Local Residual Attention Network for Classification of Breast Cancer Histopathology Images

Rakshith Subramanyam RAKSHITH.SUBRAMANYAM@ASU.EDU
Skanda Suresh SSURES40@ASU.EDU
Mark Naufel MNAUFEL@ASU.EDU
The Luminosity Lab, Arizona State University

Spring Berman SPRING.BERMAN@ASU.EDU
Autonomous Collective Systems Laboratory, Arizona State University

Pavan Turaga PTURAGA.EDU
Geometric Media Lab, Arizona State University

Editors: Under Review for MIDL 2022

Abstract

Automatic classification of breast histopathological images is a challenging task, as subtle changes in morphometric features can result in misclassifications. To increase broader adoption and trust in deep-learning based solutions, we need methods that produce the right results for the right reasons, while still maintaining high performance. To make progress toward these goals, we propose a novel Local Residual Attention Network (LRAN), that improves the predictive performance of a base-network by attending to class relevant regions of an image. LRAN follows an encoder-decoder architecture with local attention enforced on the skip connections between the encoder and decoder and global attention on the feature maps of the base-network. Our experiments demonstrate that the inclusion of attention mechanisms increases the classification accuracy by 5-8% points over the base-network. Our LRAN with ReseNet-18 as the base-network produces a classification accuracy of 91.83% on the ICIAR 2018 BreAst Cancer Histology (BACH 2018) dataset, which is comparable to the performance on this dataset by a top-performing classification networks.

Keywords: Local Residual Attention, Breast Histopathology Classification

1. Introduction

Among all cancers that affect women, breast cancer is still the leading cause of death (Ahmad, 2019). According to the World Health Organization (WHO) report, in 2018, 15% of all cancer-related deaths in women worldwide were due to breast cancer (wor, 2018). Early detection of breast cancer improves the prognosis of recovery, which motivates the development of more accurate screening and diagnostic techniques. Current screening methods entail a human assessment of subtle morphological changes in breast biopsies and the assignment of discrete Nottingham histologic grades (Elston and Ellis, 1991). Analyzing a tissue slide for morphometric features of the disease is a strenuous and time-consuming process that is subject to variation among pathologists, even highly experienced ones (Mittal et al., 2019). The increasing incidence rate of breast cancer produces a higher diagnostic workload that puts further pressure on pathologists (Williams et al., 2017), which may result in an incorrect diagnosis.

© 2022 R. Subramanyam, S. Suresh, M. Naufel, S. Berman & P. Turaga.

Recent breakthroughs in deep learning techniques have enabled their extensive use in breast histopathological image classification (Chennamsetty et al., 2018; Araújo et al., 2017). Many of these techniques achieve high predictive performance by incorporating an ensemble of neural networks or by increasing the depth of the network. In real-world applications, and especially in medical diagnostics, neural networks should not only consistently exhibit high predictive performance, but also make right predictions for the right reasons. Incidentally, it has been shown in the literature that attention mechanisms learn to focus on relevant regions of an input to make informed decisions for different downstream tasks. Many such solutions have benefited the field of natural language processing (NLP), and image processing.

Inspired by these advances, we propose a novel Local Residual Attention Network (LRAN) to improve the predictive performance of a base-network for breast histopathological image classification. LRAN utilizes an attention network to captures class relevant information on coarse and fine granularity. The attention network is implemented using an encoder-decoder architecture that imposes fine grain local attention through the skip connections and coarse grain global attention through the attention mask obtained from the output of the decoder. The attention mask weighs the base-networks feature maps through residual dot product attention to enhance class relevant features while diminishing other non relevant features. We perform extensive study to investigate the effects of using an attention mechanism on the accuracy of three different base-networks. We also investigate the efficacy of using local attention along with global attention by visualizing the Class Activation Maps and attention masks.

Key contributions of this paper:

- We propose a Local Residual Attention Network (LRAN), a end-to-end trainable attention mechanism that improves the predictive performance of classification architectures by learning to look at regions of relevance.
- Along with global attention, we introduce local attention gates in the skip connections of the attention network to improve local texture propagation and enhancement.
- We investigate the effect of attention mechanisms on classification networks' Class Activation Maps (CAMs). Our experiments show that attention mechanisms increase the classification accuracy by 5-8% points over conventional classification architectures.

2. Related Works

Conventionally histopathological images are stored as high-resolution Whole Slide Images (WSIs), which makes it is difficult to develop an end-to-end neural network for classifying these images. Although there are solutions () that are end-to-end trainable, these solutions lackMany existing solutions employ patch-based classification, in which the high-resolution WSI is divided into multiple lower-resolution patches. One such approach is (Roy et al., 2019), in which the authors train a patch-based classifier consisting of a hierarchical CNN network and use a majority voting scheme to assimilate the classification output of each patch and compute the final class for the image. They report an average patch-wise classification accuracy of 77.4% and image-wise classification accuracy of 87% on the ICIAR 2018 BreAst Cancer Histology dataset (BACH 2018) (Aresta et al., 2019). On the other

