PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Cross-domain and multi-task transfer learning of deep convolutional neural network for breast cancer diagnosis in digital breast tomosynthesis

Ravi Samala, Heang-Ping Chan, Lubomir Hadjiiski, Mark Helvie, Caleb Richter, et al.

Ravi K. Samala, Heang-Ping Chan, Lubomir Hadjiiski, Mark A. Helvie, Caleb Richter, Kenny Cha, "Cross-domain and multi-task transfer learning of deep convolutional neural network for breast cancer diagnosis in digital breast tomosynthesis," Proc. SPIE 10575, Medical Imaging 2018: Computer-Aided Diagnosis, 105750Q (27 February 2018); doi: 10.1117/12.2293412

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

Cross-domain and Multi-task Transfer Learning of Deep Convolutional Neural Network for Breast Cancer Diagnosis in Digital Breast Tomosynthesis

Ravi K. Samala,[1] Heang-Ping Chan, Lubomir Hadjiiski,
Mark A. Helvie, Caleb Richter, Kenny Cha

Department of Radiology, University of Michigan, Ann Arbor, MI 48109-5842

ABSTRACT

We propose a cross-domain, multi-task transfer learning framework to transfer knowledge learned from non-medical images by a deep convolutional neural network (DCNN) to medical image recognition task while improving the generalization by multi-task learning of auxiliary tasks. A first stage cross-domain transfer learning was initiated from ImageNet trained DCNN to mammography trained DCNN. 19,632 regions-of-interest (ROI) from 2,454 mass lesions were collected from two imaging modalities: digitized-screen film mammography (SFM) and full-field digital mammography (DM), and split into training and test sets. In the multi-task transfer learning, the DCNN learned the mass classification task simultaneously from the training set of SFM and DM. The best transfer network for mammography was selected from three transfer networks with different number of convolutional layers frozen. The performance of single-task and multi-task transfer learning on an independent SFM test set in terms of the area under the receiver operating characteristic curve (AUC) was 0.78±0.02 and 0.82±0.02, respectively. In the second stage cross-domain transfer learning, a set of 12,680 ROIs from 317 mass lesions on DBT were split into validation and independent test sets. We first studied the data requirements for the first stage mammography trained DCNN by varying the mammography training data from 1% to 100% and evaluated its learning on the DBT validation set in inference mode. We found that the entire available mammography set provided the best generalization. The DBT validation set was then used to train only the last four fully connected layers, resulting in an AUC of 0.90±0.04 on the independent DBT test set.

Keywords: breast cancer, mammography, digital breast tomosynthesis, transfer learning, multi-task, cross-domain

1. INTRODUCTION

Transfer learning is frequently used in medical imaging to overcome the limitation of large data requirements typically needed for deep learning.[1-15] It also has the advantage of training in multi-task, weakly supervised and unsupervised mode. In cross-domain transfer learning, the aim is to transfer the knowledge from the *source* task to the *target* task where the tasks are from different domains. In multi-task learning, multiple interrelated tasks from the same domain are simultaneously learned to improve generalization. In this study we used multi-task transfer learning to train from multiple auxiliary sources while transferring knowledge from non-medical imaging domain through cross-domain transfer learning to the task of classifying masses in digital breast tomosynthesis (DBT) for a computer-aided diagnosis (CAD) system. The whole process was embedded into two transfer learning stages: (a) cross-domain transfer learning from ImageNet trained DCNN to multi-task mammography trained DCNN, and (b) cross-domain transfer learning from mammography trained DCNN to DBT trained DCNN. We studied the methods for cross-domain, multi-task transfer learning and its effectiveness in overcoming the limited availability of DBT data.

2. MATERIALS AND METHODS

2.1 Data set

ImageNet DCNN was trained on 1.2 million non-medical images to classify targets into 1000 object classes seen in everyday life (~1,200 images/class on average).[16] Our mammography data set is a collection of heterogeneous cases

