Research report on Semi-supervised … Segmentation

Xinze LI

Nov.23, 2024

坏消息一则

- MICCAI2025 DDL 提前了16天

MICCAI 2025 ☆

Sep 23-27, 2025 Daejeon, Republic of Korea

International Conference on Medical Image Computing and Computer Assisted Intervention

CCF B | CORE A | THCPL B **NOTE:** abstract deadline on Feb 7, 2025.

交叉/综合/新兴

90 days 00 h 10 m 15 s

Deadline: Fri Feb 21st 2025 15:59:00 CST (2025-02-20 23:59:00 UTC-8)

website: https://conferences.miccai.org/2025/en/default.asp

- 做梦：

ICCV 2025 ☆

October 19-25, 2025 Honolulu, Hawaii

IEEE International Conference on Computer Vision

CCF A | CORE A* | THCPL A

Acc. Rate: 26.2%(2160/8260 23') 人工智能

TBD

Deadline: pull request to update

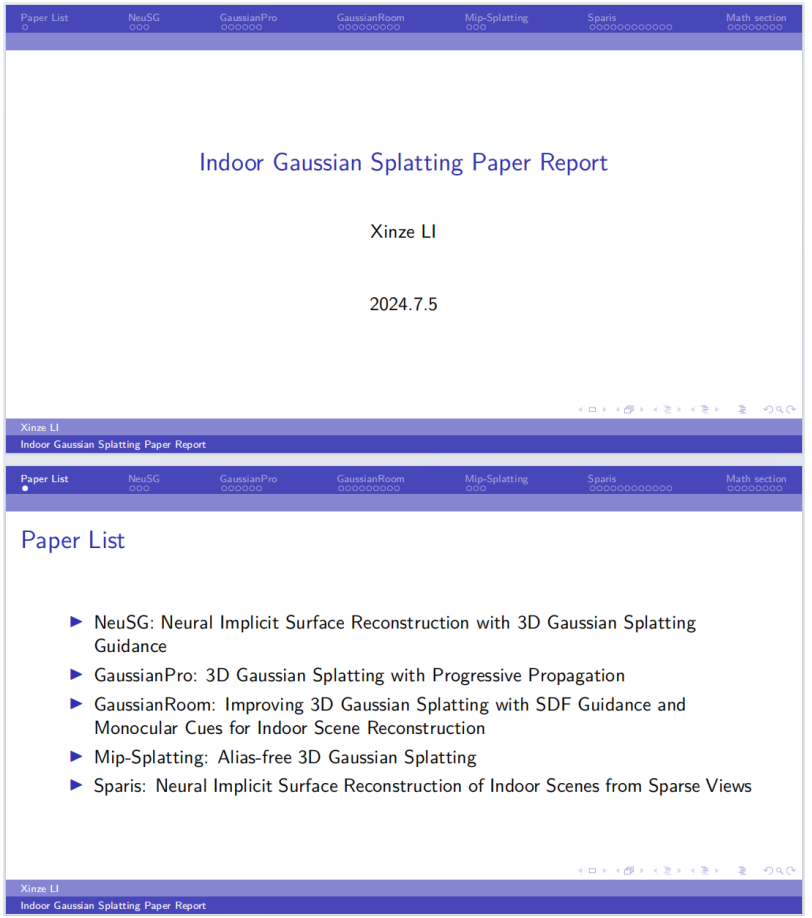
website: http://iccv2025.thecvf.com/home/

这一年研究过的 Topic

- Content:
  - 3D-Reconstruction, focusing on Indoor Scene and 3D Gaussian Splatting
  - Multimodality Image Fusion
  - Semi-supervised Medical Image Segmentation
- 坏消息：都没做成功，兜兜转转还是回到了分割
  - 也许之后有机会还是会把一些想到的点子做出来，不想浪费掉
- 好消息：
  - 拓宽了知识面
  - 提升了 Story telling 水平
  - 第一个工作居然也投中了（感谢苏老师和全组的师兄师姐）
  - 有很多 Idea 来自这些不同的领域

一个受到 Indoor Reconstruction 的启发

- 目前的场景重建技术：3D Gaussian Splatting，可以理解为模拟往物体上泼雪球，雪球碰到物体后的分布符合 Gaussian Distribution.
- 雪球泼溅得准不准->重建效果
- 室内场景边缘多：要引导模型泼到这些边缘上->SDF