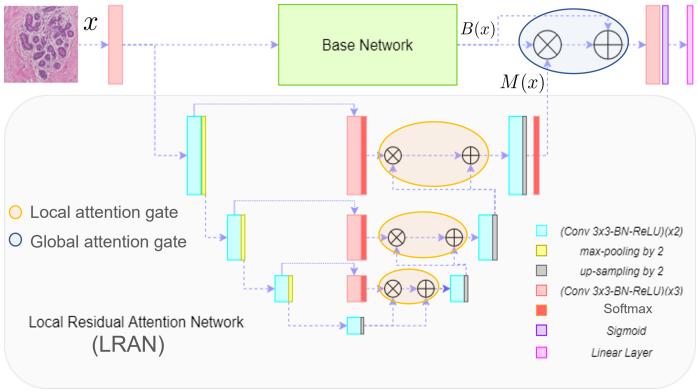

Figure 1: Illustration Local Residual Attention Network. The input image is passed through the base network and the attention network branch. Fine grain features are captured by the local attention gates in the skip connections of the attention network. The attention mask $M(x)$ weighs the feature maps $B(x)$ through the global attention gate.

hand the top performing approach (Marami et al., 2018) of the BACH 2018 utilizes an ensemble of networks trained on BACH 2018 data along with auxiliary data to produce an accuracy of 94%. Such ensemble of networks use complex architectures to achieve high predictive performance, but (Guo et al., 2017) show that complex architectures have high uncertainty in their predictions, that is, these networks gives high confidence value even for miss-classified results. It is crucial for neural networks that perform predictive analysis of histopathology images to learn morphometric features of an image to aid in classification.

Attention-guided networks (Wang et al., 2017; Vaswani et al., 2017) have been widely utilized for the purpose of directing a neural network to focus on relevant regions of the input in natural language processing and computer vision problems. Self-attention networks have gained prominence with the introduction of transformer architecture (Vaswani et al., 2017), the network uses positional encoding along with key, value and queries to direct the network to focus on relevant regions. Self-attention highly depends on effective positional encoding such as relative 2D position embeddings for vision tasks. Such positional dependency, makes identifying relevant regions in pathology tasks hard due to presence of morphological features which cannot be captured by relative positional encoding and requires more effective information extraction mechanisms. (Yang et al., 2019) employs a guided attention mechanism for breast histopathology image classification, in which the proposed network uses guided Region of Interest (ROI) proposals to aid the classification process. This network depends on segmentation masks to train the RoI proposal network and achieves an accuracy of 93% on the BACH 2018 dataset. (Zhang et al., 2017) employs a language model to explore descriptive image features in medical reports and uses this as an integrated attention mechanism for an image model. (Sun et al., 2020) introduces secondary shape streams in parallel with texture streams to capture shape-dependent texture for seg-

mentation of medical images. These methods are dependent on the requirement of metadata or rely on segmentation masks to improve the classification accuracy and interpretability.

This network depends on segmentation masks to train the RoI proposal network and achieves an accuracy of 93% on the BACH 2018 dataset. (Zhang et al., 2017) employs a language model to explore descriptive image features in medical reports and uses this as an integrated attention mechanism for an image model. (Sun et al., 2020) introduces secondary shape streams in parallel with texture streams to capture shape-dependent texture for segmentation of medical images. These methods are dependent on the requirement of metadata or rely on segmentation masks to improve the classification accuracy and interpretability.

3. Methods

The proposed method is constructed with two branches: (a) a base network branch and (b) Local Residual Attention network branch. The base network branch, which could be any traditional classification architecture acts as a feature extractor. The attention network enhances or subdues certain regions in the feature maps of the base-network.

The LRAN, depicted in Figure 1, follows an encoder-decoder architecture. In the encoder path, each layer performs a convolutional operation followed by down-sampling to reduce the feature map size. The consecutive down-sampling operation makes the encoder propagate the most important features in order to obtain an encoded representation of the input image. In the decoder path, each layer consists of an bi-linear up-sampling operation followed by convolutional operation to reconstruct the features from the encoded representation. Skip connections are introduced between each layer of the encoder and the decoder path to propagate high-frequency textural information in images.

Given an input image x, the base network outputs a set of feature maps represented by $B(x)$, and the attention network outputs a set of attention masks $M(x)$ of the same size. The attention mask $M(x)$ identifies and propagates salient regions in the image to preserve task-specific elements of the feature maps $B(x)$ using dot product attention, as shown in equation (1). Repeated application of the attention mask to $B(x)$ can decay the value of features learned in the initial layers and degrade good properties learned by the base network branch. To preserve these properties, a residual connection of the base network is made by adding the feature map to the dot product attention:

$$H_{i,c}(x) = B_{i,c}(x) \cdot M_{i,c}(x) + B_{i,c}(x), \tag{1}$$

where $i \in \{1, ..., K\}$ is the index of the spatial positions of the pixels in the image and $c \in \{1, ..., C\}$ is the index of the channels. The output of an attention network with K pixels for every channel is normalized into a probability distribution consisting of K probabilities per channel using a pixel-wise softmax function:

$$M_{i,c}(x) = \frac{e^{(x_{i,c})}}{\sum_{j=1}^{K} e^{(x_{j,c})}}. \tag{2}$$

The softmax operation in equation (2) normalizes each pixel value of the attention mask $M(x)$ to the interval $(0, 1)$, with all the pixel values for each channel summing to 1. The attention mask enhances the diagnostically relevant pixels of the feature map $B(x)$ by reducing the values of non-relevant pixels to zero.

