Image Guidance for Robot-Assisted Ankle Fracture Repair

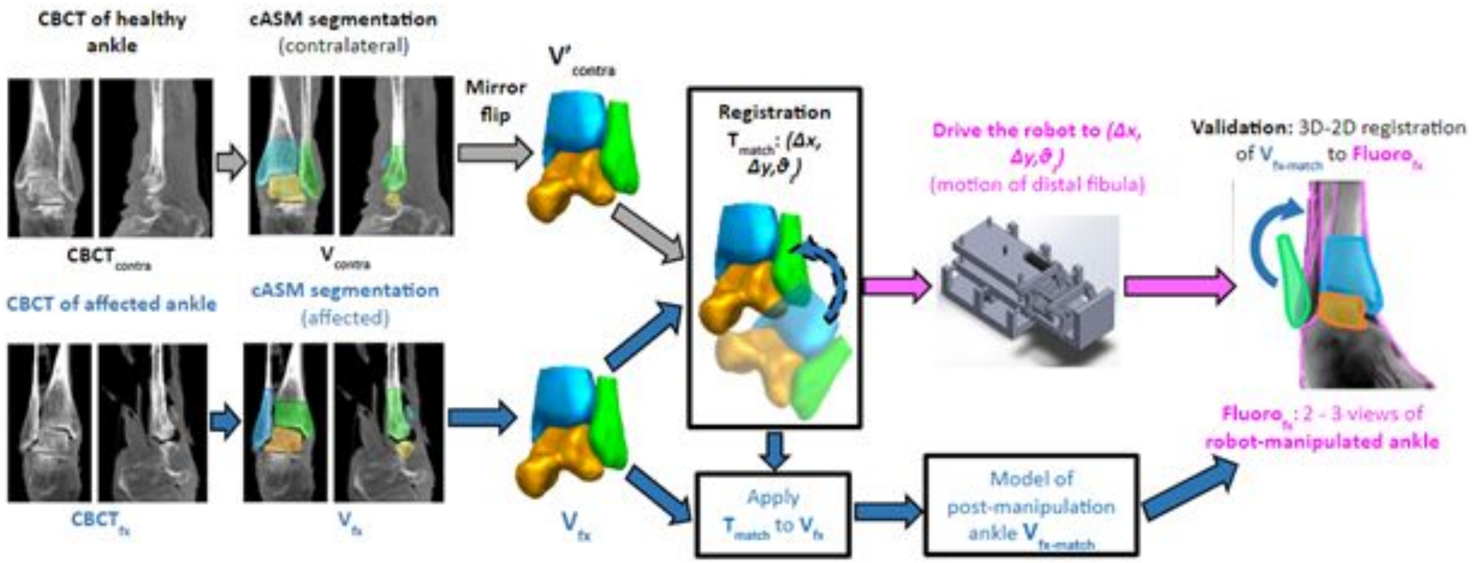
Jayaram Mandavilli
Team Members: Asef Islam, Anthony Wu
Mentors: Dr. Jeffrey H. Siewerdsen and Dr. Wojtek Zbijewski

Project Background

- Ankle fractures cause disruption of the syndesmosis
- Displacement of fibula causes tissue to shift
- During surgical repair surgeons are required to manually estimate where to shift the fibula to and where to place the fixation
- Leads to Post Traumatic Osteoarthritis in 70% of ankle fracture cases
- Requires surgical re-intervention



Technical Approach

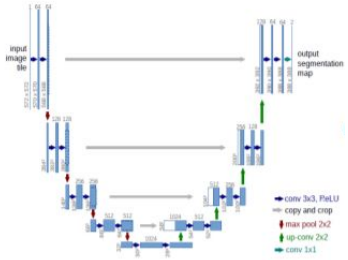


Segmentation Approaches

- Coupled Active Shape Model (cASM)
  - Image segmentation
  - Trained with consideration of multiple bones
    - Better distinguish possible overlaps in bones from ASM
  - Have worked on high resolution cone beam CT (CBCT) images
    - Have not been used on C-arm images, which are lower quality than CBCT
- Deep Learning
  - Either for initialization of cASM or conduct entire image segmentation
  - Mean segmentation error less than 2 mm for final model
  - Trying convolutional neural nets (CNN) and U-Nets

