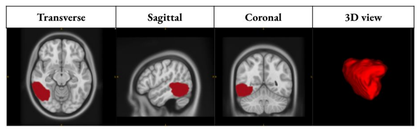
RESSEG and DeepResection: A Comparison of Classifiers Trained on Simulated and Real Resections

Joseph P. Dong, Chelsea Pan, and Campbell Arnold

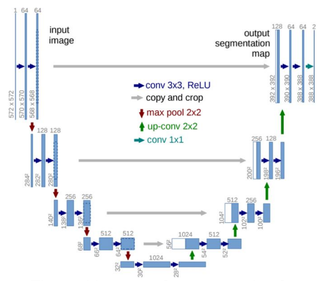
Davis Lab for Electrophysiology and Neuroimaging, University of Pennsylvania

BACKGROUND

- Resection Segmentation
  - Analyzing postoperative epilepsy patients after brain resection surgery may improve patient outcomes
  - This analysis requires segmentation, the process of labelling where the brain has been resected
  - Neural networks, in particular U-Net Convolutional Neural Networks (U-Net CNNs) can be trained to find and segment resections
- Training Data
  - Training neural networks to classify tissue resection boundaries requires labeled MRIs
    - Manually labelling takes skill and time
    - Inter-rater variability
  - Resections can be simulated to augment the training data
- Objective
  - DeepResection is a classifier trained on manual labels, and RESSEG is a classifier trained on simulated resections
  - We will compare DeepResection with RESSEG to see how we can improve the methods



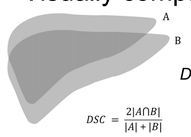
An example of a simulated resection



The architecture of one example of a U-Net CNN

METHODS

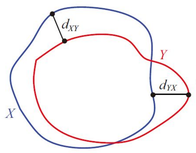
- Installed RESSEG on laboratory server
- Ran a validation study using 15 postoperative images
- Manually segmented T1-weighted MRI images for a ground truth dataset (Manual)
- Calculated Dice similarity coefficients (DSC), 100% Hausdorff distance, and 95% Hausdorff distance for classifier outputs
  - RESSEG vs Manual
  - DeepResection (DR) vs Manual
- Computed T-test scores for the references
- Visually compared failure cases for each classifier



$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

DSC: Dice similarity coefficient

Dice similarity coefficient computes overlap



Hausdorff distance calculates distance between labels

RESULTS

	RESSEG vs Manual			DR vs Manual		
MRI	*DSC*	*100% Hausdorff*	*95% Hausdorff*	*DSC*	*100% Hausdorff*	*95% Hausdorff*
1	0.8142	19.3688	9.7522	0.7234	25.5390	16.6066
2	0.8445	24.0190	13.8107	0.8950	7.2674	2.4017
3	0.0000	134.2435	130.1204	0.3883	45.0597	38.9661
4	0.4600	39.1377	34.3847	0.0000	49.1421	43.5297
5	0.5696	49.8785	40.5744	0.6835	19.0865	11.1891
6	0.0000	88.8236	85.9556	0.0000	110.7672	102.5498
7	0.7441	21.9533	13.2083	0.7528	15.7769	7.1392
8	0.6551	44.0305	38.3112	0.8585	17.3608	3.6792
9	0.7785	41.4801	31.9893	0.4992	24.0000	20.0000
10	0.7985	39.5463	29.8309	0.0000	127.4176	119.0696
11	0.6611	36.0218	29.4975	0.7988	36.3213	30.6788
12	0.8053	27.4566	13.8587	0.0000	95.4614	85.9215
13	0.7184	17.1191	13.1474	0.7714	11.0864	7.0000
14	0.6897	38.8598	31.0184	0.7514	21.1640	11.7187
15	0.8663	15.8260	4.9842	0.8285	28.2205	16.1703
Mean	0.6270	42.5176	34.6963	0.5301	42.2447	34.4414

