Disentangling Disease-related Representation from Obscure for Disease Prediction

Churan Wang [1 2] **Fei Gao** [3] **Fandong Zhang** [1] **Fangwei Zhong** [4] **Yizhou Yu** [5 6] **Yizhou Wang** [2 7]

Abstract

Disease-related representations play a crucial role in image-based disease prediction such as cancer diagnosis, due to its considerable generalization capacity. However, it is still a challenge to identify lesion characteristics in obscured images, as many lesions are obscured by other tissues. In this paper, to learn the representations for identifying obscured lesions, we propose a disentanglement learning strategy under the guidance of alpha blending generation in an encoder-decoder framework (DAB-Net). Specifically, we take mammogram mass benign/malignant classification as an example. In our framework, composite obscured mass images are generated by alpha blending and then explicitly disentangled into disease-related mass features and interference glands features. To achieve disentanglement learning, features of these two parts are decoded to reconstruct the mass and the glands with corresponding reconstruction losses, and only disease-related mass features are fed into the classifier for disease prediction. Experimental results on one public dataset DDSM and three in-house datasets demonstrate that the proposed strategy can achieve state-of-the-art performance. DAB-Net achieves substantial improvements of 3.9%∼4.4% AUC in obscured cases. Besides, the visualization analysis shows the model can better disentangle the mass and glands in the obscured image, suggesting the effectiveness of our solution in exploring the hidden characteristics in this challenging problem.

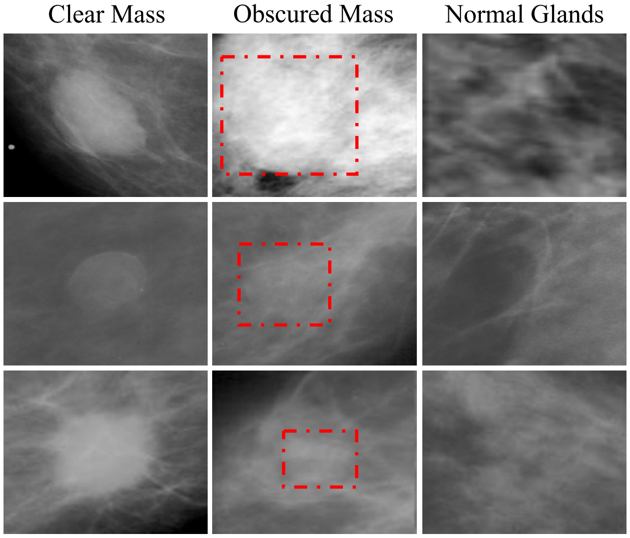

Figure 1. Masses with different obscured degrees. Column1: the mass is clearer when overlaid by fatty tissues; Column2: masses (marked by the red rectangles) are hidden in glands due to that the dense glands and the mass both look white on mammograms; Column3: the normal glands with diverse appearances and density, which can make masses with various obscured degrees.

1. Introduction

For disease benign/malignant diagnosis, exploring the disease-related representations is essential to the clinical practice. It can not only help to promote trustworthiness from patients but also to provide interpretability for clinicians (Wang et al., 2021a). However, in clinical practice, a considerable proportion of lesions are **obscured by other normal tissues** (Gassert et al., 2021; Diekmann & Bick, 2007) such as glandular and fibrous tissues, especially in the imaging modality of X-ray with the principle of projection overlay imaging as shown in the 2nd column in Fig. 1. As a result, for the task of image-based lesions (Dai et al., 2018; Liu et al., 2019) or disease characteristics identification (Li et al., 2018; Lee et al., 2019; Chen et al., 2020), a common obstacle is formed, which makes lesion attributes more difficult to be recognized. For this problem, the key is to find a way to **eliminate the effect of redundant content (interference of other tissues) and mine the essential content (disease-related features) hidden in the image**.

[1] Center for Data Science, Peking University, Beijing, China [2] Center on Frontiers of Computing Studies, School of Computer Science, Peking University, Beijing,China [3] School of Computer Science, Peking University, Beijing, China [4] School of Artificial Intelligence, Peking University, Beijing, China [5] AI lab, Deepwise Healthcare, Beijing, China [6] Department of Computer Science, The University of Hong Kong, Hong Kong [7] Inst. for Artificial Intelligence, Peking University, Beijing, China.
Correspondence to: Fangwei Zhong <zfw1226@gmail.com>.

Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022. Copyright 2022 by the author(s).

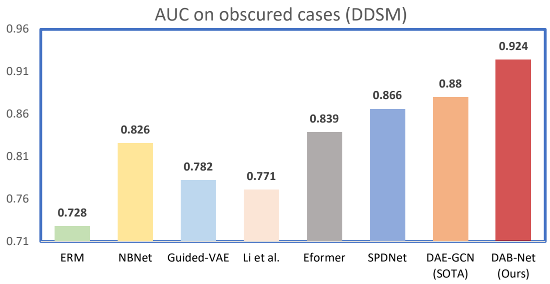

Figure 2. The results listed in the table show the considerable superiority of our method on obscured cases in mammogram benign/-malignant classification.

With the advances of computer vision, a number of methods in natural images processing, particularly for image restore (Cheng et al., 2021; Wang et al., 2021b; Luthra et al., 2021; Wu et al., 2021; Yi et al., 2021), are designed to remove the redundant noise (such as haze, rain, and *etc.*) and enhance the main content in the image. Intuitively, they provide a possible solution to address this problem with supervised paired data. However, most of these methods focus on learning the distribution of image content or redundant content rather than the hidden structured relationship between image content and noise. Moreover, as the patterns of tissues/glands are diverse and complex, it is challenging to precisely model the redundant content. Therefore, it remains difficult for these methods to explore disease-related features from dense glands.

In clinical diagnosis, doctors usually observe obscured disease-related features by *disentangling the structural relationships between lesions and glands empirically*, since obscured patterns can be diverse and the surrounding glands can be kind of dense. Taking the breast mammogram (McKinney et al., 2020; Lotter et al., 2021; Rodriguez-Ruiz et al., 2019) as an example, it is reported that dense glands are common among women, especially more than half of their breasts have dense glands in Asian women (D'Orsi et al., 2018). As shown in the 1st and the 2nd column in Fig.1, masses can be hidden behind the breast glands in varying degrees. The 3rd column shows different types of normal glands (without lesions). The density and distribution of breast glands determine the obscured degree and consequently make the identification of disease-related features with different difficulties. To disentangle disease-related features, the recent method (Wang et al., 2021a) designs a disentangle mechanism guided by attribute learning using Graph Convolutional Network (GCN). However, it remains limitations in this method for obscured lesions, as the lesion attributes are frequently hidden in the dense glands.

Motivated by the diagnosis prior to clinical practice, we propose a Network (DAB-Net) to **D**isentangle disease-related representation from obscure with **A**lpha **B**lending. It can better disentangle the obscured mass from the image and learn the disease-related features for benign/malignant classification. To realize it, we argue that the key is to design a suitable learning mechanism for the disentanglement of masses and glands. We build a mechanism based on *composite* and *disentanglement* for reconstruction. Since tissues grow naturally and are irreversible for re-scanning, there is no supervision data for de-obscure learning in the real clinical scenario. Thus, compositing obscured masses is firstly necessary for disentanglement learning. To *composite obscured mass*, we employ alpha blending (Wallace, 1981), a classic image processing method, to mix the clear mass patch and the gland patch into an image without adding any additional learning parameter. And then the composite images are encoded as composite features and disentangled into two hidden factors (h_m for masses, h_g for surrounding glands). To achieve *disentanglement*, the two hidden factors are then respectively decoded and constrained by reconstructed loss supervised by corresponding clear mass patches and gland patches. Moreover, only h_m is fed into classification layer for learning better disease-related features. In this way, the disease-related mass features and glands features are disentangled. Note that the real images are also fed into DAB-Net during training, but only used for the training of the malignancy classification instead of reconstruction, as the lack of disentangled ground truth. In this way, our DAB-Net can essentially learn to model the structural relationship between masses and glands, instead of just ostensibly reducing the redundant information as in previous works (Cheng et al., 2021; Wang et al., 2021b; Luthra et al., 2021; Wu et al., 2021; Yi et al., 2021).

