Identification of Kernels in a Convolutional Neural Network: Connections Between Level Set Equation and Deep Learning for Image Segmentation

Jonas A. Actor
Computational and Applied Mathematics
Rice University
Houston, TX USA
jonasactor@rice.edu

David Fuentes
Imaging Physics
MD Anderson Cancer Center
Houston, TX USA
dtfuentes@mdanderson.org

Beatrice Riviere
Computational and Applied Mathematics
Rice University
Houston, TX USA
riviere@rice.edu

***Abstract*—Two common techniques for image segmentation - level set methods and convolutional neural networks (CNN) - rely on alternating convolutions with nonlinearities to describe image features: neural networks with mean-zero convolution kernels can be treated as upwind finite difference discretizations of differential equations. Such a comparison provides a well-established framework for proving properties of CNNs, such as stability and approximation accuracy. We test this relationship by constructing a level set network, a CNN where forward-propagation is equivalent to solving the level set equation. The level set network achieves comparable segmentation accuracy to solving the level set equation, while not obtaining the accuracy of a CNN. We therefore analyze which convolution filters are present in our CNN, to see whether finite difference stencils are learned during training. We observe certain patterns form in the decoding layers of the network, where kernels cannot be accounted for by finite difference stencils alone.**

***Index Terms*—image segmentation; convolutional neural networks; numerical analysis; clustering**

I. AUDIENCE

This talk is an intermediate-level technical talk, for researchers and data scientists working in image analysis, computer vision, deep learning, and medical imaging. This talk requires some familiarity with neural networks and numerical methods for differential equations.

II. INTRODUCTION

Liver cancer is the sixth most common form of cancer annually; in 2018, liver cancer was the fourth most common ICD-10 cancer-related code specified for cancer-related deaths globally [1]. The majority of liver cancer cases are instances of hepatocellular carcinoma (HCC) [2]. A diagnosis of HCC relies heavily on the results of biopsy and medical imaging [3]. However, intra-observer variability of medical images complicates the diagnosis process [4]. Image segmentation addresses this issue by providing a reproducible measure to evaluate liver and tumor volumetric data.

While many methods have been employed for both automatic and semiautomatic image segmentation [5], we focus on level set methods and deep convolutional neural networks, with the aim of combining these two frameworks to provide a fast, accurate, and interpretable segmentation model. Level sets and CNNs are considered the current standards for medical image segmentation. Both level sets and CNNs rely on convolutions to detect and explain image features. Upwind finite difference approximations, such as those in the fast marching method implemented in ITK-SNAP [6], can be expressed as the convolution of finite difference stencils followed by a ReLU nonlinearity. As such, a forward Euler discretization of the level set equation can be written in the same language as a CNN: a series of convolutions followed by nonlinear activation functions. In this work, we examine which types of convolution kernels are important for image segmentation of the liver, and we compare how well the level set methods fair at liver segmentation when we replace the finite difference stencils in the level set equation with convolutions learned during training.

III. CURRENT STATE OF THE ART

Level set segmentation methods conduct image segmentation as propagating either a region or a curve within an image as to match the desired region in question. The evolution of this curve is described by the level set equation, which couples the curvature of the expanding region, image intensities and gradients to specify exactly how the curve evolves in space and time [7]. However, level set methods are only semiautomatic, requiring an intial configuration for propagation, and they are comparatively expensive to evaluate, especially for 3D imaging modalities [7].

CNN architectures have achieved remarkable accuracy in several online benchmarks and challenges; for example, the UNet [8] and ResNet [9] architectures both displayed fundamental improvements in medical image analysis, particularly for image classification. In the MICCAI LiTS Challenge 2017, many of the top-performing entrants used some type of CNNs [5]. However, CNNs are complex systems, often treated as black boxes, lacking interpretability and difficult to analyze.

IV. METHODS AND KEY RESULTS

Both level sets and CNNs rely on convolutions of fixed stencils to explain image features. For the level set equation, the chosen stencils are finite difference stencils designed to detect edges, used to weight forward and outward expansion of the level set curve. Upwind finite difference approximations, such as the in the fast marching method implemented in ITK-SNAP [6], can be expressed as finite difference convolutions followed by a ReLU-style nonlinearity. As such, a forward Euler discretization of the level set equation can be written in the same language as a CNN: a series of convolutions followed by nonlinear activation functions. We exploit this relationship between CNNs and numerical differential equations to design a neural network whose architecture and connections mirror the structure of solving the level set equation, while taking advantage of the flexibility of learning convolution kernels as in a CNN.

