



Medical Open Network for AI

open-source initiative built by academic & industry leaders to establish & standardize best practices for deep learning in healthcare imaging

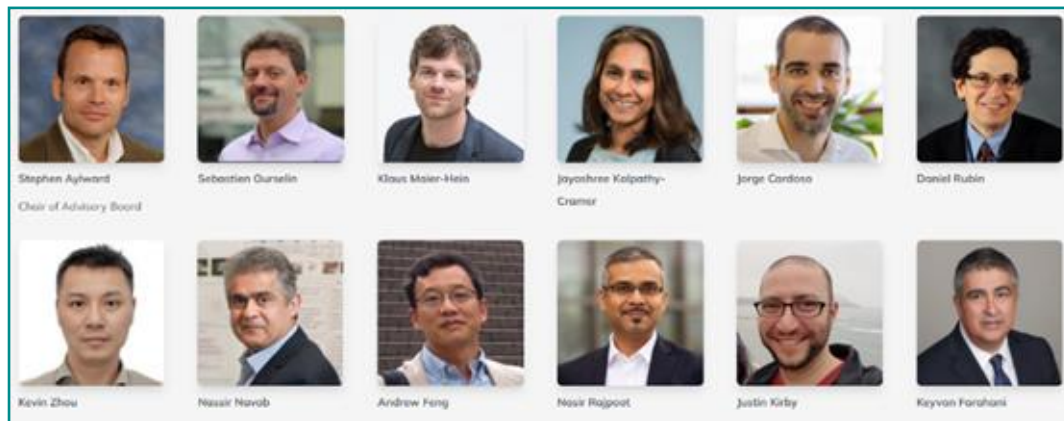
NETWORK OF AI THOUGHT LEADERS

Advisory Board: NVIDIA, KCL, CCDS, Stanford, DKFZ, TUM, CAS, Kitware, Vanderbilt, UCL, NIH/NCI and Warwick

MONAI Working Groups

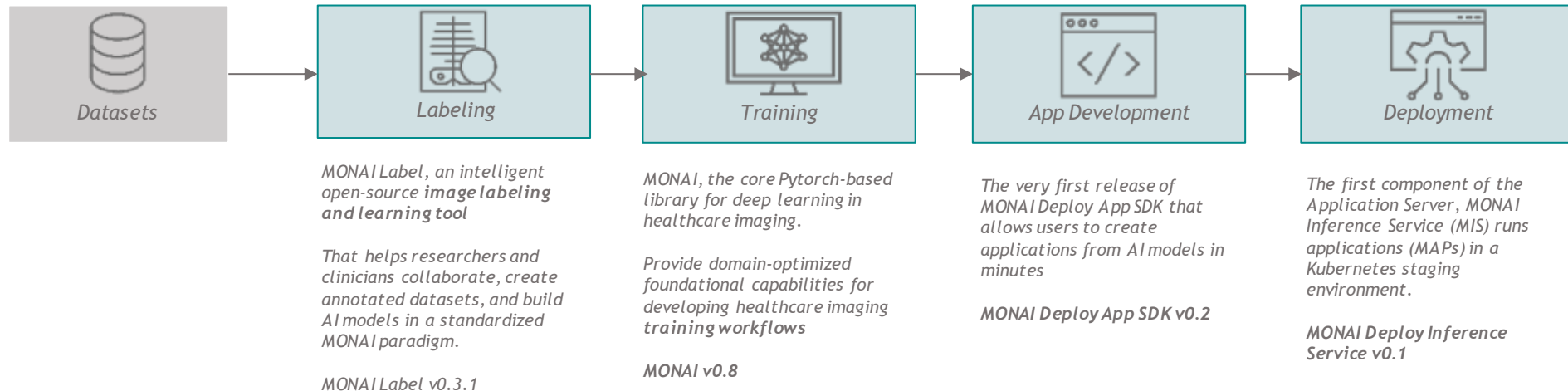
- *IMAGING I/O* - Stephen Aylward
- *DATA DIVERSITY* - Brad Genereaux
- *REPRODUCIBILITY* - Lena Maier-Hein
- *TRANSFORMATIONS* - Jorge Cardoso
- *FEDERATED LEARNING* - Jayashree Kalpathy, Daniel Rubin, Holger Roth
- *PATHOLOGY* - Nasir Rajpoot
- *ADVANCED RESEARCH* - Paul Jaeger
- *COMMUNITY ADOPTION* - Prerna Dogra, Michael Zephyr
- *DEPLOY* - Haris Shuaib and David Bericat
- *DIGITAL PATHOLOGY* - Nasir Rajpoot

Advisory Board



WHAT IS MONAI?

Accelerate Pace of Research Innovation With a Common Foundation



ADOPTION MOMENTUM

All Paths Lead to MONAI

250K

PyPI link	https://pypi.org/project/monai
Total downloads	256,896
Total downloads - 30 days	50,666
Total downloads - 7 days	12,184
PyPI link	https://pypi.org/project/monai-label
Total downloads	4,549
Total downloads - 30 days	646
Total downloads - 7 days	112
PyPI link	https://pypi.org/project/monai-deploy-app-sdk
Total downloads	3,954
Total downloads - 30 days	702
Total downloads - 7 days	134

105 Individuals
30
Institutions

Users by Country



MONAI Public

AI Toolkit for Healthcare Imaging

Python 2.6k 486

PARTNERS & GROWTH



MONAI Community

Get involved in the MONAI community today



Contribute



Engage



Adopt



AUTOMATED 3D MEDICAL IMAGE SEGMENTATION USING AUTOML AND NEURAL ARCHITECTURE SEARCH

Dong Yang

Research Scientist, NVIDIA

March 22, 2022





AGENDA

Overview

Medical image analysis and problem formulation of 3D medical image segmentation

Neural Architecture Search (NAS)

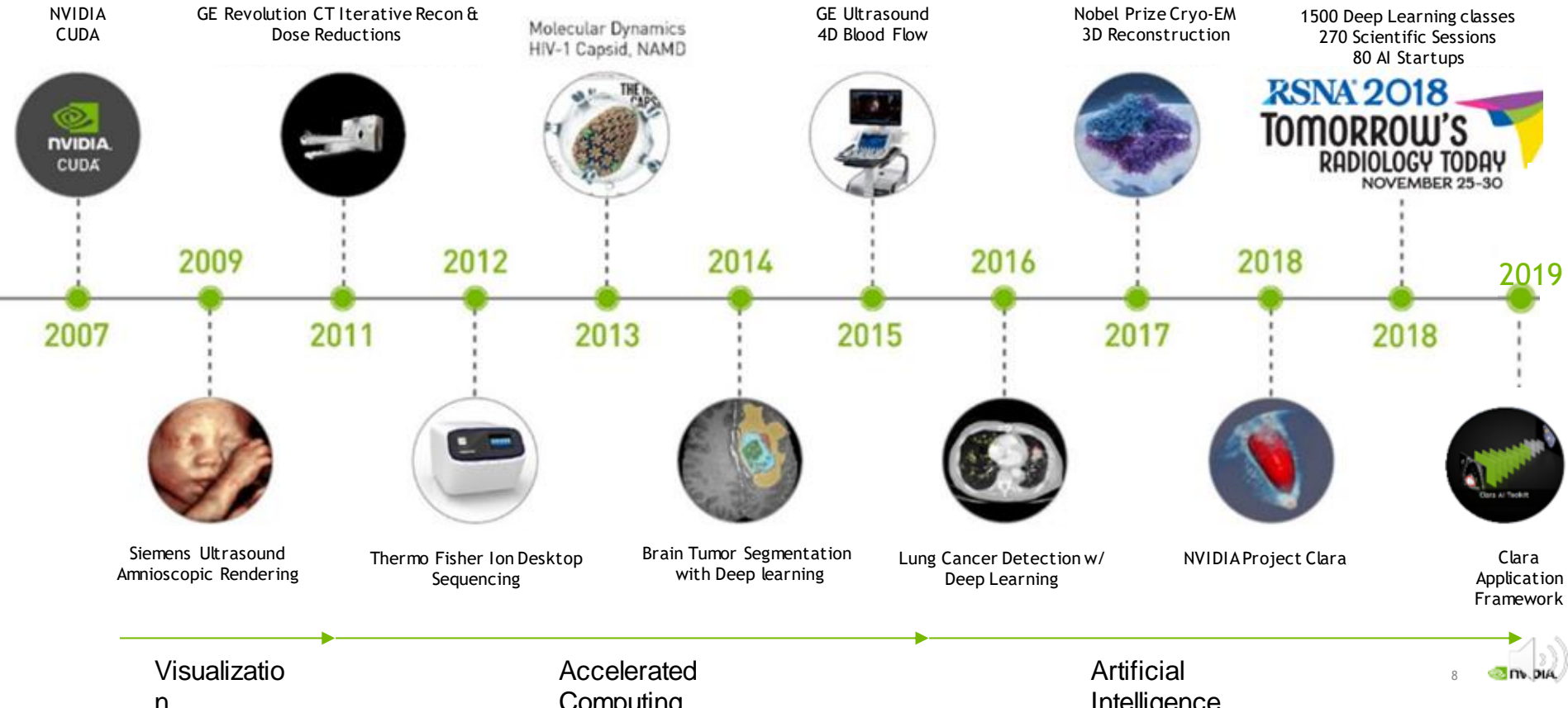
Novel algorithms using neural architecture search for large-scale medical datasets

Summary

Efficient NAS algorithm achieving state-of-the-art segmentation performance



12 YEARS OF MEDICAL IMAGING



MEDICAL IMAGE ANALYSIS

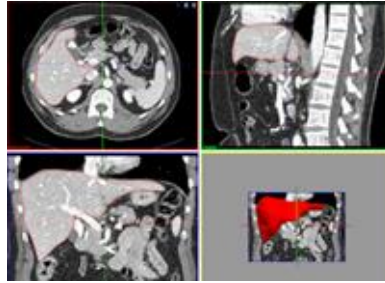
Motivations

- What?
 - To gain high-level understanding from medical images
- Why?
 - Disease diagnosis, treatment planning and surgery guidance
 - Applications
 - Accurate radiotherapy (calculating dose distribution, delineation of the tumor and normal organs)
 - “COVID-19” screening, etc.

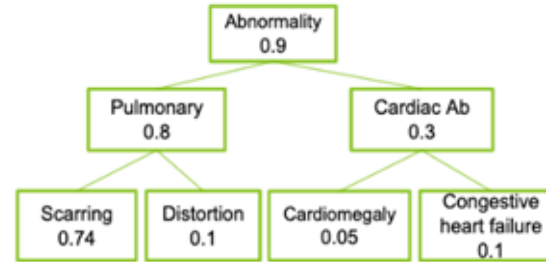
MEDICAL IMAGE ANALYSIS

What is it?

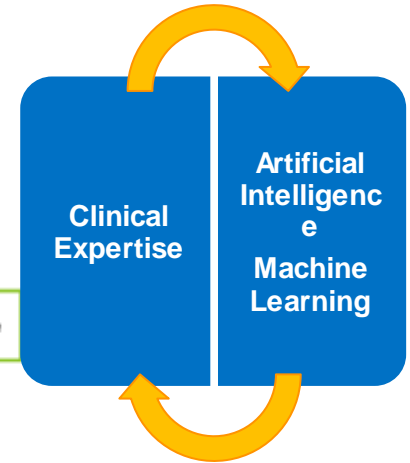
CT
MRI
Ultrasound
X-Ray
Microscopy
Pathology
PET
OCT
EHR



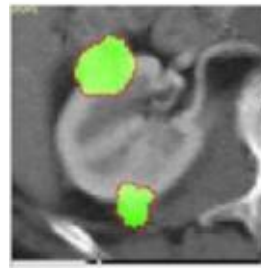
Segmentation
Anatomy Understanding



Classification
Abnormality Detection



Detection
Cancer Staging



Survival Model Prediction
Longitudinal Monitoring



Findings:
There are diffuse bilateral emphysema and several
apical bullae consistent with chronic obstructive lung
disease and bullous emphysema.
There are multiple nodules in the right lung apex,
most of which represent a calcified lesion in the left
lung apex.
There are several nodules in the right upper lobe,
most of which represent a calcified lesion.

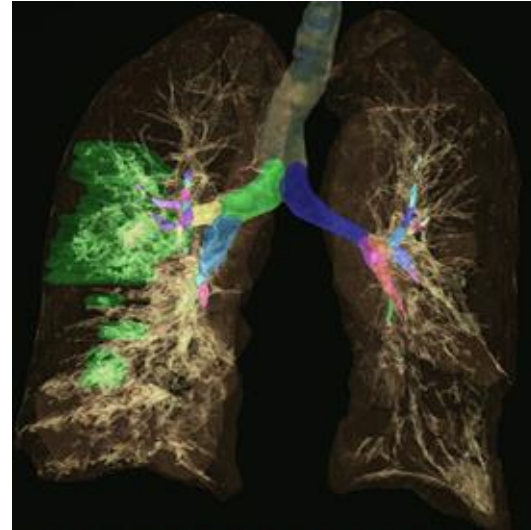
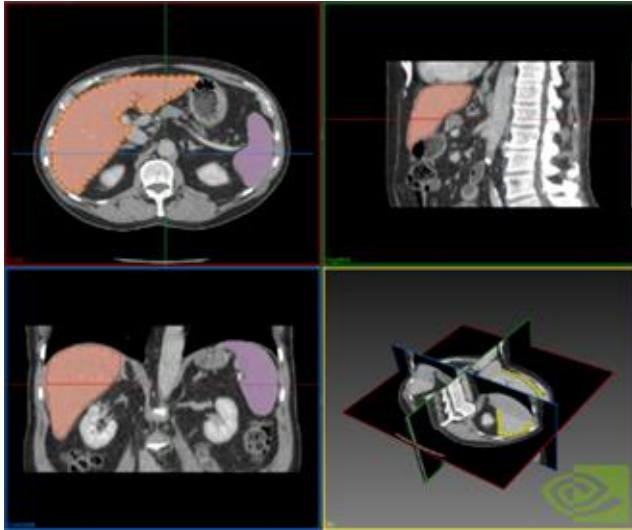
Impression:
1. Diffuse emphysema and several bullae.
2. Multiple nodules in the right apex, although
difficult to exclude a certain lesion.