U-Net

- Designed to work with less data
  - Would allow us to achieve high segmentation accuracy with our limited data
- U-Net has been applied to medical imaging projects
  - This interest in U-Nets in the medical field has been increasing
- Established architecture
  - There is a lot of documentation
  - Proof that it has been successful for certain projects
  - Lot of research building on it

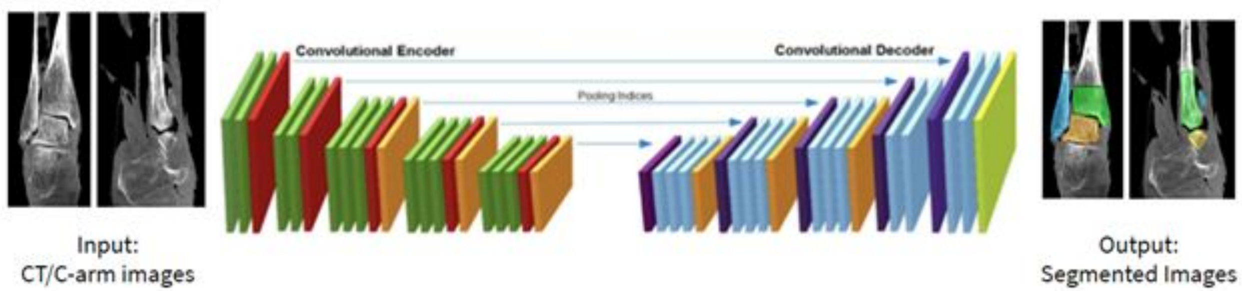


Paper Selection

- Wanted to find a neural network architecture that could be applied to our project
- Paper: NAS-Unet: Neural Architecture Search for Medical Image Segmentation
  - Variation to already established U-Net to make it even better
  - Could lead to higher segmentation accuracy than just regular U-Net in our project also
- Need to find best structure for limited data
- This paper gives us another architecture that we can build on

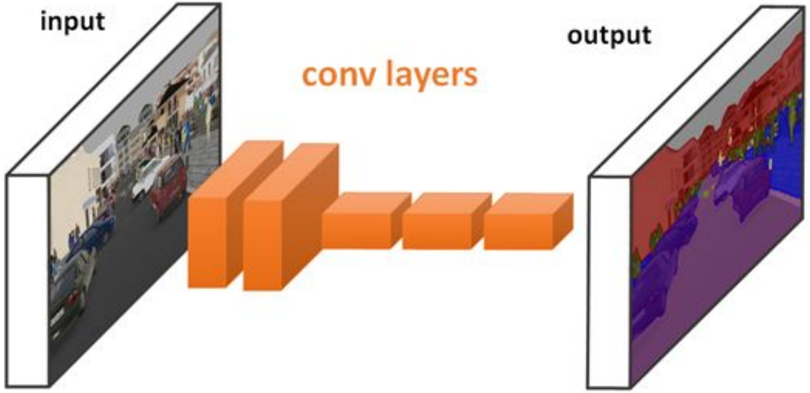
Paper Deep Learning Background

- Semantic segmentation classifies each pixel instead of identifying objects
  - As deep learning has evolved, it has moved away from semantic segmentation (outside of medical imaging more interested in objects and higher level classifications)
  - Need semantic segmentation for most medical applications, including our project
- CNN have shown to have the highest success rate for general image segmentation and are thus commonly used
- They work by taking patches around each pixel (looking at features in image) to classify each pixel and create a multi channel likelihood map
- Pooling (like max pooling) is used to alleviate the computational burden