To verify the utility and effectiveness of our DAB-Net, we conduct experiments on the mammogram mass benign/malignant classification task, a common and important medical problem (D'Orsi et al., 2018). Specifically, we apply our model on Digital Database for Screening Mammography (DDSM) (Bowyer et al., 1996)) and three in-house datasets in mammogram mass benign/malignant diagnosis. The Area Under the Curve (AUC) of receiver operating characteristic (ROC) results on obscured cases show the superiority of our DAB-Net, 4.4% improvement over the recent state-of-the-art (SOTA) method shown in Fig.2. We also calculate the AUC results on the whole dataset. Without any attribute annotations, DAB-Net still achieves comparable results with the SOTA method. When combining with the SOTA method (Wang et al., 2021a), ours can further get 1.3% to 2.9% improvements on AUC.

In summary, our contributions are mainly three-fold: **a)** we propose Disease-related Representation Disentangling Network with Alpha Blending (DAB-Net) to learn the features of obscured lesions. **b)** We develop a classification method based on disentanglement learning to achieve disease-related features learning in disease prediction. **c)** Our method achieves state-of-the-art performance on one public and three in-house datasets.

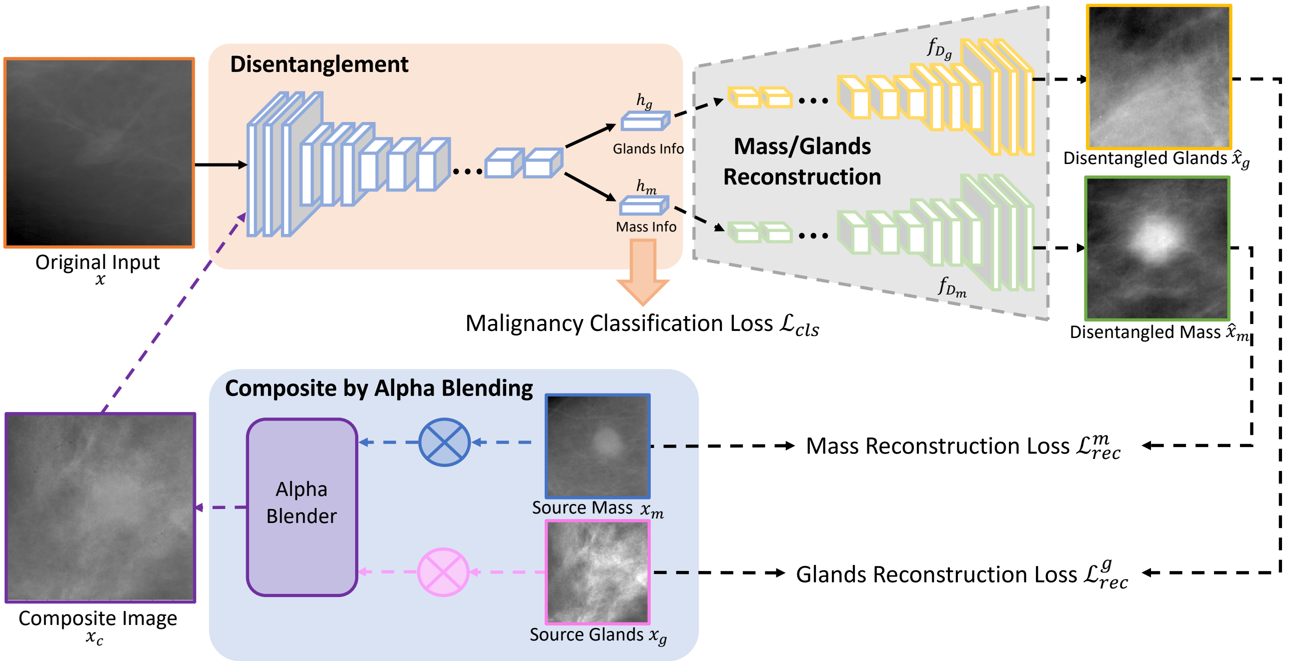
Original Input
x
Disentanglement
h_g
Glands Info
h_m
Mass Info
Malignancy Classification Loss $\mathcal{L}_{cls}$
f_{D_g}
Mass/Glands
Reconstruction
f_{D_m}
Disentangled Glands $\hat{x}_g$
Disentangled Mass $\hat{x}_m$
Composite by Alpha Blending
Alpha
Blender
Source Mass x_m
Source Glands x_g
Composite Image
x_c
Mass Reconstruction Loss $\mathcal{L}^m_{rec}$
Glands Reconstruction Loss $\mathcal{L}^g_{rec}$

Figure 3. The schematic overview of our method. The whole framework is based on the encoder-decoder structure and disentanglement mechanism. The composite obscured mass images are generated by alpha blending using the clear (margin) mass and glands images. Then the composite and real images are encoded into hidden factor h_c and disentangled into h_m, h_g. The two-branch decoders decode h_m, h_g of composite images to reconstruct the clear mass and glands with $\mathcal{L}^m_{rec}$ and $\mathcal{L}^g_{rec}$. The classifier conducts the disease prediction via $\mathcal{L}_{cls}$ by using h_m of both composite and real data. Note the dashed lines represent procedures only in the training phase, *i.e.* the alpha blending generation and two-branch reconstruction. The solid lines denote procedures both in training and inference phases.

2. Related Work

2.1. Learning Representation for Disease Prediction

For disease prediction (McKinney et al., 2020; Lotter et al., 2021; Rodriguez-Ruiz et al., 2019), feature representation learning from region of interests is very important (Zhao et al., 2018; Lei et al., 2020). Since during diagnosis, radiologists are more tending to identify the local features from the most discriminated regions rather than the full image. Take mammogram as an example, previous approaches used to learn feature representation for mammogram mass benign/malignant classification are roughly categorized into three classes: (i) the GAN-based methods, *e.g.,* Li *et al.* (Li et al., 2019). They propose an adversarial generation to augment training data for better prediction. However, lacking the guidance of medical knowledge descents their performance. (ii) the attribute-guided methods, *e.g.,* Chen *et al.* (Chen et al., 2019), ICADx (Kim et al., 2018). Attribute prior (D'Orsi et al., 2018) is considered into their methods and can provide little benefit for obscured mass classification. (iii) the disentangle-based methods, *e.g.,* Guided-VAE (Ding et al., 2020), DAE-GCN (Wang et al., 2021a). The disentangle mechanism provides effective disease-related representation learning when lesions are relatively visible. However, in obscured cases they lack the mechanism to disentangle lesion features under the dense glands. In summary, above all methods do not consider tackling the problem of hard cases of obscured lesions, which will cause their performances directly drop. Motivated by above, we attempt to realize a disentangling mechanism for learning representation about obscured mass. To do this, we composite obscured masses for better learning disease-related lesion features on obscured cases to further improve benign/malignant classification performance.