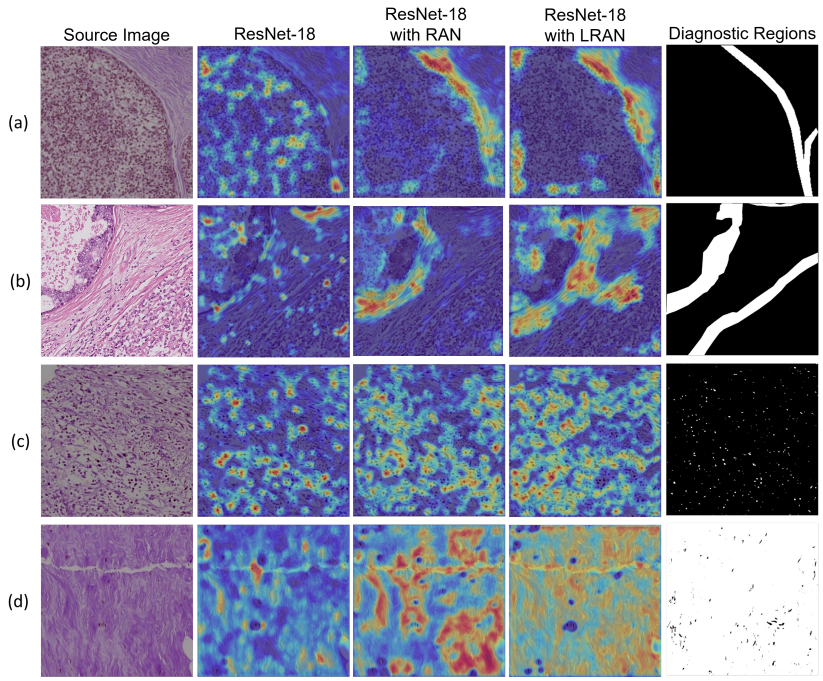

Figure 2: *First column:* Patches of source images from the BACH 2018 dataset, classified as (a) and (b) *In situ carcinoma*, (C) *Invasive carcinoma*, and (d) *Normal* tissue. *Second, Third and fourth columns:* Class Activation Maps (CAMs), superimposed on the source images, that were output by ResNet and R-LRAN. *Fifth column:* Regions of diagnostic relevance annotated by a pathologist

Along with attending to salient coarse grain information of an image, it is also crucial to look at fine textural details before making a prediction. During consecutive down-sampling in the encoder path coarse features are propagated while some fine grain informations are lost. Adding a local residual attention gate on the skip connections propagates the fine grain textural information from the encoder path to the decoder path to enforce better object representation. Along with filtering the feature activation in the forward pass of the network, the attention gates also degrade the gradients from the background regions and prevent noise in the image from influencing the classification.

4. Experiments

4.1. Datasets

We trained our networks on the BreAst Cancer Histology (BACH 2018) image dataset (Aresta et al., 2019). This dataset consists of 400 H&E stained breast histology microscopy images divided into four classes: *Normal*, *Benign*, *In situ carcinoma*, and *Invasive carcinoma* with 100 images per class. In addition, the dataset includes 10 Whole Slide Images (WSIs), each a digital representation of an entire tissue sample which are pixel-wise annotated for the

four classes. The top performing methods in BACH 2018 challenge augment the original dataset either by using images from other datasets (non-BACH) or extract more classification images from the WSI dataset. Similarly we also utilized the pixel-wise annotations to obtain more classification images to create a dataset with 16729 images for *Normal*, 479 images for *Benign*, 166 images for *In situ carcinoma*, and 4857 for *Invasive carcinoma*. We used 90% of the images for training and the remaining for testing.

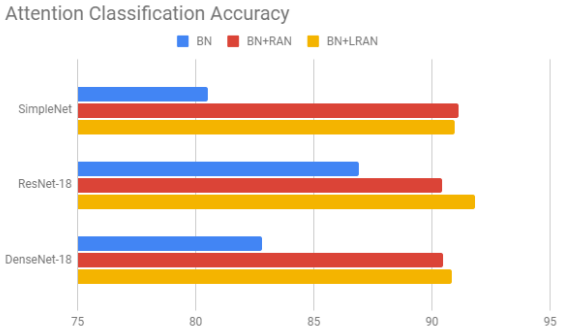

Figure 3: Classification accuracy (%) of different combinations of base networks (BN) with a Residual Attention Network (RAN) and a Local Residual Attention Network (LRAN), evaluated using the test dataset.

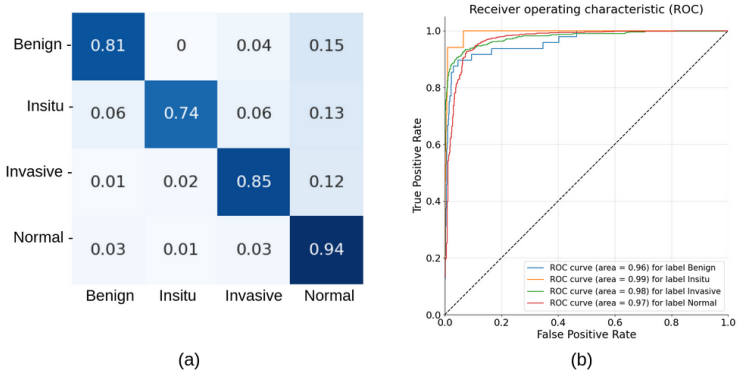

Figure 4: Confusion matrix and Receiver Operating Characteristics (ROC) curve of the R-LRAN network, evaluated using the test dataset.