Paper Deep Learning Background Continued

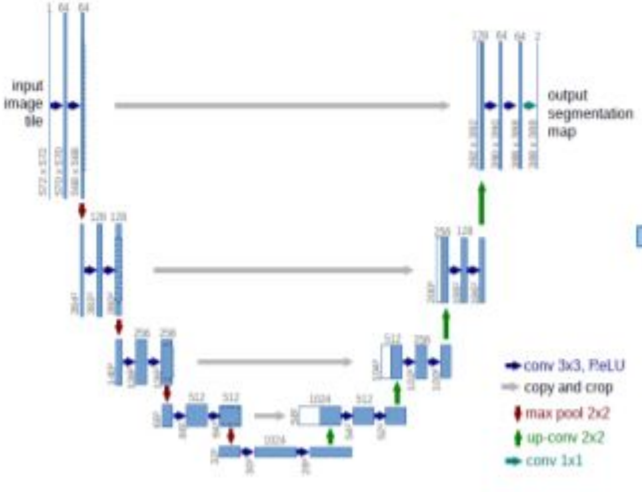
- Fully Convolutional Networks (FCNs) were looked into as alternatives to classic CNNs
  - FCNs do not have any fully connected layers (while CNNs have some convolutional and some fully connected layers), replace these with up sampling layers
  - They can take in any sized input images due to no fully connected layers
  - Fully connected layers require a lot more computation

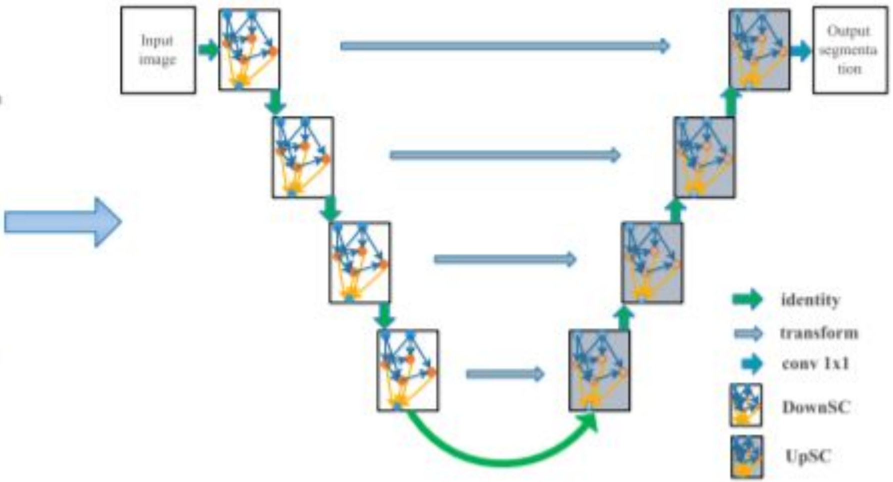


Paper Summary (Deep Learning Background) Continued

- U-Net is a type of CNN similar to FCN
  - FCN only up samples once but U-Net has multiple up sampling layers
  - U-Net uses skip connections to connect each pair of down sampling and up sampling layers
    - This makes spatial information connect to deeper layers
    - Leading to higher segmentation accuracy
- Neural Architecture Search (NAS) is a variation of U-Net and used in this paper
  - Has the “U” backbone
  - Use different primitive operation sets for up scaling and down scaling (so not all just convolutional steps in each scale)
  - They searched for what to use in up scaling and down scaling layers and found the best option using Neural Architecture Search (NAS)
  - They also replaced the skip connections with cweight operations (assigns weights to feature channels to show which features are redundant and which ones are useful)

Unet vs NAS-Unet

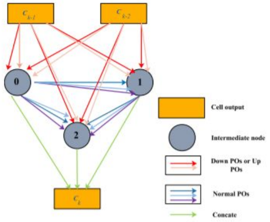




Experiment Overview

- Cell based architecture, search for best architecture in each cell
  - Still using the U-Net backbone
  - Replace all blocks with cells searched by NAS
  - The cell architecture found is shared throughout the network
- Selecting primitive operations
  - No repetitions in the same cell
  - Use less parameters than U-Net by using depthwise separable convolution
- Search strategy
  - Over parameterized structure where each edge is a mixed operation