2.2. Disentanglement Learning

Disentangled representation learning aims to identify and disentangle the underlying explanatory factors (Bengio et al., 2013; Burgess et al., 2018; Ridgeway, 2016). Peng *et al.* (Peng et al., 2017) propose to disentangle identity and pose information in the latent feature space. Lin *et al.* (Lin et al., 2019) explore to obtain domain-specific features and domain-independent features by using feature disentanglement. Recent classification-related works such as Ding *et al.* (Ding et al., 2020) and Wang *et al.* (Wang et al., 2021a), attempt to learn a transparent representation by introducing the guidance to the latent variables in VAE/AE (Doersch, 2016). They prove the effectiveness of the disentangle mechanism in learning transparent features and disease-related features in malignancy classification. Inspired by this, in this work, we further enhance the disentangle mechanism with alpha blending.

3. Methodology

As analysis in Sec.1, it is crucial to learn disease-related representation for benign/malignant classification. However, the disease-related lesions are often obscured by other tissues, e.g., glands, which are of diverse patterns in appearance. Thus it is helpful but difficult to identify actual characteristics of obscured lesions for disease diagnosis. Motivated by this, we focus on discovering the de-obscured features hidden in the glands.

In this section, we introduce a network to **Disentangle Disease-related Representation from Obscure with Alpha Blending (DAB-Net)** for interpretable disease diagnosis. Fig. 3 outlines the overall network architecture of our framework. There are mainly three steps in the training phase. (i) (Sec. 3.2), we blend the clear mass patches and the normal gland patches based on alpha blending to obtain realistic composite obscured images. Note that the obscured masses, the clear masses and the normal gland are annotated by the doctors following the guidelines (D'Orsi et al., 2018). Through blending, we can provide paired images for disentanglement learning and learn the structural relationship between the masses and the glands in the obscured cases. (ii) (**Encoder** and **Two-Branch Decoders** in Sec. 3.3), composite images and original real training data are both fed into disentangle network to learn disentangled mass features and glands features. (iii) (**Disease Classifier** in Sec. 3.3), the disentangled mass features from all source data are fed into the classification layer for benign/malignant classification. The whole network is trained in an end-to-end manner. In the inference stage, DAB-Net not only can output a high-accuracy classification result for diagnosis but also provide disentangled features of the input image for better interpretability.

Next, we will briefly introduce the problem setup and notations at first, then provide the details of frameworks in the following, including 1) how to composite obscured images with alpha blending, and 2) how to build and train the DAB-Net to learn the disentangled disease-related representations from obscure.

3.1. Problem Setup & Notations.

Denote $x \in \mathcal{X}, y \in \mathcal{Y}$ respectively as the images and disease benign/malignant labels in the real dataset. We collect the training data $\{x_n, y_n\}_{n \in [N]}$ with $[N] := \{1, ..., N\}$. Our goal is to learn a prediction model $f : \mathcal{X} \to \mathcal{Y}$ based on disentangled disease-related features h_m from their surrounding glands h_g no matter the mass is clear or obscured for interpretable disease benign/malignant diagnosis.

3.2. Compositing by Alpha Blending

As for the particularity of medical tissues, de-obscured masses can not be obtained by re-scanning (Ancuti et al., 2019) or adjustment of imaging parameters (Anaya & Barbu, 2018; Plotz & Roth, 2017) as in natural scenes since the masses are obscured inherently (D'Orsi et al., 2018). This problem results in that the real data can not provide supervision in disentanglement learning. For learning disease-related features of obscured masses, we try to composite obscured data and get disentangle training supervision meanwhile. Blending mass and glands directly (Zhang et al., 2018) is rigid, and GAN-based methods (Li et al., 2020) have large additional parameters on blending generation. Thus we employ a parsimonious mechanism, alpha blending, for obscured mass generation. The comparison among different blending methods will be discussed in Sec. 4.4.

Alpha blending is a classic image processing technique that is more likely to be used in computer graphics and visualization (Wallace, 1981; Alashkar et al., 2017; Li et al., 2020). Generally, it combines two colors to produce a new blending color for transparency effects.Under our task, we attempt to utilize disentanglement learning to explore the disease-related features hidden in the glands. Alpha blending can provide a parsimonious composite way for obscured cases and can be used to construct a learning mechanism for disentanglement. The composite results are employed to learn to disentangle the masses under dense glands supervised by the inputs of alpha blending.

Specifically, the alpha blending combines two images and produces a new blended image. The value of the alpha channel is range from 0.0 to 1.0, representing the transparency of the pixel, , 0.0: fully transparent, 1.0: fully opaque. We manipulate the alpha values, A^{ij}, of the image by a 2D Gaussian distribution, as (Najibi et al., 2018). The Gaussian distribution is located at the center of the bounding box, which annotates the region of lesions. The co-variance of the Gaussian distribution is determined by the bounding box's height and width. It acts as spatial weighting to the region of lesions, so that the region can be highlighted in the composited image. To this end, the composited image x_c is generated by alpha blending acting on clear mass image x_m and glands image x_g as:

$$x_c^{ij} = x_g^{ij}(1 - A^{ij}) + x_m^{ij} A^{ij} \tag{1}$$

where x^{ij} is a vector of pixel values in position (i, j) and A^{ij} is a scalar alpha value of the same position.

3.3. Learning Disentanglement with Composite

To better explore the disease-related features, we introduce disentangle mechanism. Specifically, there contains the following components: (**Encoder**) an encoding network f_E to encode the whole image into two hidden factors, h_m denoting for de-obscured mass features and h_g denoting for obscuring glands features; (**Two-Branch Decoder**) two decoding networks for reconstruction of de-obscured masses and glands and named f_{D_m} and f_{D_g} respectively to construct disentangle learning; (**Disease Classifier** & **Training**) a disease classifier f_C to prompt the disentanglement of disease-related features and for final disease classification.

Encoder. In the encoder, both composite and original real data are as inputs. The input image is encoded into hidden factor h_c including mixed mass and glands features through the encoder f_E with parameters θ_E. To extract the effective disease-related features for benign/malignant classification, we disentangle h_c into two parts h_m and h_g with the same size to capture the features of de-obscured masses and glands. We construct three constraints for such disentanglement learning. The constraints are based on two image reconstruction branches and one disease classification branch respectively employing different hidden factors. The details of the training strategy will be introduced in the last paragraph of this section.

Two-Branch Decoder. In disentangle training, h_m and h_g aim to represent features of de-obscured masses and obscuring glands respectively. Thus, h_m and h_g need to have the ability of reconstructing de-obscured masses and glands. Based on above, two-branch decoder is designed, which contains decoder f_{D_m} for disentangled mass reconstruction: $\widehat{x}_m = f_{D_m}(h_m)$ with parameters θ_{D_m} and decoder f_{D_g} for disentangled glands reconstruction: $\widehat{x}_g = f_{D_g}(h_g)$ with parameters θ_{D_g}. The reconstruction losses $\mathcal{L}^m_{rec}$, $\mathcal{L}^g_{rec}$ of the mass and the glands are as:

$$\mathcal{L}^m_{rec}(\theta_E, \theta_{D_m}) := \|x_m - \widehat{x}_m(\theta_E, \theta_{D_m})\|_1 \qquad (2)$$

$$\mathcal{L}^g_{rec}(\theta_E, \theta_{D_g}) := \left\|x_g - \widehat{x}_g(\theta_E, \theta_{D_g})\right\|_1 \qquad (3)$$

Disease Classifier. The purpose of our design is to explore actual disease-related features from obscuring glands to boost the disease diagnosis performance. Therefore, the disentangled mass features h_m is also supervised by benign/-malignant classification label. h_m is used as the input to classification branch f_C for disease prediction with a binary cross-entropy loss $\mathcal{L}_{cls}(\theta_E, \theta_C)$.