4.2. Implementation details

All networks were trained for 50 epochs with 24 images per batch. To mitigate the class imbalance in the training data, we employed stratified sampling for each batch. All images

Table 1: Classification accuracy of R-LRAN and other networks from the literature

Network	Test Accuracy (%)	Parameters	Dataset Used
(Roy et al., 2019)	87.00	28,157	BACH 2018
(Kwok, 2018)	87.00	442,148	BACH 2018
(Chennamsetty et al., 2018)	87.5	12,332,740	BACH 2018
(Yan et al., 2020)	91.30	24,000,000	BACH 2018 + proprietary dataset
R-LRAN (this paper)	91.83	5,777,444	BACH 2018 + WSI
(Yang et al., 2019)	93.00	7,024,844	BACH 2018 with proprietary segmentation masks
(Marami et al., 2018)	94.00	23,885,392	BACH 2018 + WSI + BreakHis

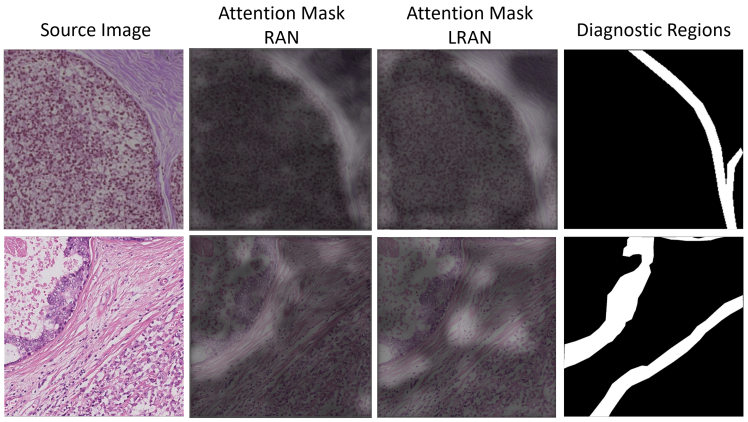

Figure 5: *In situ carcinoma* images overlayed with attention mask of RAN and LRAN compared against regions of diagnostic relevance annotated by a pathologist. LRAN attends to more class relevant information when compared to RAN

in each batch were reshaped to size 256×256. The training loss was computed with a cross-entropy function. The networks were optimized using the Adam optimizer with an initial learning rate of $1e-4$, which was reduced to $1e-5$ after the 10^{th} epoch and to $5e-6$ after the 20^{th} epoch.

4.3. LRAN improves the predictive performance of base-network

We evaluated the performance of 9 different networks: three base networks (without attention), SimpleNet, ResNet, and DenseNet, and combinations of each base network with the two attention networks, RAN (global attention) and LRAN (global and local attention). We use the performance of the three base networks as a baseline for evaluating improvements in performance that result from incorporating the attention architectures. The base networks were trained independently in order to isolate the effect of the attention mechanisms on the networks classification accuracy. From Figure 3, it can be seen that the inclusion of an attention network improves the classification accuracy of each base network by at least 5%. This demonstrates the ability of attention mechanisms to enhance the performance of the base-network. For complex architectures like ResNet-18 and DenseNet-18 LRAN had an improved accuracy over RAN. We attribute this improvement in performance to the ability

of LRAN to attend to local and global information. The combination of ResNet-18 as the base-network and LRAN (R-LRAN) yielded the highest classification accuracy of all the tested networks, 91.83%.

In Table 1, we compare the accuracy of R-LRAN, our best-performing network, to the accuracy of other classification networks from the literature. (Yan et al., 2020) achieved an accuracy of 91.3% with a hybrid CNN-RNN network that uses LSTM to improve attention on the images. (Yang et al., 2019) achieved an accuracy of 93% using a guided attention network trained on both the BACH 2018 dataset and segmentation masks that is not publicly available. (Marami et al., 2018) uses an ensemble of network trained on images from BACH 2018 and BreskHis(Spanhol et al., 2015) dataset to obtain an accuracy of 94%. As Table 1 shows, the R-LRAN network achieved comparable classification accuracy to the top performing models, without using extra memory footprint exhibited by the ensemble of networks or addition segmentation masks like in (Yang et al., 2019). Figure 4(a) displays the confusion matrix of R-LRAN. The classification accuracy was lowest for the *In situ carcinoma* class, 74%, due to the low number of samples in this class relative to the other classes. The Receiver Operating Characteristics (ROC) curve in Figure 4(b) shows that the R-LRAN exhibits a high true positive rate and a low false positive rate, and that all four classes have an Area Under the Curve (AUC) exceeding 0.96.

4.4. Improved visual interpretability with LRAN

Figure 2 shows that the inclusion of the LRAN attention network improves the Class Activation Maps (CAMs) output of the ResNet architecture. For evaluating the efficacy of the CAMs we got few images annotated by an expert pathologist on the regions they focus during diagnostic classification. *In situ carcinoma* is a cancer type that is contained, with a distinct boundary; the attention network was able to enhance the pixels near the boundaries while the remaining pixels were suppressed, as demonstrated by the attention mask in Figures 2(a) and (b). The attention network's enhancement of the feature map of the base network is also evident in Figure 2(c), where the source image shows normal tissue, and the attention network spreads the attention over the entire image (i.e., it does not identify particular regions that are associated with cancer). Figure 5 shows that inclusion of local attention along with global attention focuses on more class relevant details when compared to using only global attention (RAN).

5. Conclusion

This paper presents a new method for breast histopathological image classification using a combination of a base-network and a local residual attention network. The proposed method improves the predictive performance of the base-network and generates feature-rich Class Activation Maps (CAMs) that can be used to interpret the classification results and identify diagnostically relevant regions in an image. With extensive experimentation, we showed that the integration of an local residual attention network with a small base-network improved the classification accuracy by 5-8% points. We also showed that local residual attention masks were largely in agreement with the diagnostically relevant regions marked by an expert pathologist. Finally towards future work, we hypothesise that a pre-trained attention network can be used to learn semantic segmentation effectively with few examples.