Natural Language Processing
Medical Report

CASE STUDY

3D Medical Image Segmentation

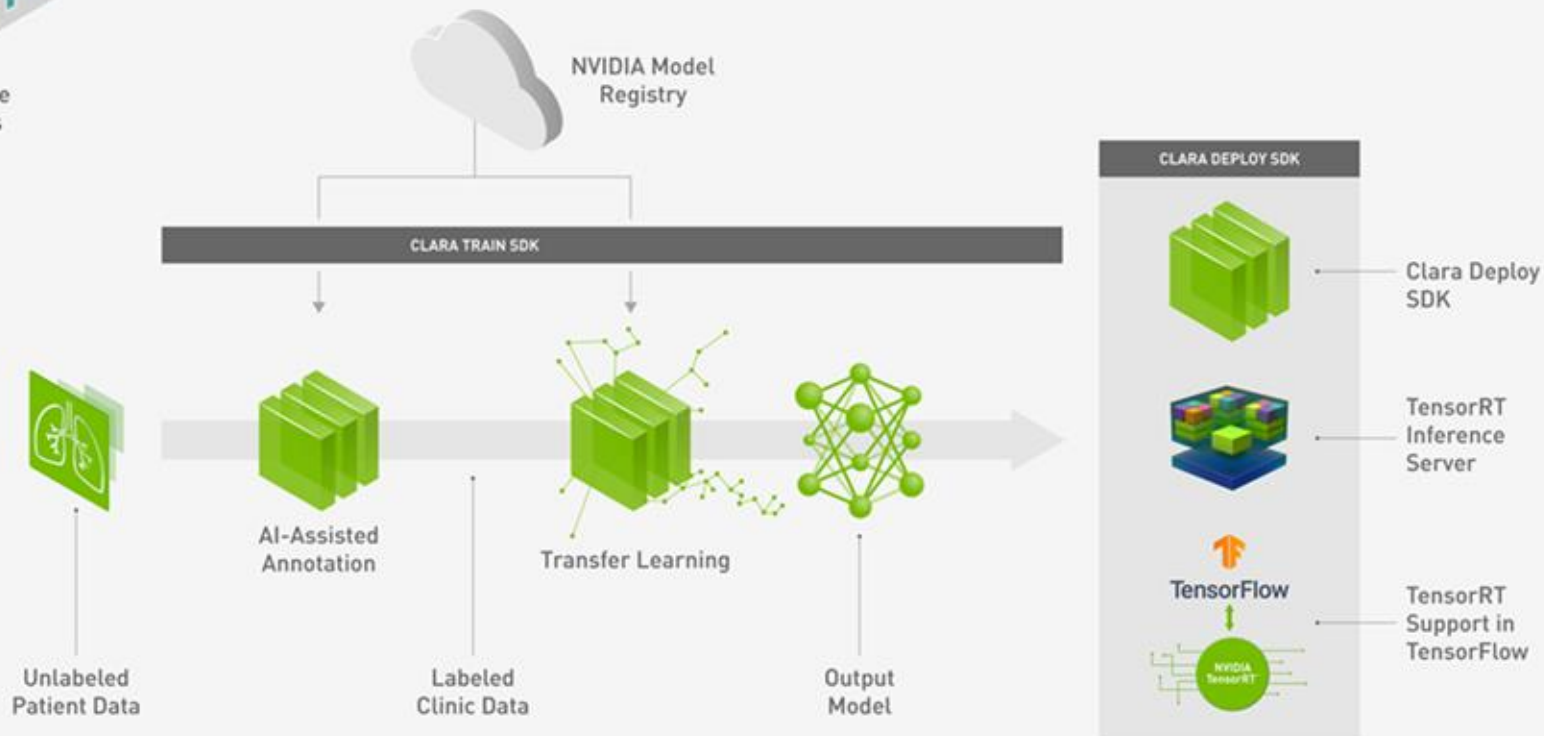
- Given 3D volumes (e.g., CT, MRI) as input, to extract 3D structures of organs or tumors



<https://devblogs.nvidia.com/annotation-transfer-learning-clara-train/>
<https://www.nvidia.com/en-us/clara/>



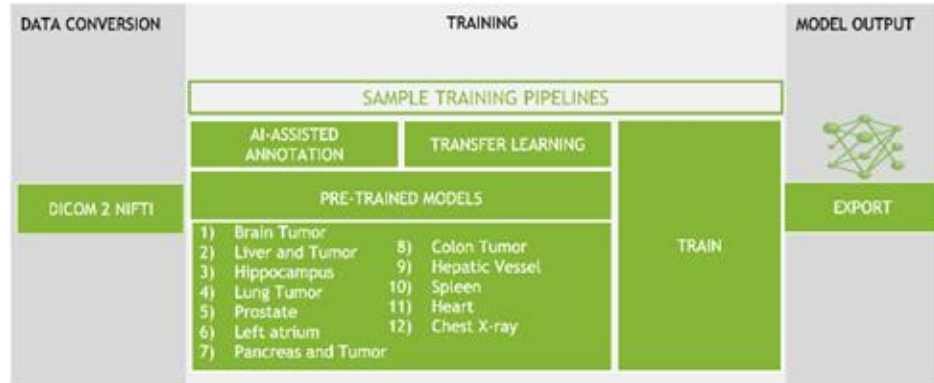
Healthcare
Institutes



OVERVIEW

Deep Learning Solution

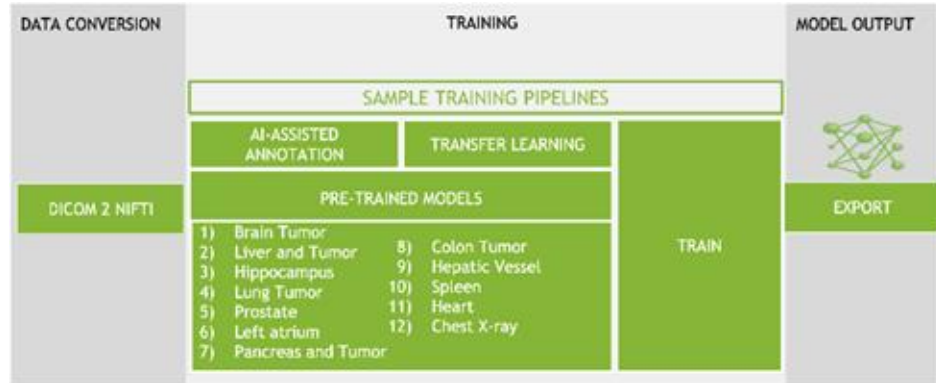
- Traditional deep learning solutions require expertise to define and train the pipeline
 - Various tasks with different image modalities, scanning protocols, qualities, etc.



OVERVIEW

Deep Learning Solution

- Traditional deep learning requires expertise to define and train the pipeline
 - Various tasks with different image modalities, scanning protocols, quality, etc.

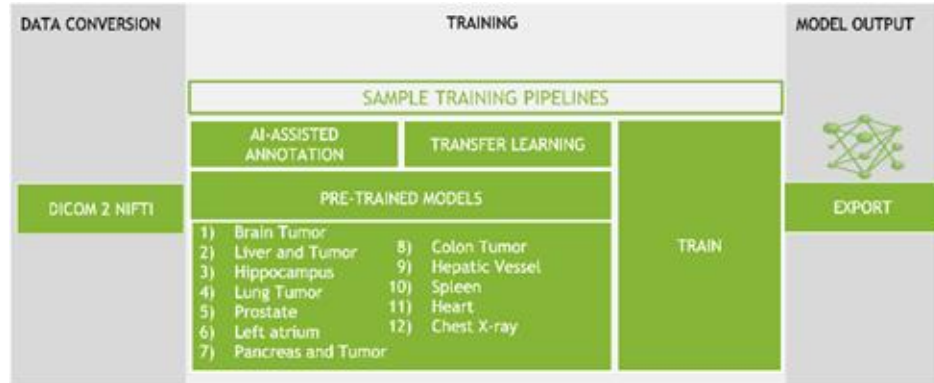


- What is the optimal solution?
 - Optimal components: data pre-processing, neural network architecture, etc.

OVERVIEW

Deep Learning Solution

- Traditional deep learning requires expertise to define and train the pipeline
 - Various tasks with different image modalities, scanning protocols, quality, etc.



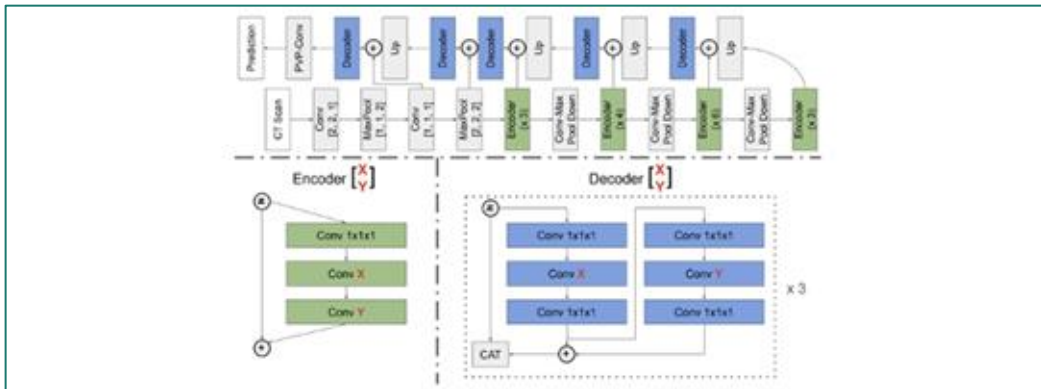
- What is the optimal solution?
 - Optimal components: data pre-processing, neural network architecture, etc.
- Automated Machine Learning (AutoML)
 - leveraging large-scale computing resource with less human effort

OVERVIEW

Automated Machine Learning (AutoML)

- AutoML
 - Automated data preparation
 - Automated feature engineering
 - Automated model selection
 - Hyperparameter optimization
 - Automated pipeline selection
- Scenario - Automated Deep Learning (AutoDL)
 - Customized AutoML algorithms for deep learning
 - For instance, **automated model selection** → **neural architecture search**

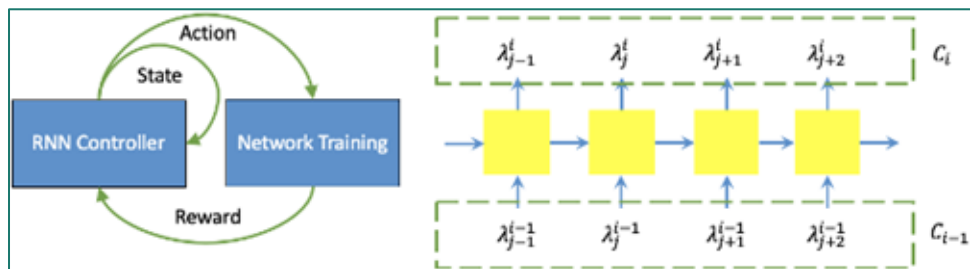
WHAT HAVE WE DONE IN AUTOML SO FAR?