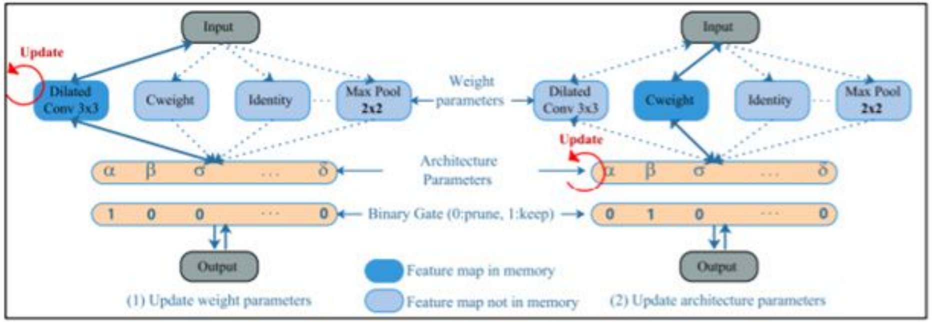
$$MixO(x) = \sum_{(i=1)}^{N} w_i o_i(x)$$



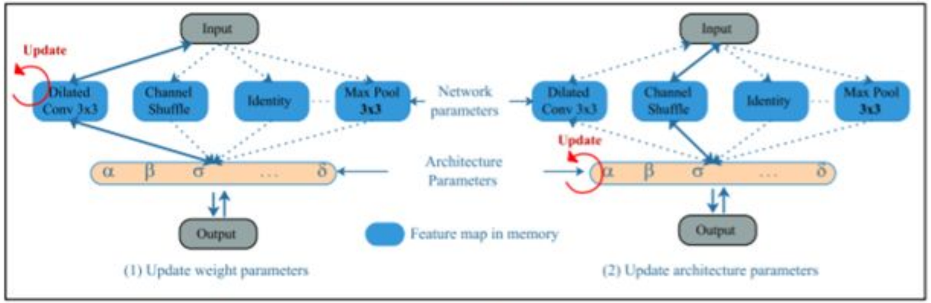
  - They found a much more efficient operation parameter update strategy

Experiment Search Strategy

POs type	Down POs	Up POs	Normal POs
1	average pooling	up cweight	identity
2	max pooling	up depth conv	cweight
3	down cweight	up conv	dilation conv
4	down dilation conv	up dilation conv	depth conv
5	down depth conv	-	conv
6	down conv	-	-



(a)



(b)

Experiments Conducted

- 3 different experiments
  - Each one utilized a different data set
  - The data sets varied by both image format and anatomical features
- Tested with variety of medical image formats
  - MRI
  - CT
  - Ultrasound
- Tested multiple architectures including established ones to use as baseline
  - U-Net
  - FC-densenet
  - NAS-Unet

Results - Experiment 1

- 50 Cases, 1250 slices of MRI prostate images

Model Type	mIoU	DSC	Train Time	GM
U-Net	0.978	0.938	**6h**	**1.3**
FC-Densenet	0.982	0.956	23h	2.7
NasUnet	**0.983**	**0.9737**	8h	1.5

Results - Experiment 2

- Used database of Abdominal CT and MRI images
- Segmented liver in CT, and other abdominal organs in MRI images
- 40 patient CT images, 2874 slices for training and 1408 slices for testing
- 1594 MRI slices for training and 1537 MRI slices for testing

