V-NAS: Neural Architecture Search for Volumetric Medical Image Segmentation

Zhuotun Zhu[1], Chenxi Liu[1], Dong Yang[2], Alan L. Yuille[1], Daguang Xu[2]
[1]Johns Hopkins University [2]NVIDIA Corporation

3DV 2019 International Conference on 3D Vision
Québec City, Canada
September 16 - 19, 2019

Abstract

Deep learning algorithms, in particular 2D and 3D fully convolutional neural networks (FCNs), have rapidly become the mainstream methodology for volumetric medical image segmentation. However, 2D convolutions cannot fully leverage the rich spatial information along the third axis, while 3D convolutions suffer from the demanding computation and high GPU memory consumption. In this paper, we propose to automatically search the network architecture tailoring to volumetric medical image segmentation problem. Concretely, we formulate the structure learning as differentiable neural architecture search, and let the network itself choose between 2D, 3D or Pseudo-3D (P3D) convolutions at each layer. We evaluate our method on 3 public datasets, i.e., the NIH Pancreas dataset, the Lung and Pancreas dataset from the Medical Segmentation Decathlon (MSD) Challenge. Our method, named V-NAS, consistently outperforms other state-of-the-arts on the segmentation tasks of both normal organ (NIH Pancreas) and abnormal organs (MSD Lung tumors and MSD Pancreas tumors), which shows the power of chosen architecture. Moreover, the searched architecture on one dataset can be well generalized to other datasets, which demonstrates the robustness and practical use of our proposed method.

Background

Pancreatic cancer is a major killer of human beings. In United States of 2019,
- 56,770 new cases will be diagnosed
- 45,750 will die from the disease
- 5-year survival rate is 9% (2008 ~ 2014)

Challenges

2D v.s. 3D
- 2D >> Well-pretrained models☺
- 2D >> Less memory-demanding☺
- 3D >> Spatial information☺

Volumetric Image Segmentation
- 2D Based Segmentation: UNet, 2D C2F[7], FCN+LSTM [4]
- 3D Based Segmentation: 3D UNet [8], VNet [9], ResDSN [5]
- 2D/3D Fused Segmentation: VFN [3], H-DenseUNet, nnUNet

Proposed Method

- **Neural Architecture Search for Volumetric Medical Image Segmentation**
  - One of the first to explore the idea of NAS/AutoML in medical imaging field
  - Adopt the differentiable NAS[11] directly on volumetric data

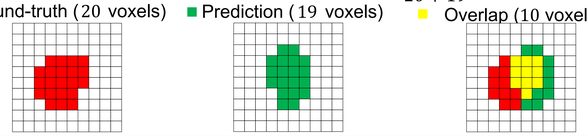
Algorithm 1: V-NAS

Partition the whole labeled dataset $\mathcal{S}$ into the **disjoint** $\mathcal{S}_{\text{train}}$, $\mathcal{S}_{\text{val}}$ and $\mathcal{S}_{\text{test}}$;
Create the mixed operations $\bar{O}_e^l$ and $\bar{O}_d^b$ parametrized by α_i^l and β_i^b, respectively;
while *training not converged* **do**
  1. Update weights w by descending $\nabla_w \mathcal{L}_{\text{train}}(w, \boldsymbol{\alpha}, \boldsymbol{\beta})$
  2. Update $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by descending $\nabla_{\boldsymbol{\alpha},\boldsymbol{\beta}} \mathcal{L}_{\text{val}}(w, \boldsymbol{\alpha}, \boldsymbol{\beta})$

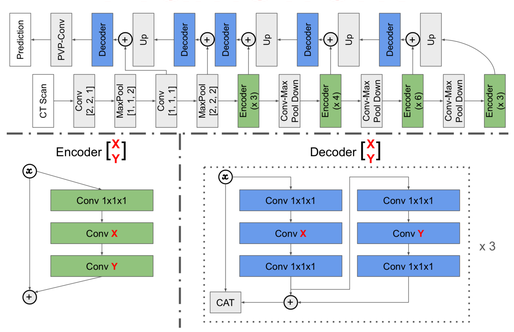
Replace the relaxed operation $\bar{O}_e^l$ with $O_e^l = E_i, i = \texttt{argmax}_k \exp(\alpha_k^l) / \sum_{j=0}^{2} \exp(\alpha_j^l)$;
Replace the relaxed operation $\bar{O}_d^b$ with $O_d^b = D_i, i = \texttt{argmax}_k \exp(\beta_k^b) / \sum_{j=0}^{2} \exp(\beta_j^b)$;
Retrain the discretized architecture on the $\mathcal{S}_{\text{trainval}}$.