什么是 SDF

- Signed Distance Function

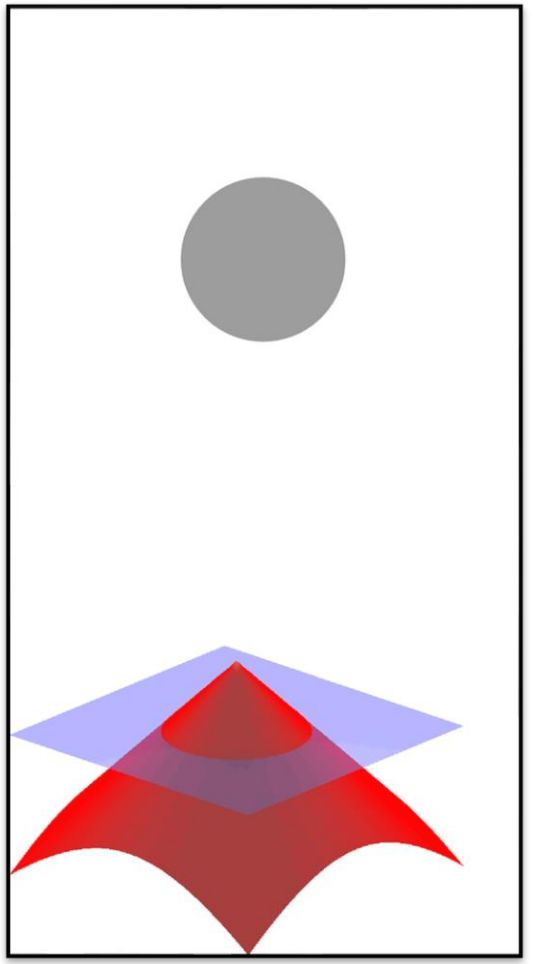
Let Ω be a subset of a metric space X with metric d, and $\partial\Omega$ be its boundary. The distance between a point x of X and the subset $\partial\Omega$ of X is defined as usual as

$$d(x, \partial\Omega) = \inf_{y \in \partial\Omega} d(x, y),$$

where $\inf$ denotes the infimum.

The *signed distance function* from a point x of X to Ω is defined by

$$f(x) = \begin{cases} d(x, \partial\Omega) & \text{if } x \in \Omega \\ -d(x, \partial\Omega) & \text{if } x \notin \Omega. \end{cases}$$



一个被遗忘的东西

1 **Mutual information in coupled multi-shape model for medical image segmentation**

在医学图像分割中，多个符号距离函数被用作多个形状类别的隐式表示，有效地捕捉不同形状之间的共同变化。

Highly Cited

2004 · **163** Citations · A. Tsai et al. · Medical image analysis

3 **A shape-based approach to the segmentation of medical imagery using level sets**

训练数据的符号距离表示用于基于形状的医学图像分割方法，处理多维数据并且对噪声和初始轮廓位置具有鲁棒性。

Highly Cited

2003 · **994** Citations · A. Tsai et al. · IEEE Transactions on Medical Imaging

Same author

短暂的文艺复兴

6 A novel segmentation model for medical images with intensity inhomogeneity based on adaptive perturbation

所提出的分割模型在使用符号距离函数和自适应扰动的噪声鲁棒性和分割精度方面优于最先进的方法。

2018 · 0 Citations · Haiping Yu et al. · Multimedia Tools and Applications

无人问津

5 Shape-Aware Organ Segmentation by Predicting Signed Distance Maps

从医学扫描中预测符号距离图 (SDM) 可以提高分割性能，并且具有更好的平滑度和形状的连续性，从而实现更好的器官分割。

Highly Cited

2019 · 94 Citations · Yuan Xue et al. · ArXiv

AAAI2020

文艺复兴的小高潮

8 SimCVD: Simple Contrastive Voxel-Wise Representation Distillation for Semi-Supervised Medical Image Segmentation

SimCVD 是一个简单的对比蒸馏框架，在使用较少标记数据的医学图像分割中获得了 90.85% 的平均 Dice 分数。

Highly Cited

2021 · **110** Citations · Chenyu You et al. · IEEE Transactions on Medical Imaging

