated from U Gabor filters of size $H \times H$ with V learned kernels of size $H \times H$, where U indicates the number of orientations of Gabor filters. We have:

$$\hat{G}_{v,u} = S_v \circ G_u, \quad (3)$$

where G_u is the Gabor filter with orientation u , and $\hat{G}_{v,u}$ is the adaptive Gabor filter corresponding to the u^{th} orientation and the v^{th} scale. For DGConvs, different layers share the same Gabor filters $G = (G_1, \dots, G_U)$ with various orientations but are adjusted with different information from the corresponding deformable convolution features.

If the dimensions of the weights in traditional convolution are $M_0 \times N_0 \times H \times H$, the dimensions of the learned convolution filters are $M \times N \times U \times H \times H$ in DGConv, where U represents the number of additional orientation channels, $N(N_0)$ and $M(M_0)$ represent the channel number of the input and output respectively. In DGConv we set $N = N_0/\sqrt{U}$ and $M = M_0/\sqrt{U}$ to maintain similar amount of parameters with traditional convolution. Additional parameters in DGConv include $V \times H \times H$ parameters of mask and $(2 \times H \times H) \times N \times U \times H \times H$ parameters for offset learning, where $2 \times H \times H$ is the channel of offset fields and means that each position of input feature corresponds to an offset of size $2 \times H \times H$ for deformable convolution.

In DGConv, the number of orientation channels in the input and output feature needs to be U . So the number of orientation channels in the first input feature must be extended to U . For example, if the dimension of original input feature is $1 \times N \times W \times W$ where $W \times W$ is the size of input feature, it will be $U \times N \times W \times W$ after duplicating and concatenating. Thus the new module is light weight and can easily be implemented with a small number of additional parameters.

3.2. Forward Propagation

We use deformable Gabor convolutions (DGConvs) to produce deformable Gabor features. Given the input features F , the output Gabor features \hat{F} are denoted:

$$\hat{F} = \text{DGConv}(F, \hat{D}, \hat{G}), \quad (4)$$

*The subscript is omitted for easy presentation.

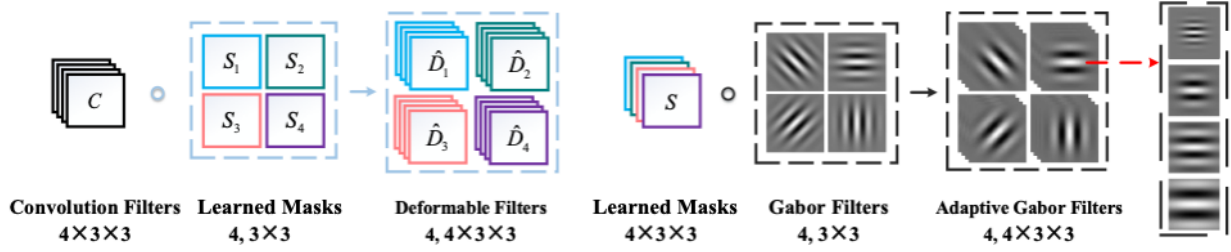


Figure 2. The modulation process of deformable filters and adaptive Gabor filters. The left shows how convolution filters are modulated by learned masks to generate deformable filters. The right illustrates the generation of adaptive Gabor filters. For illustration convenience, we set the number of learned masks as $V=4$ and the orientation channel of convolution filters and Gabor filters as $U=4$.

where DGConv is the operation which includes deformable convolution filters \hat{D} and adaptive Gabor filters \hat{G} . So the deformable features $E_v^{(m)}$ and the deformable Gabor features $\hat{F}_u^{(m)}$ are obtained by:

$$E_v^{(m)} = \sum_{n,u} F_u^{(n)} \odot \hat{D}_{u,v}^{(n,m)}, \quad \hat{F}_u^{(m)} = \sum_v E_v^{(m)} \otimes \hat{G}_{v,u}, \quad (5)$$

where \otimes denotes the traditional convolution, \odot denotes the deformable convolution shown in Eq. (2), and n and m denote the number of channels in the input and output features respectively. $E_v^{(m)}$ represents the deformable feature with v^{th} modulation in the m^{th} channel. u indicates $\hat{F}_u^{(m)}$ being the u^{th} orientation response of the deformable Gabor features $\hat{F}^{(m)}$. Figure 1 shows that deformable Gabor feature maps reveal better spatial detection results of lesions after the adaptive Gabor convolutions.

