EUROScholars
European Undergraduate Research Opportunities

UZ LEUVEN

KU LEUVEN

The Creation of a Breast Cancer Voxel Model Database for Virtual Clinical Trials in Digital Breast Tomosynthesis

Risha Das

Arizona State University

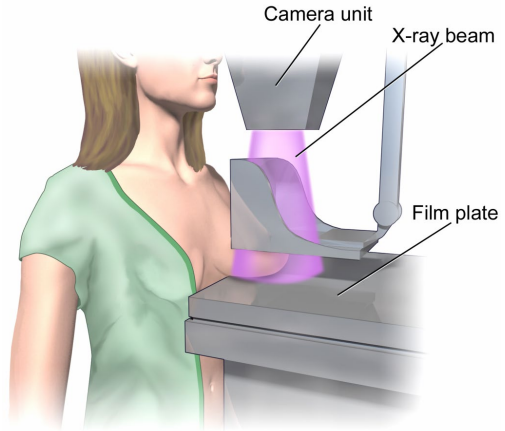
EUROScholars
European Undergraduate Research Opportunities

About Me

- Double Major in Mechanical Engineering & Math at ASU Barrett
- Bachelor's: Class of 2024
- Loves all things space & NASA
- Reader, Singer, Writer, Traveler, & Foodie
- Bengali & 1st Generation American Citizen

Introduction: What is 2D & Pseudo-3D mammography?

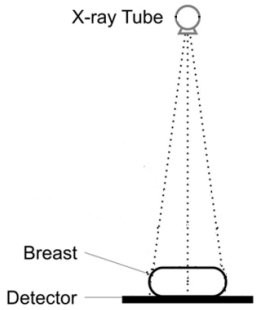
- **Mammography** - usage of low doses of radiation to take an image of a breast
- **How it Works:**
  - Breast tissue compressed between two plates
  - Breast tissue is made of 2 tissues:
    - **Fibroglandular**, which appear white on mammograms
    - **Fatty**, which appear gray on mammograms
    - Radiologists look for high density regions of the breast to detect breast cancer early on



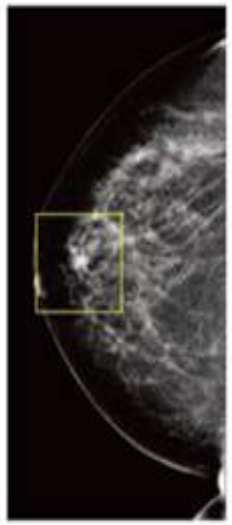
Mammogram

- Belgium mostly relies on 2D mammography but pseudo-3D mammography is much more effective at cancer detection

Introduction: Digital Breast Tomosynthesis (DBT)



2D Planar digital mammography



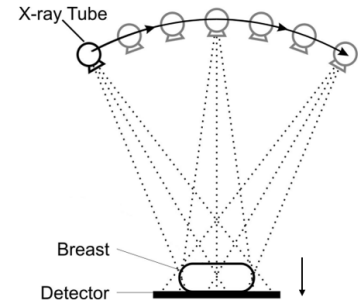
2D "conventional" mammogram

2D

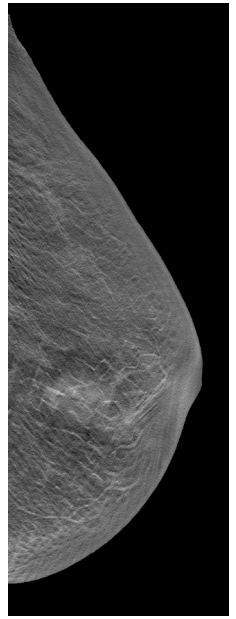
- Takes one image of the breast & one from side, creating images with overlapping tissue
- Hard to distinguish between fibroglandular tissues or cancer because **both appear white** on mammograms as cancer cells *only* grow in fibroglandular tissue
- Especially difficult to detect cancer lesions in very dense breasts (high concentration of fibro. tissue)

