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

Jonas A. Actor[a], David T. Fuentes[b], and Beatrice Riviere[a]

[a]Rice University, Houston, TX USA
[b]University of Texas MD Anderson Cancer Center, Houston, TX USA

ABSTRACT

Medical image segmentation remains a difficult, time-consuming task; currently, liver segmentation from abdominal CT scans is often done by hand, requiring too much time to construct patient-specific treatment models for hepatocellular carcinoma. Image segmentation techniques, such as level set methods and convolutional neural networks (CNN), rely on a series of convolutions and nonlinearities to construct image features: neural networks that use strictly mean-zero finite difference stencils as convolution kernels can be treated as upwind discretizations of differential equations. If this relationship can be made explicit, one gains the ability to analyze CNN using the language of numerical analysis, thereby providing a well-established framework for proving properties such as stability and approximation accuracy. We test this relationship by constructing a level set network, a type of CNN whose architecture describes the expansion of level sets; forward-propagation through a level set network 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 common CNN architecture. We therefore analyze which convolution filters are present in a standard CNN, to see whether finite difference stencils are learned during training; we observe certain patterns that form at certain layers in the network, where the kernels cannot be accounted for by finite difference stencils alone.

Keywords: image segmentation; convolutional neural networks; numerical analysis; clustering

1. INTRODUCTION - DESCRIPTION OF PURPOSE

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

Send correspondence to JAA at jonasactor@rice.edu

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.

2. METHODS

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 advantange 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 tehe context of adjoint equations for dynamical systems[7].[8] 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 attempt to construct a neural network that approaches this specific PDE. In this sense, the correct curve evolution is then 'learned' by the level set network. The LSN maintains the architecture of solving the level set equation, but replaces the finite difference operators with learned convolution kernels. 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.

To test this concept, we employed three segmentation methods on the MICCAI 2017 LiTS Challenge dataset,[5] consisting of 131 abdominal contrast-enhanced CT image stacks. These methods were: UNet,[9] a type of CNN; ITK-SNAP,[6] a segmentation application using the level set equation; and our level set network (LSN). 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 to achieve the best segmentation, after which ITK-SNAP solved the level set equation until DSC no longer improved.