[1] rsamala@umich.edu

Medical Imaging 2018: Computer-Aided Diagnosis, edited by Nicholas Petrick, Kensaku Mori, Proc. of SPIE Vol. 10575, 105750Q · © 2018 SPIE · CCC code: 1605-7422/18/$18 · doi: 10.1117/12.2293412

from two sources. With approval from the Institutional Review Board (IRB), a total of 1,655 SFM views and 310 DM views were collected from the University of Michigan Health System (UM). Another set of 277 SFM views were obtained from the Digital Database for Screening Mammography (DDSM). With IRB approval, DBT cases were collected from UM and Massachusetts General Hospital (MGH) for previous studies.[6, 17, 18] Craniocaudal (CC) and mediolateral oblique (MLO) views were acquired in the UM set and only MLO views were acquired in the MGH set. The DBT-UM set consisted of 61 malignant and 118 benign masses from 186 views and the DBT-MGH set consisted of 87 malignant and 51 benign masses from 138 views. From each lesion, a 128 x 128-pixel region-of-interest (ROI) was extracted at a pixel size of 200 μm x 200 μm. Background correction was used to normalize the gray-levels across SFM, DM and DBT.[6, 19, 20] Lesions on mammography and DBT views were identified by a radiologist. For each mammography lesion, a single ROI enclosing the lesion was extracted. For each DBT lesion, five DBT ROIs were extracted from five slices approximately centered at the lesion. Each ROI was rotated in four directions and flipped to generate 8 augmented ROIs.

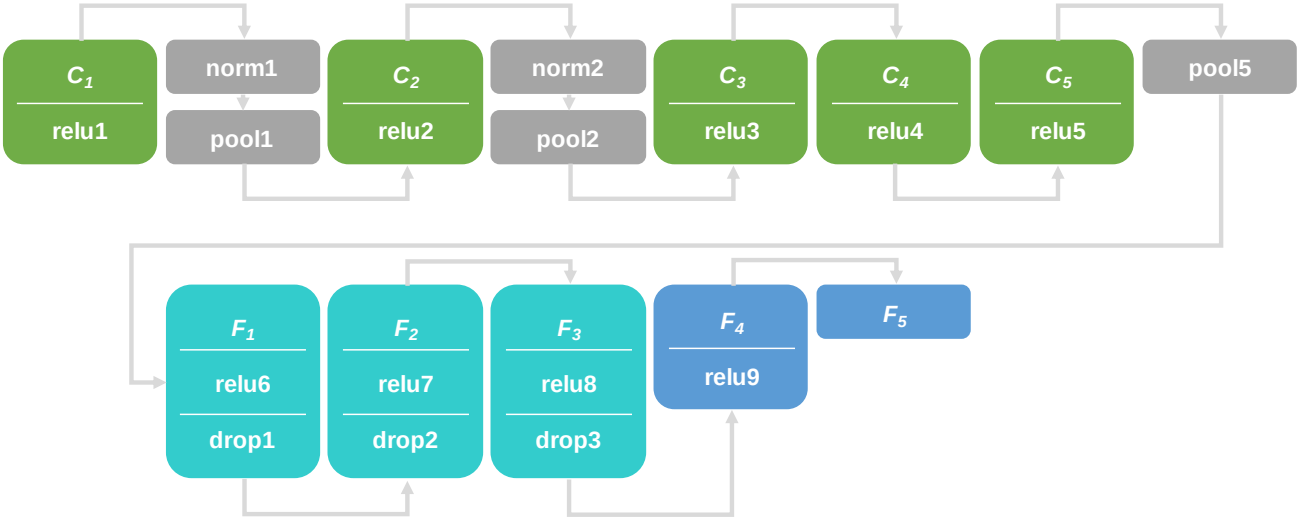

Fig. 1. DCNN structure used for cross-domain and multi-task transfer learning.

2.2 Transfer learning

Fig. 1 shows the ImageNet trained DCNN network that consists of five convolutional layers (denoted as C_i, $i=1,...,5$) with 64, 192, 384, 256, 256 filter kernels of sizes 11x11 5x5, 3x3, 3x3 and 3x3, respectively, and three fully connected layers (denoted as F_1, F_2 and F_3) with 4096, 4096 and 1000 nodes, respectively. The input ROI to the DCNN is a 128x128-pixel RGB channel. There are a total of 148K neurons and 33M tunable parameters in the DCNN. In the first stage of cross-domain transfer learning, we appended the ImageNet DCNN with two fully connected layers to reduce the large divergence between the non-medial and medical domains and also to reduce the number classes from 1000 to 2 (malignant and benign breast masses). The additional fully connected layers had (F_4) 100 and (F_5) 2 nodes as shown in fig. 1. In this transfer learning stage, the DCNN was trained using multi-task learning mode where the domain was the same between the tasks (SFM and DM) and the target was breast mass in mammography. Due to the low distribution divergence across SFM and DM, the transfer learning was expected to obtain a set of rich feature hierarchies resulting in higher performance. We have verified the training results using spectral embedding method and receiver operating characteristic (ROC) curves between the single-task (SFM only) and the multi-task (SFM, DM) modes. In this stage, we also (a) identified the best transfer network between three types: C_1 frozen, C_1 to C_3 frozen and C_1 to C_5 frozen using N-fold cross-validation of the mammography training data set, and (b) studied the mammography training data size requirements for the three mammography trained DCNNs by evaluating the inference ability of the trained DCNNs on the classification of masses in the DBT validation data. We observed that the classification performance on DBT masses increased as the mammography training set size increased without reaching a plateau for the available training data, we therefore chose the mammography transfer network trained on the entire mammography data set for the second stage of cross-domain transfer learning. In this stage, the mammography trained DCNN was transfer-trained on the DBT validation