3.3. Backward Propagation

During the back propagation in the DGConv, we need to update the kernels C and S , which can be jointly learned. The loss function of the network \mathcal{L} is differentiable within a neighborhood of a point, which will be described in the next section. We design a novel back propagation (BP) scheme to update parameters:

$$\delta_S = \frac{\partial \mathcal{L}}{\partial S} = \frac{\partial \mathcal{L}}{\partial \hat{G}} \circ \sum_{u=1}^U G_u, \quad S \leftarrow S - \eta_1 \delta_S, \quad (6)$$

where G_u is the Gabor filter with orientation u and η_1 denotes the learning rate for S . We then fix S and update parameters D of deformable convolution filters:

$$\delta_C = \frac{\partial \mathcal{L}}{\partial C} = \frac{\partial \mathcal{L}}{\partial \hat{D}} \circ \sum_{v=1}^V S_v, \quad D \leftarrow C - \eta_2 \delta_C, \quad (7)$$

where S_v is the v^{th} learned kernel and η_2 denotes the learning rate of convolution parameters.

4. Biomedical Image Analysis

There are many different ways to formulate problems in biomedical image analysis. Two of the most common are to classify an entire image as either having a particular condition or not (a binary-label task) and to associate the image with several labels (a multi-label task). To test our deformable Gabor feature network (DGFN), we have identified two representative datasets, the INbreast dataset [18] and the ChestX-ray14 dataset [27].

4.1. The INbreast Dataset

The INbreast Dataset [18] is a dataset of mammogram images consisting of 410 images from a total of 115 cases, of which 90 cases are from women with both breasts (4 images per case) and 25 cases are from mastectomy patients (2 images per case) [18]. The dataset includes four types of lesions: masses, calcifications, asymmetries, and distortions. We focus on mass malignancy classification from mammograms.

For mammogram classification, the equivalent problem is that if there exists a malignant mass, the mammogram I should be classified as positive. Likewise, a negative mammogram I should not have any malignant masses. If we treat each patch Q_k of I as an instance, the mammogram classification is a standard multi-instance learning problem. For a negative mammogram, we expect all the malignant probabilities p_k to be close to 0. For a positive mammogram, at least one malignant probability p_k should be close to 1.

4.2. The ChestX-ray14 Dataset

As one of the largest publicly available chest x-ray datasets, ChestX-ray14 consists of 112,120 frontal-view x-ray images scanned from 32,717 patients including many patients with advanced lung diseases [27]. Each image is labeled with one or multiple pathology keywords, such as atelectasis, or cardiomegaly. This dataset consists of complicated diseases which may have interrelations which can be challenging for the classification task. The ChestX-ray14

dataset has fourteen different labels, so the image classification problem is to associate each instance with a subset of those labels. This is a multi-instance, multi-label classification problem.

4.3. Our Approach

We use the proposed Gabor module to extract highly representative features and design a multi-instance learning method to deal with deformable Gabor features. In this section, we describe the structure of the deformable Gabor feature networks (DGFNs) for these two problems.

4.3.1 Multi-Instance Learning for Mammograms

After multiple DGConv layers and rectified linear units, we acquire the last deformable Gabor features F with multiple channels. $F_{i,j,:}$ is the feature map for patch $Q_{i,j}$ of the input image, where i and j denote the spatial index of the row and column respectively, and $:$ denotes the channel dimension. We employ a logistic regression model with weights shared across all the patches of the output feature map. A sigmoid activation function for nonlinear transformation is then applied along channels for each element of the output feature map $F_{i,j,:}$ and we slide it over all the pixel positions to calculate the malignant probabilities. The malignant probability of pixel (i, j) in feature space is:

$$p_{i,j} = \text{sigmoid}(\mathbf{w} \cdot \mathbf{F}_{i,j,:} + b), \quad (8)$$

where \mathbf{w} is the weight in the logistic regression, b is the bias, and \cdot is the inner product of the two vectors \mathbf{w} and $\mathbf{F}_{i,j,:}$. \mathbf{w} and b are shared for different pixel positions (i, j) . $\mathbf{p} = (p_{i,j})$ is flattened into a one-dimensional vector as $\mathbf{p} = (p_1, p_2, \dots, p_K)$ corresponding to flattened patches (Q_1, Q_2, \dots, Q_K) , where K is the number of patches.

