JULY 12–16 2020 VIRTUAL JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

# Normal Tissue and Tumor Segmentation Using V-Net Regularized by YOLO

C HSU[1], C MORIN[2], T KIRBY[3], M METZGER[4], J FLERLAGE[5], S KASTE[2], M KRASIN[1], B SHULKIN[2], J T LUCAS JR.[1]

(1) Department of Radiation Oncology, St. Jude Children's Research Hospital, Memphis, TN
(2) Department of Diagnostic Imaging, St. Jude Children's Research Hospital, Memphis, TN
(3) Department of Computer Science, Rhodes College, Memphis , TN
(4) Department of Global Pediatric Medicine, St. Jude Children's Research Hospital, Memphis, TN
(5) Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN

St. Jude Children's Research Hospital

## INTRODUCTION

Deep learning approaches have been successfully applied to tumor segmentation in medical imaging. Specifically, V-Net was introduced for 3D image segmentation with end-to-end capabilities. In a recent work, a variational encoder (VAE) branch was attached to V-Net for improving segmentation results in brain gliomas [1]. VAE regularizes the encoder by adding image reconstruction as additional constraints.

Besides segmentation networks, object detection networks also show performance improvement by adding segmentation as an auxiliary task. YOLO is an object detection network that transforms object detection into a regression problem [2]. YOLO optimizes object location, height and width of the bounding box, confidence of object, and class of the detected object. By doing so, YOLO can detect objects efficiently and reliably.

Automated tumor segmentation from Fluorodeoxyglucose ($^{18}$FDG)-positron emission tomography and computed tomography ($^{18}$FDG-PET/CT) have the potential to aid diagnosis and treatment response evaluations. Manual segmentation is time consuming and prone to bias. Conventional segmentation methods are often obscured by radiotracer uptake from normal tissues (brain, heart, liver, bilateral kidneys, & bladder), some of which are variably avid (heart, liver, and kidneys).

New tools are needed for reliable quantification of FDG-avid disease relative to FDG-avid normal tissues.

## AIM

To propose a deep learning model for improved segmentation accuracy in PET/CT images.

## METHOD

• Pre-treatment $^{18}$FDG-PET/CT images from 34 pediatric patients with Hodgkin lymphoma were collected with Institutional Review Board (IRB) approval. CT and PET images were acquired sequentially. Rigid image registration was performed to ensure alignment of the images. Figure 1 shows an example of the segmented tissue masks overlay on Pet and CT.

• **Patient Dataset:** The patients were randomly separated into a training set of thirty patients and validation set of four patients. No data augmentation was performed during training.

• **Network Structure:** we propose a semantic segmentation network based on autoencoder architecture reinforced by an object detection branch for segmenting normal tissue and tumor from 3D $^{18}$FDG-PET/CT. The autoencoder network consists of symmetrical encoder and decoder with ResNet blocks and skip connections. The object detection branch is derived from You only look once (YOLO) -real time object detection which is added at the end of the encoder as shown in Figure 2.

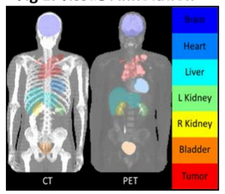
• **2D-->3D Adaptation**: We expand the utilization of YOLO beyond its original application to 2D object detection by adapting it for 3D images. This is accomplished by adding Z location, depth (D), and class probabilities to the existing prediction parameters (location (X, Y), height (H), width (W), and confidence (C)). Therefore, our YOLO inference layer outputs 8x8x16 with 15 filters (X,Y,Z,W,H,D,C, 8 classes). A single convolutional layer with kernel size 1 was used to convert the encoder output to YOLO branch. Sigmoid activations were applied to XYZ and confidence, while exponential was applied to WHD. Instead of sigmoid activation, we used softmax activation for class probabilities.

• **Loss function:** Included the segmentation loss and the object detection loss. For segmentation loss, we used categorical cross entropy. Object detection loss is the sum of four terms, including location (XYZ), box dimension (WHD), confidence, and. Mean square error was used for XYZ loss and WHD loss. Categorical cross entropy was used for confidence loss and class loss.