TMI 高引：228

9 Shape-aware Semi-supervised 3D Semantic Segmentation for Medical Images

形状感知半监督分割通过使用多任务深度网络联合预测物体表面的语义分割和符号距离图（SDM），改善医学图像中的形状估计。

Highly Cited

2020 · **172** Citations · Shuailin Li et al. ·

MICCAI2020

文艺复兴是一种错觉

4 Score-Based Generative Models for Medical Image Segmentation using Signed Distance Functions

医学图像分割中的符号距离函数（SDF）可以对统计量进行更自然的扭曲和评估，增强预测鲁棒性。

2023 · 3 Citations · L. Bogensperger et al. · ArXiv

效果不佳，方法极度不自然；
无名会议（德国CV）闻所未闻；

10 An Instance Segmentation Algorithm Based On Signed Distance Field

提出的基于符号距离场的实例分割算法提高了定位精度，解决了医学图像处理中的粗边界分割问题。

2023 · 0 Citations · Hongbao Sun et al. · 2023 42nd Chinese Control Conference (CCC)

无人问津，无人引用

早期的MICCAI就是这么朴实无华

- *Shape-aware Semi-supervised 3D Semantic Segmentation for Medical Images*

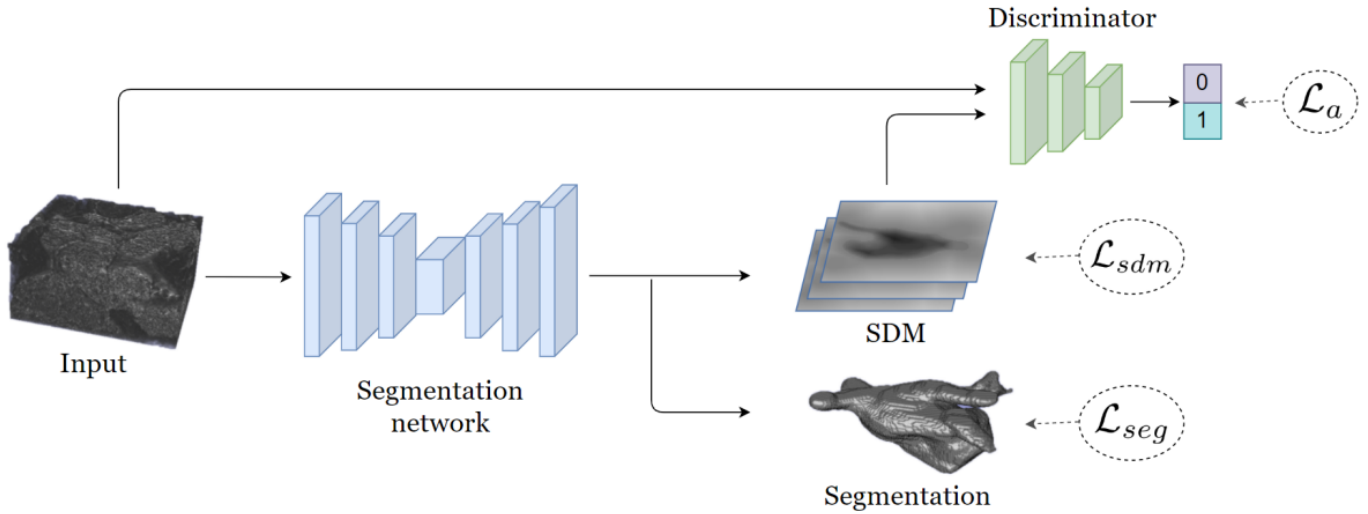


Fig. 1: Overview of our method. Our network takes as input a 3D volume, and predicts a 3D SDM and a segmentation map. Our learning loss consists of a multi-task supervised term and an adversarial loss on the SDM predictions.