Thus, it is natural to use the maximum component of \mathbf{p} as the malignant probability of the mammogram \mathbf{I} :

$$\begin{aligned} p(y = 1|\mathbf{I}) &= \max\{p_1, p_2, \dots, p_K\}, \\ p(y = 0|\mathbf{I}) &= 1 - p(y = 1|\mathbf{I}). \end{aligned} \quad (9)$$

The cross entropy-based cost function can be defined as:

$$\mathcal{L} = - \sum_{n=1}^N \log(p(y = y_n|\mathbf{I}_n)), \quad (10)$$

where N is the total number of mammograms, and $y_n \in \{0, 1\}$ is the true label of malignancy for mammogram \mathbf{I}_n in the training. Typically, a mammogram dataset is imbalanced, where the proportion of positive mammograms is much smaller than negative mammograms, about 1/5 for the INbreast dataset. We therefor introduce a weighted loss:

$$\mathcal{L} = - \sum_{n=1}^N w(y_n) \log(p(y = y_n|\mathbf{I}_n)), \quad (11)$$

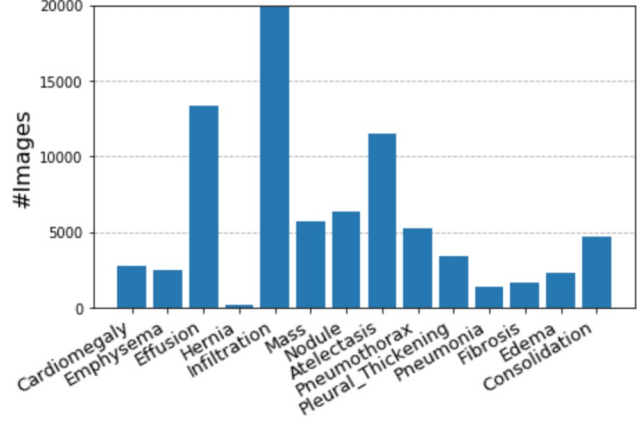


Figure 3. Histogram of label frequencies on ChestX-ray14 dataset. The ChestX-ray14 dataset is imbalanced.

where $w(c) = \frac{N}{\sum_{n=0}^N \mathbb{I}(y_n=c)}$ and $\mathbb{I}(\cdot)$ is an indicator function for y_n being label c .

4.3.2 Multi-Instance Multi-Label Learning for Chest X-Rays

In our DGFNs for Chest X-Rays dataset, we define a fourteen-dimensional label vector $\mathbf{y}_n = [y_n^1, y_n^2, \dots, y_n^C]$ for n^{th} image \mathbf{I}_n , where $C = 14$ with binary values, representing either the absence (0) or the presence (1) of a pathology. The y_n^c indicates the presence of an associated pathology in the n^{th} image where $c = \{1, 2, \dots, C\}$, while a zero vector $[0, 0, \dots, 0]$ represents the current x-ray image without any pathology. We consider each pathology as an independent multi-instance learning problem, which is the same as the mammogram classification, to solve the weakly supervised multi-label classification problem. We consider each patch as an instance and the problem can be formulated using equation (10). If there is no explicit priors on these labels, we can derive the loss function as:

$$\mathcal{L} = - \sum_{n=1}^N \sum_{c=1}^C \log(p(y = y_n^c|\mathbf{I}_n)), \quad (12)$$

where N is the total number of x-ray images on training set. As a multi-label problem, we treat all labels equally by defining C binary cross-entropy loss functions. As the dataset is highly imbalanced as illustrated in Figure 3, we incorporate weights within the loss function based on the label frequency:

$$\mathcal{L} = - \sum_{n=1}^N \sum_{c=1}^C w^c(y_n^c) \log(p(y = y_n^c|\mathbf{I}_n)), \quad (13)$$

where $w^c(0) = \frac{N}{\sum_{n=0}^N \mathbb{I}(y_n^c=0)}$ and $w^c(1) = \frac{N}{\sum_{n=0}^N \mathbb{I}(y_n^c=1)}$.