2018

V-NAS (3DV)

First work ever of neural architecture search for volumetric medical image segmentation



2018

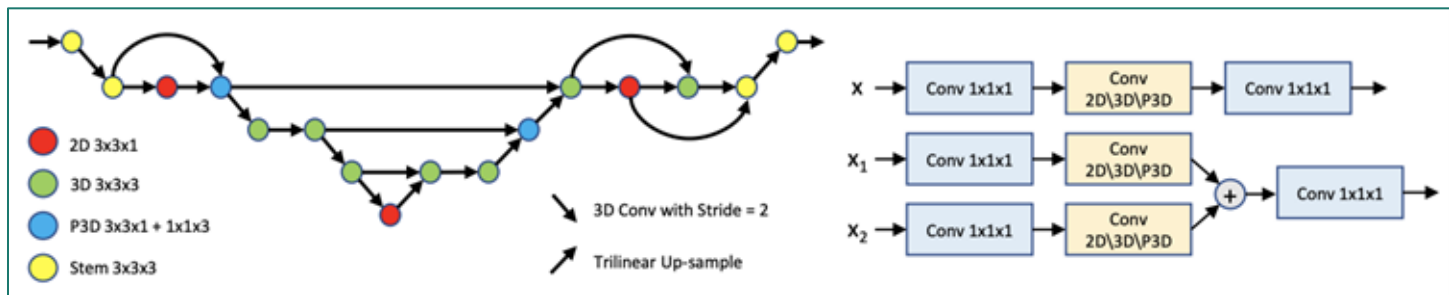
V-NAS (3DV)

First work ever of neural architecture search for volumetric medical image segmentation

2019

Reinforcement Learning Controller (MICCAI)

Searching optimal training configuration (hyper-parameters)



2018

V-NAS (3DV)

First work ever of neural architecture search for volumetric medical image segmentation

2019

Reinforcement Learning Controller (MICCAI)

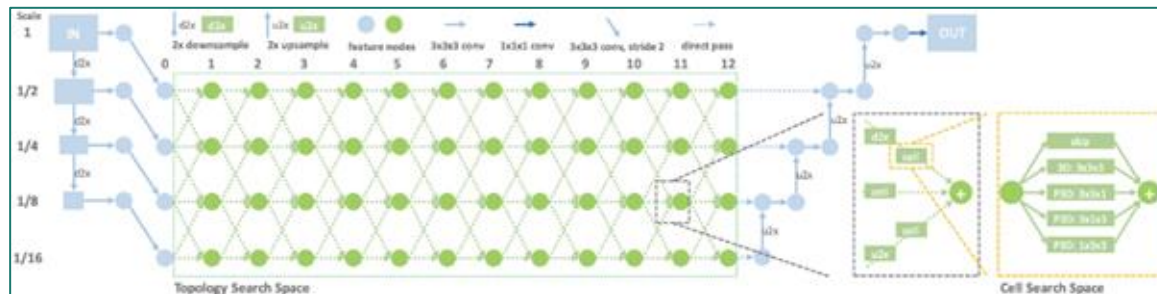
Searching optimal training configuration (hyper-parameters)

2020

C2FNAS (CVPR)

State-of-the-art of neural architecture search for volumetric medical image segmentation

#1 on MSD in 2020



2018

V-NAS (3DV)

First work ever of neural architecture search for volumetric medical image segmentation

2019

Reinforcement Learning Controller (MICCAI)

Searching optimal training configuration (hyper-parameters)

2020

C2FNAS (CVPR)

State-of-the-art of neural architecture search for volumetric medical image segmentation

#1 on MSD in 2020

2021

DiNTS (CVPR)

New state-of-the-art of neural architecture search for medical image segmentation with **great efficiency**

#1 on MSD in 2021

Federated Neural Architecture Search (MICCAI 2021)

Dr. Holger Roth

Adapting different neural architectures to local clients



Training-Free Neural Architecture Search (WIP)

Dr. Vishwesh Nath and Intern

Reducing searching cost to "zero"

Proxy Data and Network (MICCAI 2021)

Dr. Vishwesh Nath

Reducing searching cost with the most representative data and network models



NEURAL ARCHITECTURE SEARCH (NAS)

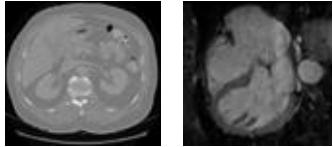


OVERVIEW

Deep Learning Solution

Data

2D/3D Images
CT / MRI



Data Pre-Processing

- Intensity normalization
- Intensity clipping
- Re-sampling
- ROI cropping, etc.



Neural Network

- 2D-3D AH-Net
- Res-UNet
- V-Net
- U-Net, etc.



Data Post-Processing

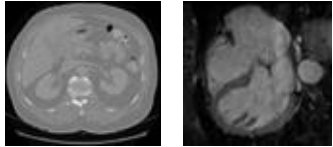
- Connected-component analysis
- Filtering,
- Test-time augmentation, etc.

OVERVIEW

Deep Learning Solution

Data

2D/3D Images
CT / MRI



Data Pre-Processing

- Intensity normalization
- Intensity clipping
- Re-sampling
- ROI cropping, etc.



Neural Network

- 2D-3D AH-Net
- Res-UNet
- V-Net
- U-Net, etc.



Data Post-Processing

- Connected-component analysis
- Filtering,
- Test-time augmentation, etc.

Search for Optimal Neural Network Architecture Published in CVPR 2020

C2FNAS: Coarse-to-Fine Neural Architecture Search for 3D Medical Image Segmentation

Qihang Yu^{1*} Dong Yang² Holger Roth²
Yutong Bai¹ Yixiao Zhang^{1*} Alan L. Yuille¹ Daguang Xu²

¹ The Johns Hopkins University ² NVIDIA

Abstract

3D convolution neural networks (CNN) have been proved very successful in parsing organs or tumours in 3D medical images, but it remains sophisticated and time-consuming to choose or design proper 3D networks given different task contexts. Recently, Neural Architecture Search (NAS) is proposed to solve this problem by searching for the best network architecture automatically. However, the inconsistency between search stage and deployment stage often exists in NAS algorithms due to memory constraints and large search space, which could become more serious when applying NAS to some memory and time-consuming tasks, such as 3D medical image segmentation. In this paper, we propose a **coarse-to-fine neural architecture search (C2FNAS)** to automatically search a 3D segmentation network from scratch without inconsistency on network size or input size. Specifically, we divide the search procedure into two stages: 1) the coarse stage, where we search the macro-level topology of the network, i.e. how each convolution module is connected to other modules; 2) the fine stage, where we search at micro-level for operations in each cell based on previous searched macro-level topology. The coarse-to-fine manner

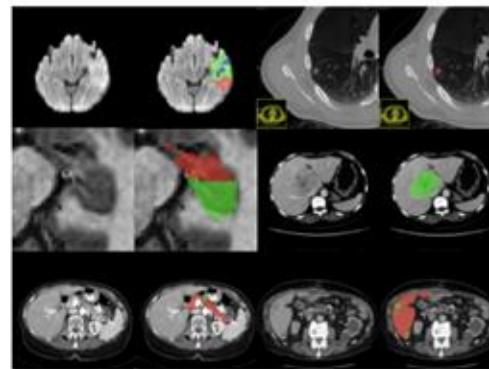


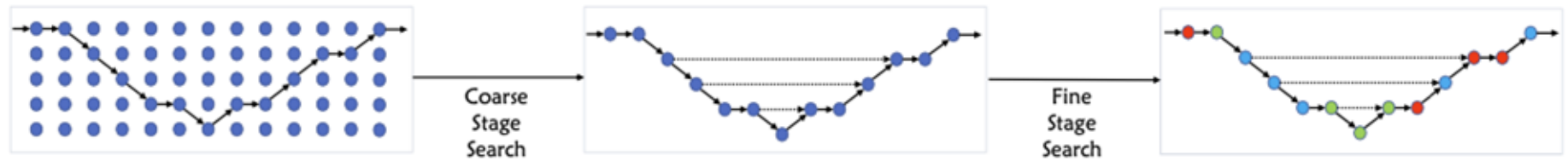
Figure 1. Image and mask examples from MSD tasks (from left to right and top to bottom): brain tumours, lung tumours, hippocampus, hepatic vessel and tumours, pancreas tumours, and liver tumours, respectively. The abnormalities, texture variance, and anisotropic properties make it very challenging to achieve satisfying segmentation performance. Red, green, and blue correspond to labels 1, 2 and 3, respectively, of each dataset.

to get satisfying segmentation for some challenging structures, which could be extremely small with respect to the



C2FNAS

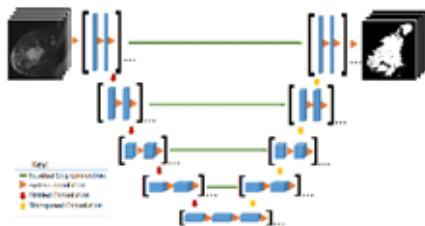
Framework



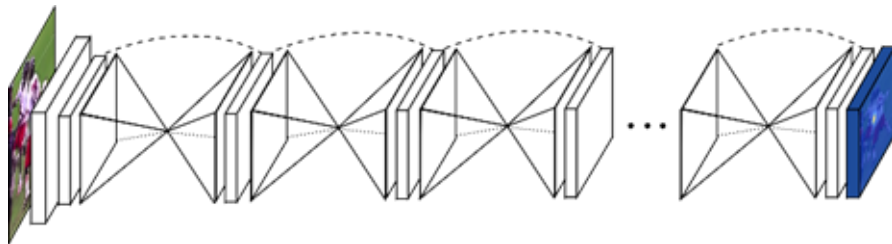
SEARCH SPACE - STEP 1/3

Search at Macro-Level

- Network shape
 - According to the order of **down-sample** and **up-sample layers**, networks can be divided into U-Net-kind and Stacked-Hourglass-kind
- Layer assignment
 - Different from a symmetric U-Net design, search for different assignments of layers, which make it **asymmetric**



U-Net



Stacked-Hourglass

SEARCH SPACE - STEP 2/3

Search at Micro-Level

- ▶ Searching for a replacement for **operation** in each cell, each **operation** can be selected from
 - $3 \times 3 \times 3$ 3D convolution
 - $5 \times 5 \times 5$ 3D convolution
 - $3 \times 3 \times 1$ pseudo 3D convolution
 - $5 \times 5 \times 1$ pseudo 3D convolution
 - $3 \times 3 \times 3$ 3D convolution with dilation 2
 - $5 \times 5 \times 5$ 3D convolution with dilation 2

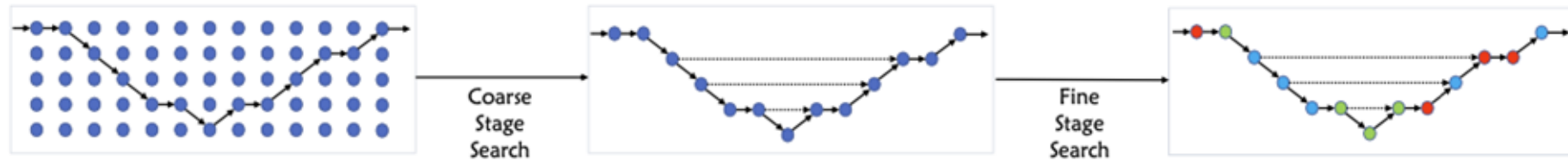
SEARCH SPACE - STEP 3/3

Compound Scaling

- To better balance the performance and model size, scale the patch size, block numbers, and filter numbers
 - Inspired by EfficientNet (state-of-the-art algorithm on ImageNet classification)

C2FNAS

Algorithms

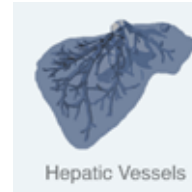
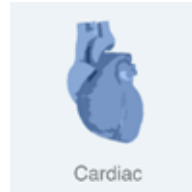
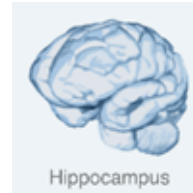
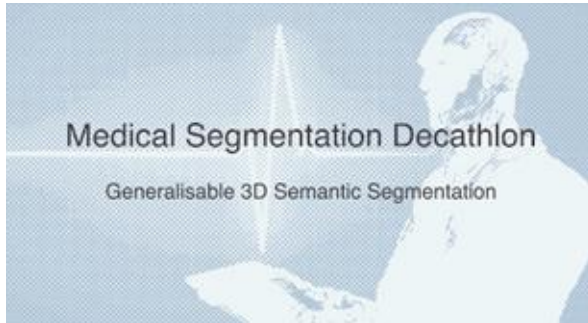


- Step 1 - Macro-level
 - Evolutionary algorithm (EA)
- Step 2 - Micro-level
 - Super-Net training (differentiable neural architecture search)
 - Gradient-based optimization
- Step 3 - Compound Scaling
 - Grid search

DATASETS

Medical Segmentation Decathlon (MSD)

- Public challenge with 10 different applications of 3D medical image segmentation



RESULTS

Test Set - Dice Score

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

Task	Brain				Liver			Pancreas			Prostate		
Class	1	2	3	Avg	1	2	Avg	1	2	Avg	1	2	Avg
CerebriuDIKU [22]	69.52	43.11	66.74	59.79	94.27	57.25	75.76	71.23	24.98	48.11	69.11	86.34	77.73
Lupin	66.15	41.63	64.15	57.31	94.79	61.40	78.10	75.99	21.24	48.62	72.73	87.62	80.18
NVDLMED [34]	67.52	45.00	68.01	60.18	95.06	71.40	83.23	78.42	38.48	58.45	69.36	86.66	78.01
K.A.V.athlon	66.63	46.62	67.46	60.24	94.74	61.65	78.20	74.97	43.20	59.09	73.42	87.80	80.61
nnU-Net [12]	67.71	47.73	68.16	61.20	95.24	73.71	84.48	79.53	52.27	65.90	75.81	89.59	82.70
C2FNAS-Panc	67.62	48.56	69.09	61.76	94.91	71.63	83.27	80.59	52.87	66.73	73.11	87.43	80.27
C2FNAS-Panc*	67.62	48.60	69.72	61.98	94.98	72.89	83.94	80.76	54.41	67.59	74.88	88.75	81.82

Task	Lung	Heart	Hippocampus			Hepatic Vessel			Spleen	Colon	Avg (Task)	Avg (Class)
Class	1	1	1	2	Avg	1	2	Avg	1	1		
CerebriuDIKU [22]	58.71	89.47	89.68	88.31	89.00	59.00	38.00	48.50	95.00	28.00	67.01	66.40
Lupin	54.61	91.86	89.66	88.26	88.96	60.00	47.00	53.50	94.00	9.00	65.61	65.89
NVDLMED [34]	52.15	92.46	87.97	86.71	87.34	63.00	64.00	63.50	96.00	56.00	72.73	71.66
K.A.V.athlon	60.56	91.72	89.83	88.52	89.18	62.00	63.00	62.50	97.00	36.00	71.51	70.89
nnU-Net [12]	69.20	92.77	90.37	88.95	89.66	63.00	69.00	66.00	96.00	56.00	76.39	75.00
C2FNAS-Panc	69.47	92.13	86.87	85.44	86.16	63.78	69.41	66.60	96.60	55.68	75.87	74.42
C2FNAS-Panc*	70.44	92.49	89.37	87.96	88.67	64.30	71.00	67.65	96.28	58.90	76.97	75.49

Table 1. Comparison with state-of-the-art methods on MSD challenge test set (number from MSD leaderboard) measured by **Dice-Sørensen coefficient (DSC)**. * denotes the 5-fold model ensemble. The numbers of tasks hepatic vessel, spleen, and colon from other teams are rounded. We also report the average on tasks and on targets respectively for an overall comparison across all tasks/targets.