Psuedo-3D or DBT

- Uses multiple 2D images to build DBT image of breast
- Arcs over breast to take 25 pictures as it moves (Siemens Tomosynthesis System)
- 3x as much radiation as 2D but still very safe
- Pros:
  - Reduces rate of false positive readings & biopsy need
  - Fewer women need to return for another mammogram

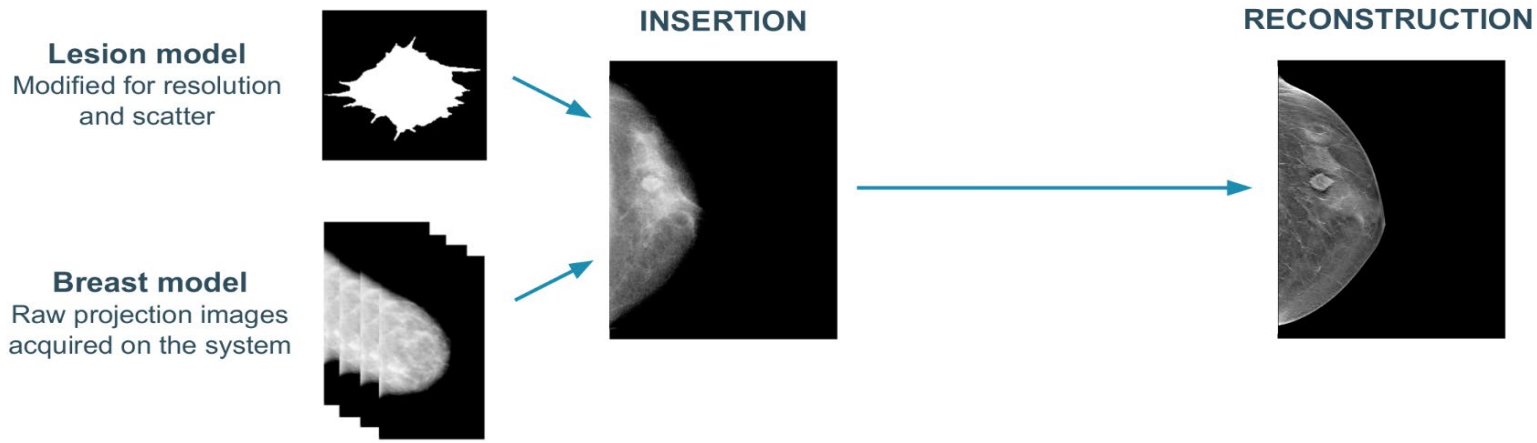


Digital Breast Tomosynthesis



Introduction: What do we do with these mammograms?

1. Manually segment the cancer lesions, using a software called ITK-Snap
2. Exported segmentations and create voxel models
3. Randomly insert voxel models in healthy breast images through a MATLAB simulation framework (this requires an appropriate location)
    - **Note**: Voxel models refer to models comprised of pixels in the 3D space



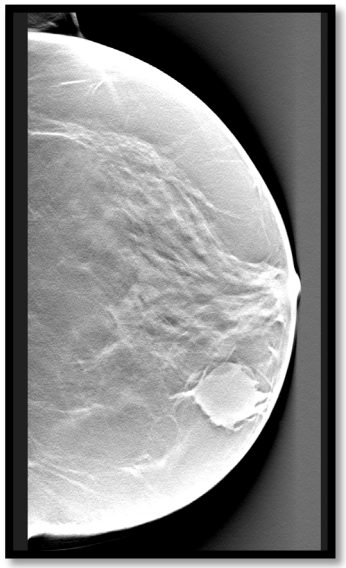
Introduction: Why does this Research Matter?