Table 1. The performance of DGFNs ($U=4$) with different V on INbreast dataset. The last line describes the average training time of one epoch with batch size of 128.

DGFNs	$V=1$	$V=2$	$V=3$	$V=4$	$V=5$
AUC (%)	79.28	80.72	81.67	82.05	82.53
Times (s)	2.96	4.03	5.87	6.85	7.92

5. Experiments

Our deformable Gabor feature networks (DGFNs) are evaluated on the two medical image datasets described above and CIFAR-10 dataset. To balance the performance and training complexity, we use traditional convolution in the first two blocks and deploy deformable Gabor feature convolution in the following high level features.

5.1. Experiments on the INbreast Dataset

To prepare the data we first remove the background of the mammograms in a pre-processing step using Otsu’s segmentation method [19]. We then resize the pre-processed mammogram to 224×224 . We use five-fold cross validation with three-fold training, one-fold validation and one-fold testing. We randomly flip the mammograms horizontally, rotate within 90 degree, shift them by 10% horizontally and vertically, and set a 50×50 box as 0 for data augmentation.

The proposed DGFNs employ AlexNet and ResNet18 as the backbones. We use the Adam optimization [5] algorithm with the initial learning rate of 0.0001 for both η_1 and η_2 and weight decay of 0.00005 in the training process. The learning rate decay is set to 10% for every 100 epochs and the total number of epochs for training is 1000.

Evaluation of U and V : We first perform the experiments on the hyper-parameters U and V to evaluate the additional channel number of orientations and scales. As shown in Table 1, given a fixed number of orientations ($U=4$), the average area under the ROC curve (AUC) increases from 79.28% to 82.53% when V is changed from 1 to 5. Additional evaluation on U shows that DGFN performs better when the number of orientations increases. In the following experiments, we choose $U=4$, $V=4$ to balance the training complexity and performance.

Deformation Robustness and Model Compactness:

To validate the networks robustness to deformation, we generate a deformable version of the dataset called INbreast-Deform by sampling 50 images with random scale and rotation for each test sample of the INbreast dataset. Scale factors are in the range $[0.5, 1.5]$, and rotation angles are in the range $[0, 2\pi)$. The results in Table 2 confirm that our DGFNs outperform CNNs even with fewer parameters by reducing the channel size of features in the network. When compared to CNNs with a similar number of parameters, DGFNs with kernel stage 8-16-32-64 and 16-32-64-128 obtain larger AUC improvements from 75.89% to 81.29% and

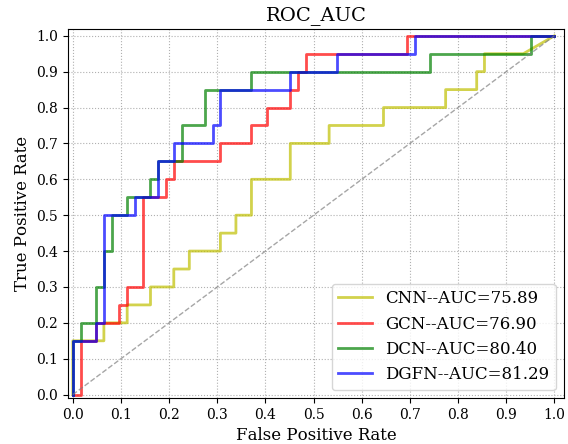


Figure 4. AUC comparison on INbreast-deform. All the networks are of similar model sizes with CNN 0.70M, GCN 0.70M, DCN 0.83M and DGFN 0.98M.

Table 2. Comparisons among CNNs, GCNs, DCNs and DGFNs on INbreast-Deform.