• **Training and Model Selection:** We trained three models including VNet, VNet with VAE (VNet-VAE), and VNet with YOLO (VNet-YOLO). The VNet is constructed using symmetrical encoder and decoders. The VNet-VAE uses the same autoencoder structure with the VAE following identical design of the decoder. All three models use categorical cross entropy for segmentation loss. Additionally, VNet-VAE included L2-norm regularization and KL divergence shown in with loss weights following suggestion in the original paper.

• **Optimization:** We used an Adam optimizer [13] with constant learning rate of 0.0001 and trained for 500 epochs with batch size of 1. Segmentation results were compared using F1 score and Hausdorff distance.

**Fig 1. Tissue Annotation**



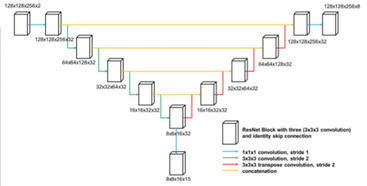




**Fig 2. Proposed network structure.**

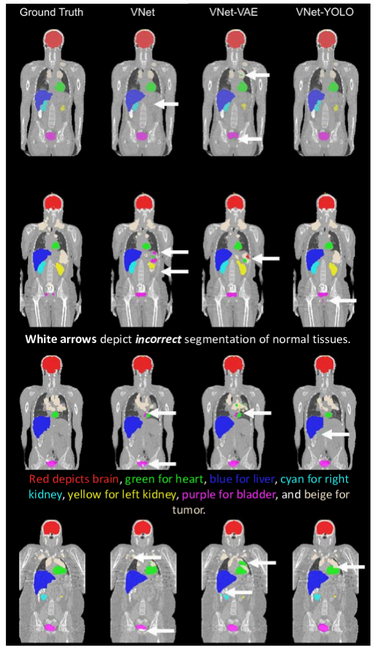
## RESULTS

**Table 1. Final F1 score for the validation set**

| Model | Patient Index | Brain | Heart | Liver | L. Kidney | R. Kidney | Bladder | Tumor |
|---|---|---|---|---|---|---|---|---|
| VNet | 1 | 0.93 | 0.95 | 0.90 | 0.62 | 0.83 | 0.94 | 0.81 |
| | 2 | 0.97 | 0.62 | 0.91 | 0.43 | 0.89 | 0.61 | 0.48 |
| | 3 | 0.98 | 0.79 | 0.91 | 0.72 | 0.72 | 0.83 | 0.68 |
| | 4 | 0.97 | 0.84 | 0.91 | 0.85 | 0.81 | 0.73 | 0.58 |
| VNet-VAE | 1 | 0.94 | 0.91 | 0.88 | 0.72 | 0.83 | 0.94 | 0.82 |
| | 2 | 0.96 | 0.74 | 0.91 | 0.71 | 0.87 | 0.80 | 0.55 |
| | 3 | 0.98 | 0.79 | 0.93 | 0.68 | 0.78 | 0.88 | 0.64 |
| | 4 | 0.97 | 0.70 | 0.92 | 0.85 | 0.77 | 0.87 | 0.60 |
| VNet-YOLO | 1 | 0.96 | 0.97 | 0.96 | 0.96 | 0.95 | 0.94 | 0.86 |
| | 2 | 0.96 | 0.96 | 0.96 | 0.97 | 0.97 | 0.86 | 0.75 |
| | 3 | 0.97 | 0.96 | 0.96 | 0.97 | 0.94 | 0.84 | 0.88 |
| | 4 | 0.97 | 0.95 | 0.96 | 0.98 | 0.96 | 0.85 | 0.83 |