1. **Once inserted,** imaging system parameters or scan settings are changed to see if there are better ways to observe the lesions.
    - i.e. dose (number of x-rays), spectrum (material/thickness filter), etc.
    - These factors are hard to adjust or change in real life to optimize the visual output of the system but in a simulation framework, it is very simple to do
2. **End Goal**: To have a reader study from where we can conclude if the **visibility & detectability of simulated lesions** is better or worse with new/different scan settings
    - Such results will allow us to conclude if these new scan settings are viable in **clinical practice**

Challenges

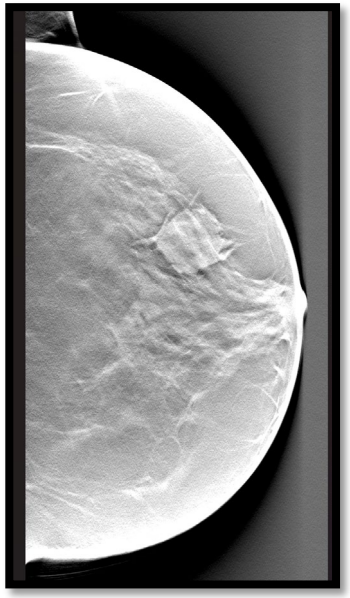
1. **Difficult** to find appropriate locations for lesions when inserted in healthy breast background images
   - Realism of simulated images are affected by **lesion shape, location, breast density, and lesion's ability to blend in** with surrounding structures of healthy breast images

The following illustrates impacts of different locations in simulations:



Notice that in both simulated images, the lesions look unrealistic:

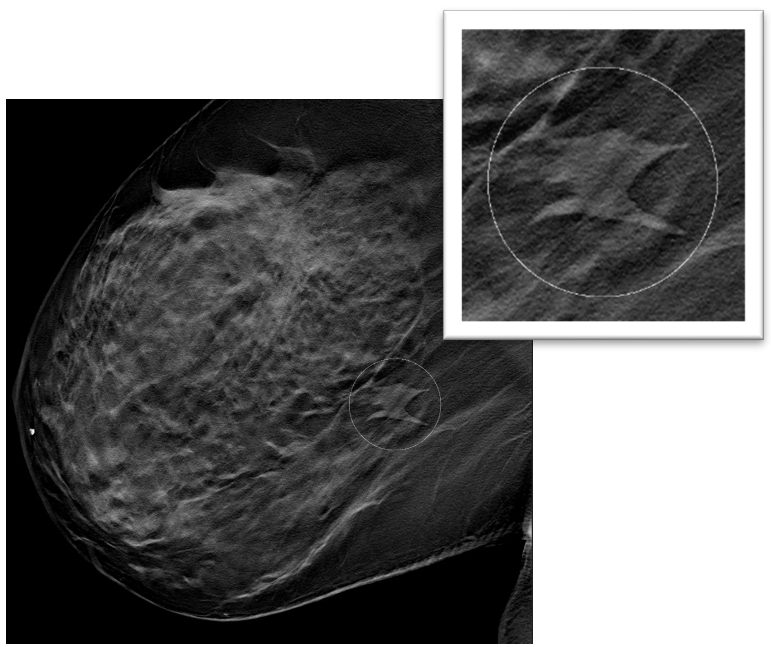
- To the left, its too close to the skin
- To the right, it doesn't flow with surrounding breast structures!