RESULTS

Pros and Cons

- Model Comparison

Model	3D U-Net	V-Net	AH-Net	nnU-Net	Ours
Params (M)	16.32	45.61	27.11	10.36	3.91
FLOPS (G)	802.9	322.5	29.5	202.25	184.8

- Computation cost

- In total: ~2480 GPU hrs = ~180 GPU days
- In practice: Using **32 GPUs** in parallel → searching is done within **100 hours** on average

Search for Optimal Neural Network Architecture

Oral Presentation in CVPR 2021

DiNTS: Differentiable Neural Network Topology Search for 3D Medical Image Segmentation

Yufan He¹ Dong Yang² Holger Roth² Can Zhao² Daguang Xu²
¹Johns Hopkins University ²NVIDIA

Abstract

Recently, neural architecture search (NAS) has been applied to automatically search high-performance networks for medical image segmentation. The NAS search space usually contains a network topology level (controlling connections among cells with different spatial scales) and a cell level (operations within each cell). Existing methods either require long searching time for large-scale 3D image datasets, or are limited to pre-defined topologies (such as U-shaped or single-path). In this work, we focus on three important aspects of NAS in 3D medical image segmentation: flexible multi-path network topology, high search efficiency, and budgeted GPU memory usage. A novel differentiable search framework is proposed to support fast gradient-based search within a highly flexible network topology search space. The discretization of the searched optimal continuous model in differentiable scheme may produce a sub-optimal final discrete model (discretization gap). Therefore, we propose a topology loss to alleviate this problem. In addition, the GPU memory usage for the searched 3D model is limited with budget constraints during search. Our Differentiable Network Topology Search scheme (DiNTS) is evaluated on the Medical Segmentation Decathlon (MSD) challenge, which contains ten challeng-

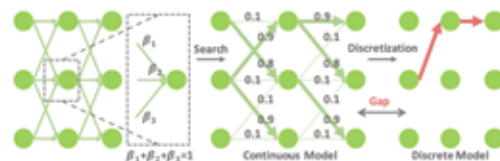


Figure 1. Limitations of existing differentiable topology search formulation. E.g. in Auto-DeepLab [21], each edge in the topology search space is given a probability β . The probabilities of input edges to a node sum to one, which means only one input edge for each node would be selected. A single-path discrete model (red path) is extracted from the continuous searched model. This can result in a large “discretization gap” between the feature flow of the searched continuous model and the final discrete model.

can vary considerably. This makes the direct application of even a successful network like U-Net [34] to a new task less likely to be optimal.

The neural architecture search (NAS) algorithms [49] have been proposed to automatically discover the optimal architectures within a search space. The NAS search space for segmentation usually contains two levels: network topology level and cell level. The network topology controls the connections among cells and decides the flow



DINTS

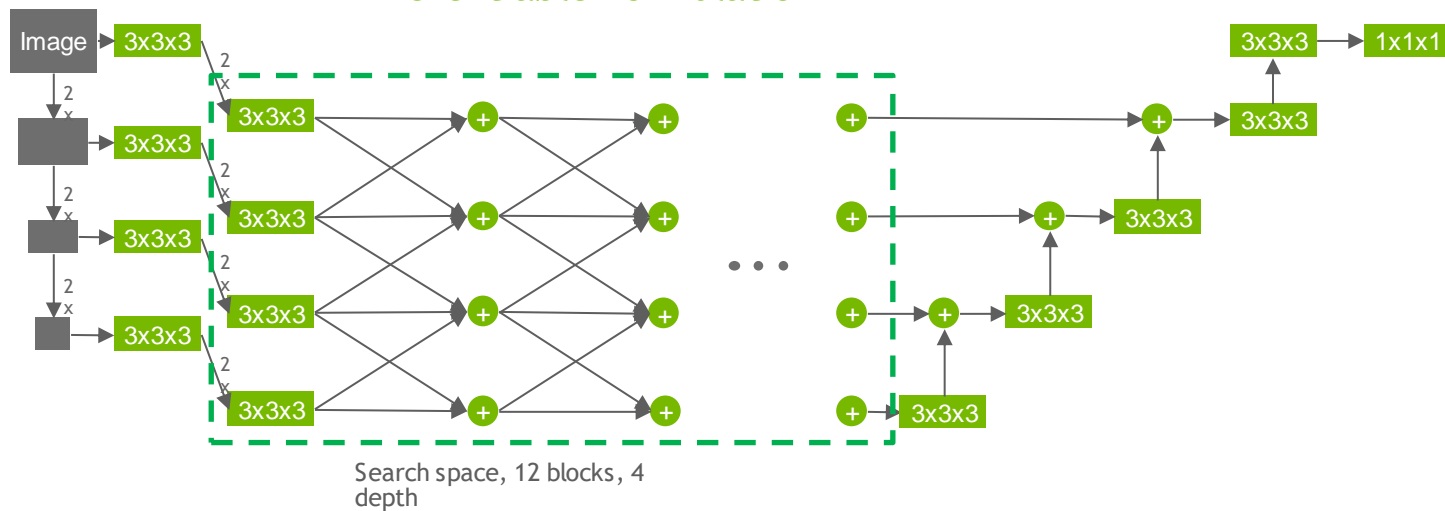
Motivation

- “C2FNAS”
 - Pros: High accuracy
 - Cons: Limited to U-Net, long searching time
- Questions
 - How to design a multi-path model that’s not using “each cell has a single input path assumption”?
 - How to solve GPU memory problem when training with 3D medical images?
 - How to reduce search time?

DINTS

Differentiable Formulation

Search space



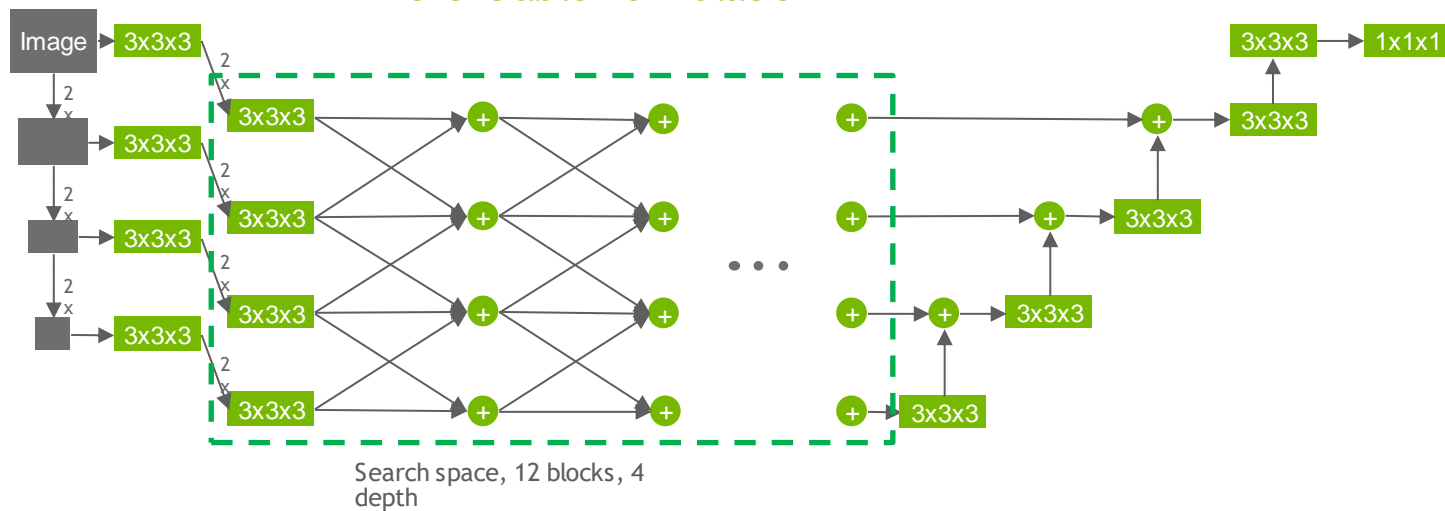
The cell operation is on the edge, the up-sample and down-sample is included in each cell

- ▶ Skip connection
- ▶ $3 \times 3 \times 3$ convolution
- ▶ $3 \times 3 \times 1 + 1 \times 1 \times 3$ convolution pseudo 3D
- ▶ $3 \times 1 \times 3 + 1 \times 3 \times 1$ convolution pseudo 3D

DINTS

Differentiable Formulation

- Search space

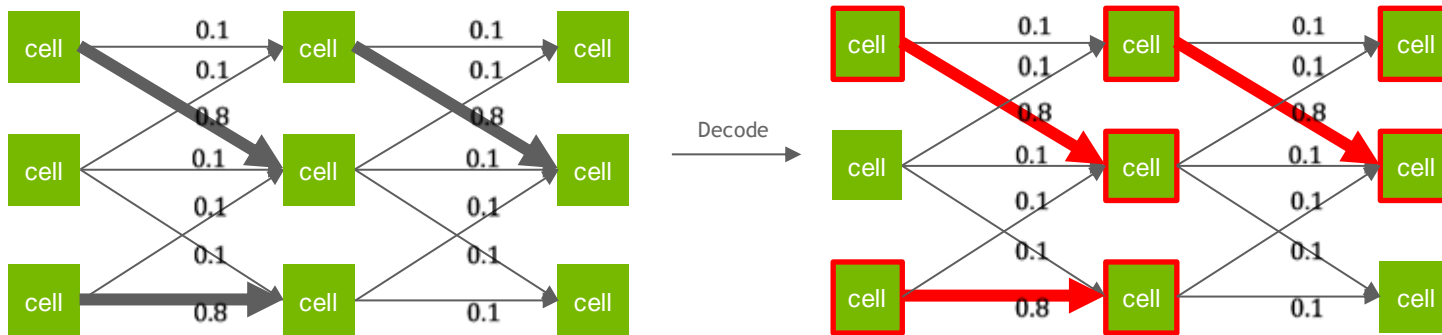


- Model/network weights
- Path/edge weights (0, 1)
- Cell/operation weights (0, 1)

DINTS

Motivation

- Potential problems in differentiable architecture search (DARTS)
 - How to **reduce the binarization gap** between searched model and final decoded model?
 - How to **solve the topology problem** if we don't have “each cell has a single input path assumption”?