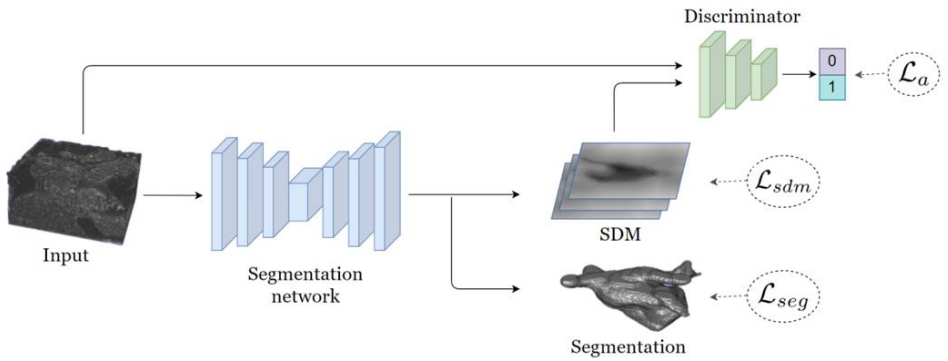
L_{sdm}

$$L_{sdm} = \frac{1}{N}\sum_{i=1}^{N} l_{mse}(f_{sdm}(X_i;\theta), Z_i)$$

- N 为 Labeled data 数量
- $X_i \in R^{H\times W\times D}$ 为 Input Image
- θ 该项参数
- Z_i 为 GT 的 Signed Distance Map
- Comment:一个非常直观的设计来约束边界

一个问题

- Problem: 还没有保证在 Unlabeled data 上生成的 Signed Distance Map 与 Labeled data 上生成的 Consistency
- Solution: 设计了一个GAN网络来维持一致性
  - Input: (X_i, S) 图像及其网络生成的SDM
  - Output: $D(X_i, S; \zeta)$ 一个类概率，表示该输入是否来自标注数据集；
  - 判别器的目标是区分来自 Labeled data 的高质量SDM和来自Unlabeled data 的SDM。



那年他们在这个数据集上还没开始比5%

Table 1: Quantitative comparisons of semi-supervised segmentation models on the LA dataset. All models use the V-Net as backbone network. Results on two different data partition settings show that our SASSNet outperforms the state-of-the-art results consistently.