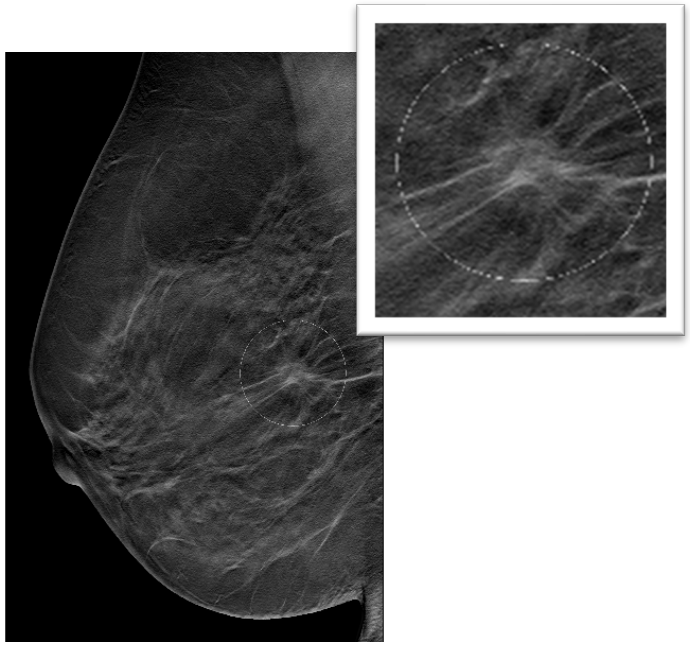


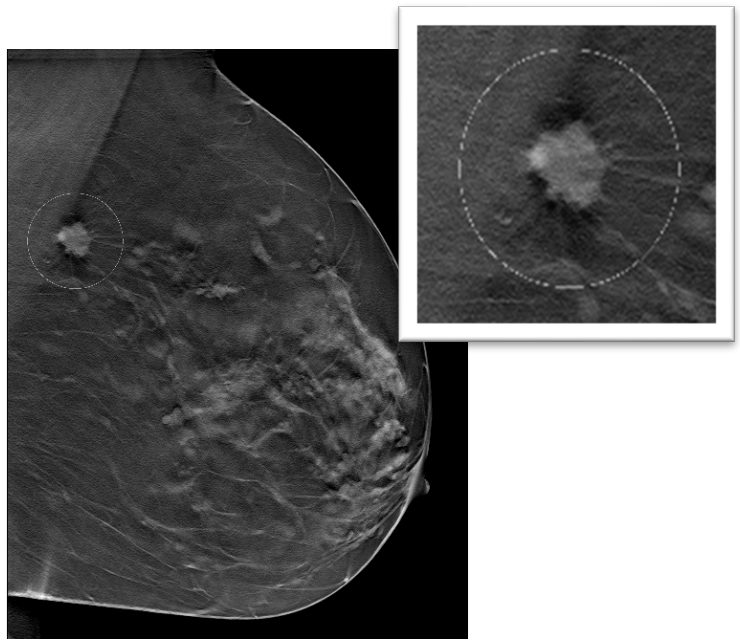
Challenges (continued)

- Can be difficult to note if simulations would be usable in clinical practice so simulations are ranked on realism by radiologists in 'visual grading' studies
- Ranking occurs on the Likert Scale from 1-5, where 1 = clearly simulated & 5 = clearly real

	nr	real				sim	Remarks
Answers:		5	4	3	2	1	
Simulated	1					x	Spics are perpendicular to breast structure -> rotating lesion OR LMLO breast, maybe spics too thick
Simulated	12	x					spics to subtle to be simulated
Real	3	x					Smooth transistion over slices + subtle and long spiculations







Challenges (continued)

2. Too lengthy of a process to manually segment all cancerous lesions & need more data for effective Virtual Clinical Trials

- Requires segmentation of **every** lesion on **every** slice of **every** viable DBT image
- DBT images' slice count varies wildly => can range from 30 to 70+
- To make the appropriate voxel model:
  - Screenshot, segment, crop, convert to raw file