set where the layers from C_1 to F_1 are frozen and only layers F_2 to F_5 were allowed to train. Most of the network could be frozen because of the similarity between the masses in mammography and in DBT and because the reduced DBT data size was too small to fine-tune the 12M additional tunable parameters if these layers were not frozen. The DBT trained DCNN was evaluated by an independent DBT test set in the inference mode and the performance was analyzed with the area under the ROC curve (AUC).

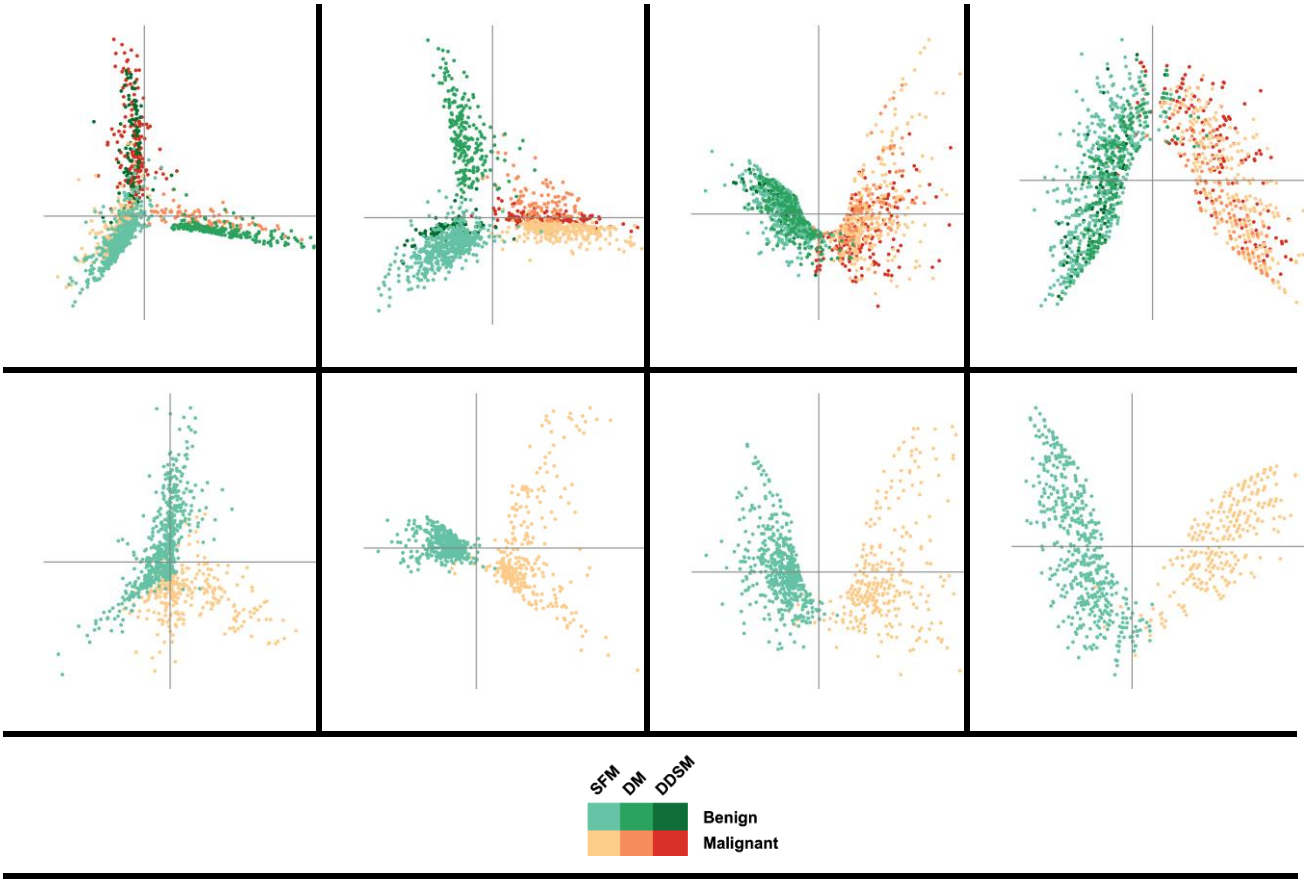

Fig. 2. Spectral embedding of features extracted at the F_1, F_2, F_3 and F_4 layers (shown along the four columns) of the single-task and multi-task mammography trained DCNNs. Top row: Multi-task (UM-SFM, DDSM, UM-DM), bottom row: Single-task (UM-SFM only).

Table 1. *N*-fold cross-validation results for three transfer networks trained with mammography data.

Transfer network	C_1 frozen	C_1 - C_3 frozen	C_1 - C_5 frozen
AUC	0.76	0.73	0.73

3. RESULTS

The *N*-fold cross-validation of the first stage transfer learning on the mammography training data set is shown in Table 1. The transfer network C_1 yielded the best AUC. For this network, the spectral embedding method was used to reduce the 4096-, 4096-, 1000- and 100-dimensional feature vectors from the four fully connected layers to two-dimensions as shown in fig. 2. The lesion-based ROC curves for the independent mammography test set are shown in fig. 3. For the same network, the inference ability on the DBT validation set was studied by varying the mammography training