Backbone	Kernel Stages	AUC (%)	#Params (M)
ResNet18	16-32-64-128	75.89	0.70
	32-64-128-256	78.26	2.80
ResNet18 (GCNs)	8-16-32-64	76.90	0.70
	16-32-64-128	79.16	2.80
ResNet18 (DCNs)	16-32-64-128	80.40	0.83
	32-64-128-256	82.03	3.05
ResNet18 (DGFNs)	8-16-32-32	77.59	0.53
	8-16-32-64	81.29	0.98
	16-32-64-128	83.30	3.40

from 78.26% to 83.30% respectively. Figure 4 is the comparison of the average area under the ROC curve (AUC) of CNN, GCN, DCN and DGFN with similar sizes around 0.70-0.98M. DGFNs also achieve better performance than baseline methods including GCNs and DCNs. Thus DGFN enhances the robustness to spatial variations widely existing in biomedical images and largely reduces the complexity and redundancy of the network.

On the INbreast dataset, we combine DGFN with the multi-instance loss explained in section 4.3.1. As shown in Figure 5, our designed method can extract features and pinpoint the malignant region effectively. DGFNs with AlexNet and ResNet18 are compared with previous state-of-the-art approaches based on sparse multi-instance learning (Sparse MIL) [33]. As shown in Table 3, DGFNs have enhanced representative ability and achieve better AUC than previous approaches.

5.2. Experiments on the ChestX-ray14 Dataset

We resize the x-ray images from 1024×1024 to 224×224 to reduce the computational cost and normalize them

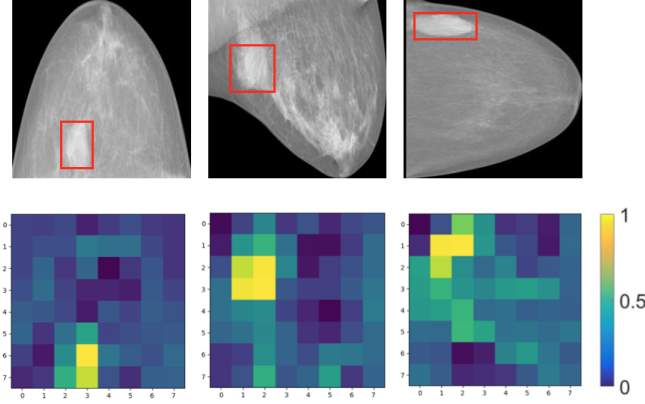


Figure 5. Malignant probability of each patch on INbreast dataset. The feature map has 8×8 patches.

Table 3. Comparisons on INbreast dataset. DGFN with ResNet18 yields the best performance.

Methods	Acc (%)	AUC (%)
AlexNet+Label Assign. MIL [33]	84.16	76.90
AlexNet+ DGFN+ MIL	86.22	78.12
ResNet18+ DGFN+ MIL	88.61	82.19
Pretrained AlexNet+Sparse MIL [33]	90.00	85.86
Pretrained AlexNet+ DGFN + MIL	91.34	87.22
Pretrained ResNet18 + DGFN + MIL	93.18	88.05

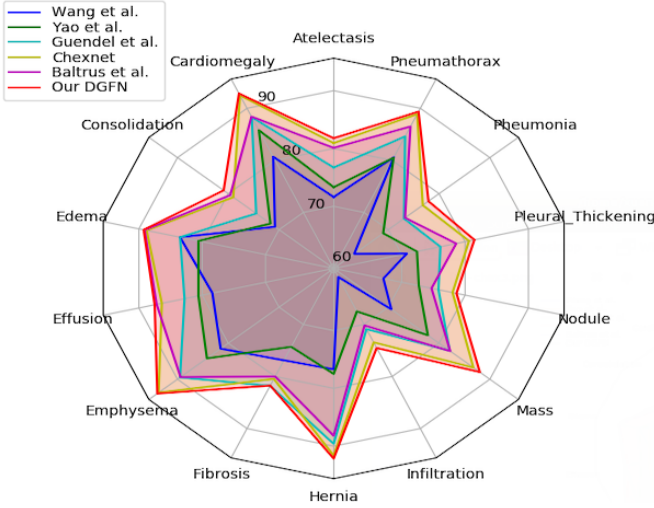


Figure 6. AUC (%) comparisons of our best model with state-of-art methods on ChestX-ray14 dataset.