Purpose

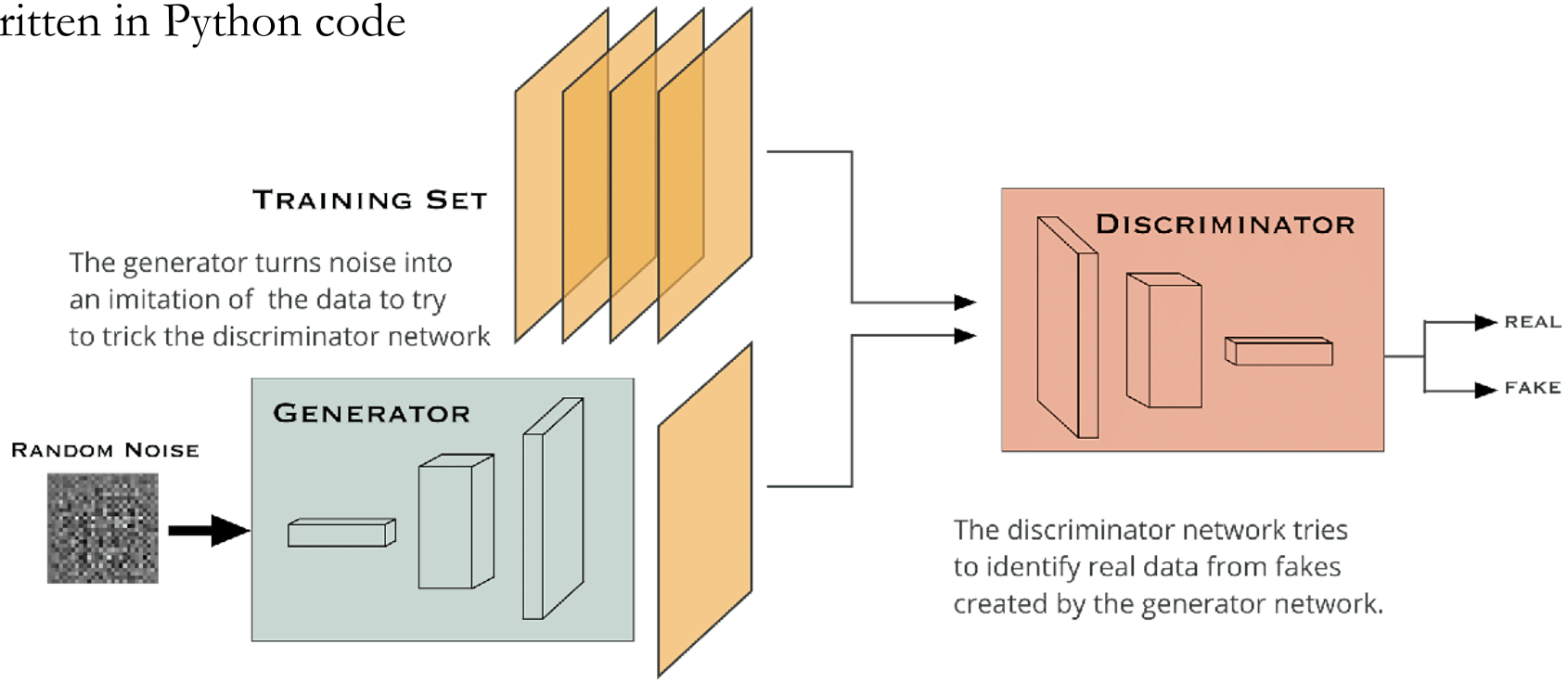
- Simply put, the goal is **to create realistic voxel models of cancerous lesions for virtual clinical trials to validate & optimize new breast imaging techniques**

- We plan to do this by **creating & training two 3D Generative Adversarial networks (GAN) to generate extra segmentations of spiculated & non-spiculated breast cancer lesions**

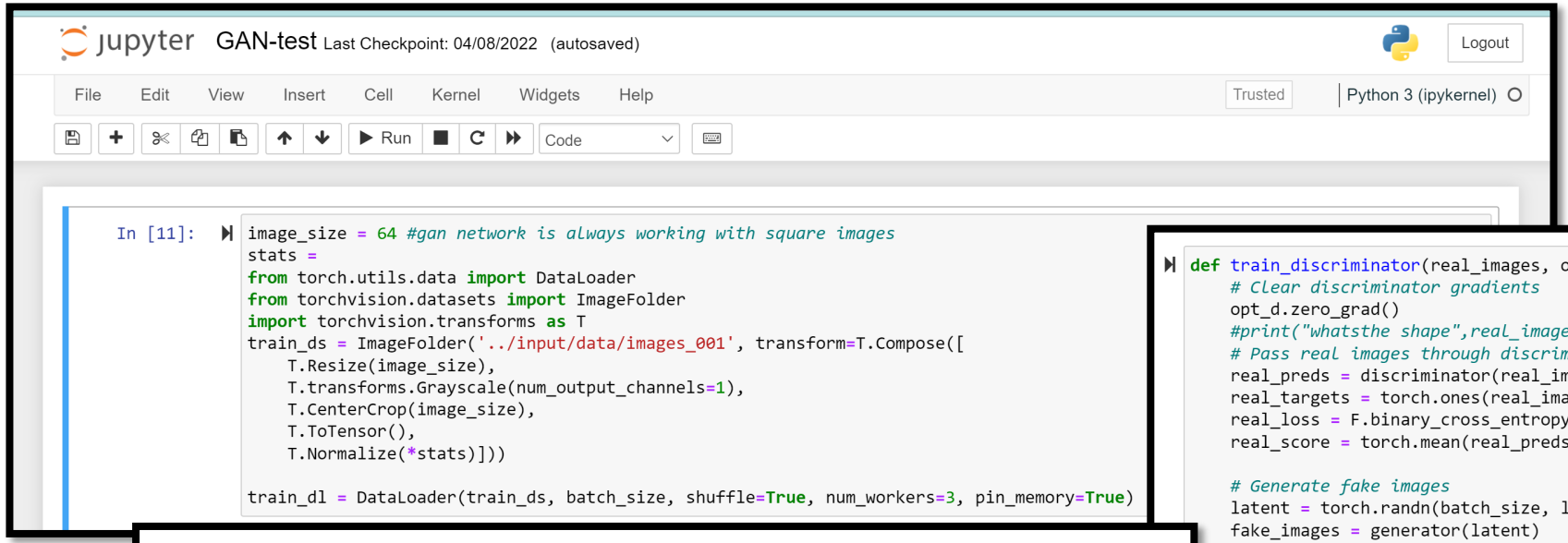
- To understand the purpose, we must also understand what a GAN network is