References

Breast cancer. https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/, Sept 2018. World Health Organization.

Aamir Ahmad. Breast cancer statistics: recent trends. In *Breast Cancer Metastasis and Drug Resistance*, pages 1–7. Springer, 2019.

Teresa Araújo, Guilherme Aresta, Eduardo Castro, José Rouco, Paulo Aguiar, Catarina Eloy, António Polónia, and Aurélio Campilho. Classification of breast cancer histology images using convolutional neural networks. *PloS one*, 12(6):e0177544, 2017.

Guilherme Aresta, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan, et al. BACH: Grand challenge on breast cancer histology images. *Medical Image Analysis*, 56:122–139, 2019.

Sai Saketh Chennamsetty, Mohammed Safwan, and Varghese Alex. Classification of breast cancer histology image using ensemble of pre-trained neural networks. In *International Conference on Image Analysis and Recognition*, pages 804–811. Springer, 2018.

Christopher W Elston and Ian O Ellis. Pathological prognostic factors in breast cancer. i. the value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, 19(5):403–410, 1991.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1321–1330. JMLR. org, 2017.

Scotty Kwok. Multiclass classification of breast cancer in whole-slide images. In *International Conference on Image Analysis and Recognition*, pages 931–940. Springer, 2018.

Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan, Gerardo Fernandez, and Jack Zeineh. Ensemble network for region identification in breast histopathology slides. In *International Conference on Image Analysis and Recognition*, pages 861–868. Springer, 2018.

Shachi Mittal, Catalin Stoean, Andre Kajdacsy-Balla, and Rohit Bhargava. Digital assessment of stained breast tissue images for comprehensive tumor and microenvironment analysis. *Frontiers in Bioengineering and Biotechnology*, 7:246, 2019.

Kaushiki Roy, Debapriya Banik, Debotosh Bhattacharjee, and Mita Nasipuri. Patch-based system for classification of breast histology images using deep learning. *Computerized Medical Imaging and Graphics*, 71:90–103, 2019.

Fabio A Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, 63(7):1455–1462, 2015.

Jesse Sun, Fatemeh Darbehani, Mark Zaidi, and Bo Wang. SAUNet: shape attentive U-Net for interpretable medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 797–806. Springer, 2020.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164, 2017.

Bethany Jill Williams, David Bottoms, and Darren Treanor. Future-proofing pathology: the case for clinical adoption of digital pathology. *Journal of Clinical Pathology*, 70(12): 1010–1018, 2017.

Rui Yan, Fei Ren, Zihao Wang, Lihua Wang, Tong Zhang, Yudong Liu, Xiaosong Rao, Chunhou Zheng, and Fa Zhang. Breast cancer histopathological image classification using a hybrid deep neural network. *Methods*, 173:52–60, 2020.

Heechan Yang, Ji-Ye Kim, Hyongsuk Kim, and Shyam P Adhikari. Guided soft attention network for classification of breast cancer histopathology images. *IEEE transactions on medical imaging*, 39(5):1306–1315, 2019.

Zizhao Zhang, Yuanpu Xie, Fuyong Xing, Mason McGough, and Lin Yang. MDNet: A semantically and visually interpretable medical image diagnosis network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6428–6436, 2017.

Appendix A. BACH dataset K-Fold cross validation

To further investigate the effectiveness of our proposed method we performed K-Fold cross validation on the entire curated dataset (Section 4.1). We divided the dataset into 5 folds and followed the implementation explained in section 4.2. Table 3 demonstrates the ability of LRAN to improve the classification accuracy of Resnet on all the different folds of validation.

Table 2: K Fold classification accuracy of Resnet and Resnet with LRAN

Method	Fold1	Fold 2	Fold 3	Fold 4	Fold 5
Resnet	88.17	89.92	89.65	90.00	90.35
Resnet + LRAN	92.40	92.25	92.84	92.45	92.18

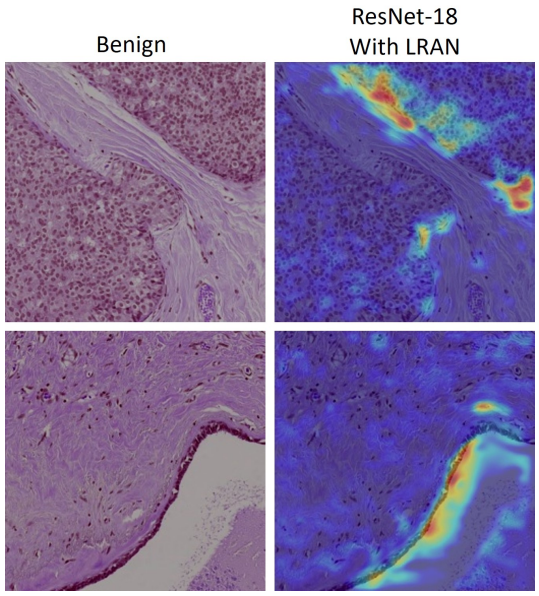

Figure 6: **Additional example for benign images:** LRAN is able to focus on regions of *benign* cells to direct the base network

A.1. Miss-classified images of BACH dataset using Resnet-LRAN

Figure 10 shows examples where LRAN missclassified an image. The first image from top belongs to in-situ carcinoma class, but was classified as benign. The CAM shows that the model did not identify the defining boundary of in-situ carcinoma. Likewise the second image belongs to invasive carcinoma class but was categorized as benign, as the class specific morphological features are not prominently spread across the image. The third image was categorized as in-situ carcinoma when it belongs to the normal class, the definitive boundary on the lower half of the image made LRAN to categorize it as in-situ carcinoma.