based on the mean and standard deviation of images from the ImageNet training set [4]. In our experiments, we employ a DenseNet121 [12] as the backbone of our DGFN on ChestX-ray14 dataset. We resize the images to 224×224 and further augment the training data with random rotation and horizontal flipping. During training we use stochastic gradient descent (SGD) with momentum 0.9 and batch size 16. We use initial learning rates of 0.001 that are decayed

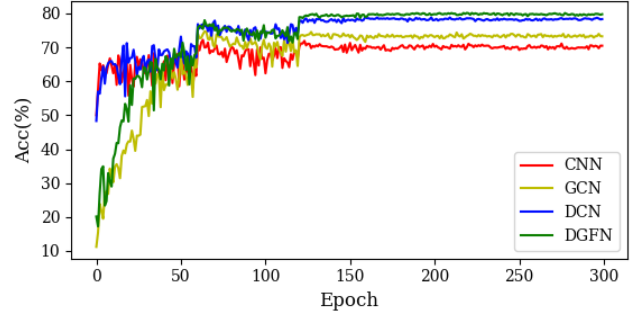


Figure 7. Comparisons of accuracy on CIFAR-10-Noisy. Note that the four models are of similar size with CNN 2.80M, GCN 2.80M, DCN 3.05M and DGFN 3.40M.

by a factor of 10 each time when the validation loss has no improvement.

We used the official split released by Wang *et al.* [27] with 70% training, 20% testing and 10% validation. While Yao *et al.* [30] and Chexnet [21] randomly split the dataset and ensure that there is no patient overlap between the splits. Yao *et al.* [30] noted that there is insignificant performance difference with different random splits. Thus it is a fair comparison. We divide the compared methods into Fine-Tune (FT) and Off-The-Shelf (OTS) based on whether it used additional data for training. Guendel *et al.* [11] used another fully annotated dataset-PLCO Dataset [9] to facilitate training. While our DGFN and other comparable fine-tuned methods [21, 27, 1] are initialized with ImageNet. Table 4 demonstrates that among the group labeled fine-tune, DGFN with DenseNet121 outperforms [21, 27, 1] on all fourteen pathologies from the ChestX-ray14 dataset. Among the group labeled off-the-shelf, DGFN achieves average AUC of 78.39% and performs better on 11 out of 14 pathologies than other methods [30, 1]. Figure 6 illustrative effectiveness of DGFN to enhance variant representations, which is potentially of great help on automated biomedical image analysis.

5.3. Experiments on the CIFAR-10 Dataset

To verify the effectiveness of DGFN on the natural image dataset, we conduct extensive experiments on CIFAR-10 as well as CIFAR-10 with noise. We generate a noisy version of CIFAR-10 called CIFAR-10-Noisy by replacing the pixel value with 255 at a probability of 1% percentage to test the network’s robustness to random Gaussian noise. We train on CIFAR-10 with random flipping and crop as augmentation. We test on CIFAR-10 and CIFAR-10-Noisy respectively. We use ResNet18 as the backbone and use SGD optimization with the initial learning rates as 0.05. The batch size is set as 128 and the total number of training epochs is 300. Figure 7 is the comparison of test accuracy on CIFAR10-Noisy with CNN, GCN, DCN and DGFN of similar sizes. Table 5 shows that the proposed DGFNs

Table 4. AUC (%) comparisons of DGFN with Off-The-Shelf (OTS) and Fine-Tune (FT) state-of-art methods on ChestX-ray14 dataset. Bold text emphasizes the highest value among each group.