Training Strategy. Finally, the whole network is trained in an end-to-end manner by combining $\mathcal{L}_{cls}$ and $\mathcal{L}^m_{rec}$, $\mathcal{L}^g_{rec}$ as follows:

$$\mathcal{L} = \mathcal{L}^m_{rec}(\theta_E, \theta_{D_m}) + \mathcal{L}^g_{rec}(\theta_E, \theta_{D_g}) + \mathcal{L}_{cls}(\theta_E, \theta_C) \quad (4)$$

Note that the training data consists of composite obscured mass images (by alpha blending) and real images (from real data). In the training stage, only composite obscured images participate in the computation of $\mathcal{L}^m_{rec}$ and $\mathcal{L}^g_{rec}$. That is because the reconstruction supervision is only provided from the inputs of alpha blending, while real data does not have such disentangled supervision. For the loss of $\mathcal{L}_{cls}$, all data among composite obscured images and real images are utilized. The classification labels of composite obscured images are the same as the disease labels of corresponding clear masses while blending.

4. Experiments

4.1. Datasets & Implementation

To evaluate the effectiveness of our model, take mammogram mass benign/malignant classification as an example, we consider both the public dataset *DDSM* (Bowyer et al., 1996) and three in-house datasets (*Inh1*, *Inh2*, *Inh3*). For fair comparison, all settings of datasets are the same as (Wang et al., 2021a), *i.e.,* the region of interests (ROIs) (malignant/benign masses) are cropped based on the annotations of radiologists the same as (Kim et al., 2018; Wang et al., 2021a) and the number of ROIs and patients and the data division of each dataset are the same as (Wang et al., 2021a), which are already shared by (Wang et al., 2021a). More details of each dataset we use are shown in Appendix. A.

We implement Adam to train our model. For a fair comparison, all methods are conducted under the same setting and share the same encoder backbone, *i.e.*, ResNet34 (He et al., 2016). Area Under the Curve (AUC) is used as evaluation metrics in image-wise. To remove the randomness, we run for ten times and report the average value of them. For disentangle learning, we first use only composite obscured data to pretrain for 500 epochs while the number of samples is the same as the real dataset. Then we add composite obscured data to original real data for classification training, and the number of added composite data is 30% of the original datasets. For implementation of compared baselines, we directly load the published codes of ERM (He et al., 2016), Chen *et al.* (Chen et al., 2019), NBNet (Cheng et al., 2021), Uformer (Wang et al., 2021b), SPDNet (Yi et al., 2021), AECRNet (Wu et al., 2021) and shared code of DAE-GCN (Wang et al., 2021a) during test; while we re-implement methods of Guided-VAE (Ding et al., 2020), Eformer (Luthra et al., 2021), ICADx (Kim et al., 2018) and Li *et al.* (Li et al., 2019) for lacking published source codes.

4.2. Comparison with Baselines

Compared Baselines. In this section, we conduct informative experiments to verify the effectiveness of DAB-Net. Firstly, we compare with several representation learning methods including SOTA patch level mammogram benign/-malignant classification method (He et al., 2016; Wang et al., 2021a), other related disentangle-based method (Ding et al., 2020) and the attribute-based method (Chen et al., 2019) that can be extended to our task. Secondly, GAN-based methods (Li et al., 2019; Kim et al., 2018) are also compared with our method. Thirdly, the SOTA algorithms related to image restore (denoising (Cheng et al., 2021; Wang et al., 2021b; Luthra et al., 2021), dehazing (Wu et al., 2021) and

Table 1. The AUC evaluation on public DDSM (Bowyer et al., 1996) and three in-house datasets. The first column notes the methods we compared. The second column represents the AUC on overall testing sets. We additionally report results on the *obscured cases* that appeared in the testing sets. Results based on our methods are **boldfaced** and the best results among baselines are underlined.

Methodology	AUC				AUC only on *obscured cases*			
	Inh1	Inh2	Inh3	DDSM	Inh1	Inh2	Inh3	DDSM
ERM (He et al., 2016)	0.888	0.847	0.776	0.847	0.739	0.707	0.630	0.728
Chen *et al.* (Chen et al., 2019)	0.924	0.878	0.827	0.871	0.790	0.748	0.669	0.777
Guided-VAE (Ding et al., 2020)	0.921	0.867	0.809	0.869	0.782	0.751	0.673	0.782
DAE-GCN (Wang et al., 2021a)	0.963	0.901	0.857	0.919	0.871	0.837	0.783	0.880
Li *et al.* (Li et al., 2019)	0.908	0.859	0.828	0.875	0.767	0.726	0.648	0.771
ICADx (Kim et al., 2018)	0.911	0.871	0.816	0.879	0.801	0.793	0.665	0.782
NBNet (Cheng et al., 2021)	0.912	0.875	0.824	0.877	0.839	0.821	0.749	0.826
Uformer (Wang et al., 2021b)	0.923	0.879	0.832	0.872	0.845	0.813	0.757	0.834
Eformer (Luthra et al., 2021)	0.928	0.883	0.838	0.875	0.849	0.815	0.760	0.839
SPDNet (Yi et al., 2021)	0.908	0.862	0.814	0.866	0.823	0.791	0.739	0.816
AECRNet (Wu et al., 2021)	0.911	0.870	0.826	0.870	0.846	0.818	0.752	0.825
DAB-Net **(Ours)**	**0.956**	**0.907**	**0.849**	**0.913**	**0.910**	**0.878**	**0.826**	**0.924**
DAB-Net(**Ours**) + (Chen et al., 2019)	**0.964**	**0.913**	**0.861**	**0.920**	**0.916**	**0.883**	**0.835**	**0.934**
DAB-Net(**Ours**) + (Ding et al., 2020)	**0.959**	**0.903**	**0.855**	**0.918**	**0.913**	**0.882**	**0.829**	**0.932**
DAB-Net(**Ours**) + (Wang et al., 2021a)	**0.976**	**0.930**	**0.878**	**0.943**	**0.921**	**0.891**	**0.847**	**0.945**

deraining (Yi et al., 2021)) are compared with our method to demonstrate that it is crucial to model and disentangle the structural relationship between image content and interference instead of only image content. Finally, integrating our strategy to the first group methods is also compared to indicate the extensibility and additional value.

These methods are briefly introduced as following: **a)** ERM (He et al., 2016) directly trains the classifier via ResNet34 by Empirical Risk Minimization (ERM); **b)** Chen *et al.* (Chen et al., 2019) achieves multi-label classification with GCN; **c)** Guided-VAE (Ding et al., 2020) also implements disentangle network but lacks the medical prior knowledge of attributes during learning; **d)** Li *et al.* (Li et al., 2019) improves performance by generating more benign/-malignant images via adversarial training; **e)** ICADx (Kim et al., 2018) proposes the adversarial learning method and additionally introduces shape/margins information for reconstruction; **f)** DAE-GCN (Wang et al., 2021a) also develops a disentanglement learning framework with graph neural network to boost benign/malignant classification performance **g)** NBNet (Cheng et al., 2021) proposes a denoising network with subspace projection; **h)** Uformer (Wang et al., 2021b) constructs a U-shaped Transformer for image restoration; **i)** Eformer (Luthra et al., 2021) additionally introduces an edge enhancement in medical image denoising; **j)** SPDNet (Yi et al., 2021) introduces a residue channel prior for rain removal; **k)** AECRNet (Wu et al., 2021) develops a compact dehazing network based on contrastive learning.