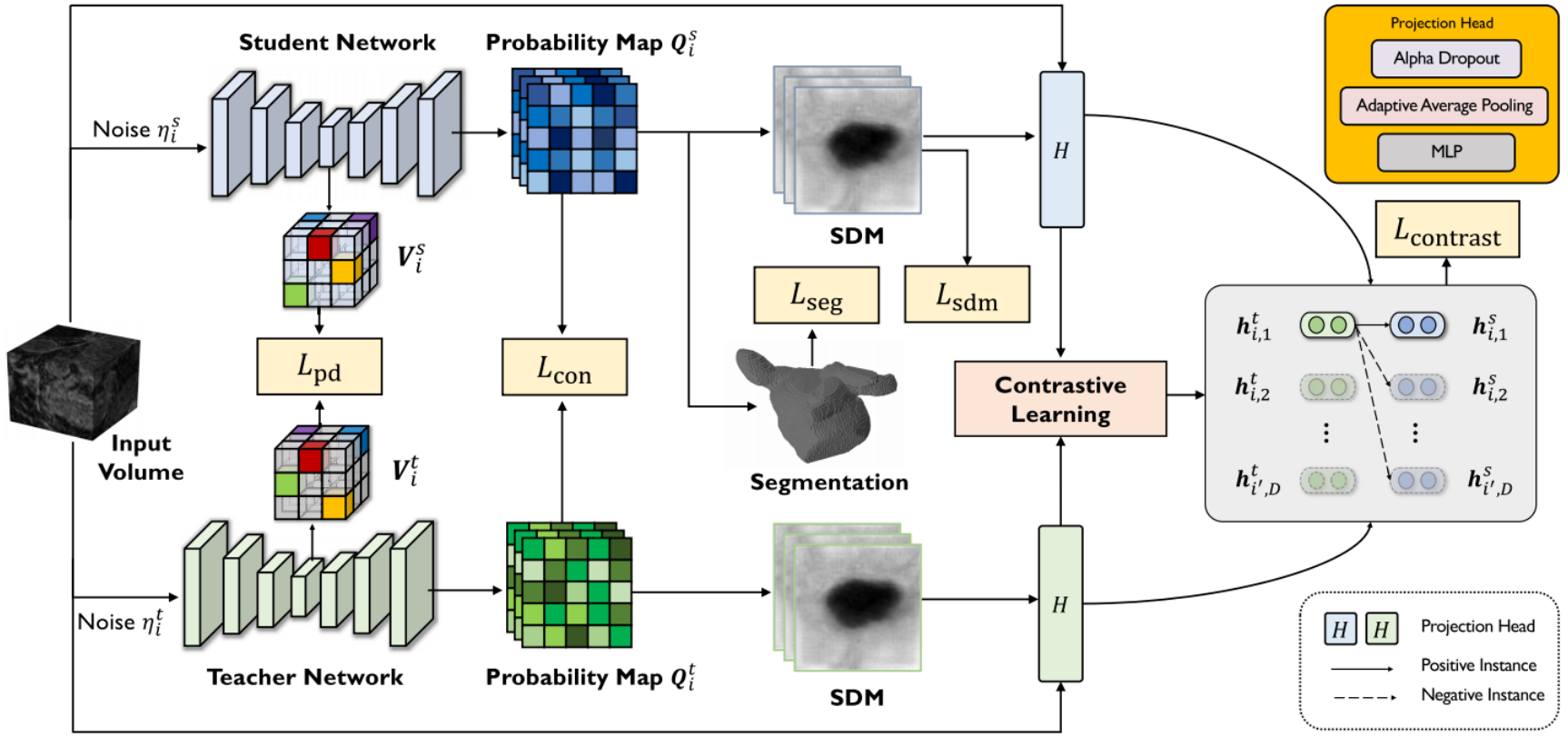
Method	# scans used		Metrics			
	Labeled	Unlabeled	Dice[%]	Jaccard[%]	ASD[voxel]	95HD[voxel]
V-Net	80	0	91.14	83.82	1.52	5.75
V-Net	16	0	86.03	76.06	3.51	14.26
DAP [20]	16	64	87.89	78.72	2.74	9.29
ASDNet [11]	16	64	87.90	78.85	**2.08**	9.24
TCSE [9]	16	64	88.15	79.20	2.44	9.57
UA-MT [18]	16	64	88.88	80.21	2.26	7.32
UA-MT(+NMS)	16	64	89.11	80.62	2.21	**7.30**
SASSNet(ours)	16	64	89.27	80.82	3.13	8.83
SASSNet(+NMS)	16	64	**89.54**	**81.24**	2.20	8.24
V-Net	8	0	79.99	68.12	5.48	21.11
DAP [20]	8	72	81.89	71.23	3.80	15.81
UA-MT [18]	8	72	84.25	73.48	3.36	13.84
UA-MT(+NMS)	8	72	84.57	73.96	2.90	12.51
SASSNet(ours)	8	72	86.81	76.92	3.94	12.54
SASSNet(+NMS)	8	72	**87.32**	**77.72**	**2.55**	**9.62**

TABLE I

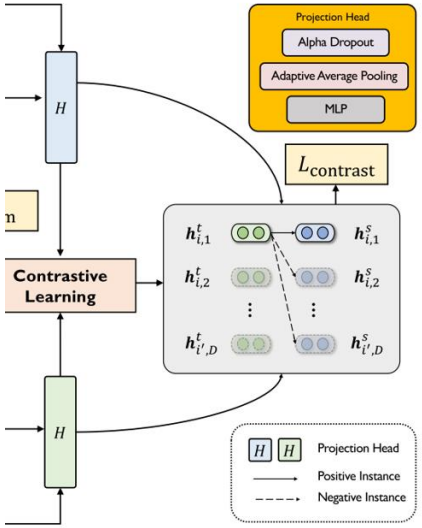
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE LA DATASET.