Appendix B. Qualitative Performance on Invasive Duct Carcinoma dataset

To investigate the efficacy of LRAN we experimented on Invasive Duct Carcinoma (IDC) images [1], a common sub-type of breast cancers. The dataset contains 194668 images for training and 2000 images for testing, with IDC negative and IDC positive classes. We followed the experimental setup given in Section 4.2. The performance on IDC duct carcinoma follows similar trends to BACH dataset with attention improving the accuracy of base network and producing rich CAM.

1. https://www.kaggle.com/paultimothymooney/predicting-idc-in-breast-cancer-histology-images

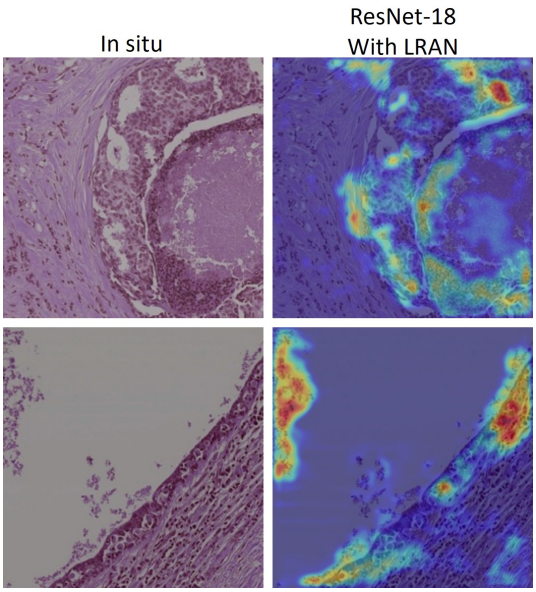

Figure 7: **Additional example for *in situ* images:** LRAN is able to focus on boundary region of *in situ carcinoma*

Table 3: Classification accuracy of Resnet and Resnet with LRAN on IDC duct carcinoma

Method	Test Accuracy
Resnet	88.25
Resnet + LRAN	90.90

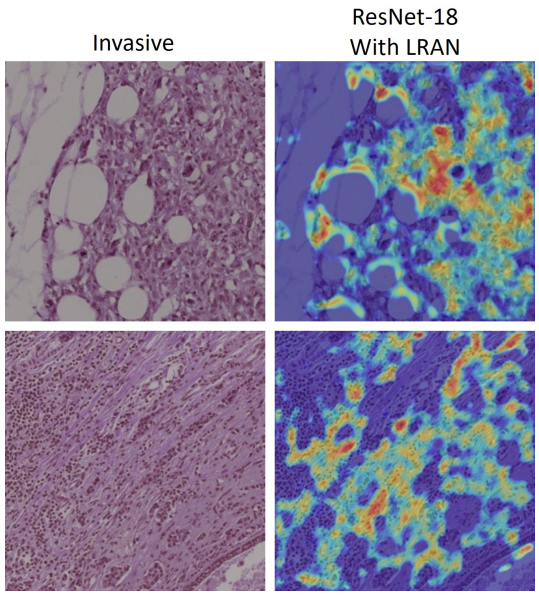

Figure 8: **Additional example for *invasive* images:** LRAN is able to focus on cells spread over the image. In the first image the attention reduces the focus on the fat globules, which is non cancerous.

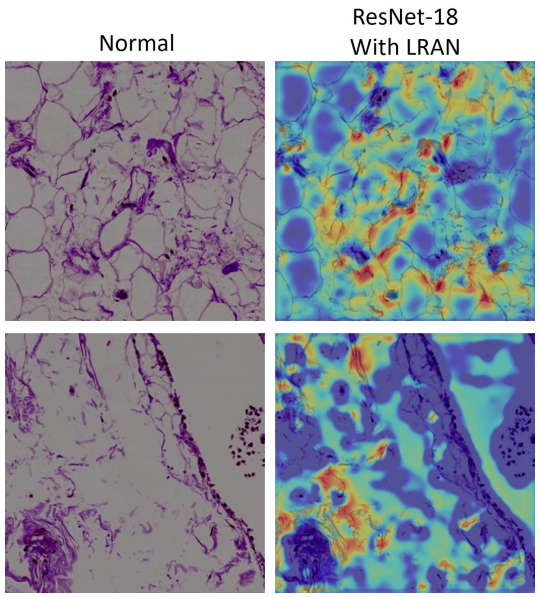

Figure 9: **Additional example for *normal* images:** LRAN focuses on the entire image to classify the images into normal class