To do so, we unrolled a numerical method for solving the level set equation, creating a *level set network* (LSN). In this framework, each timestep becomes a layer in a CNN. This concept, of treating layers in a CNN as a system of differential equations, has gained recent attention using the ResNet architecture in the context of dynamical systems [10]. However, these neural network formulations do not assume that the differential equation in question has a specific form. As image segmentation has been accomplished using the level set equation, it is intuitive to construct a neural network that approaches this specific PDE. The LSN maintains the architecture of solving the level set equation, but replaces the finite difference operators with learned convolution kernels.

To test this concept, we employed three segmentation methods on the MICCAI 2017 LiTS Challenge dataset

TABLE I: DSC scores for each fold, from training the level set network.

K-Fold	ITK-SNAP	UNET	LSN Test	LSN Validation
0	0.736	0.912	0.837	0.619
1	0.600	0.919	0.847	0.729
2	0.483	0.874	0.116	0.005
3	0.730	0.895	0.827	0.606
4	0.643	0.915	0.831	0.596
Avg	0.640	0.903	0.692	0.511
Avg$\setminus\{2\}$	0.604	0.911	0.837	0.638

[5], consisting of 131 abdominal contrast-enhanced CT image stacks. These methods were: UNet [8], a type of CNN; ITK-SNAP [6], a segmentation application using the level set equation; and our level set network (LSN). Our UNet architecture is sketched in Figure 1. For UNet and LSN, we performed a 5-fold cross validation, training via the Adadelta optimizer until saturation (40 epochs for UNet, 20 for LSN), with the Dice Similarity Coefficient (DSC) as the loss function. For ITK-SNAP, which allows users to hand-tune a few parameters, we allotted 10 min for a user to alter these parameters, after which ITK-SNAP solved the level set equation until DSC no longer improved.

Using the three methods described above, we obtained the DSC scores listed in Table I. It is somewhat unsurprising that the results from the LSN are roughly on-par with those of ITK-SNAP, as these two methods approach segmentation using the same framework, of a curve propagating outwards, with its expansion rate determined by the underlying image topology. The superior performance of UNets suggest that finite difference kernels do not explain the power of convolutional networks alone.

We confirmed this insight by plotting the convolution kernels obtained from training our UNet, as seen in Figure 2. We first flattened our 3x3 convolution kernels into a vector in $\mathbb{R}^9$, and we then performed clustering with $n = 3$ clusters using k-means, using Euclidean distance in 9-dimensional space. To visualize our results, we projected the kernels using PCA onto the 2-dimensional subspace spanned by the eigenvectors with greatest variation. Onto this projection, we superimposed various hand-crafted kernels that describe common image processing features, such as a Gaussian blurring kernel, and various edge detection kernels; these edge detection kernels are the finite difference stencils used by the level set equation.

Our clustering results, as illustrated in Figure 2, suggest that for many layers in this UNet, there is no clear distinction between various types of kernels.