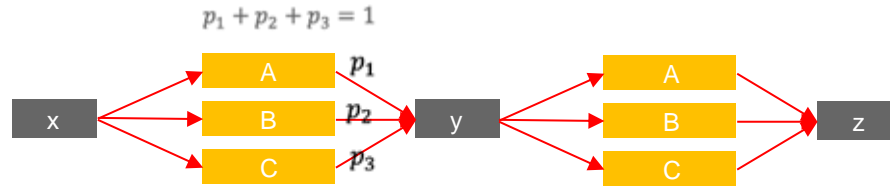


Binarization gap and topology problem

DINTS

Entropy loss

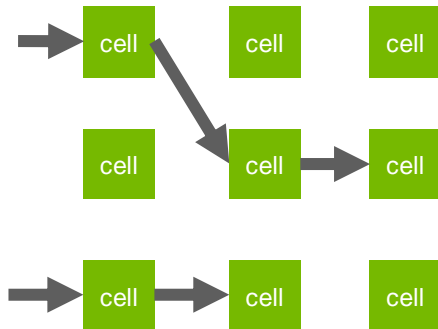
- ▶ Reducing the binarization gap between searched model and final decoded model
 - ▶ Entropy loss for both connections and cell operations: $-\sum_i p_i \log(p_i)$



DINTS

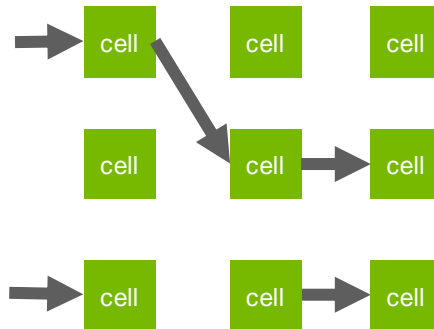
Topology Loss

- Solving the topology problem if we don't have “each cell has a single input path assumption”
 - Topology problem: Connection in block i and block i+1 are related
 - Topology feasibility: Cell with input path must have output path, Cell without input path cannot have output path



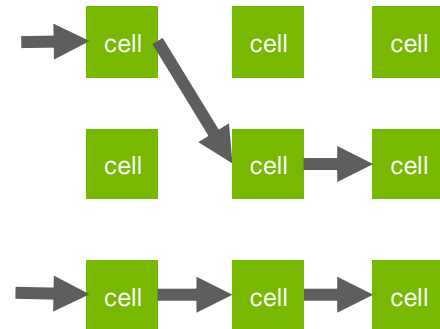
Block i Block i+1

Not
Allowed



Block i Block i+1

Not
Allowed



Block i Block i+1

Allowed

DINTS

Memory Loss

- ▶ How to solve GPU memory problem when training with 3D medical images?
 - ▶ GPU memory loss: $L_1 (\sum_{edge} p_{edge} * (\sum_i memory(O_i) * p_{o_i}) / (\sum_{edge} memory(O_i) * p_{o_i}) - \sigma)$
 - ▶ Measuring GPU consumptions of candidate operations/connections
 - ▶ Operation-wise $memory(O_i)$ is estimated by tensor bit size
 - ▶ σ is the hyper-parameter to control searched model memory usage

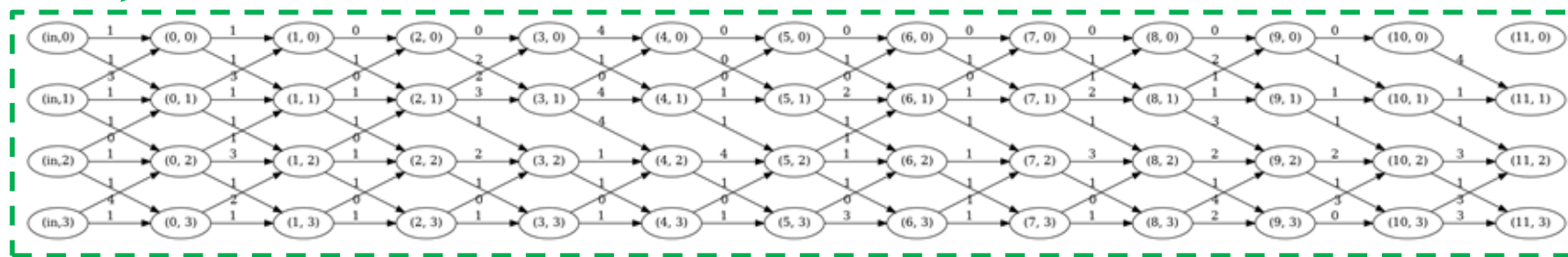
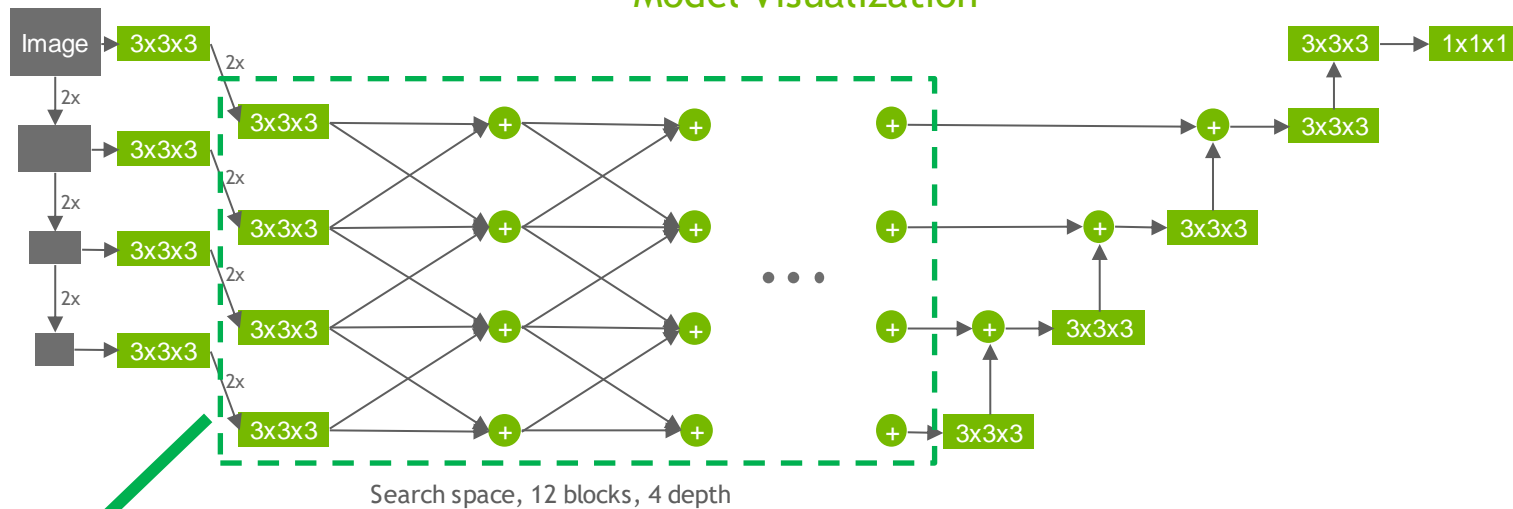
DINTS

Innovations

- **A multi-path model** by utilizing a super-node idea, without relying on a “each cell has a single input path” assumption
 - A novel framework/viewpoint for segmentation network search
- **Entropy loss** to mitigate “binarization problem” in differentiable NAS
 - A novel loss that has been used by only a few concurrent work
- **Topology loss** to solve topology problem in a fully flexible segmentation search space
 - A novel solution for maintaining correct network topology
- **Memory loss** to limit the required memory usage of the searched model
 - A more important computation constraints for medical image analysis than FLOPS

DINTS

Model Visualization



Medical Segmentation Decathlon








[Home](#)[Join](#)[Challenge
Leaderboard](#)

Challenge Leaderboard

Search:

Additional metrics ▾

Hide additional metrics

↑										
#	User (Team)	Created	Mean Position	BRATS 1_Dice (Position)	BRATS 1_Dice (Position)	BRATS 1_NSD (Position)	BRATS 2_Dice (Position)	BRATS 2_NSD (Position)	BRATS 3_Dice (Position)	BR. 3_NSD (Position)
1st	 heyufan1995	30 Oct. 2020	2.9	0.69 (1)	0.69 (1)	0.89 (1)	0.49 (1)	0.73 (1)	0.70 (1)	0.9
2nd	 lsensee	6 Dec. 2019	3.1	0.68 (4)	0.68 (4)	0.88 (7)	0.47 (7)	0.72 (4)	0.68 (3)	0.9
3rd	 fhaghigh (JLiangLab_TransVW)	2 Dec. 2020	3.7	0.68 (5)	0.68 (5)	0.88 (4)	0.47 (4)	0.72 (5)	0.68 (6)	0.9
4th	 JLiangLab	15 Nov. 2020	3.8	0.68 (5)	0.68 (5)	0.88 (4)	0.47 (4)	0.72 (5)	0.68 (6)	0.9
5th	 yucornetto	13 Dec. 2019	5.1	0.68 (7)	0.68 (7)	0.88 (3)	0.49 (2)	0.73 (2)	0.70 (2)	0.9
6th	 ShufanYang	27 April 2020	5.2	0.68 (3)	0.68 (3)	0.88 (6)	0.47 (6)	0.72 (3)	0.68 (3)	0.9
7th	 curtis.abcd (Kakao Brain)	29 Aug. 2019	7.0	0.67 (9)	0.67 (9)	0.87 (11)	0.46 (8)	0.72 (8)	0.68 (8)	0.9

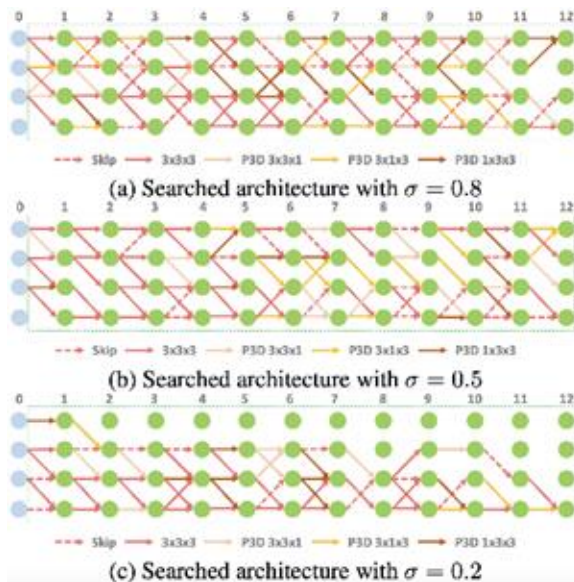
DINTS
5.8 GPU days

C2FNAS
180 GPU days

DINTS

Ablation Study

- Searched models with different GPU memory constraints





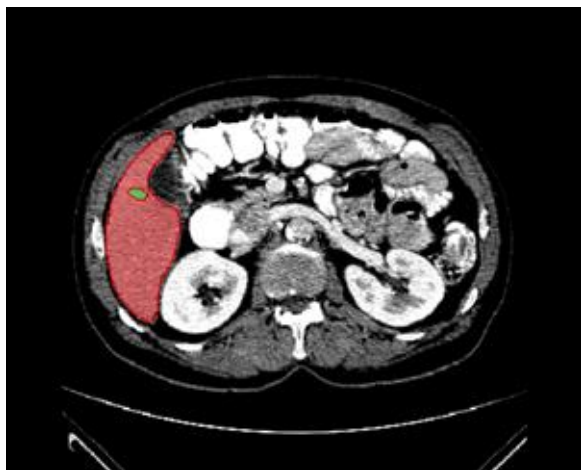
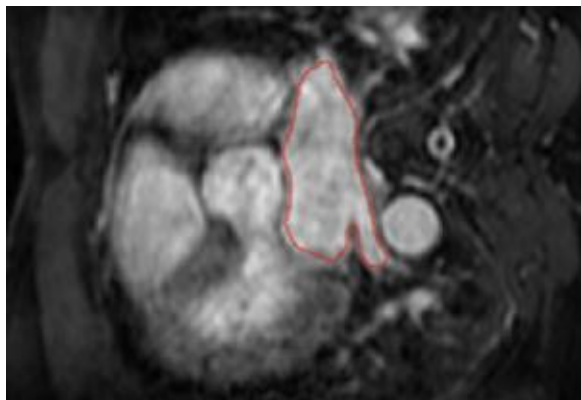
SUMMARY



SUMMARY

Efficient and Effective NAS in 3D Medical Imaging

- We developed efficient NAS algorithms to search in a highly flexible search space
 - Achieving state-of-the-art accuracy in 3D medical image segmentation
 - Searching time is **5.8 GPU days** with large improvement (from 180 GPU days in the baseline C2FNAS)
 - Capable to search different model with varying memory consumptions



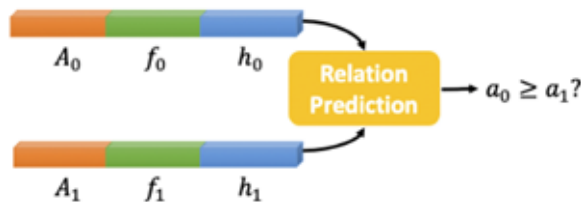
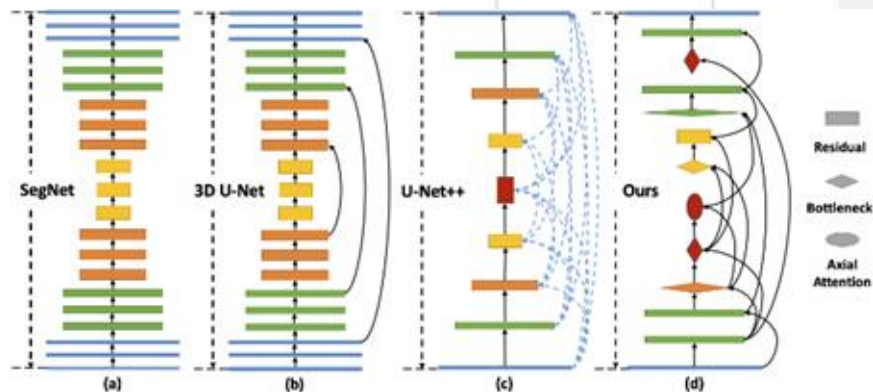
SUMMARY

AutoML in Medical Imaging

- Grand vision
 - Explanation
 - “*What are the critical components in the best solutions?*”
 - Possible, such as using attention mechanism in algorithm
 - Better understanding than human heuristics
 - Better understanding → more efficient algorithms and more effective models
 - Discovery
 - “*What is the new thing that people didn't know before?*”
 - New optimization scheme
 - New machine learning logic
 - New findings in medical imaging

T-AUTOML: AUTOMATED MACHINE LEARNING FOR LESION SEGMENTATION USING TRANSFORMERS IN 3D MEDICAL IMAGING

Dong Yang, Andriy Myronenko, Xiaosong Wang,
Ziyue Xu, Holger R. Roth, Daguang Xu
NVIDIA