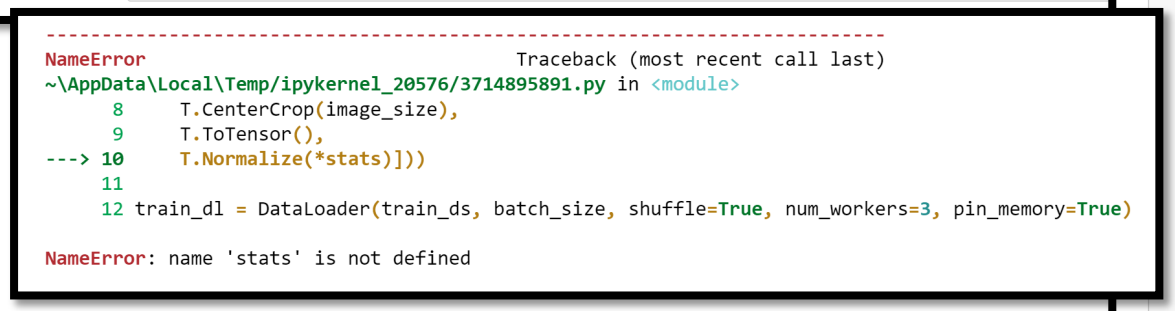
Background: What is a GAN network?

- AI networks are meant to automate processes without explicit instructions by detecting patterns => GAN network will generate accurate segmentations from a training data set of manual segmentations
- Requires a large data set, from anywhere around 100 minimum images to thousands
- Network written in Python code



Setting Up the GAN Network





- Dedicated two to three weeks learning everything I could about Python coding from virtually starting out with nothing

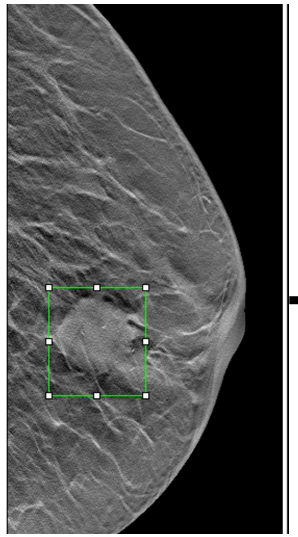
- Involved lots of incredibly interesting applications of Object Oriented Programming, overlapped with my pre-existing Java knowledge, & required/still requires much debugging

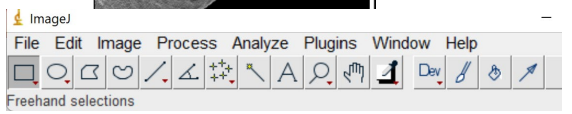
EUROScholars
European Undergraduate Research Opportunities