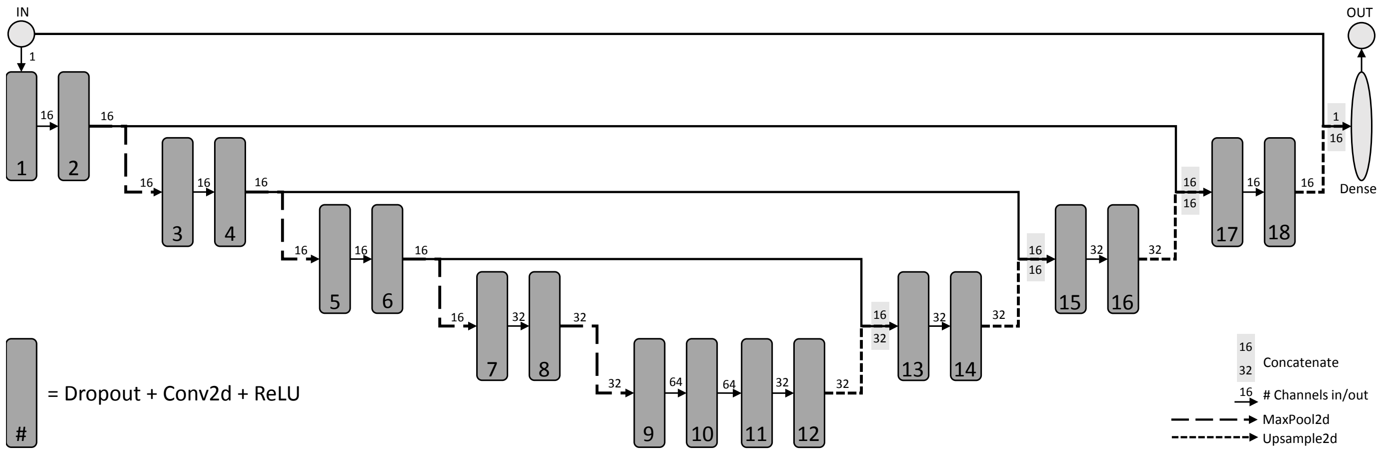
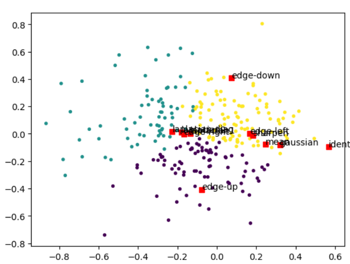
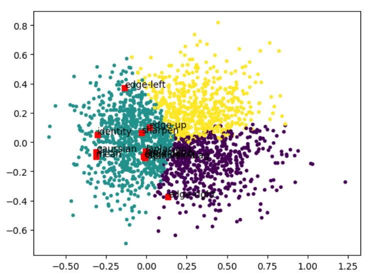


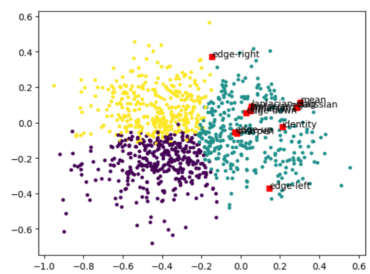
Fig. 1: Schematic of UNet used in kernel experiments.



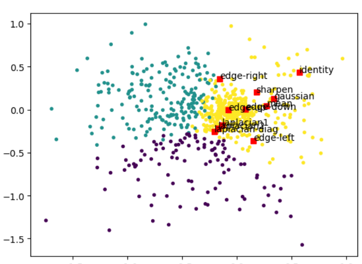
(a) Convolution Layer 2



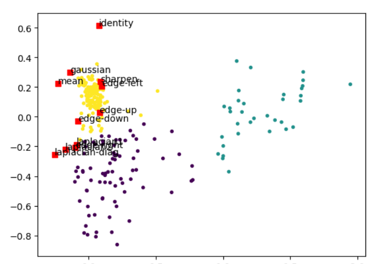
(b) Convolution Layer 11



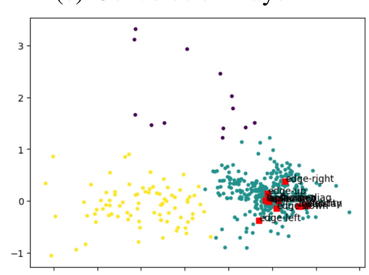
(c) Convolution Layer 12



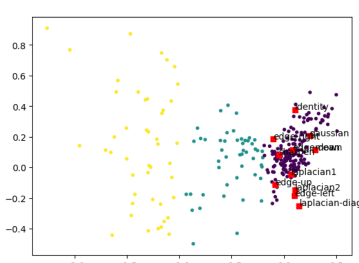
(d) Convolution Layer 13



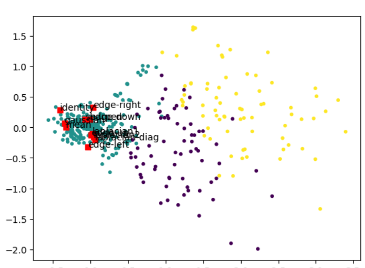
(e) Convolution Layer 14



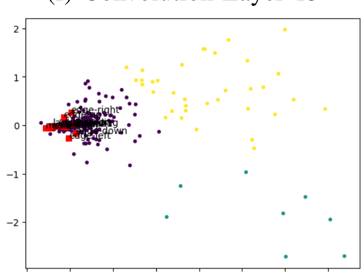
(f) Convolution Layer 15



(g) Convolution Layer 16



(h) Convolution Layer 17



(i) Convolution Layer 18

Fig. 2: Visualization of 3x3 convolution kernels from selected layers of a UNet with a depth of 4. Colors correspond to K-means cluster assignment. Layer numbers correspond to thoes in Figure 1. Layers 1,3-10 (not shown) are similar to Convolution Layer 11.