TRANSFER LEARNING TOOLKIT

Contributions

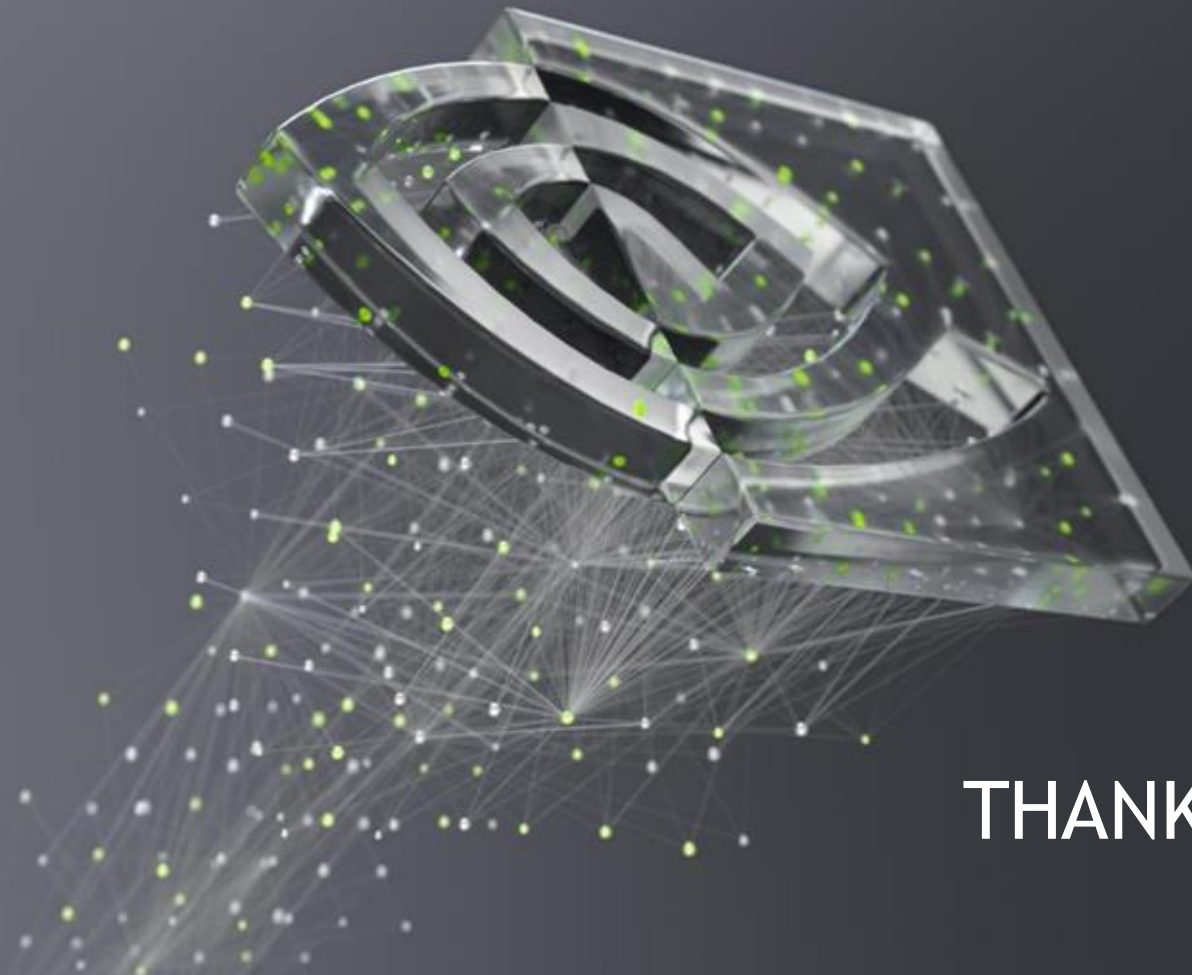
1. Predictor based AutoML algorithm covering all possible components;
2. Searching under reasonable computing budgets;
 - a. 300 GPU-hours for a new 3D large-scale data set;
3. Efficient predictor training algorithm;
4. Flexible architecture design.

	lesion	liver	mean
RA-U-Net [75]	0.5950	0.9610	0.7780
AH-Net [76]	0.6340	0.9630	0.7895
FCN [77]	0.6610	0.9510	0.8060
3D-DenseUNet [78]	0.6960	0.9620	0.8290
H-DenseUNet [20]	0.7220	0.9610	0.8415
LW-HCN [79]	0.7300	0.9650	0.8475
U ³ -Net [80]	0.7369	0.9638	0.8504
Cascade U-ResNets [81]	0.7520	0.9490	0.8505
VolumetricAttention [82]	0.7410	0.9610	0.8510
MA-Net [83]	0.7490	0.9600	0.8545
DistanceMetric [74]	0.7640	0.9650	0.8645
nnU-Net [68]	0.7630	0.9670	0.8650
T-AutoML (ours)	0.7650	0.9670	0.8660

Table 2. LiTS challenge test-set performance evaluation for lesion and liver segmentations in terms of the average Dice score per case. The metrics of our method are copied from the LiTS leaderboard, and the metrics of the other methods are copied from their respective publications and the leaderboard entries.



2021 **ICCV** OCTOBER 11-17
VIRTUAL



THANK YOU!

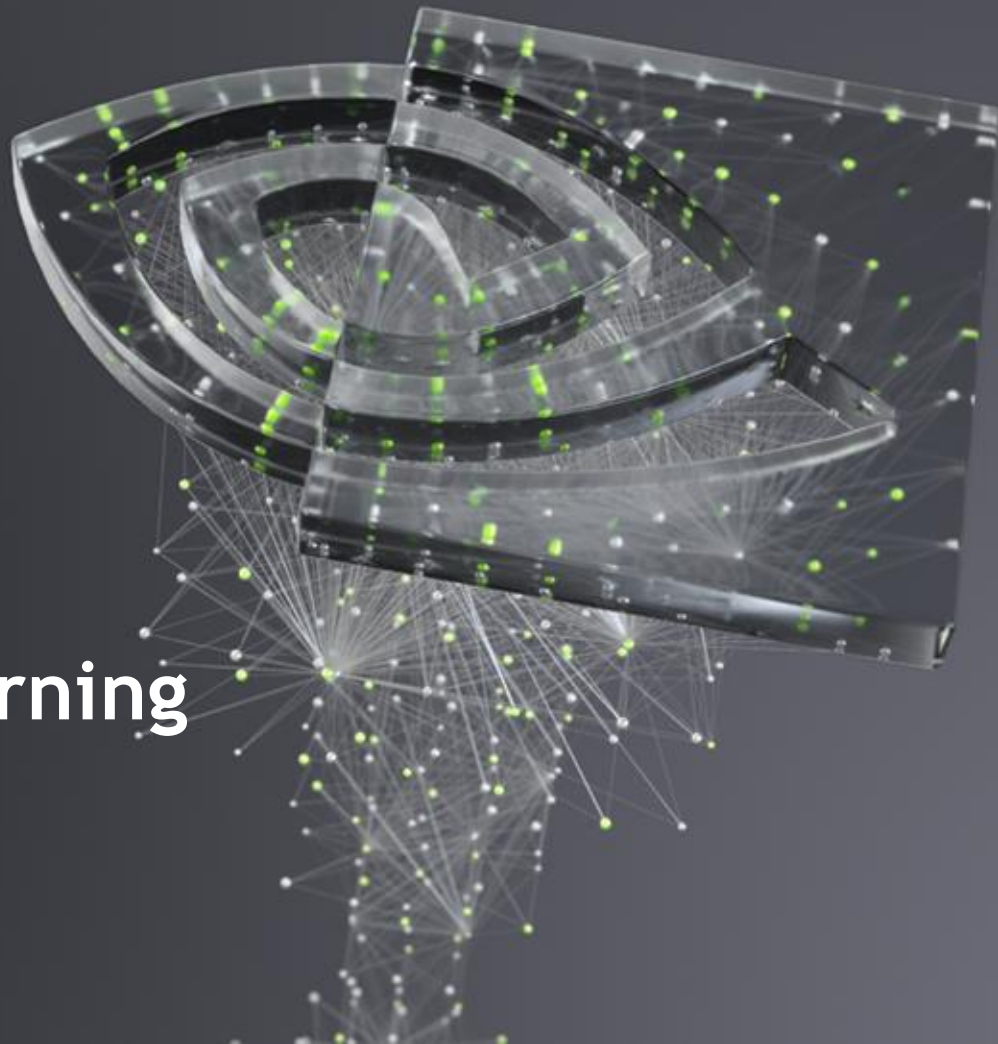




DEEP
LEARNING
INSTITUTE

Self-Supervised Learning Using Transformers

Vishwesh Nath
Applied Research Scientist
Date: 03/23/2022





AGENDA

What is Self-Supervised Learning? What are Transformers?

How to use techniques such as advanced augmentations and contrastive loss for self-supervised learning.

SSL Method Breakdown & Applicability towards downstream tasks

How to use the pre-trained model from SSL for 3D segmentation, classification, detection etc

Hands On Live Demo

Demo for using self-supervised learning and then using the pretrained model for a downstream 3D segmentation task



Self-Supervised Learning & Transformers

What is Self-Supervised Learning?

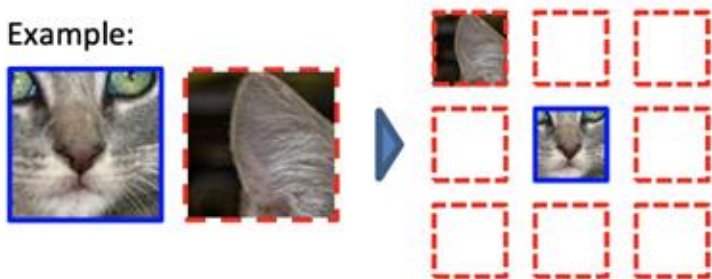
“After all, without labels, it is not even clear what should be represented. How can one write an objective function to encourage a representation to capture, for example, objects, if none of the objects are labeled?” -Efros et. al, ICCV 2015

“A prominent paradigm is the so-called self-supervised learning that defines an annotation free pretext task, using only the visual information present on the images or videos, in order to provide a surrogate supervision signal for feature learning.” -Gidaris et. al, ICLR 2017

Self-Supervised Learning

Unsupervised Visual Representation Learning by Context Prediction

Example:



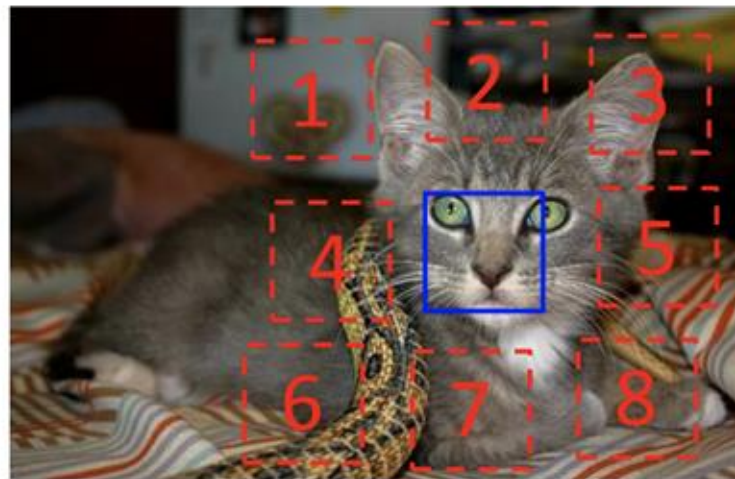
Question 1:



Question 2:



[Efros et. Al 2015]



$$X = \left(\begin{array}{c} \text{cat face} \\ \text{cat ear} \end{array} \right); Y = 3$$

This was just the beginning of the many self-supervised methods that would be proposed ...