data from 1% to 100% by randomly selecting 10 subsets at each desired percentage from the original training data set as shown in fig. 4. The box plot demonstrates (a) the generalization ability of the single-stage mammography trained DCNN to DBT for classification of masses and (b) the gradual increase in the validation AUC as the size of the mammography training data increased. The data requirement plot was also studied for the two other transfer networks (Table 1) as shown in fig. 5. A similar trend of increasing AUC with increasing mammography training set size was observed for all three transfer networks. The DCNN with only C_1 frozen and transfer-trained on the entire mammography data achieved an AUC of 0.88±0.02 in the inference mode for the DBT validation data set (fig.6). For the second stage cross-domain transfer learning, the mammography trained DCNN was transfer-trained on the DBT validation set by freezing all the convolutional layers and the first fully connected layer (i.e., C_1 to F_1). The performance of the DCNN after the second stage transfer learning achieved an ROC curve with an AUC of 0.90±0.04 for the independent DBT test set as shown in fig. 6.

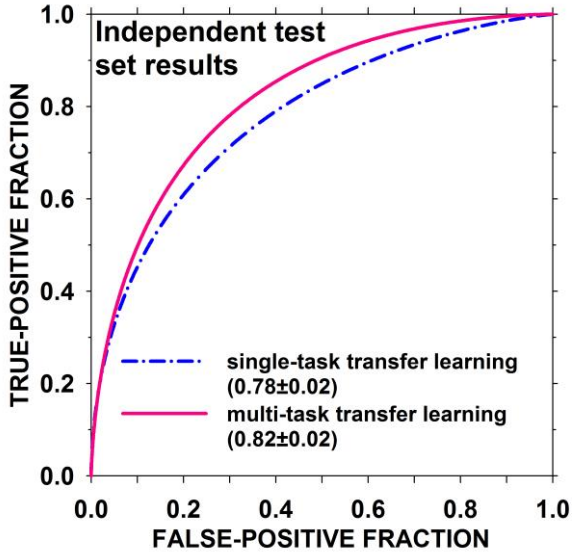

Fig. 3. Lesion-based ROC curves of single-task and multi-task mammography trained DCNN for the independent SFM test set.