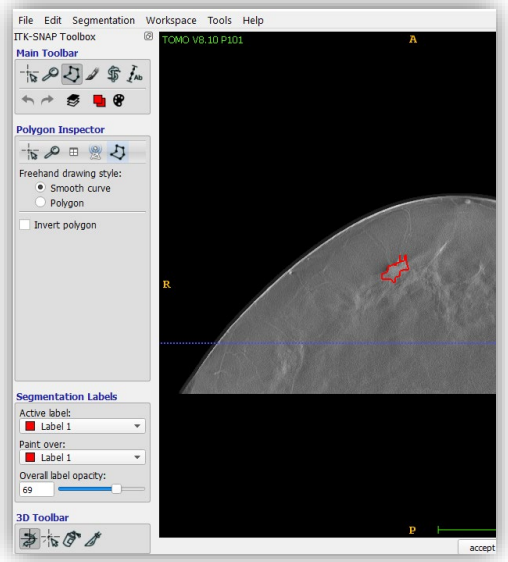
Summary of my Work

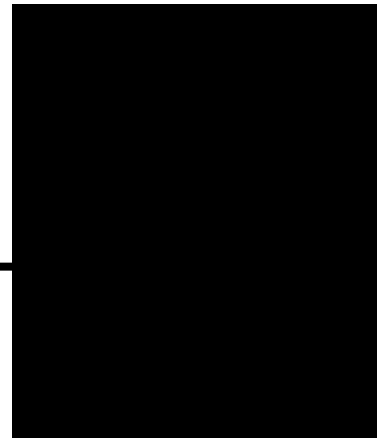
1. Literature Review of existing Research Hospital team has already completed
2. Download Images (.dcm file type & opened in ImageJ) from IMPAX Medical Image database from Hospital that meet certain criteria
3. Catalogue all images in Running Excel Sheet by various parameters
4. Screenshot all cases to identify locations of lesions prematurely
5. Consult with Radiologist on cases that we were unsure of to further define border
6. Manually segment lesions in ITK-Snap and export as cropped raw files for voxel models (segmented about 100 myself)
7. Classify voxel models as spiculated or non-spiculated
8. Train two GAN networks to generate accurate voxel models (In-Progress)

Patient info		Anatomo-pathology						Tumor location	DBT			
EADnr	PACSnr	IDA	ILA	DCIS	LCIS	APO other	- or multifo	Laterality	Date of imaging	Patient Name	PACS nr	Station name
[redacted]		FALSE	TRUE	FALSE	FALSE		multi	left	11/30/2016	Choul	[redacted]	7
		FALSE	FALSE	TRUE	FALSE	Invasief micropapillair	uni	right	12/14/2016	Van Dessel		4
		TRUE	FALSE	TRUE	FALSE		uni	left	11/24/2016	Nijhof		7









```
def train_generator(opt_g):
    # Clear generator gradients
    opt_g.zero_grad()

    # Generate fake images
    latent = torch.randn(batch_size, latent_size, 1, 1, device=device)
    fake_images = generator(latent)

    # Try to fool the discriminator
    preds = discriminator(fake_images)
    targets = torch.ones(batch_size, 1, device=device)
    loss = F.binary_cross_entropy(preds, targets)

    # Update generator weights
    loss.backward()
    opt_g.step()

    return loss.item()
```

Results & Reflections

- Most segmentations have been completed & GAN networks are still being perfected/debugged
  - The challenge right now is focusing on generating lesion images with the appropriate shapes
- Despite focus of project being completely out of the scope of my field, I was enthralled to be working on a project with health cross-applications
- Enjoyed being challenged to learn entirely new concepts & work from ground up with little initial knowledge
- Gained understanding of what research demands, i.e. PhD research practices => even more excited to pursue my Master's & PhD in mechanical engineering now
- AI applications are incredibly relevant to engineering & I am looking forward to seeing the cross applications of GAN networks/coding play out in college/my career
- I gained a thorough understanding of a new coding language, so am now proficient in 3 coding languages: MATLAB, Python, & Java!
- It meant a lot to me personally, as a woman and having lost family to breast cancer before, to work directly on a project that is devoted to optimizing detection of breast cancer

EUROScholars
European Undergraduate Research Opportunities

UZ LEUVEN

KU LEUVEN

Thank you for listening!

Please let me know if you have any questions!

Arizona State University