- **Brain segmentation:** All 3 models demonstrated reliable performance with F1 scores >0.9 for all 4 patients. Brain location & FDG-avidity were consistent and thus, regularization using VAE or YOLO did not result in significant improvement.
- **Heart Segmentation:** VNet failed to segment heart in patient 2 with F1 of 0.62. VNet-VAE consistently segmented heart with F1 score above 0.7. VNet-YOLO has F1 score >0.9 for all 4 cases. Segmentation was not impeded by inconsistent signals from PET & seemed to benefit from closely approximating ribs & lungs.
- **Liver Segmentation:** Liver was consistently segmented by all 3 models with F1 scores >0.9 despite lack of signal in PET This could be due to that liver is adjacent to the lungs which is filled with air & great contrast was provided by CT between air & soft tissue.
- **L. Kidney Segmentation:** L kidney was the most challenging to segment. VNet can only segment 2 patients with F1 score above 0.7. VAE regularization improved L. kidney segmentation with only 1 patient <0.7 for F1 score. VNet-YOLO exhibited no difficulties in segmenting left kidney. This could be because lack of intensity contrast in surrounding tissues on CT & inconsistent signal in PET for left kidney. On the other hand, due to adjacency to liver, the models could still segment R. kidney reliably with F1 scores all >0.7.
- **Bladder segmentation:** VNet struggled with patient 2 (0.61 F1). Both VNet-VAE & VNet-YOLO demonstrated reliable segmentation with F1 scores all >0.8. Bladder location & intensity was consistent on PET & is surrounded by pelvis with great contrast in CT.
- **Tumor Segmentation:** Tumor was the most difficult to segment. VNet segmented 3 patients with F1 score <0.7. VNet-VAE did not improve tumor segmentation significantly. VNet-YOLO improved segmentation performance (all >0.7 in F1 score).

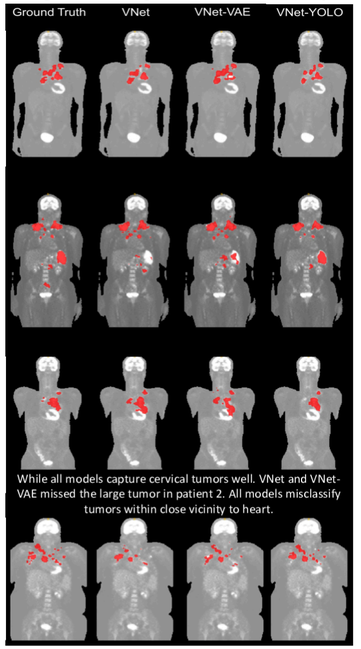
**Fig 3. Normal Tissue Segmentation - CT overlay**




**White arrows** depict ***incorrect*** segmentation of normal tissues.

Red depicts brain, green for heart, blue for liver, cyan for right kidney, yellow for left kidney, purple for bladder, and beige for tumor.

**Fig 4. Tumor Segmentation - PET overlay**




While all models capture cervical tumors well. VNet and VNet-VAE missed the large tumor in patient 2. All models misclassify tumors within close vicinity to heart.

**Table 2. Final Hausdorff distance for the validation set**

| Model | Patient Index | Brain | Heart | Liver | L. Kidney | R. Kidney | Bladder | Tumor |
|---|---|---|---|---|---|---|---|---|
| VNet | 1 | 1.73 | 87.30 | 5.10 | 52.21 | 2.24 | 1.41 | 99.58 |
| | 2 | 91.03 | 86.28 | 16.52 | 18.63 | 3.16 | 100.29 | 9.70 |
| | 3 | 78.61 | 39.31 | 3.16 | 31.95 | 66.83 | 130.07 | 56.76 |
| | 4 | 90.59 | 46.05 | 56.44 | 65.15 | 79.32 | 121.41 | 68.48 |
| VNet-VAE | 1 | 2.24 | 1.73 | 5.39 | 6.40 | 2.24 | 1.41 | 121.53 |
| | 2 | 90.13 | 22.87 | 20.52 | 11.18 | 2.24 | 99.78 | 13.04 |
| | 3 | 73.71 | 8.60 | 3.16 | 2.00 | 57.08 | 92.46 | 17.72 |
| | 4 | 43.46 | 37.75 | 4.69 | 30.43 | 3.00 | 120.31 | 11.58 |
| VNet-YOLO | 1 | 2.00 | 1.00 | 14.32 | 1.00 | 1.41 | 2.83 | 1.00 |
| | 2 | 1.41 | 1.41 | 2.83 | 1.41 | 1.41 | 9.49 | 1.73 |
| | 3 | 2.24 | 1.00 | 25.71 | 1.41 | 13.45 | 11.36 | 1.41 |
| | 4 | 4.24 | 1.41 | 17.23 | 1.00 | 1.41 | 30.77 | 2.00 |