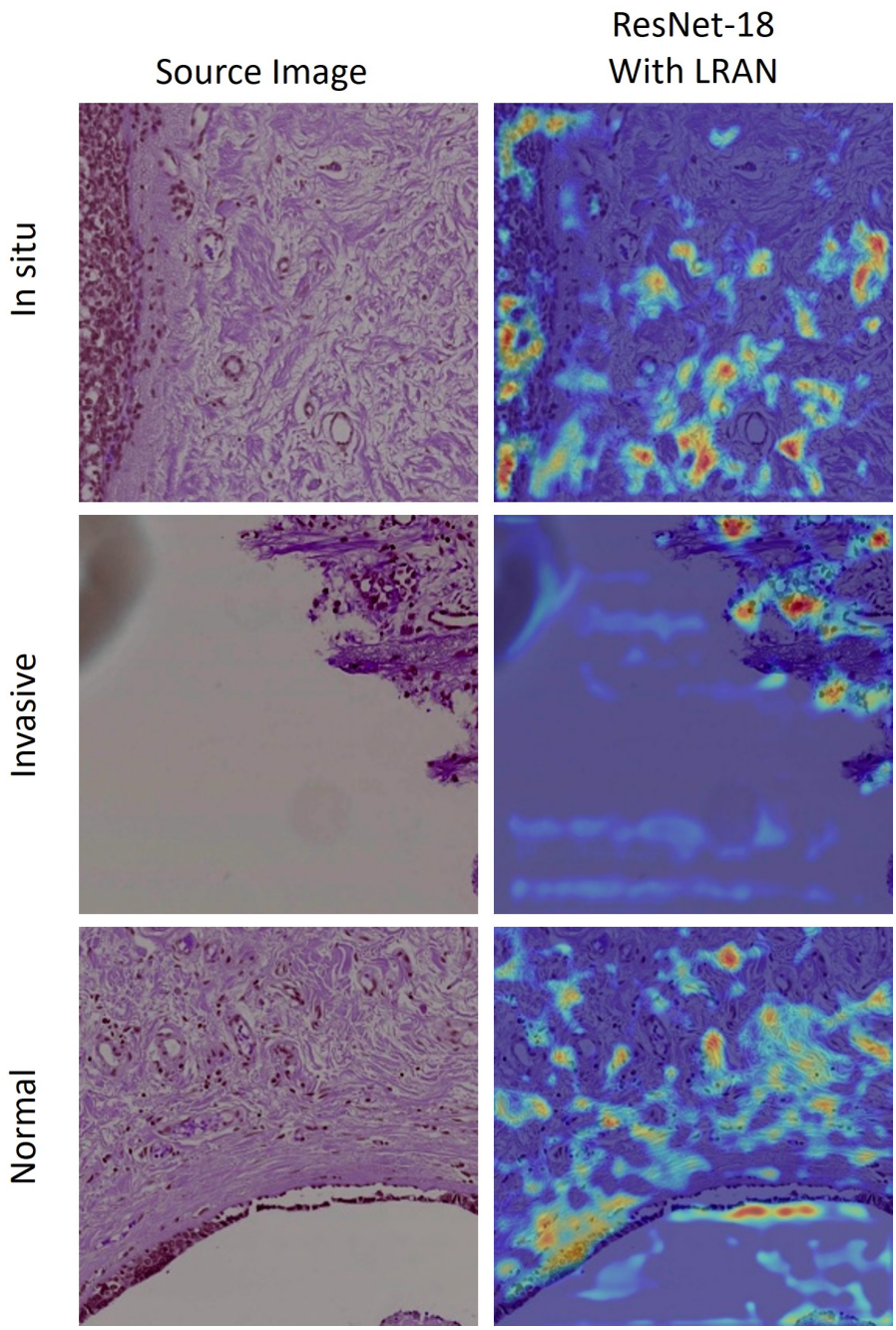

Figure 10: *First column:* Patches of source images from the BACH 2018 dataset *Second column:* Class Activation Maps (CAMs), superimposed on the source images, that were output by R-LRAN. First image is In-situ carcinoma but R-LRAN classified as benign, second image is invasive carcinoma classified as benign, and the third image is normal class which was classified as in-situ.