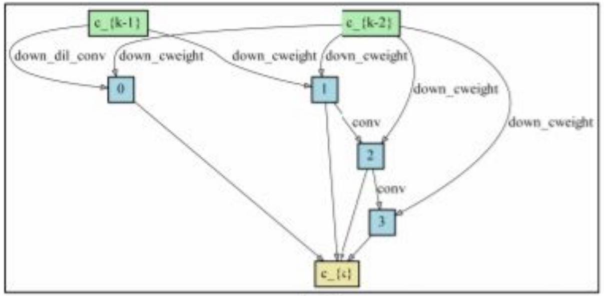
Model Type	mIoU	DSC	Train Time	GM
U-Net (CT)	0.982	0.937	**1d-6h**	**2.4**
FC-Densenet (CT)	0.983	0.965	3d-4h	6.84
NasUnet (CT)	**0.985**	**0.974**	1d-15h	3.5
U-Net (MR)	0.46	0.682	**9h**	**1.3**
FC-Densenet (MR)	0.51	0.734	21h	2.7
NasUnet (MR)	**0.54**	**0.76**	11h	1.5

Results - Experiment 3

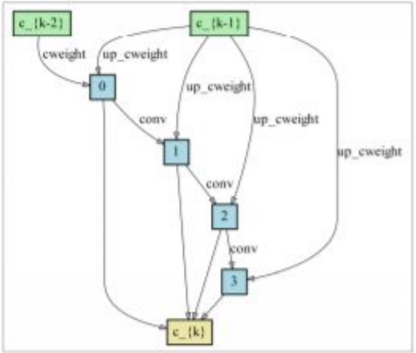
- Brachial plexus (for segmentation of nerves) ultrasound images, 5635 training images an 5508 testing images

Model Type	mIoU	DSC	Train Time	GM
U-Net	0.989	0.74	**18h**	**2.3**
FC-Densenet	0.989	0.844	2d-3h	6.85
NasUnet	**0.992**	**0.881**	1d-3h	3.4

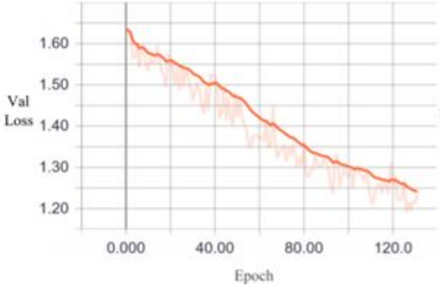
Overall Results

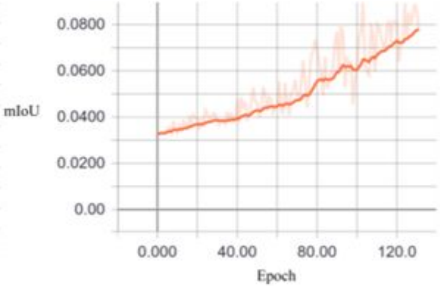


(a)



(b)





Conclusions and Relavence

- They were able to prove that for various image types, their NAS-Unet had the highest segmentation accuracy for each image type (mean intersection over union)

$$IoU = 100 * \frac{true-positives}{true-positives+false-negatives+false-positives}$$

- Their architecture took significantly longer to train than the U-Net
- High accuracy suggests we can apply this to our project
  - Might need to slightly modify so it is better suited to bones instead of organs
  - Also need to modify so that have separate output matrices for each bone in ankle
  - Can use same basis (so U-Net bakbone and cell based architecture) but to find the best cell architecture could either use theirs or do what they did and use a memory saving search algorithm to find the best cell architecture

Paper Pros

- Looked at multiple image formats (including CT dicom images which is very relevant to this project)
- Compared to established structures like FC and U-Net
- Showed methods to find best cell architectures
  - Can use this to find a better suited cell architecture for our project
- All code is open source

Paper Cons

- Though very relevant to project, the images were not of bones so might need to edit architecture
- Used different number of slices for each image type, so difficult to compare how the architectures worked for each image type
- Should have reported metrics on training data accuracy (to show no overfitting)

References

[1] Weng, Y., Zhou, T., Li, Y., & Qiu, X. (2019). NAS-Unet: Neural Architecture Search for Medical Image Segmentation. *IEEE Access*, *7*, 44247–44257. doi: 10.1109/access.2019.2908991