Results & Analysis. As shown in Tab. 1, the second to the fifth lines are the representation learning methods including the attribute-based and disentangle-based. The next two lines are the GAN-based methods. The eighth to the twelfth lines are related to image restore, such as denoising, deraining and dehazing. The final four lines are our method and our method combining different representation learning methods. We calculate AUC performance not only on overall testing sets (Tab. 1-AUC) but also on only obscured cases in the testing sets (Tab. 1-AUC only *on obscured cases*).

Specifically, Li *et al.* (Li et al., 2019) generate more data to improve the performance compared with ERM (He et al., 2016), but the generated data does not consider hard cases, *i.e.,* obscured masses and is distributed close to real clearer data which limits its diversity and performance. The advantage of Chen *et al.* (Chen et al., 2019) predominantly lies in the attributes modeling of GCN. ICADx (Kim et al., 2018) also uses the information of attributes and combines with GAN to make image enhancement. Nevertheless, GCN is also limited when mass characteristics are hidden due to lacking modeling obscured masses. Guided-VAE (Ding et al., 2020) can find disease-related features relying on the capability of disentanglement learning. By integrating attributes learning via GCN into disentanglement learning, DAE-GCN (Wang et al., 2021a) further boosts the performance of representation learning and helps to explore more powerful disease-related features. Whereas, it is still challenging to extract obscured mass features. Compared with the methods mentioned above, the critical point of our method is smart modeling of structural relationship of masses and glands utilizing disentanglement learning, which is lacking in those methods even with the powerful disentanglement learning and GCN. The results of only DAB-Net (the fourth-to-last line) show a substantial improvement

Table 2. Ablation Studies: overall AUC evaluation on public dataset DDSM (Bowyer et al., 1996) and three in-house datastes.

Alpha Blending	Disentangle	DAE-GCN	Inh1	Inh2	Inh3	DDSM
×	×	×	0.888	0.847	0.776	0.847
AD	×	×	0.886	0.843	0.771	0.836
Simple	✓	×	0.936	0.882	0.839	0.893
GAN-based	✓	×	0.950	0.906	0.847	0.902
✓	*No Mass Decoder*	×	0.921	0.878	0.818	0.869
✓	*No Glands Decoder*	×	0.925	0.884	0.824	0.873
✓	*One branch*	×	0.939	0.887	0.831	0.889
✓	✓	×	**0.956**	**0.907**	**0.849**	**0.913**
✓	✓	✓	**0.976**	**0.930**	**0.878**	**0.943**

over these representation learning methods except for DAE-GCN that uses extra attributes information effectively. And even compared with this strong competitor DAE-GCN, our method shows a comparable performance.

With regard to the methods related to image restore, the mass characteristics can be emphasized for disease prediction. The targets of denoising (Cheng et al., 2021; Wang et al., 2021b; Luthra et al., 2021), dehazing (Wu et al., 2021) and deraining (Yi et al., 2021) are kind of similar to our task, each of which can be generalized as interference removal, and we try to extend them to our task. In principle, compared with our proposed method, these methods show the same disadvantage of lacking medical prior knowledge, *i.e.* modeling the structural relationship between image content and interference.

The combination of our method with the representation learning methods mentioned above (Chen et al., 2019; Ding et al., 2020; Wang et al., 2021a) (the last four lines) demonstrate that our method can be easily embedded into representation learning frameworks and show considerable additional performance improvements which achieve SOTA performances.

Tab. 1-AUC only on *obscured cases* can further reveal the superiority of our method. Our method obtains state-of-the-art performance and it shows a large margin gap even compared with the strong competitor DAE-GCN which uses extra attributes. This proves the attribute representation learning fails in invisible or obscured data and our disentangle representation mechanism from obscure is effective. Moreover, combined with effective representation learning, our method can get further improvement. These results further demonstrate the tremendous advantage of our method on obscured masses due to the effective design.

4.3. Ablation Study

To verify the effectiveness of each component in our model, we evaluate some variant models. Tab. 2 shows the ablation study results on DDSM and three in-house datasets. Here are some interpretations for the variants: *AD* means simply adding generated data into original training data and increasing the number of inputs in ERM (He et al., 2016); *Simple* denotes using blending mechanism (Zhang et al., 2018) without using the alpha map as pixel weight; *No Mass Decoder* denotes the setting without mass decoder f_{D_m} and thus without loss of $\mathcal{L}^m_{rec}$. *No Glands Decoder* denotes the setting without glands decoder f_{D_g} and thus without loss of $\mathcal{L}^g_{rec}$; *One branch* denotes Mass Decoder and Glands Decoder share the same weights; *GAN-based* denotes using GAN-based method replace alpha blending for data generation (Li et al., 2020).

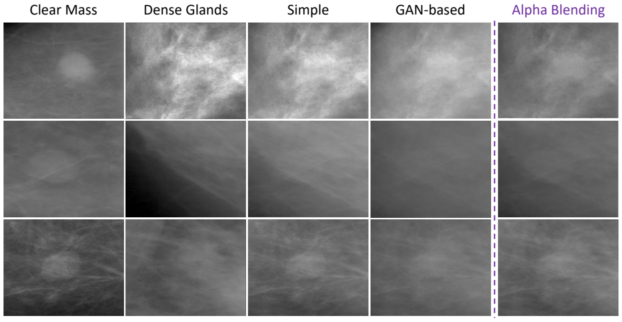

Figure 4. The blending results of three different methods. Each row indicates different instances. The first column shows the original clear masses from DDSM dataset (Bowyer et al., 1996) and the second column shows the dense glands. The third to the fifth columns show the blending results combining the first and the second columns by using simple add (Zhang et al., 2018), GAN-based method (Li et al., 2020) and alpha blending respectively.

The results demonstrate that alpha blending is more powerful than all other methods and the disentanglement learning of both masses and glands shows the best performance among different disentanglement settings. Specifically, the second row is to verify that the improvement of DAB-Net is not simply by adding more training data. It shows that *AD* even degrades performance with more training data due to a large amount of hard cases in training which is harmful to learning. The third and the fourth rows compared with the