- **Evaluation Matrix**
  - Dice-Sørensen Coefficient
  - If the set of ground-truth pancreas voxels is A and the set of predicted pancreas voxels is B, then the accuracy of segmentation, or the Dice-Sørensen coefficient (DSC), is computed as $2\times|A \cap B|/(|A| + |B|)$ $\quad \text{DSC} = \frac{2\times 10}{20+19} = 51.28\%$

■ Ground-truth (20 voxels) ■ Prediction (19 voxels) ■ Overlap (10 voxels)



Framework



Encoder Search Space and Operation Relaxation

$$\mathcal{E} = \{\underbrace{\text{Encoder}\begin{bmatrix}3\times3\times1\\1\times1\times1\end{bmatrix}}_{E_0:\text{ 2D}}, \underbrace{\text{Encoder}\begin{bmatrix}3\times3\times3\\1\times1\times1\end{bmatrix}}_{E_1:\text{ 3D}}, \underbrace{\text{Encoder}\begin{bmatrix}3\times3\times1\\1\times1\times3\end{bmatrix}}_{E_2:\text{ P3D}}\}$$

$$x^{l+1} = O_e^l(x^l) \approx \bar{O}_e^l(x^l) = \sum_{i=0}^{2} \frac{\exp(\alpha_i^l)}{\sum_{j=0}^{2}\exp(\alpha_j^l)} E_i(x^l), l = 1, \ldots, L.$$

Decoder Search Space and Operation Relaxation

$$\mathcal{D} = \{\underbrace{\text{Decoder}\begin{bmatrix}3\times3\times1\\3\times3\times1\end{bmatrix}}_{D_0:\text{ 2D}}, \underbrace{\text{Decoder}\begin{bmatrix}3\times3\times3\\3\times3\times3\end{bmatrix}}_{D_1:\text{ 3D}}, \underbrace{\text{Decoder}\begin{bmatrix}3\times3\times1\\1\times1\times3\end{bmatrix}}_{D_2:\text{ P3D}}\}$$

$$x^{b+1} = O_d^b(x^b) \approx \bar{O}_d^b(x^b) = \sum_{i=0}^{2} \frac{\exp(\beta_i^b)}{\sum_{j=0}^{2}\exp(\beta_j^b)} D_i(x^b), b = 1, \ldots, B.$$

Results

Table 1. Evaluation of different methods on the NIH pancreas dataset[2]. Our proposed one-stage framework outperforms previous two-stage state-of-the-arts[3][5].

DSC	Ours V-NAS	Ours Mix	Xia et al[3]	Zhu et al[5]	Yu et al[6]	Cai et al[4]	Zhou et al[7]	Roth el al[2]
Average	**85.15±4.55**	84.36±5.25	84.63±5.07	84.59±4.86	84.50±4.97	82.40±6.70	82.37±5.68	71.42±10.11
Max	91.18	91.29	**91.57**	91.45	91.02	90.10	90.85	86.29
Min	**70.37**	67.20	61.58	69.62	62.81	60.00	62.43	23.99

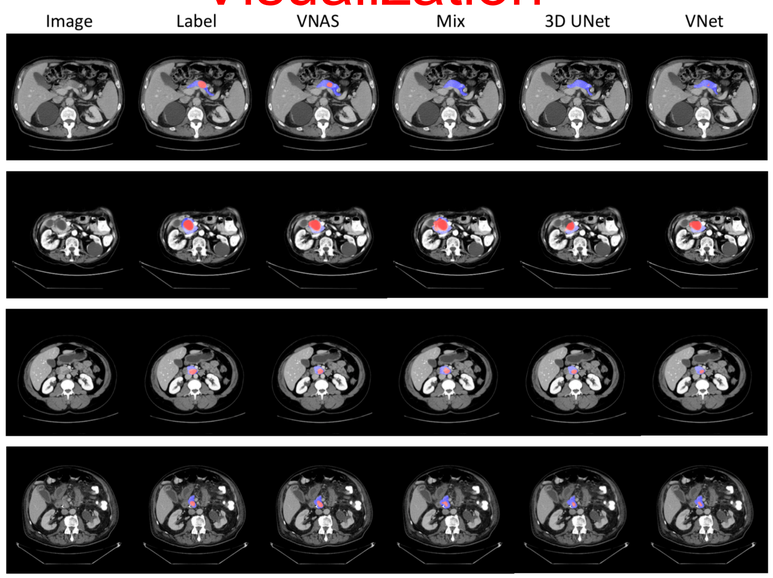
Table 2. Evaluations of the MSD Lung tumors dataset.