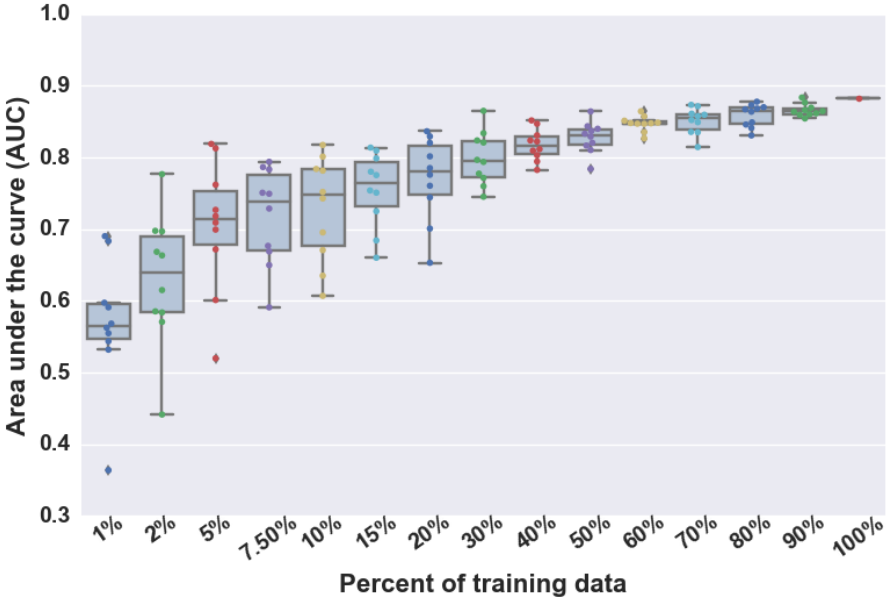

Fig. 4. Dependence on percentage of mammography training data vs ROI-based AUC on the DBT validation set for the single-stage C_1-frozen transfer-learned DCNN.

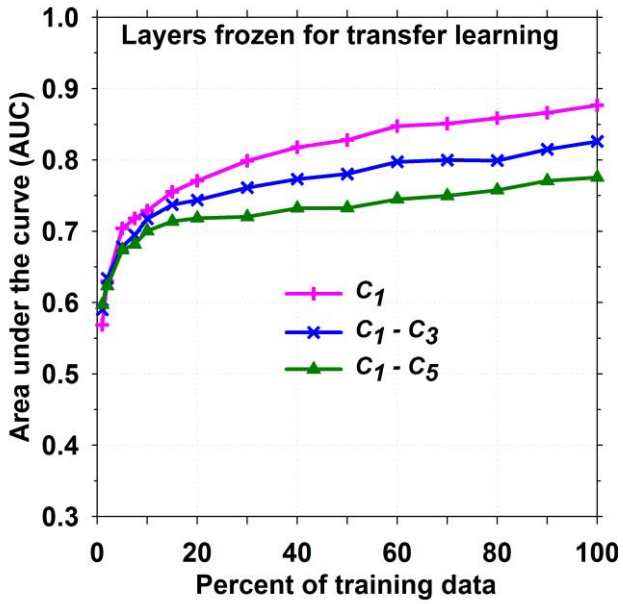

Fig. 5. Dependence on percentage of mammography training data vs ROI-based AUC on the DBT validation set for three transfer-learned DCNNs.

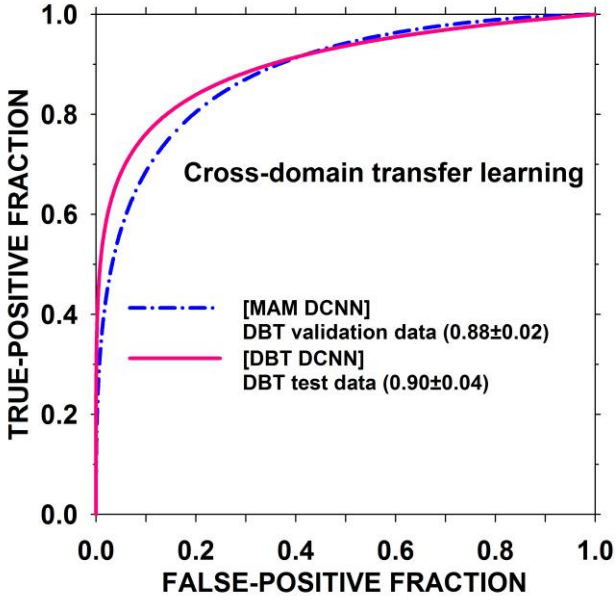

Fig. 6. ROC curves for DBT validation set using single-stage mammography transfer trained DCNN with C_1 frozen, and for the independent DBT test set using two-stage mammography-DBT trained DCNN where C_1 was frozen in the first stage and C_1-F_1 were frozen in the second stage transfer training.