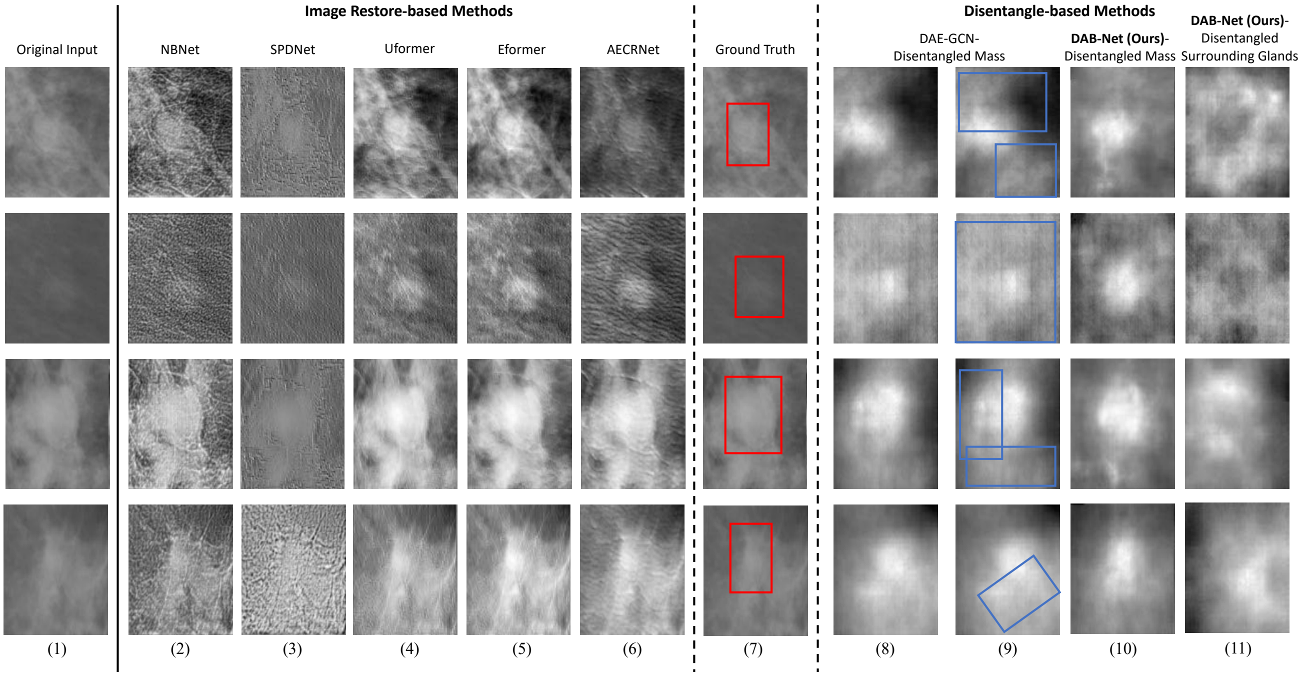

Figure 5. Restored/Disentangled masses visualization. Each row denotes different cases. The first column is original obscured masses from real data; the second to the sixth columns are restored masses by using recent SOTA image restore-based methods: NBNet (Cheng et al., 2021), SPDNet (Yi et al., 2021), Uformer (Wang et al., 2021b), Eformer (Luthra et al., 2021), AECRNet (Wu et al., 2021); the seventh column is the ground truth of masses location; the eighth to the tenth columns are the disentangled mass results by using disentangle-based methods. The eighth column is the results using the SOTA method DAE-GCN(Wang et al., 2021a). The redundant glands disentangled by DAE-GCN are marked by blue rectangles in the ninth column. The tenth and the eleventh columns are the masses and the surrounding glands disentangled from the obscured cases by using DAB-Net respectively. See Fig. 8 for more examples.

eighth row can investigate the effectiveness of alpha blending and suggest that alpha blending is more effective. The fifth and the sixth rows indicate only modeling of masses or glands is limited to explore the hidden mass features due to the diversity and complexity of masses and glands. The simultaneous modeling of masses and glands is the essence of our method. This result also implies the drawbacks of image restore-based methods. The seventh row indicates that the decoders of mass and glands are to learn different capabilities. Channels of decoders in *One Branch* are doubled to eliminate the effect of different learnable parameters. The last two rows demonstrate DAE-GCN integrating attributes modeling can bring additional performance improvements, which is also indicated in Tab. 1.

4.4. Visualization

To further evaluate the validity of our DAB-Net on the blending mechanism and the ability of exploring disease-related features from dense glands, we visualize the blending results (Fig. 4) and the explored disease-related results (Fig. 5).

Visualization of Blending. We visualize three different blending methods as shown in Fig. 4. Blending with simple add (Zhang et al., 2018) (the 3rd column) shows rigid results and the structural relationship between masses and glands is unreal. When glands are kind of dense, the mass is hard to identify and it is unusual challenging to learn disease-related features. When glands are clearer, simple add does not provide an effective occlusion. With weighting map around lesion, the adopted alpha blending (the 5th column) shows better fusion effect. GAN-based blending (Li et al., 2020) (the 4th column) shows the comparable results with ours alpha blending and achieve the purpose of occlusion. However, GAN-based method needs large parameters to learn as shown in Tab. 3 while our alpha blending do not need parameters during generation.

Table 3. Parameters of different blending methods.

Blending Methodology	Parameters
Simple (Zhang et al., 2018)	0M
GAN-based (Li et al., 2020)	100.848M
Alpha Blending(Ours)	0M

Visualization of Disease-related Features. We also visualize the explored disease-related results in Fig. 5. The explored results based on image restore-based methods (Cheng et al., 2021; Wang et al., 2021b; Luthra et al., 2021; Wu et al., 2021; Yi et al., 2021) are shown in the 2nd to the 6th columns. The results they restore are more like edge enhancement for all information or increased contrast of the image. Though such restore operation helps to provide partial features that are not easy to identify with the naked eye, they does not disentangle real disease-related features and it remains much other unnecessary redundant information. Based on disentangle mechanism, disentangle-based methods (Wang et al., 2021a) and ours are trying to explore only disease-related features for diagnosis in the 8th to the 11th columns. However, without considering to learn de-obscured features, DAE-GCN(Wang et al., 2021a) learns redundant glands in disease-related features as shown in the blue bounding boxes in the 9th column.

5. Conclusion

We propose a novel framework for Disentangling Disease-related Representation from Obscure with Alpha Blending (DAB-Net) for medical diagnosis. Obscured cases are common in clinical practice and limit the performance of recent image-based disease diagnosis methods. We design disentangle mechanism with alpha blending to help explore interpretable disentangling de-obscured lesion features. We evaluate our method on both public and in-house datasets for mammogram mass benign/malignant classification. Potential results demonstrate the effectiveness of our DAB-Net, especially on obscured cases. In the future, we will try to generalize our framework to other medical imaging problems such as lung cancer, liver cancer, and *etc*.

6. Acknowledgement

This work was supported by NSFC-62061136001, MOST-2018AAA0102004, China National Postdoctoral Program for Innovative Talents (No. BX2021008), National Key Research & Development Program (No. 2019YFB1404804, No. 2019YFC0118101), and ZheJiang Province Key Research Development Program (No. 2020C03073).

References

Alashkar, T., Jiang, S., Wang, S., and Fu, Y. Examples-rules guided deep neural network for makeup recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

Anaya, J. and Barbu, A. Renoir-a benchmark dataset for real noise reduction evaluation. *Journal of Visual Communication and Image Representation*, pp. 144–154, 2018.

Ancuti, C. O., Ancuti, C., Sbert, M., and Timofte, R. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *2019 IEEE International Conference on Image Processing*, pp. 1014–1018. IEEE, 2019.

Bengio, Y., Courville, A., and Vincent, P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35 (8):1798–1828, 2013.

Bowyer, K., Kopans, D., Kegelmeyer, W., Moore, R., Sallam, M., Chang, K., and Woods, K. The digital database for screening mammography. In *Third international Workshop on Digital Mammography*, volume 58, pp. 27, 1996.

Burgess, C. P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., and Lerchner, A. Understanding disentangling in β-vae. *arXiv preprint arXiv:1804.03599*, 2018.

Chen, H., Wang, Y., Zheng, K., Li, W., Chang, C.-T., Harrison, A. P., Xiao, J., Hager, G. D., Lu, L., Liao, C.-H., et al. Anatomy-aware siamese network: Exploiting semantic asymmetry for accurate pelvic fracture detection in x-ray images. In *European Conference on Computer Vision*, pp. 239–255. Springer, 2020.