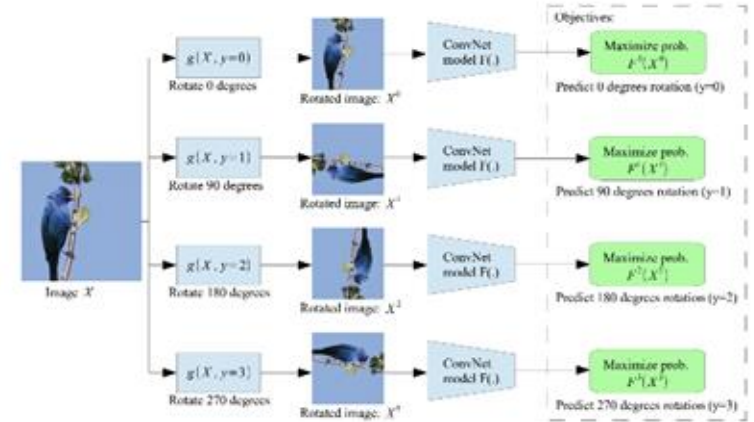
Self-Supervised Learning

Colorful Image Colorization



[Efros et. AI 2016]

Image Rotation Prediction



[Gidaris et. AI 2017]

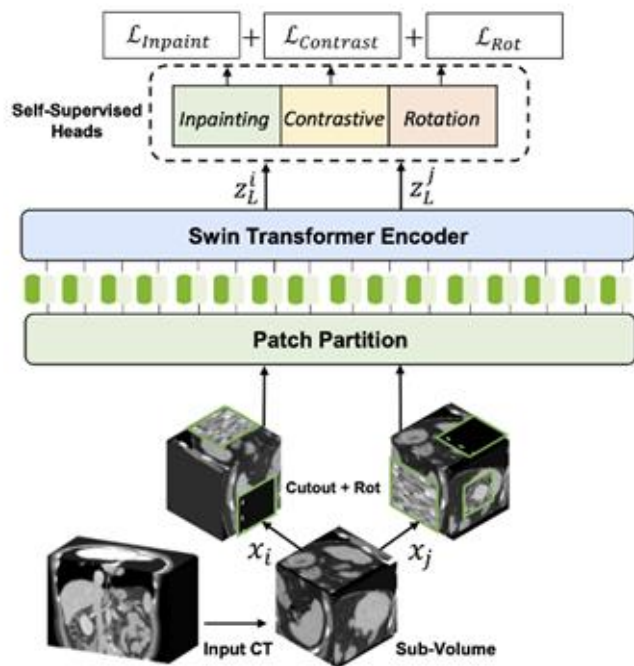
Solving Jigsaw Puzzles



[Favaro et. AI 2016]

Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis (CVPR 2022)

Yucheng Tang, Dong Yang, Wenqi Li, Holger R. Roth, Bennett A. Landman, Daguang Xu, Vishwesh Nath, Ali Hatamizadeh



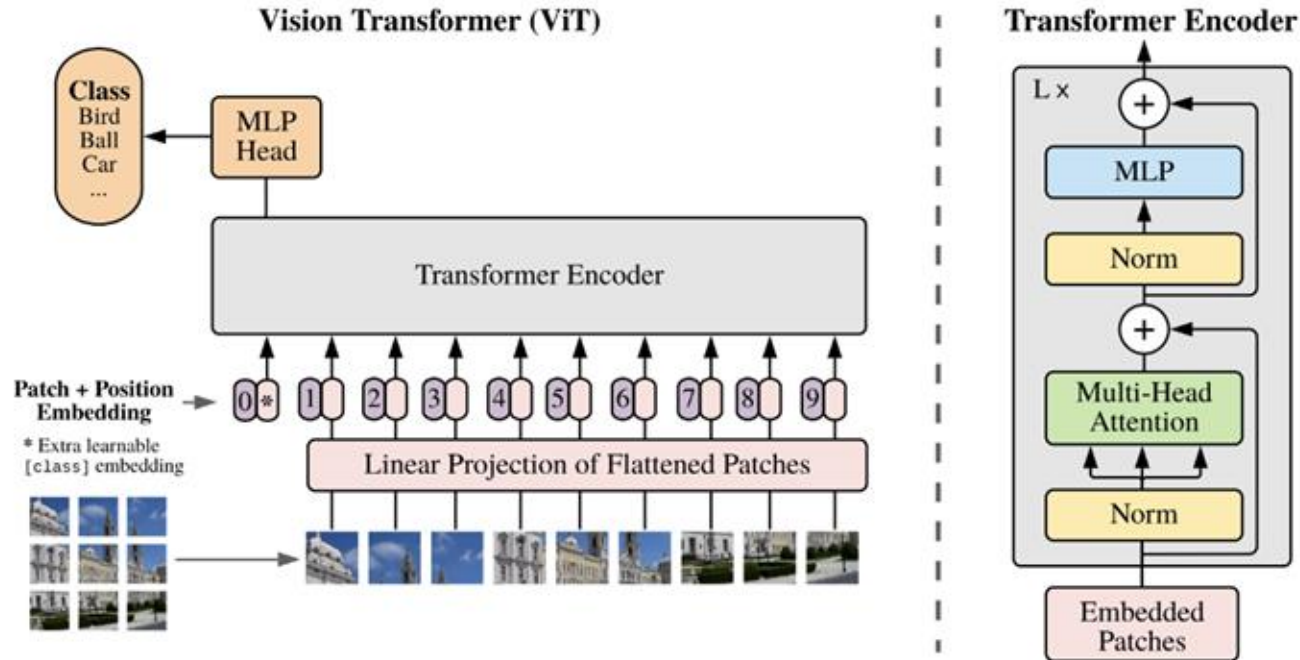
Methods	Spl	RKid	LKid	Gall	Eso	Liv	Sto	Aor	IVC	Veins	Pan	AG	Avg.
SETR NUP [61]	0.931	0.890	0.897	0.652	0.760	0.952	0.809	0.867	0.745	0.717	0.719	0.620	0.796
SETR PUP [61]	0.929	0.893	0.892	0.649	0.764	0.954	0.822	0.869	0.742	0.715	0.714	0.618	0.797
SETR MLA [62]	0.930	0.889	0.894	0.650	0.762	0.953	0.819	0.872	0.739	0.720	0.716	0.614	0.796
ASPP [9]	0.935	0.892	0.914	0.689	0.760	0.953	0.812	0.918	0.807	0.695	0.720	0.629	0.811
TransUNet [7]	0.952	0.927	0.929	0.662	0.757	0.969	0.889	0.920	0.833	0.791	0.775	0.637	0.838
CoTr* [55]	0.943	0.924	0.929	0.687	0.762	0.962	0.894	0.914	0.838	0.796	0.783	0.647	0.841
CoTr [55]	0.958	0.921	0.936	0.700	0.764	0.963	0.854	0.920	0.838	0.787	0.775	0.694	0.844
RandomPatch [52]	0.963	0.912	0.921	0.749	0.760	0.962	0.870	0.889	0.846	0.786	0.762	0.712	0.844
PaNN [64]	0.966	0.927	0.952	0.732	0.791	0.973	0.891	0.914	0.850	0.805	0.802	0.652	0.854
nnUNet [28]	0.967	0.924	0.957	0.814	0.832	0.975	0.925	0.928	0.870	0.832	0.849	0.784	0.888
UNETR [24]	0.972	0.942	0.954	0.825	0.864	0.983	0.945	0.948	0.890	0.858	0.852	0.812	0.891
Swin UNETR	0.975	0.942	0.954	0.826	0.864	0.985	0.946	0.948	0.893	0.894	0.868	0.840	0.908

Loss Function	Average Accuracy	
	Dice \uparrow	HD \downarrow
Scratch	83.43	42.36
\mathcal{L}_{rot}	83.56	36.19
$\mathcal{L}_{contrast}$	83.67	38.81
$\mathcal{L}_{inpaint}$	83.85	28.94
$\mathcal{L}_{inpaint} + \mathcal{L}_{rot}$	84.01	26.06
$\mathcal{L}_{inpaint} + \mathcal{L}_{contrast}$	84.45	24.37
$\mathcal{L}_{inpaint} + \mathcal{L}_{contrast} + \mathcal{L}_{rot}$	84.72	20.03

What are Transformers?

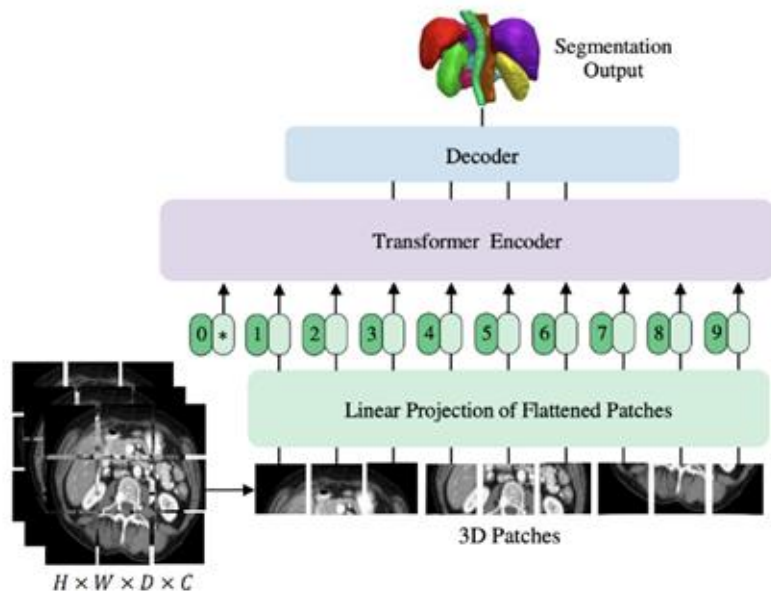
- Transformers became the de-facto for NLP
- Vision Transformer (ViT) was one of the first ones created for computer vision tasks

[Dosovitsky et. al 2021]

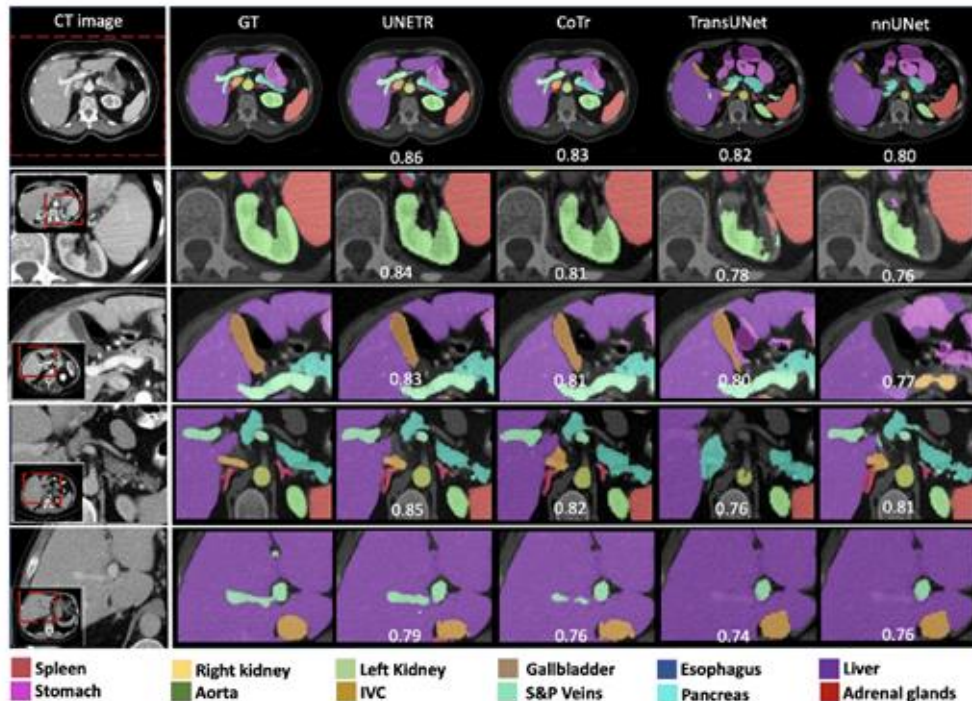


UNETR: Transformers for 3D Medical Image Segmentation (WACV 2021)

Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett A. Landman, Holger R. Roth, Daguang Xu



One of the first bodies of work to adapt Transformer for medical image segmentation