4. DISCUSSION AND CONCLUSION

We propose a cross-domain transfer learning framework through supervised training from non-medical domain to medical imaging domain with additional enhancement of generalization capabilities from multi-task learning. In the two-stage transfer learning approach, we provide a framework to transfer the knowledge from ImageNet trained DCNN to mammography trained DCNN and to DBT trained DCNN. The original DCNN with 33M parameters is pre-trained using 1.2 million ImageNet images. At the first-stage transfer learning, where the DCNN is partially frozen, 20K mammography ROIs are used for training. At the second-stage transfer learning, where most layers of the DCNN are frozen, 9K DBT ROIs are used for training. We attempt to scale the amount of DCNN allowed to train based on the available training data. Single-stage mammography transfer trained DCNN is validated through *N*-fold cross-validation and data size dependent validation strategies. Further work is planned to collect additional mammography and DBT data and extend the data requirement experiments to determine, with larger training sets, whether the observed performance of the transfer-trained DCNN at each stage can be improved and whether reducing the number of DCNN layers being frozen during the second stage transfer learning may increase the performance.

ACKNOWLEDGEMENTS

This work is supported by National Institutes of Health award numbers R01 CA151443 and R01 CA214981.

REFERENCES

[1] K. H. Cha, L. M. Hadjiiski, H.-P. Chan, R. K. Samala, R. H. Cohan, E. M. Caoili, C. Paramagul, A. Alva, and A. Z. Weizer, "Bladder cancer treatment response assessment using deep learning in CT with transfer learning." 10134, 1013404.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical Image Analysis, 42, 60-88 (2017).

[3] B. Cheng, M. Liu, D. Shen, Z. Li, D. Zhang, and A. s. D. N. Initiative, "Multi-domain transfer learning for early diagnosis of alzheimer's disease," Neuroinformatics, 15(2), 115-132 (2017).

[4] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," Journal of Medical Imaging, 3(3), 034501-034501 (2016).

[5] R. Paul, S. H. Hawkins, Y. Balagurunathan, M. B. Schabath, R. J. Gillies, L. O. Hall, and D. B. Goldgof, "Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma," Tomography: a journal for imaging research, 2(4), 388 (2016).

[6] R. K. Samala, H. P. Chan, L. Hadjiiski, M. A. Helvie, J. Wei, and K. Cha, "Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography," Medical physics, 43(12), 6654-6666 (2016).

[7] P. M. Cheng, and H. S. Malhi, "Transfer Learning with Convolutional Neural Networks for Classification of Abdominal Ultrasound Images," Journal of digital imaging, 30(2), 234-243 (2017).

[8] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," IEEE transactions on medical imaging, 35(5), 1299-1312 (2016).

[9] T. Kooi, B. Ginneken, N. Karssemeijer, and A. Heeten, "Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network," Medical physics, 44(3), 1017-1027 (2017).

[10] R. Zhang, Y. Zheng, T. W. C. Mak, R. Yu, S. H. Wong, J. Y. Lau, and C. C. Poon, "Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain," IEEE journal of biomedical and health informatics, 21(1), 41-47 (2017).

[11] N. Dhungel, G. Carneiro, and A. P. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention," Medical image analysis, 37, 114-128 (2017).

[12] X. Han, "MR-based synthetic CT generation using a deep convolutional neural network method," Medical Physics, 44(4), 1408-1419 (2017).

[13] R. K. Samala, H.-P. Chan, L. M. Hadjiiski, M. A. Helvie, K. Cha, and C. Richter, “Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms,” Physics in Medicine and Biology, 62(8894-8908), (2017).
[14] R. K. Samala, H.-P. Chan, L. Hadjiiski, K. Cha, and M. A. Helvie, “Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis,” SPIE medical imaging. International Society for Optics and Photonics, 97850Y-97850Y (2016).
[15] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. Guttmann, F.-E. de Leeuw, C. M. Tempany, and B. van Ginneken, "Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation." 516-524.
[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, 1097-1105 (2012).
[17] H.-P. Chan, J. Wei, Y. H. Zhang, M. A. Helvie, R. H. Moore, B. Sahiner, L. Hadjiiski, and D. B. Kopans, “Computer-aided detection of masses in digital tomosynthesis mammography: Comparison of three approaches,” Medical Physics, 35(9), 4087-4095 (2008).
[18] H.-P. Chan, M. A. Helvie, L. Hadjiiski, D. O. Jeffries, K. A. Klein, C. H. Neal, M. Noroozian, C. Paramagul, and M. A. Roubidoux, “Characterization of Breast Masses in Digital Breast Tomosynthesis and Digital Mammograms: An Observer Performance Study,” Academic radiology, 24(11), 1372-1379 (2017).
[19] B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, “Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images,” IEEE Transactions on Medical Imaging, 15, 598-610 (1996).
[20] H.-P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsitt, and N. Petrick, “Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space,” Physics in Medicine and Biology, 40, 857-876 (1995).