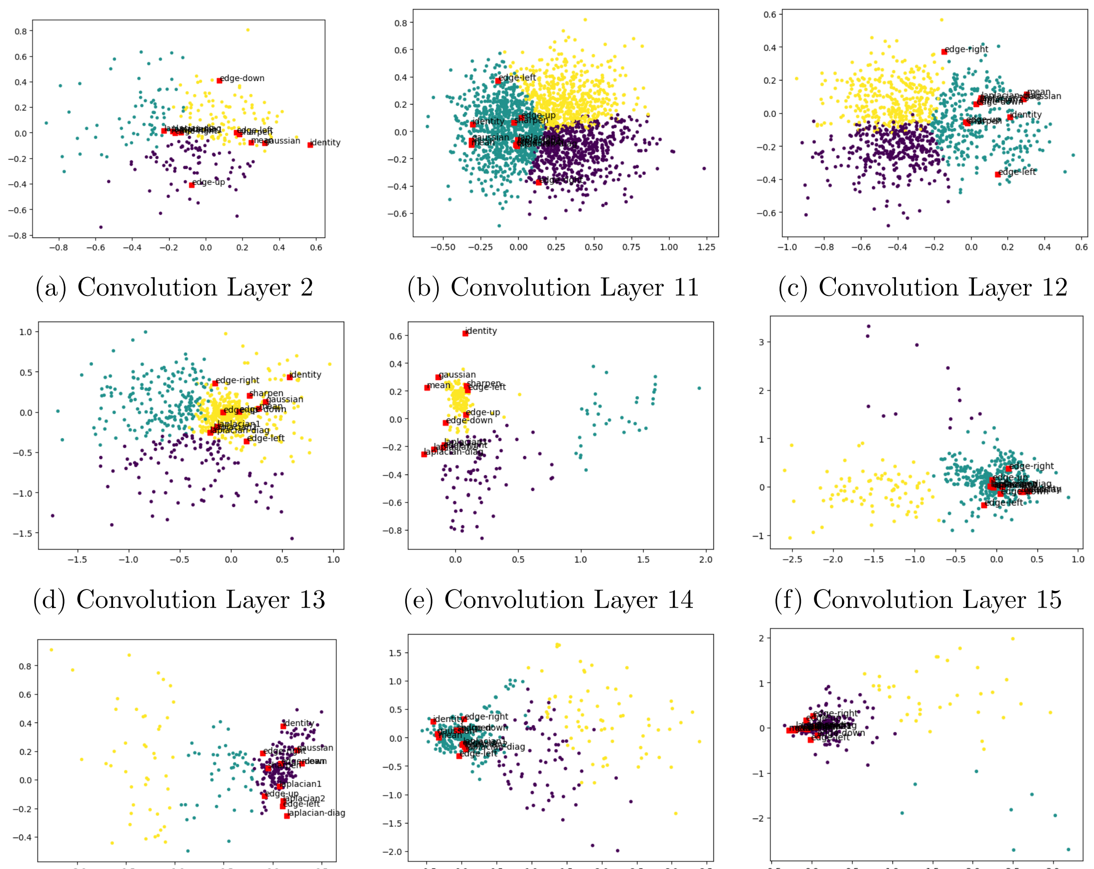
3. RESULTS

Using the three methods described above, we obtained the DSC scores listed in Table 1. 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 on their own.

We confirmed this insight by plotting the convolution kernels obtained from training our UNet. 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.

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\{2}	0.604	0.911	0.837	0.638

Table 1: DSC scores for each fold, from training the level set network. While all methods struggled with fold $k = 2$, LSN fared particularly poorly during validation, due to its initialization procedure. When discarding this fold, LSN compares more favorably to the other methods.



(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

Figure 1: Visualization of 3x3 convolution kernels from selected layers of a UNet with a depth of 4. Colors correspond to K-means cluster assignment. Layers 1,3-10 (not shown) are similar to Convolution Layer 11. Layers 1-9 belong to the encoder portion of the UNet; layers 10-18 belong to the decoder. Observe that the further along the net, the more clustered the kernels become.

Our clustering results, as illustrated in Figure 1, suggest that for many layers in this UNet, there is no clear distinction between various types of kernels. However, on the decoder (upsampling) side of the UNet architecture, several 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 otherwiseuninterpretable 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 insight from above: finite difference kernels alone cannot explain the predictive power of our UNet.

4. CONCLUSIONS AND NOVELTY

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. Our comparison between LSN, UNet, and the level set equation illustrates that finite difference kernels combined with ReLU activation functions are not sufficient on their own to obtain high-accuracy image segmentation for liver CT data; there are certain convolution kernels learned by a UNet that cannot be explained using commonly-invoked image features and filters. Developing an understanding of the bottom layer and decoder portions of a CNN is a crucial step to being able to explain the predictive power of CNNs for image segmentation, which would enable an interpretation of convolution kernels in clinical terms.

5. SUBMISSION INFORMATION

Work on the numerical analysis of level set networks has been accepted for a poster at AMIA Annual Conference 2019, and will be submitted to Rice Data Science Conference 2019.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M., Znaor, A., Soerjomataram, I., and Bray, F., "Global cancer observatory: Cancer today," tech. rep., International Agency for Research on Cancer (2018).

[2] McGlynn, K. A. and London, W. T., "Epidemiology and natural history of hepatocellular carcinoma," *Best practice & research Clinical gastroenterology* **19**(1), 3–23 (2005).

[3] Bruix, J., Reig, M., and Sherman, M., "Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma," *Gastroenterology* **150**(4), 835–853 (2016).

[4] Schmitt, P., Mandonnet, E., Perdreau, A., and Angelini, E. D., "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* **112**(2), 165–172 (2013).

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

[6] Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., and Gerig, G., "User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage* **31**(3), 1116–1128 (2006).

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

[8] Chen, T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K., "Neural ordinary differential equations," in [*Advances in neural information processing systems*], 6571–6583 (2018).

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