Multiple-Instance and Transfer Learning for Detecting Breast Cancer

Nathan Dass, Kais Kudrolli, David Liang
{ndass, kudrolli, dwliang} @ stanford.edu

Introduction

1 in 8 women are affected by breast cancer and over 40,000 women are expected to die from it in 2018 alone [1]. Early detection using mammograms is very important as it can open up treatment options and improve survival rate to 93% [2] State-of-the-art CNNs have been used previously in breast cancer detection. We aim to extend previous work by applying transfer learning and multiple-instance learning techniques to CNNs.

Data

DDSM Dataset [3]

- 2555 cases, 4 **mammogram** images each, different views of same patient
- **Labeled** as normal (688 cases), benign (814), benign without callback (140), cancer (913 cases)
- Images came as **Grayscale** in various sizes up to 7000x7000
- 255 cases (10%) held out for **validation**

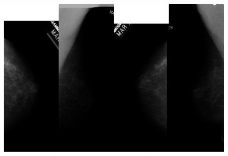


Fig 1: Normal case

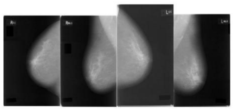


Fig 2: Cancer case

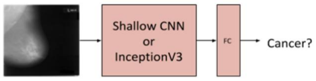
Data Preprocessing

- **Reshaped** to 299x299 grayscale for stacked MIL, RGB otherwise
- **Normalized** each image
- Data **augmentation**: rotation, flip, and zoom on each view

Models and Results

Transfer Learning

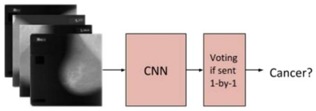
- **Baseline**: shallow network with 3 CONV + 2 FC
- **Transfer:** pre-trained InceptionV3: only train FC, all layers trainable
- Train/predict on an image-by-image basis



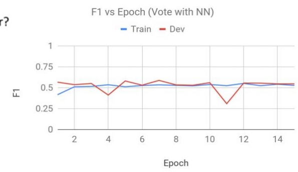
Model	Precision	Recall	F1
Shallow (Baseline)	16.4%	37.5%	21.9%
Transfer (Inception frozen)	**24.9%**	50.3%	31.4%
Transfer (Inception trainable)	23.8%	**55.1%**	**32.1%**

Multiple-Instance Learning (MIL)

- **Combine the views** in a case to try and get a better prediction
- Can **stitch** images together in 2x2 grid or **stack** each view as a channel
- Can also send 1-by-1 and take a **vote** with **max** or **mean** of output or feed through **NN**



Model	Precision	Recall	F1
Stack	44.1%	87.5%	58.2%
Vote w/ Max	**47.2%**	67.7%	55.5%
Vote w/ NN + Transfer	44.6%	63.9%	52.4%
Vote w/ NN	44.6%	**88.3%**	**58.3%**



Discussion

- Our models are unable to balance recall and precision. The **high recall** suggests that we catch most of the cancer, but **low precision** indicates that we predict cancer too often for normal or benign (**high false positive rate**).
- As expected, **transfer learning shows improvement** compared to our baseline shallow network.
- The **combined MIL + transfer** learning model **did not perform the best**. This was **unexpected** because we thought having both more info per input and the low-level feature extraction of Inception would lead to better performance. However, we need to train it for more epochs or change the FC architecture to confirm this.
- Low precision could be due to **class imbalance**. We tried to use an **error rate multiplier** to weight the loss of a true positive and noticed some improvements. We could similarly **weight the loss of false positives**.

Future Work

- **Immediate**: train each model for longer
- **Immediate**: standardize evaluations between models that use MIL and do not use MIL to give a better idea of how much MIL helps
- **Visualize** what the network focuses on when predicting
- Use **attention** or object **segmentation** to focus on breast/tumor portion of image

References

[1] Breastcancer.org. (2018). *U.S. Breast Cancer Statistics | Breastcancer.org*. [online] Available at: https://www.breastcancer.org/symptoms/understand_bc/statistics
[2] Carolmilgardbreastcenter (2018). *Early Detection is Key | Carol Milgard Breast Center*. [online] Available at: http://www.carolmilgardbreastcenter.org/early-detection
[3] Marathon.csee.usf.edu. (2018). *Computer Vision and Pattern Recognition Group*. [online] Available at: http://marathon.csee.usf.edu/Mammography/Database.html