Pathology	Off-The-Shelf			Fine-Tune				
	Yao et al. (2017)	Baltruschat et al. (2019)	DGFN (Ours)	Wang et al. (2017)	Guendel et al. (2018)	Chexnet (2018)	Baltruschat et al. (2019)	DGFN (Ours)
Atelectasis	73.3	73.2	78.04	71.6	76.7	80.94	80.1	81.78
Cardiomegaly	85.8	75.9	89.01	80.7	88.3	92.48	88.4	92.84
Consolidation	71.7	75.3	79.09	70.8	74.5	79.01	79.6	80.91
Edema	80.6	85.7	87.21	83.5	83.5	88.78	89.1	89.25
Effusion	80.6	80.6	86.89	78.4	82.8	86.38	87.2	87.51
Emphysema	84.2	79.8	81.96	81.5	89.5	93.71	89.4	93.97
Fibrosis	74.3	73.9	76.08	76.9	81.8	80.47	80.0	81.75
Hernia	77.5	81.9	77.83	76.7	89.6	91.64	88.2	92.15
Infiltration	67.5	67.0	68.49	60.9	70.9	73.45	70.2	74.52
Mass	77.8	68.6	76.32	70.6	82.1	86.76	82.2	88.03
Nodule	72.7	66.5	67.19	67.1	75.8	78.02	74.7	78.65
Pleural_Thickening	72.4	70.8	73.32	70.8	76.1	80.62	78.6	81.47
Pneumonia	69.0	68.3	72.83	63.3	73.1	76.80	73.3	77.91
Pneumothorax	80.5	79.1	83.17	80.6	84.6	88.87	86.5	89.36
Average	76.1	74.8	78.39	73.8	80.7	84.17	82.0	85.01

Table 5. Comparisons among CNNs, GCNs, DCNs and DGFNs on CIFAR-10 and CIFAR-10-Noise.

Methods	Kernel Stages	Acc (%)	Acc with noise (%)	#Params (M)
ResNet18	32-64-128-256	90.74	70.72	2.80
ResNet18	8-16-32-64	88.3	72.81	0.70
(GCNs)	16-32-64-128	89.37	74.69	2.80
ResNet18	16-32-64-128	88.92	74.30	0.83
(DCNs)	32-64-128-256	89.79	78.96	3.05
ResNet18	8-16-32-64	89.59	76.75	0.98
(DGFNs)	16-32-64-128	91.03	80.12	3.40

outperform the baseline on CIFAR-10-Noise. With a similar number of parameters, DGFN with kernel stage 16-32-64-128 achieves a 2% accuracy improvement beyond DCN, demonstrating its own superior robustness to random Gaussian noise common on natural images.

6. Conclusion

We have presented a deformable Gabor feature network (DGFN) to improve the robustness and interpretability for weakly supervised biomedical image classification. DGFN integrates adaptive Gabor filters into deformable convolutions, thus sufficiently characterizes spatial variations in objects and extracts discriminative features for various categories. Experiments show the DGFN is resistant to Gaussian noise and the architecture is both efficient and compact. DGFN is easily integrated into multi-instance, multi-label learning to facilitate the classification of biomedical image with great variations of sizes and shapes of the lesions. Extensive experiments demonstrate the effectiveness of DGFNs on both the INbreast dataset and the ChestX-ray14 dataset.

Acknowledgements

Baochang Zhang is the corresponding author. This study was supported by Grant NO.2019JZZY011101 from the Key Research and Development Program of Shandong Province to Dianmin Sun.

References

- [1] Ivo M. Baltruschat, Hannes Nickisch, Michael Grass, Tobias Knopp, and Axel Saalbach. Comparison of deep learning approaches for multi-label chest x-ray classification. *Scientific Reports*, 9(6381), 2019.
- [2] Hakan Bilen and Andrea Vedaldi. Weakly supervised deep detection networks. In *IEEE Conference of Computer Vision and Pattern Recognition*, pages 2846–2854, Las Vegas, NV, USA, 2016.
- [3] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *The IEEE International Conference on Computer Vision*, pages 764–773, 2017.
- [4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.
- [5] Jimmy Ba Diederik P. Kingma. Adam: A method for stochastic optimization. In *International Conference for Learning Representations*, 2015.
- [6] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115, 2017.
- [7] Dennis Gabor. Theory of communication: The analysis of information. *Journal of the Institution of Electrical Engineers: Radio and Communication Engineering*, 93(26):429–441, 1946.