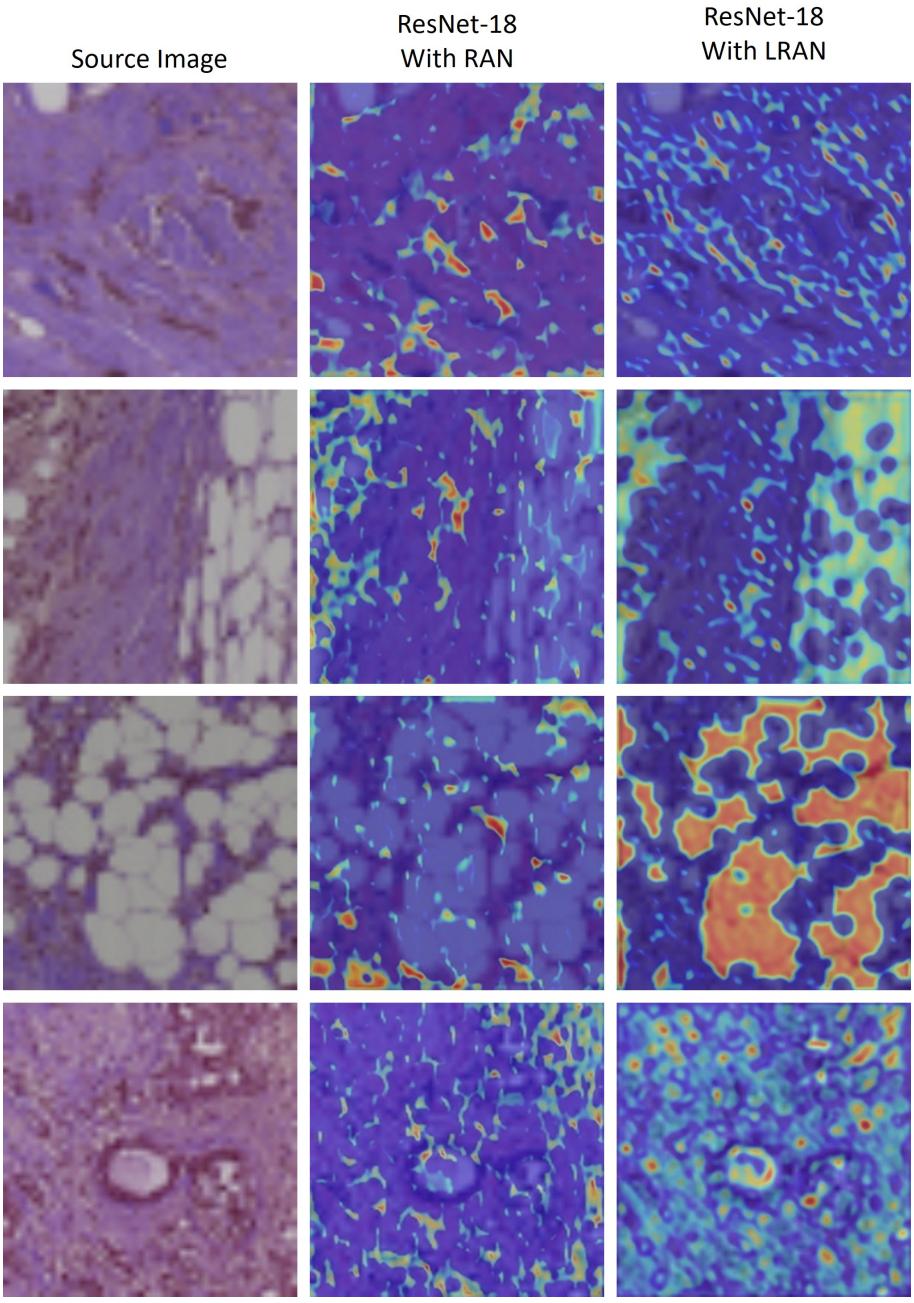

Figure 11: Qualitative performance on IDC duct carcinoma benign class

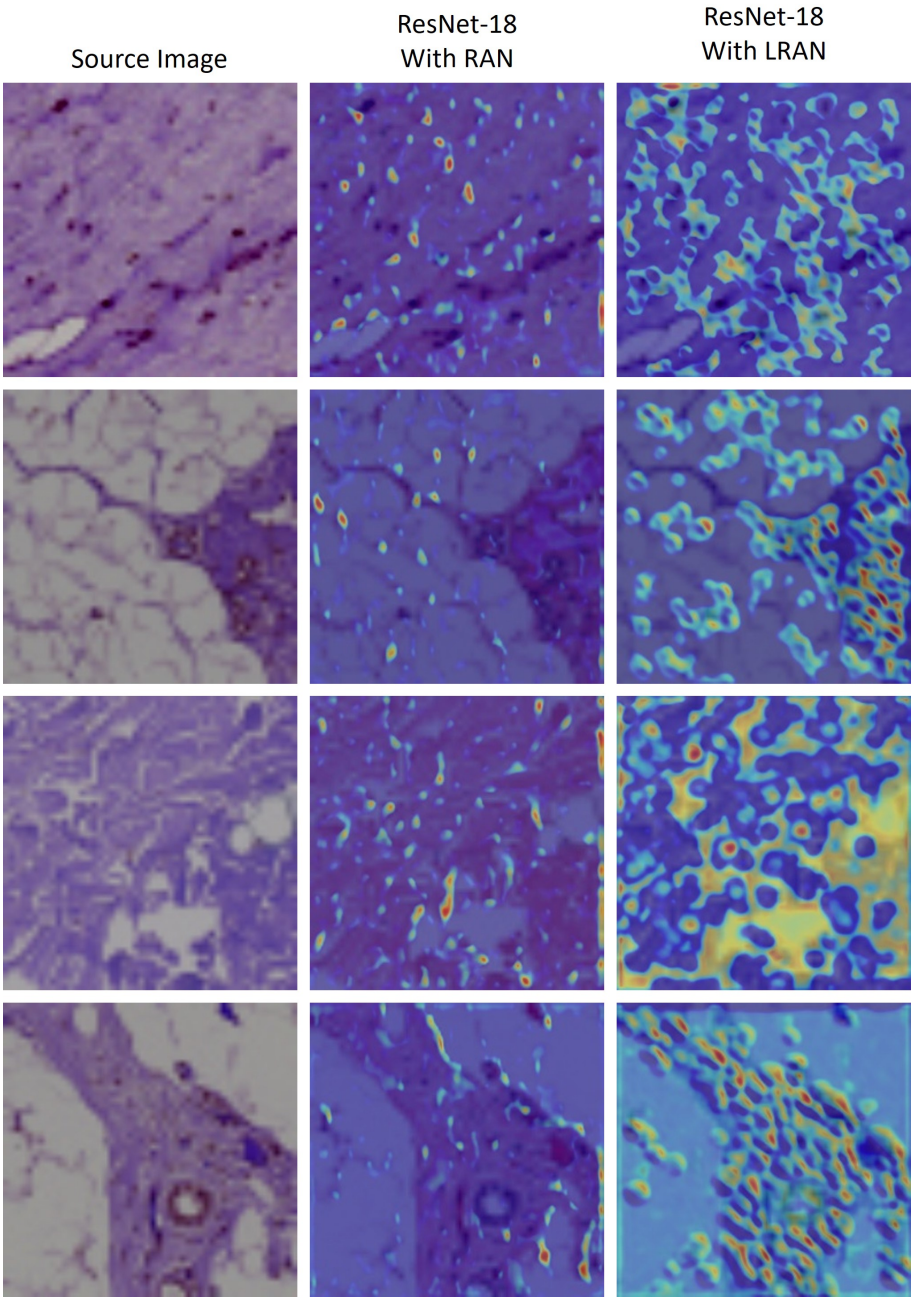

Figure 12: Qualitative performance on IDC duct carcinoma malignant class