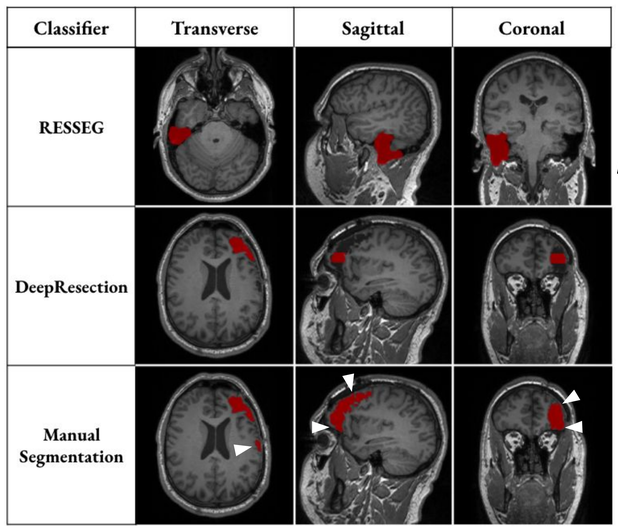
A table of all DSC, 100% Hausdorff, and 95% Hausdorff values for RESSEG and DR relative to the manual dataset. Hausdorff distance in millimeters. Green indicates agreement with manual labels, red indicates classifier error

	Dice Similarity Coefficient	100% Hausdorff distance	95% Hausdorff distance
RESSEG vs DeepResection T-Test	*P* = 0.30	*P* = 0.98	*P* = 0.95

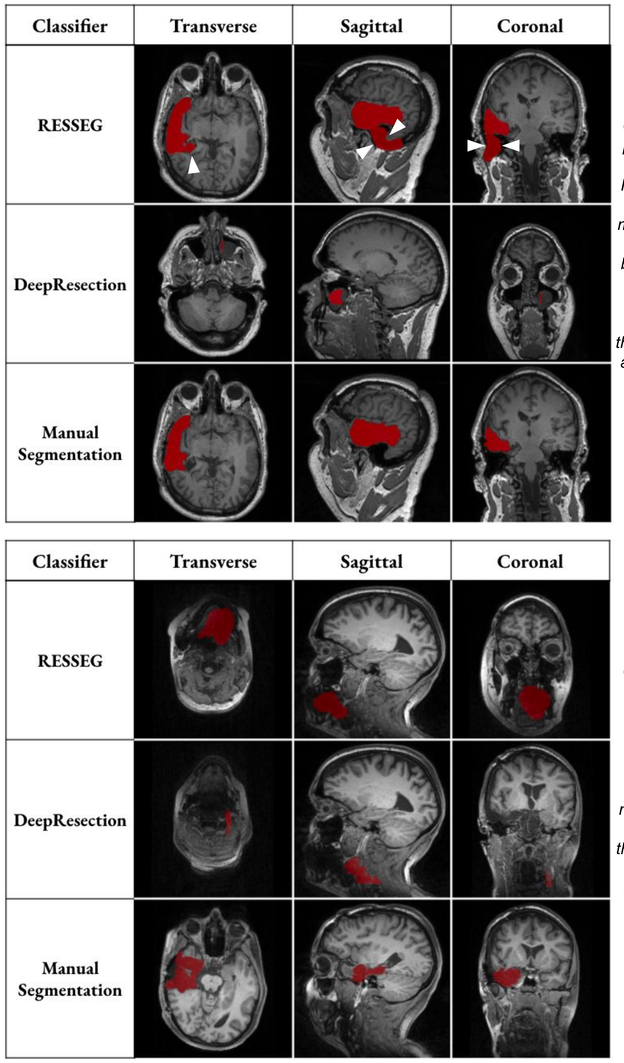
RESSEG and DR are not significantly different from each other relative to Manual

	100% Hausdorff vs 95% Hausdorff T-test
RESSEG	*P* < .001
DR	*P* < .001

The 100% Hausdorff and 95% Hausdorff distance metric are significantly different



Classifier outputs for patient 3. Frontal lobe resection. RESSEG labeled a space inferior to the patient's brain. DR labeled accurately in the transverse view but incompletely in the sagittal and coronal views. DR misclassified the areas labeled with arrows.



Classifier outputs for patient 10. Temporal lobe resection. RESSEG labeled the resection but misclassified adjacent ventricle and space below brain (arrows). DR misclassified a region in the nose texturally similar to that of a resection. DR also classified mostly along the sagittal plane

Classifier outputs for patient 6. Temporal lobe resection with heavy bias field artifact. RESSEG mislabeled the oral cavity. DR misclassified a region texturally similar to that of a resection due to bias field artifact. DR classified mostly along the sagittal plane

CONCLUSIONS

- Failure cases for RESSEG and DR were analyzed and compared, with the end goal of improving upon DR's methods
- Future Direction and Improvements
  - Model Architecture
    - Implement a 3D classifier
  - Data Augmentation
    - Increased image intensity range
    - Image cropping for simulated partial FOV images
  - Image Preprocessing
    - Skull stripping and brain extraction
    - Bias field correction