Method	Scans used		Metrics			
	Labeled	Unlabeled	Dice(%) ↑	Jaccard(%) ↑	95HD (voxel) ↓	ASD (voxel) ↓
V-Net	4	0	43.32	31.43	40.91	12.13
V-Net	8	0	79.87	67.60	26.65	7.94
V-Net	16	0	85.94	75.99	16.70	4.80
V-Net	80	0	90.98	83.61	8.58	2.10
DTC [4] (AAAI'21)	4 (5%)	76 (95%)	80.14	67.88	24.08	7.18
SS-Net [22] (MICCAI'22)			83.33	71.79	15.70	4.33
MC-Net+ [23] (MedIA'22)			83.23	71.70	14.92	3.43
CAML [24] (MICCAI'23)			86.78	75.97	10.34	2.89
Ours			**88.23**	**79.12**	**9.51**	**1.73**
DTC [4] (AAAI'21)	8 (10%)	72 (90%)	84.55	73.91	13.80	3.69
SS-Net [22] (MICCAI'22)			86.56	76.61	12.76	3.02
MC-Net+ [23] (MedIA'22)			87.68	78.27	10.35	1.85
CE-MT [25] (BIBM'23)			87.34	77.91	8.79	2.12
CAML [24] (MICCAI'23)			89.33	80.17	8.82	2.08
Ours			**89.70**	**81.45**	**7.26**	**1.62**
DTC [4] (AAAI'21)	16 (20%)	64 (80%)	87.79	78.52	10.29	2.50
SS-Net [22] (MICCAI'22)			88.19	79.21	8.12	2.20
MC-Net+ [23] (MedIA'22)			90.60	82.93	6.27	1.58
CE-MT [25] (BIBM'23)			89.67	80.53	7.21	1.99
CAML [24] (MICCAI'23)			90.37	82.56	5.64	1.72
Ours			**91.22**	**83.96**	**5.34**	**1.54**

SimCVD: Simple Contrastive Voxel-Wise Representation Distillation for Semi-Supervised Medical Image Segmentation



一个有意思的设计



- 开门见山：应用 Dropout 作为 Mask;然后做对比学习
- 下面以 $L_{contrast}$ 的设计为例
- 给定 Input volume X_i，学生/教师的 SDM 记为 $Q_i^{s,sdm}$/ $Q_i^{t,sdm}$
- 首先进行了一个特征融合：$Q_i^{t/s,ba} = X_i + Q_i^{t/s,sdm}$
  - 本人解读：融合距离和强度信息
- 然后输入到 Projection Head
- 接着：使用两个独立的 dropout mask：z_i^s / z_i^t，使用 InfoNCE 损失进行正负样本对比

好的设计：对比学习

- 来自两个边界感知特征的相同切片视为正样本，而来自不同位置或不同输入的切片视为负样本。
  - **正样本**是从同一对象或图像中提取的两个视图
  - **负样本**是来自不同对象或图像的视图
- 细节（见图），不细说了
- 本人解读：换了一种方式实现了

上一篇文章的 GAN 的功能

Denoting the output of the projection head as $\mathbf{H}_i^s = \mathcal{H}(\mathbf{Q}_i^{s,\text{ba}}; z_i^s)$, $\mathbf{H}_i^t = \mathcal{H}(\mathbf{Q}_i^{t,\text{ba}}; z_i^t)$, and the jth row of $\mathbf{H}_i$ as $\mathbf{h}_{i,j}$, the InfoNCE loss [57] is defined by:

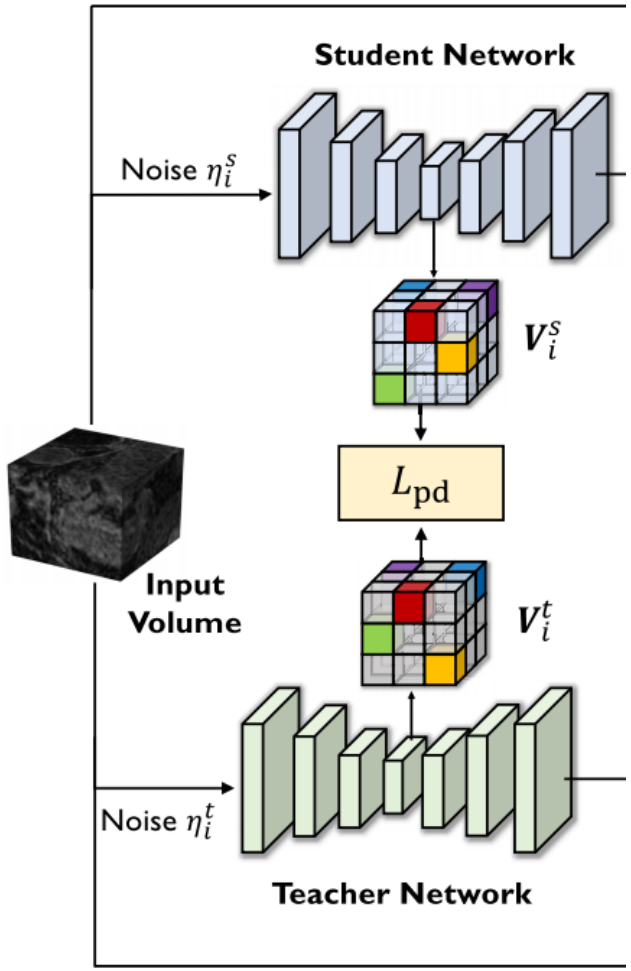
$$\mathcal{L}(\mathbf{h}_{i,j}^t, \mathbf{h}_{i,j}^s) = -\log \frac{\exp(\mathbf{h}_{i,j}^t \cdot \mathbf{h}_{i,j}^s / \tau)}{\sum_{k,l} \exp(\mathbf{h}_{i,j}^t \cdot \mathbf{h}_{k,l}^s / \tau)}, \quad (2)$$