DSC	V-NAS-Lung (Ours)	V-NAS-NIH (Ours)	Baseline (Ours)	3D/3D	2D/2D	P3D/P3D	3D UNet[8]	VNet[9]
Average	**55.27±31.18**	54.01±31.39	52.27±31.40	53.74±30.66	52.01±31.50	51.48±32.46	52.94±31.28	50.47±31.37
Max	90.32	92.17	89.57	91.44	92.58	92.40	93.58	**93.85**
Median	66.95	**68.93**	61.71	60.55	63.27	63.89	61.08	57.82

Table 3. Evaluations of the MSD Lung tumors dataset.

Method	Categorization	Pancreas Tumors DSC			Pancreas DSC		
		Mean	Max	Median	Mean	Max	Median
V-NAS(Ours)	Search	**37.78±32.12**	92.49	**38.32**	**79.94±8.85**	**92.24**	36.99
Baseline(Ours)	Mix	30.10±31.40	92.95	18.05	78.41±9.40	92.21	40.08
3D UNet[8]	3D	35.61±32.20	**93.66**	32.23	79.20±9.43	91.95	**40.72**
VNet[9]	3D	35.99±31.27	92.95	35.91	79.01±9.44	92.05	28.15

Visualization



Ablations

Ablation 1

Encoder\Decoder	3D	2D	P3D
3D	84.09%	83.77%	84.20%
2D	83.66%	83.29%	84.08%
P3D	84.32%	84.69%	**84.75%**

Table 4. Performance ("Mean DSC") of different encoder and decoder configurations on NIH dataset evaluated by the same 4-fold cross validation. The architecture is manually set with different choices from 2D, 3D and P3D. Ours obtains 85.15% in Table 1.

Ablation 2

NIH Normal Pancreas dataset
- Encoder, [0 0 0, 0 0 0 1, 2 0 2 0 2 2, 0 0 0]
- Decoder, [0 0 1 0 1]

MSD Lung Tumors dataset
- Encoder, [**0 0 0**, 1 2 **0 1**, **2 1 2 0** 0 0, **0 0 0**]
- Decoder, [**0 0** 2 1 **1**]
- Share 68% (11 out of 16 Encoder cells) Encoders
- Share 60% (3 out of 5 Decoder blocks) Decoders

MSD Abnormal Pancreas Dataset
- Encoder, [0 2 2, 2 0 0 0, 2 2 1 2 1 1, 0 1 1]
- Decoder, [1 0 2 0 1]
- More P3D and 3D convolutions

References

[1] American Cancer Society, https://www.cancer.org/cancer, accessed on 04/12/2019
[2] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In MICCAI, 2015.
[3] Y. Xia, L. Xie, F. Liu, Z. Zhu, E. K. Fishman, and A. L. Yuille. Bridging the gap between 2d and 3d organ segmen- tation. In MICCAI, 2018.
[4] J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang. Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function. In MICCAI 2017.
[5] Z. Zhu, Y. Xia, W. Shen, E. Fishman, and A. L. Yuille. A 3d coarse-to-fine framework for volumetric medical image segmentation. In 3DV, 2018.
[6] Q. Yu, L. Xie, Y. Wang, Y. Zhou, E. K. Fishman, and A. L. Yuille. Recurrent saliency transformation network: Incorpo- rating multi-stage visual cues for small organ segmentation. In CVPR, 2018.
[7] Y. Zhou, L. Xie, W. Shen, Y. Wang, E. K. Fishman, and A. L. Yuille. A fixed-point model for pancreas segmentation in abdominal CT scans. In MICCAI, 2017.
[8] O Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3D u-net: learning dense volumetric seg- mentation from sparse annotation. In MICCAI, 2016.
[9] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 3DV, 2016.
[10] C. Liu, L.-C. Chen, F. Schroff, H. Adam, W. Hua, A. Yuille, and L. Fei-Fei. Auto-deeplab: Hierarchical neural archi- tecture search for semantic image segmentation. In CVPR, 2019.
[11] H. Liu, K. Simonyan, and Y. Yang. Darts: Differentiable architecture search. In ICLR, 2019.