Chen, Z.-M., Wei, X.-S., Wang, P., and Guo, Y. Multi-label image recognition with graph convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5177–5186, 2019.

Cheng, S., Wang, Y., Huang, H., Liu, D., Fan, H., and Liu, S. Nbnet: Noise basis learning for image denoising with subspace projection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4896–4906, 2021.

Dai, Y., Yan, S., Zheng, B., and Song, C. Incorporating automatically learned pulmonary nodule attributes into a convolutional neural network to improve accuracy of benign-malignant nodule classification. *Physics in Medicine & Biology*, 63(24):245004, 2018.

Diekmann, F. and Bick, U. Tomosynthesis and contrast-enhanced digital mammography: recent advances in digital mammography. *European Radiology*, 17(12):3086–3092, 2007.

Ding, Z., Xu, Y., Xu, W., Parmar, G., Yang, Y., Welling, M., and Tu, Z. Guided variational autoencoder for disentanglement learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7920–7929, 2020.

Doersch, C. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.

D'Orsi, C., Bassett, L., Feig, S., et al. Breast imaging reporting and data system (bi-rads). *Breast imaging atlas, 4th edn. American College of Radiology, Reston*, 2018.

Gassert, F. T., Urban, T., Frank, M., Willer, K., Noichl, W., Buchberger, P., Schick, R., Koehler, T., von Berg, J., Fingerle, A. A., et al. X-ray dark-field chest imaging: qualitative and quantitative results in healthy humans. *Radiology*, 301(2):389–395, 2021.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Kim, S. T., Lee, H., Kim, H. G., and Ro, Y. M. Icadx: interpretable computer aided diagnosis of breast masses. In *Medical Imaging 2018: Computer-Aided Diagnosis*, volume 10575, pp. 1057522. International Society for Optics and Photonics, 2018.

Lee, H., Yune, S., Mansouri, M., Kim, M., Tajmir, S. H., Guerrier, C. E., Ebert, S. A., Pomerantz, S. R., Romero, J. M., Kamalian, S., et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nature Biomedical Engineering*, 3(3):173–182, 2019.

Lei et al., Y. Shape and margin-aware lung nodule classification in low-dose ct images via soft activation mapping. *Medical Image Analysis*, 60:101628, 2020.

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

Li, X., Makihara, Y., Xu, C., Yagi, Y., and Ren, M. Gait recognition invariant to carried objects using alpha blending generative adversarial networks. *Pattern Recognition*, 105:107376, 2020.

Li, Z., Wang, C., Han, M., Xue, Y., Wei, W., Li, L.-J., and Fei-Fei, L. Thoracic disease identification and localization with limited supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8290–8299, 2018.

Lin, J., Chen, Z., Xia, Y., Liu, S., Qin, T., and Luo, J. Exploring explicit domain supervision for latent space disentanglement in unpaired image-to-image translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(4):1254–1266, 2019.

Liu, L., Dou, Q., Chen, H., Qin, J., and Heng, P.-A. Multitask deep model with margin ranking loss for lung nodule analysis. *IEEE Transactions on Medical Imaging*, 39(3): 718–728, 2019.

Lotter, W., Diab, A. R., Haslam, B., Kim, J. G., Grisot, G., Wu, E., Wu, K., Onieva, J. O., Boyer, Y., Boxerman, J. L., et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nature Medicine*, 27(2): 244–249, 2021.

Luthra, A., Sulakhe, H., Mittal, T., Iyer, A., and Yadav, S. Eformer: Edge enhancement based transformer for medical image denoising. *arXiv preprint arXiv:2109.08044*, 2021.

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., et al. International evaluation of an ai system for breast cancer screening. *Nature*, 577(7788): 89–94, 2020.

Najibi, M., Yang, F., Wang, Q., and Piramuthu, R. Towards the success rate of one: Real-time unconstrained salient object detection. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1432–1441. IEEE, 2018.

Peng, X., Yu, X., Sohn, K., Metaxas, D. N., and Chandraker, M. Reconstruction-based disentanglement for pose-invariant face recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1623–1632, 2017.

Plotz, T. and Roth, S. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition*, pp. 1586–1595, 2017.

Ridgeway, K. A survey of inductive biases for factorial representation-learning. *arXiv preprint arXiv:1612.05299*, 2016.

Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Broeders, M., Gennaro, G., Clauser, P., Helbich, T. H., Chevalier, M., Tan, T., Mertelmeier, T., et al. Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. *JNCI: Journal of the National Cancer Institute*, 111(9):916–922, 2019.

Wallace, B. A. Merging and transformation of raster images for cartoon animation. In *Proceedings of the 8th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 253–262, 1981.

Wang, C., Sun, X., Zhang, F., Yu, Y., and Wang, Y. Dae-gcn: Identifying disease-related features for disease prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 43–52. Springer, 2021a.

Wang, Z., Cun, X., Bao, J., and Liu, J. Uformer: A general u-shaped transformer for image restoration. *arXiv preprint arXiv:2106.03106*, 2021b.

Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., and Ma, L. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10551–10560, 2021.

Yi, Q., Li, J., Dai, Q., Fang, F., Zhang, G., and Zeng, T. Structure-preserving deraining with residue channel prior guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4238–4247, 2021.

Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*, 2018.

Zhao et al., W. 3d deep learning from ct scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas. *Cancer research*, 78(24):6881–6889, 2018.

A. Details about Each Dataset

The details of training/validating/testing data we use for experiments are shown in Fig. 7. And the details of the number of the obscured/non-obscured masses in each dataset we use are shown in Fig. 6. All settings are the same as (Wang et al., 2021a) for fair comparison.

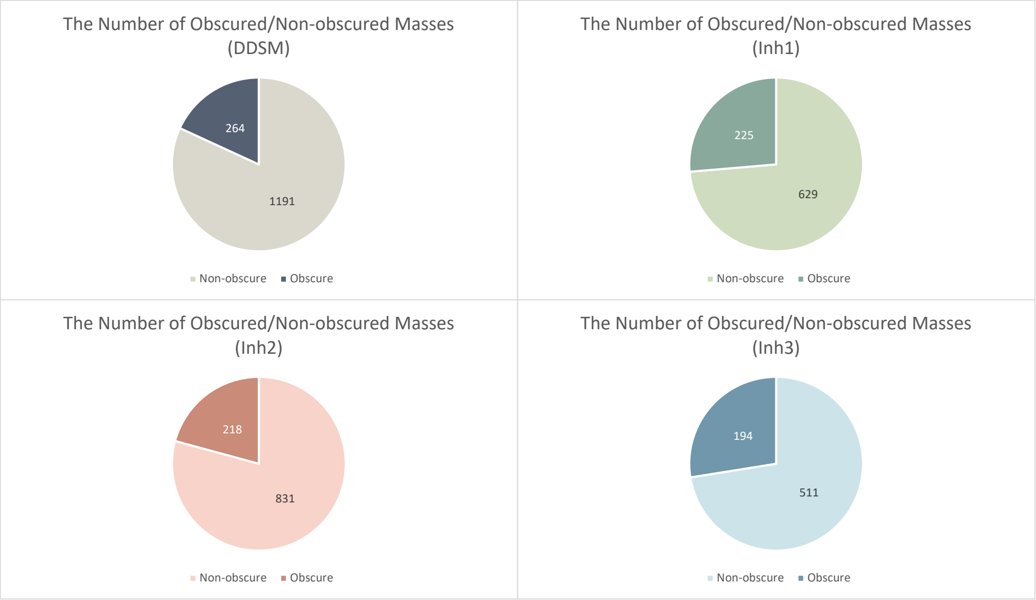

Figure 6. The number of Obscured/Non-obscured Masses in each dataset. Each pie chart denotes a each dataset.