where τ is a temperature hyperparameter. The indices k and l in the denominator are randomly sampled from a mini-batch of images such that B 2D slices are sampled in total. i and j denote the 3D image index and slice index, respectively. The $\mathbf{h}_{k,l}^s$'s in the denominator that are not $\mathbf{h}_{i,j}^s$ are called negative samples. Inspired by the recent success [24], our boundary-aware contrastive loss is defined as:

$$\mathcal{L}_{\text{contrast}} = \frac{1}{|\mathcal{N}^+|} \sum_{\forall (i,j) \in \mathcal{N}^+} [\mathcal{L}(\mathbf{h}_{i,j}^t, \mathbf{h}_{i,j}^s) + \mathcal{L}(\mathbf{h}_{i,j}^s, \mathbf{h}_{i,j}^t)], \quad (3)$$

另外一个不错的点

- 此处的L_{pd}设计了一个知识蒸馏
- Voxel 2 voxel 的 Pair-wide 蒸馏，可以关注样本之间的结构关系
- 本人解读：关注全局也要关注细节



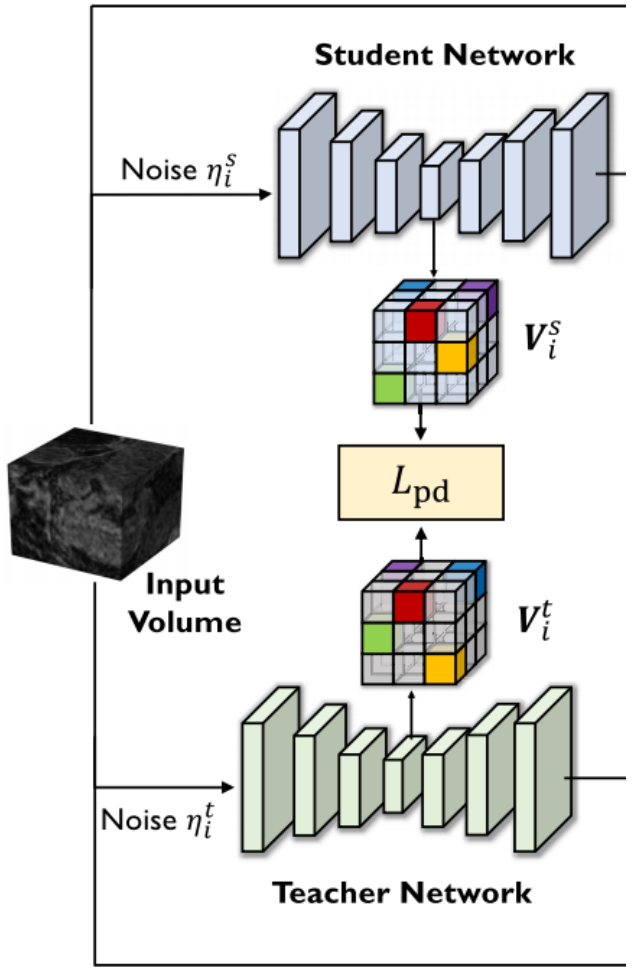
另外一个不错的点

- Hidden Pattern Size: $H' \times W' \times D' \times D_e$ (D_e 为特征图深度)
- 记$\mathrm{V}_t^{t/s} \in R^{H'W'D' \times D_e}$ 是teacher/student 的 Encoder 的 flattened hidden patterns.