- [8] Maryellen L Giger, Nico Karssemeijer, and Julia A Schnabel. Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. *Annual review of biomedical engineering*, 15:327–357, 2013.
- [9] John K. Gohagan, Philip C. Prorok, Richard B. Hayes, and Barnett-S. Kramer. The prostate, lung, colorectal and ovarian (plco) cancer screening trial of the national cancer institute: history, organization, and status. *Controlled Clinical Trials*, 21:251S–272S, 2000.
- [10] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [11] Sebastian Guendel, Sasa Grbic, Bogdan Georgescu, Kevin Zhou, Ludwig Ritschl, Andreas Meier, and Dorin Comaniciu. Learning to recognize abnormalities in chest x-rays with location-aware dense networks. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 757–765, 2018.
- [12] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.
- [13] Maximilian Ilse, Jakub M Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International Conference on Machine Learning*, 2018.
- [14] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. In *Conference on Neural Information Processing Systems*, 2015.
- [15] Bogdan Kwolek. Face detection using convolutional neural networks and gabor filters. In *International Conference on Artificial Neural Networks*, pages 551–5566, 2005.
- [16] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.
- [17] Shangzhen Luan, Chen Chen, Baochang Zhang, Jungong Han, and Jianzhuang Liu. Gabor convolutional networks. *IEEE Transactions on Image Processing*, 27(9):4357–4366, 2018.
- [18] Inês C Moreira, Igor Amaral, Inês Domingues, António Cardoso, Maria João Cardoso, and Jaime S Cardoso. Inbreast: toward a full-field digital mammographic database. *Academic radiology*, 19(2):236–248, 2012.
- [19] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.
- [20] George Papandreou, Liang-Chieh Chen, Kevin Murphy, and Alan L. Yuille. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In *International Conference on Computer Vision*, pages 1742–1750, 2015.
- [21] Pranav Rajpurkar, Jeremy Irvin, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.
- [22] Syed Shakib Sarwar, Priyadarshini Panda Panda, and Kaushik Roy. Gabor filter assisted energy efficient fast learning convolutional neural networks. In *IEEE/ACM International Symposium on Low Power Electronics and Design*, pages 1–6, 2017.
- [23] Liangchen Song, Cheng Wang, Lefei Zhang, Bo Du, Qian Zhang, Chang Huang, and Xinggang Wang. Unsupervised domain adaptive re-identification: Theory and practice. *Pattern Recognition*, 102:107173, 2020.
- [24] Liangchen Song, Yonghao Xu, Lefei Zhang, Bo Du, Qian Zhang, and Xinggang Wang. Learning from synthetic images via active pseudo-labeling. *IEEE Transactions on Image Processing*, 2020.
- [25] C Varela, S Timp, and N Karssemeijer. Use of border information in the classification of mammographic masses. *Physics in Medicine and Biology*, 51(2):425, 2006.
- [26] Fang Wan, Pengxu Wei, Jianbin Jiao, Zhenjun Han, and Qixiang Ye. Min-entropy latent model for weakly supervised object detection. *The IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10), 2019.
- [27] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *CVPR*, pages 3462–3471, 2017.
- [28] Jialian Wu, Liangchen Song, Tiancai Wang, Qian Zhang, and Junsong Yuan. Forest r-cnn: Large-vocabulary long-tailed object detection and instance segmentation. In *ACM International Conference on Multimedia*, pages 1570–1578, 2020.
- [29] Jiajun Wu, Yinan Yu, Chang Huang, and Kai Yu. Deep multiple instance learning for image classification and auto-annotation. In *CVPR*, pages 3460–3469, 2015.
- [30] Li Yao, Eric Poblentz, et al. Learning to diagnose from scratch by exploiting dependencies among labels. *Computing Research Repository*, 1710.10501, 2017.
- [31] Zhuoyao Zhong and Lianwen Jin. High performance offline handwritten chinese character recognition using googlenet and directional feature maps. In *International Conference on Document Analysis and Recognition*, pages 846–850, 2015.
- [32] Wentao Zhu, Yufang Huang, Liang Zeng, Xuming Chen, Yong Liu, Zhen Qian, Nan Du, Wei Fan, and Xiaohui Xie. Anatomynet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical physics*, 2018.
- [33] Wentao Zhu, Qi Lou, Yeeleng Scott Vang, and Xiaohui Xie. Deep multi-instance networks with sparse label assignment for whole mammogram classification. In *MICCAI*, pages 603–611, 2017.
- [34] Wentao Zhu, Yeeleng S Vang, Yufang Huang, and Xiaohui Xie. Deepem: Deep 3d convnets with em for weakly supervised pulmonary nodule detection. In *MICCAI*, 2018.