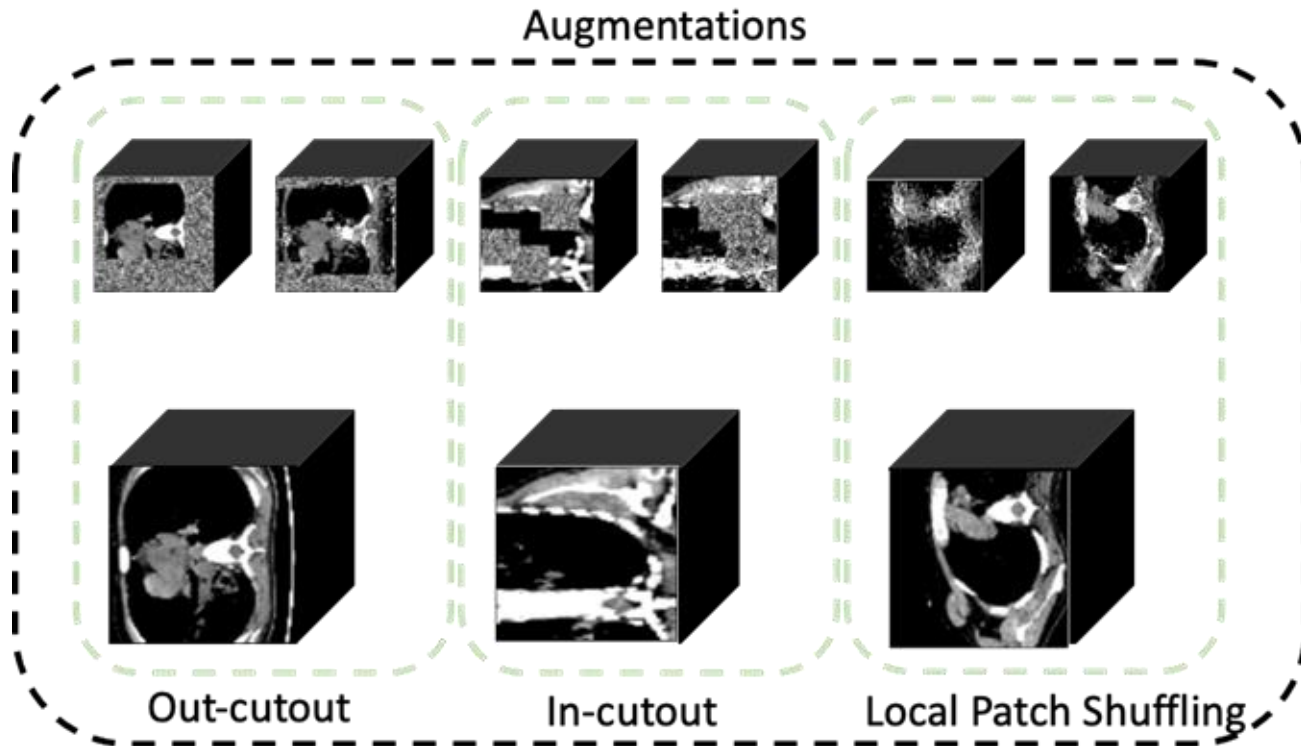




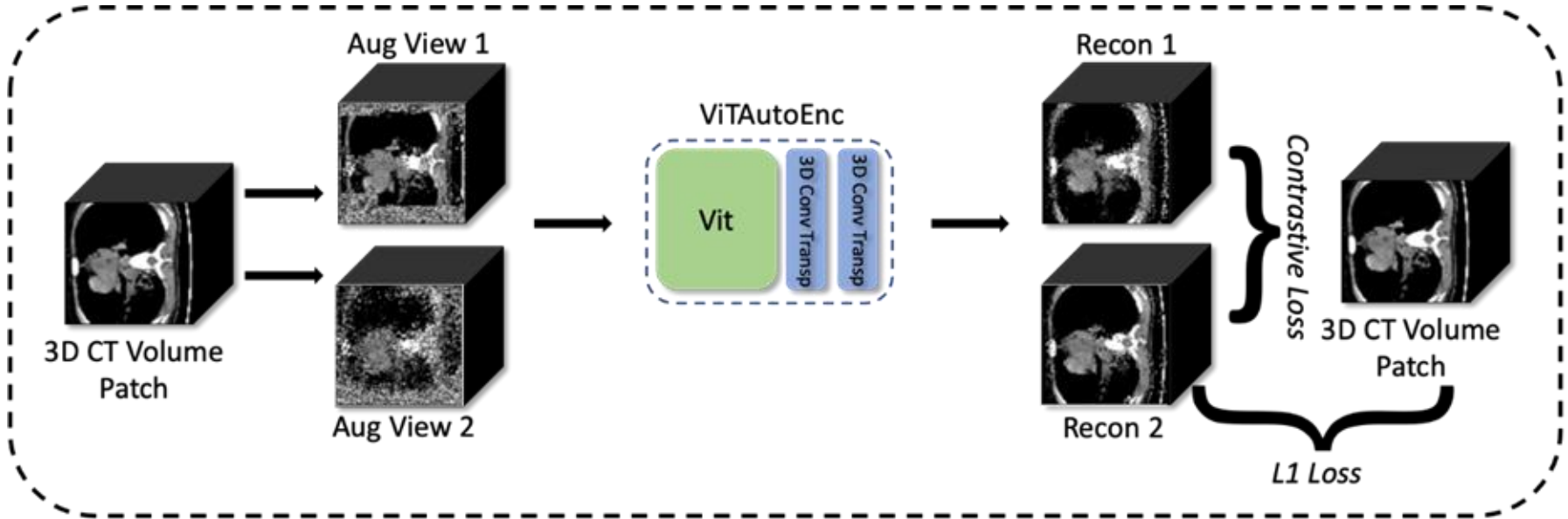
Self-Supervised Learning Method Overview

Self-Supervised Pretext Tasks

Annotation free pretext tasks suited towards computer vision and medical imaging

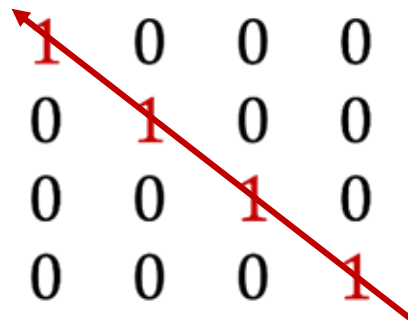


Contrastive Learning with Self-Supervised Learning



How does Contrastive Loss Work?

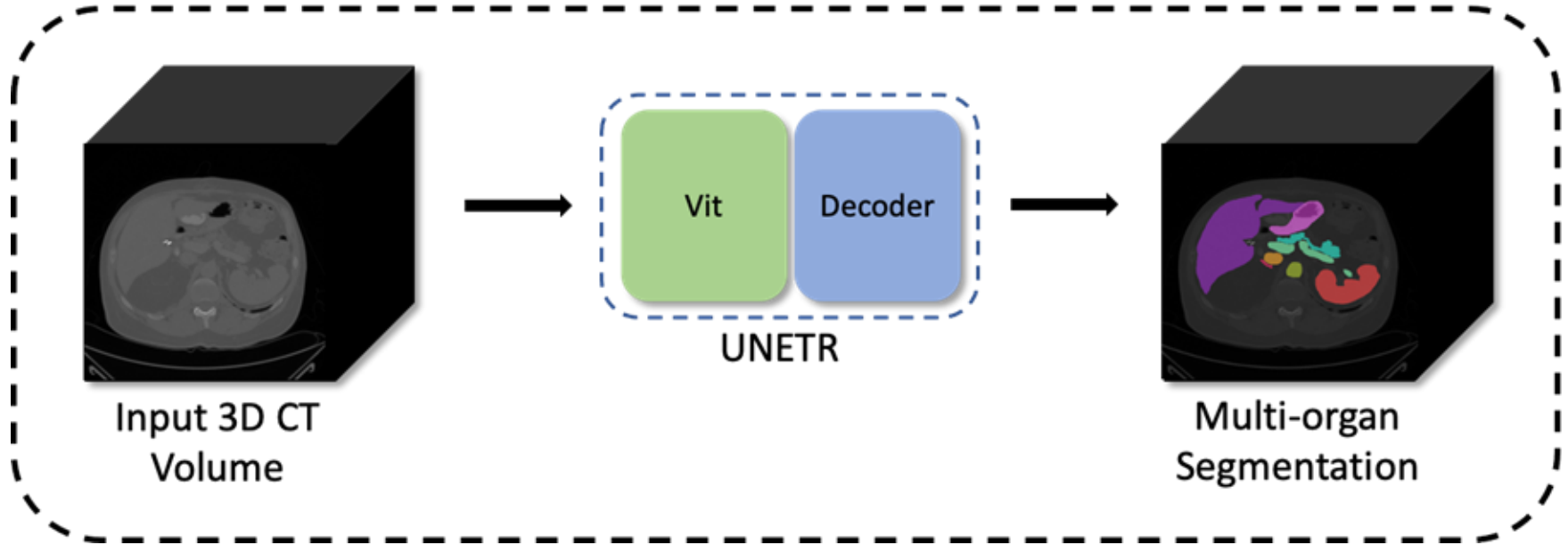
- Given a batch of a pairs of images, when inter-pairs are formed the resultant is a matrix
- Diagonal elements are positive pairs
- All off-diagonal elements are negative pairs
- Tau denotes temperature to control the effect of learning from positive and negative pairs



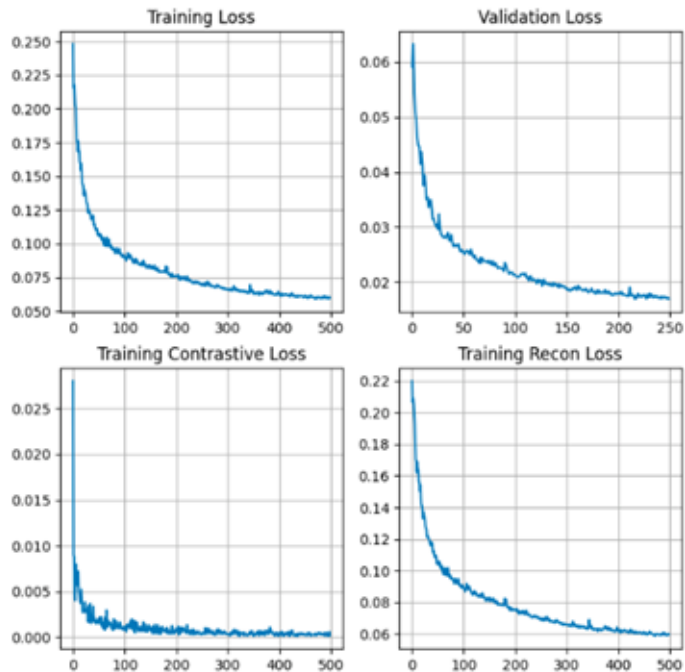
$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)},$$

[Hinton et. Al 2020]

Downstream Task: Multi-Organ 3D Segmentation



Expected Results: SSL



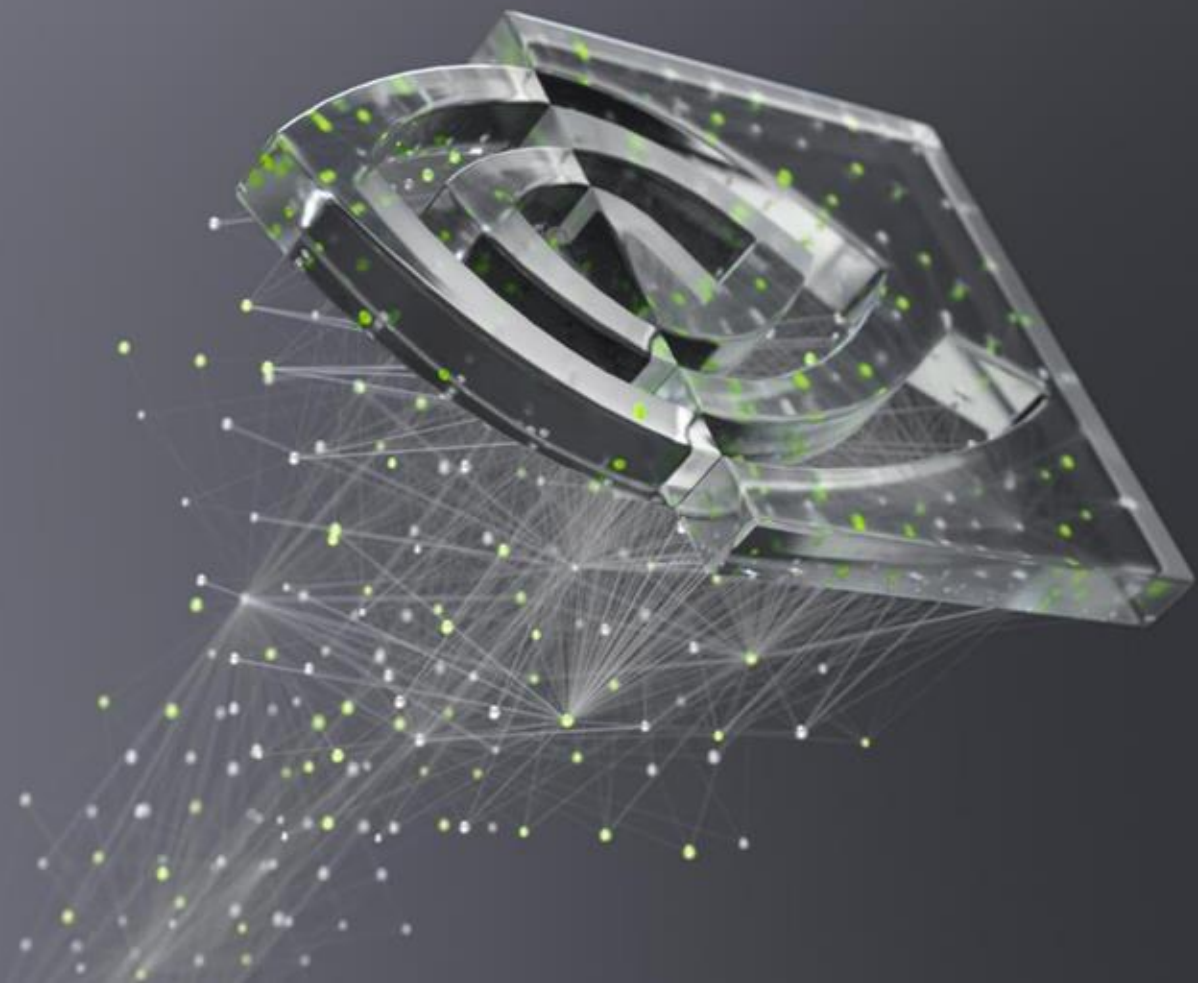
- These results are based on when using all TCIA-Covid 19 data which is ~771 3D volumes
- Smooth downward trending curves indicate that the model is training well
- A similar observation can be made for validation loss (top right corner)

Expected Results: 3D Multi-organ Segmentation

Training Volumes	Validation Volumes	Random Init Dice score	Pre-trained Dice Score	Relative Performance Improvement
6	6	63.07	70.09	~11.13%
12	6	76.06	79.55	~4.58%
24	6	78.91	82.30	~4.29%

Exercise for Self-Supervised Learning

- Contrastive Loss is extremely sensitive to the assigned loss weight and also the temperature hyper-parameter
- For experimentation purposes and also to get an intuition of how to tune hyper-parameters, we will try the following:
 - Modification of the weight of the CL loss term
 - Try to find the contrastive loss term and replace its weight with what you think is the right hyper-parameter.
 - It can range anywhere from $[0, 1]$. Some starting examples to try are $1e-3$, $1e-4$.



DEEP
LEARNING
INSTITUTE