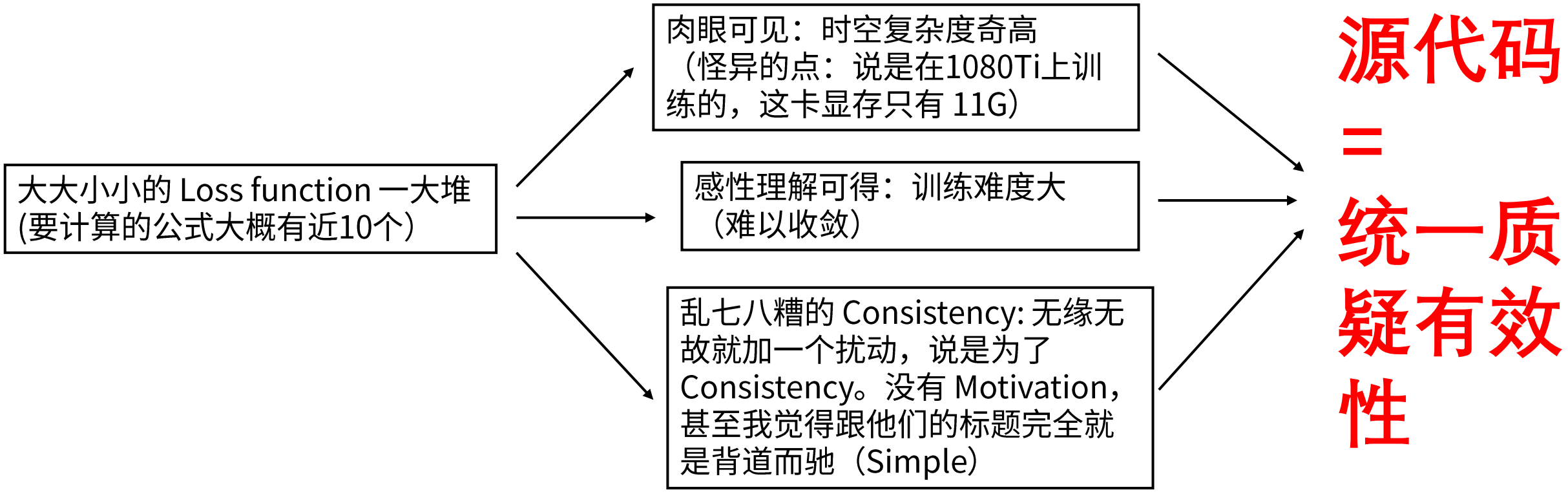
$$\mathcal{L}_{\mathrm{pd}} = -\frac{1}{M} \sum_{i=N+1}^{N+M} \sum_{j=1}^{H'W'D'} \log \frac{\exp(s(\mathbf{v}_{i,j}^s, \mathbf{v}_{i,j}^t))}{\sum_k \exp(s(\mathbf{v}_{i,j}^s, \mathbf{v}_{i,k}^t))}, \quad (4)$$

where $s(\mathbf{v}_1, \mathbf{v}_2) = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|}$ measures the cosine of the angle between two $\mathbf{v}$'s as their similarity. Note again that this loss also involves all the unlabeled data.

- 讲道理是这里用 Cosine similarity 很合理
- ~~不讲道理是自引使用我们的 CORAL-similarity~~



一些批评


大大小小的 Loss function 一大堆
(要计算的公式大概有近10个)
肉眼可见：时空复杂度奇高
（怪异的点：说是在1080Ti上训练的，这卡显存只有 11G）
感性理解可得：训练难度大
（难以收敛）
乱七八糟的 Consistency: 无缘无故就加一个扰动，说是为了 Consistency。没有 Motivation，甚至我觉得跟他们的标题完全就是背道而驰（Simple）
没有开源代码
=
统一质疑有效性