However, on the decoder side of the UNet, patterns begin to emerge, even if the data do not cleanly fall into clusters: there are several layers, specifically towards the bottom of the UNet and later, where otherwise-uninterpretable convolution features are frequent. We note from these images that there is no clear cluster among the UNet kernels around first-order finite difference stencils i.e. up-down or left-right edge detection kernels; this observation reinforces our previous insight: finite difference kernels alone cannot explain the predictive power of our UNet.

With this work, we have made two main contributions:

1) We have developed a theoretical framework for translating convolutions + ReLUs to upwind finite difference schemes, bridging a gap between numerical analysis and data science. This contribution extends to other problems, such as surrogate modeling or other numerical simulations, where scientific computing methods are employed in conjunction with neural networks. Additionally, we are the first (to our knowledge) to incorporate the nonlinearity of the ReLU function into this treatment of PDEs-as-NNs by using an upwind finite difference scheme, providing a more stable numerical discretization to this interpretation.
2) We have clustered the kernels present in a CNN. Previous efforts employed clustering for CNN model reduction [11], but they do not examine nor report the clusters they produce, only caring for the cluster centroids. This method of clustering kernels is a novel approach to visualize the features learned by a CNN, and produces a product that can subsequently be shown to an imaging scientist or radiologist to provide clinical insight.

V. CONCLUSION

We demonstrate a flexible framework for using numerical analysis to provide insight into CNNs: we interpret upwind finite difference schemes as a convolution layers with ReLU activation functions. However, this alone is not sufficient to explain why CNNs are as accurate as they are. Finite difference kernels combined with ReLU activation functions are not sufficient on their own to obtain high-accuracy image segmentation for liver CT data. Future work will quantitatively measure how changes in kernels effect the network prediction accuracy, and will outline a clinical interpretation to the patterns in decoder kernel behavior as seen above.

VI. PRESENTER BIO

Jonas A. Actor is a current PhD student at Rice University, and an NLM Training Fellow in Biomedical Informatics and Data Science. JAA's research lies at the intersection of computational mathematics, numerical analysis, medical imaging, and machine learning.

ACKNOWLEDGEMENTS

JAA is supported by a training fellowship from the Gulf Coast Consortia, on the NLM Training Program in Biomedical Informatics & Data Science (T15LM007093), with supplement from the Ken Kennedy Institute Computer Science & Engineering Enhancement Fellowship, funded by the Rice Oil & Gas HPC Conference.

REFERENCES

[1] J. Ferlay, M. Ervik, F. Lam, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram, and F. Bray, "Global Cancer Observatory: Cancer Today," International Agency for Research on Cancer, Tech. Rep., 2018.

[2] K. A. McGlynn and W. T. London, "Epidemiology and natural history of hepatocellular carcinoma," *Best Practice & Research Clinical Gastroenterology*, vol. 19, no. 1, pp. 3–23, 2005.

[3] J. Bruix, M. Reig, and M. Sherman, "Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma," *Gastroenterology*, vol. 150, no. 4, pp. 835–853, 2016.

[4] P. Schmitt, E. Mandonnet, A. Perdreau, and E. D. Angelini, "Effects of slice thickness and head rotation when measuring glioma sizes on MRI: in support of volume segmentation versus two largest diameters methods," *Journal of neuro-oncology*, vol. 112, no. 2, pp. 165–172, 2013.

[5] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser *et al.*, "The liver tumor segmentation benchmark (lits)," *arXiv preprint arXiv:1901.04056*, 2019.

[6] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[7] T. F. Chan and J. J. Shen, *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*. SIAM, 2005, vol. 94.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] L. Ruthotto and E. Haber, "Deep neural networks motivated by partial differential equations," *arXiv preprint arXiv:1804.04272*, 2018.

[11] S. Son, S. Nah, and K. Mu Lee, "Clustering convolutional kernels to compress deep neural networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 216–232.