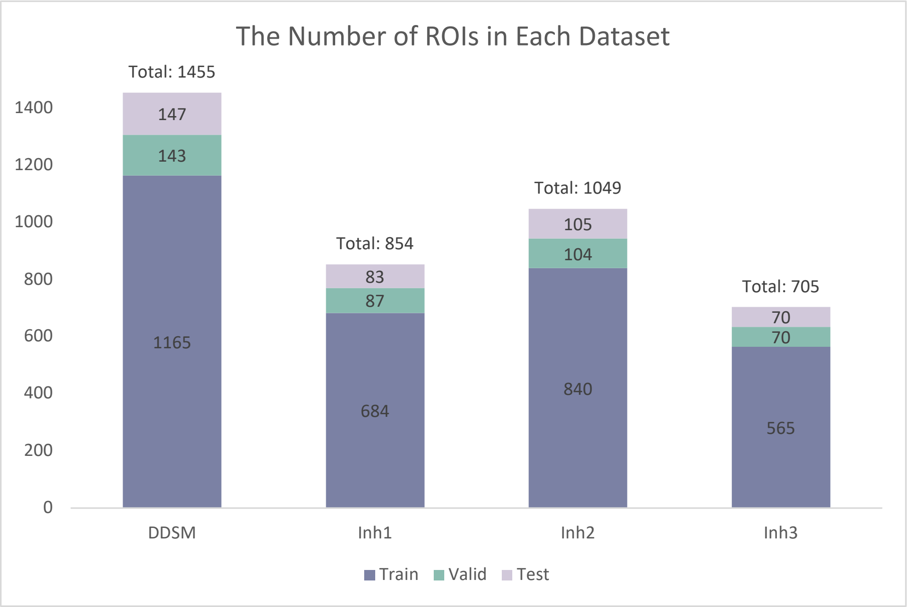

Figure 7. The number of ROIs in each dataset. Each histogram denotes a each dataset and each dataset is divided into training, validating and testing by 8:1:1.

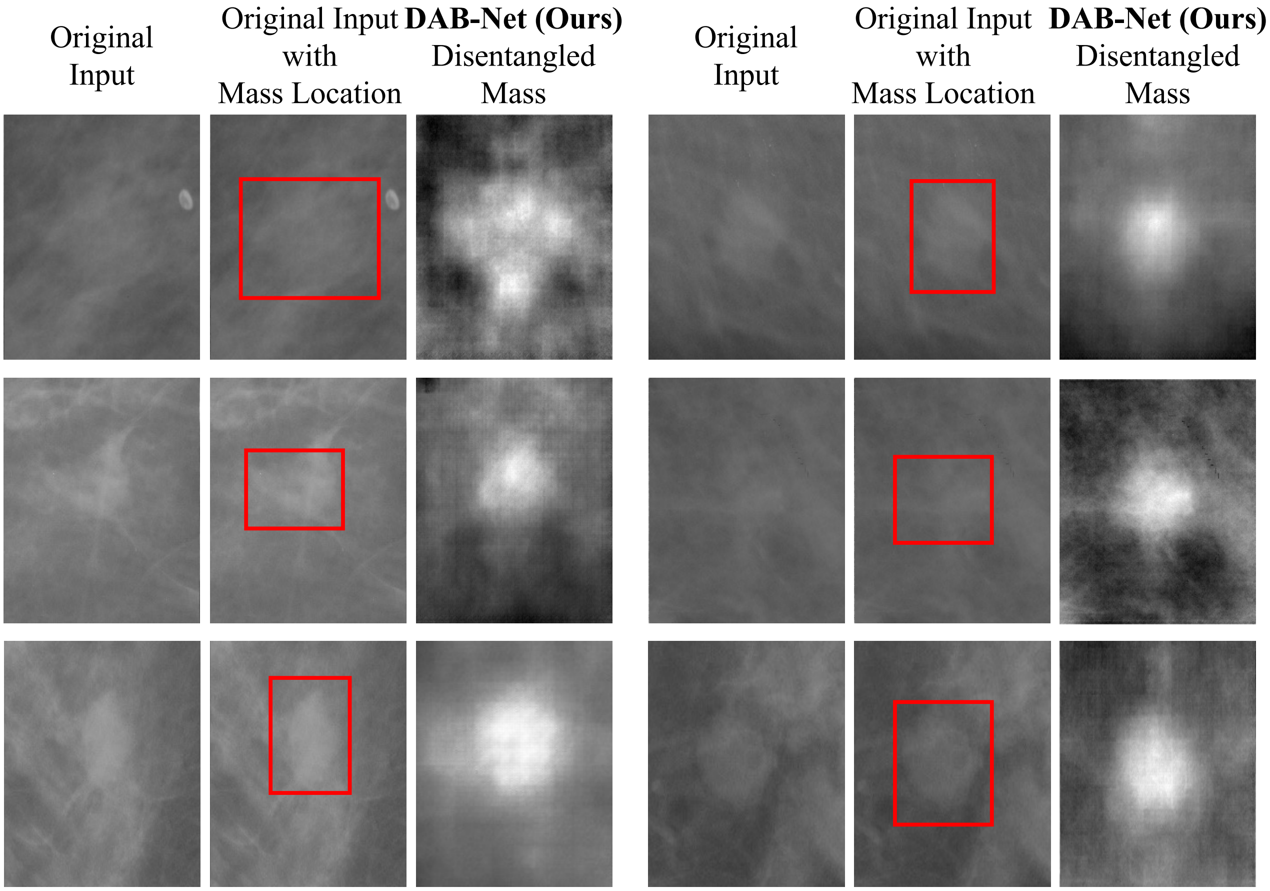
Original Input
Original Input with Mass Location
DAB-Net (Ours) Disentangled Mass
Original Input
Original Input with Mass Location
DAB-Net (Ours) Disentangled Mass

Figure 8. Disentangled masses visualization. Each row denotes a different case. The 1st column: the original obscured masses from real data; the 2nd column: the ground truth of masses location; the 3rd column: the masses disentangled from the obscured cases by DAB-Net.

B. More Visualization

We visualize more disentangled masses from the obscured images by our DAB-Net in Fig. 8. Each case contains three images, the original input, the ground truth of mass location marked by the red rectangle and the mass disentangled from the obscured image. The original inputs are the obscured cases. As we can see, the masses are obscured by glands in different degrees. Though some glands are around the mass, our method can still disentangle the true mass from the glands.

C. More Implementation Details

We aim at mammogram mass classification. The inputs are resized into 224 × 224 with random horizontal flips and fed into networks. We implement all models with PyTorch. For all experiments, we select the best model on the validation set for testing. The final results are the average results over ten times for each image. The proportion of added composite obscured images is 30% of the original dataset, since the proportion of obscured images in original data is roughly 30%.

If the proportion of added composite obscured images is too large, it will influence the original data distribution, while if this proportion is too small it will not be capable to learn from the hard cases. The proportion of the obscured masses in the real datasets we use is around 20∼30%, thus we add composite data by nearly the same amount.