- **Brain segmentation:** VNet and VNet-VAE had three HDs above 20 voxels and patient 1 within 5 voxels. VNet-YOLO segments brain in all four patients within 5 voxels.
- **Heart segmentation:** VNet resulted in all four patients with HD above 20 voxels. VNet-VAE improved the HD with two patients within 10 voxels. VNet-YOLO segments all four heart within 5 voxels.
- **Liver segmentation**: All 3 models have one patient above 20 voxels. VNet-VAE segments liver better than VNet-YOLO with the other three patients within 5 voxels.
- **Left Kidney segmentation**: L. kidney remains challenging for VNet with HD all above 10 voxels. VNet-VAE improved patient 3 to within 5 voxels. Regularization with YOLO proves to be effective with all four HDs under 5 voxels.
- **R. Kidney segmentation:** Using VNet for right kidney segmentation resulted in two patients above 20 voxels. Both regularization techniques improved the segmentation results with only one patient above 10 voxels.
- **Bladder Segmentation: A**ppeared to be the most challenging normal tissue for VNet- and VNet-VAE with three patients above 50 voxels. YOLO regularization improved the segmentation results for all patients to within 30 voxels.
- **Tumor segmentation:** VNet resulted in three patients above 10 voxels. VNet-VAE improved average performance but had all four patients above 10 voxels. VNet-YOLO improved segmentation for all patients to within 5 voxels.

## CONCLUSIONS AND FUTURE DIRECTIONS

**Contributions:** Combining branches to V-Net for improving segmentation results have been demonstrated [1]. Others have used V-Net with classification branches for improving classification performances [3]. We combined object detection with segmentation and demonstrate both feasibility & improve segmentation performance.

**Novelty:**

-Compared to other medical image segmentation projects, PET/CT segmentation resembles more of an object detection problem as the annotations are whole organs compared to tumor subregions in BRATS [4], single instance but multiple class compared to medical image decathlon [5]. Object detection networks with segmentation capabilities may also be good candidates for PET/CT segmentation [6].

**Challenges:**

-While imposing a location regularization for the tissues seems to make the model less generalizable in an object detection problem, we believe that accurate segmentation is worth the compromise & that this constraint is acceptable given the application to anatomic imaging.

-Location regularization results in masks which obey anatomic constraints. Automatic detection & segmentation of normal tissues & tumors by combining segmentation & regularization networks demonstrated in this work substantially improves the feasibility of using deep learning methods for clinical detection & treatment response evaluation of PET & CT images.

-PET images are usually acquired for whole body & the anatomy location may vary significantly between individuals. Even though our study population consists of different body sizes, we cannot rule out overfitting of the location regularization.

**Future Directions:**

-Performance of small lesion detection might be improved by introducing multi-scale networks as described in [7]. Furthermore, shape priors such as anchor boxes may lead to further improvement of segmentation accuracy [8].

-Overfitting of the location regularization may be improved by incorporating manual scaling or image registration to a body template.

## REFERENCES

[1] Myronenko, A. 3D MRI brain tumor segmentation using autoencoder regularization. in International MICCAI Brainlesion Workshop. 2018. Springer.
[2] Redmon, J., et al. You only look once: Unified, real-time object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[3] Mehta, S., et al. Y-net: Joint segmentation and classification for diagnosis of breast biopsy images. in International Conference on Medical Image Computing and Computer-Assisted Intervention. 2018. Springer.
[4] Menze, B.H., et al., The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging, 2015. **34**(10): p. 1993.
[5] Simpson, A.L., et al., A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063, 2019.
[6] He, K., et al. Mask r-cnn. in Proceedings of the IEEE international conference on computer vision. 2017.
[7] Lin, T.-Y., et al. Feature pyramid networks for object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
[8] Ren, S., et al. Faster r-cnn: Towards real-time object detection with region proposal networks. in Advances in neural information processing systems. 2015.

## ACKNOWLEDGEMENTS

The authors thank Dr. Thomas Merchant, department head of Radiation Oncology for his full support to this project. And Summer Program of Department of Computer Science at Rhodes College for their full support.

## CONTACT INFORMATION

chih-yang.hsu@stjude.org
john.lucas@stjude.org
@johnthomas75
john-lucas-5106628/