Cross-Modality Domain Adaptation for Medical Image Segmentation and Classification

Sun, September 18

Reuben Dorent

Program

3:40pm - **Challenge presentation** *(25 minutes)*

4:05pm – **Oral Session Task 1** *(40 minutes)*

MAI: *"Multi-view Cross-Modality MR Image Translation for Vestibular Schwannoma and Cochlea Segmentation"*

Bogyeong Kang, Hyeonyeong Nam), Ji-Wung, Keun-Soo Heo, Tae-Eui Kam

ne2e: *"Unsupervised Domain Adaptation in Semantic Segmentation Based on Pixel Alignment and Self-Training (PAST)"*

Hexin Dong, Fei Yu, Mingze Yuan, Jie Zhao, Bin Dong, Li Zhang, Luyi Han, Yunzhi Huang, Tao Tan, Ritse Mann

LaTIM: *"Tumor blending augmentation using one-shot generative learning for vestibular schwannoma and cochlea cross-modal segmentation"*

Guillaume Sallé, Pierre-Henri Conze, Julien Bert, Nicolas Boussion, Ulrike Schick, Dimitris Visvikis, Vincent Jaouen

4:45pm – **Sponsor presentation: NVIDIA** *(5 minutes)*

4:50pm – **Task 1: Evaluation design and results announcement** *(20 minutes)*

5:10pm – **Oral Session Task 2** (20 minutes)

Super Polymerization: *"Unsupervised Cross-Modality Domain Adaptation for Vestibular Schwannoma Segmentation and Koos Grade Prediction based on Semi-Supervised Contrastive Learning"*

Luyi Han, Yunzhi Huang, Tao Tan, Ritse Mann

SJTU_EIEE_2-426Lab: *"Image Translation-Based Unsupervised Cross-Modality Domain Adaptation for Medical Image Segmentation"*

Tao Yang, Lisheng Wang

5:40pm – **Task 2: Evaluation design and results announcement** (10 minutes)

5:50pm - **Conclusion**

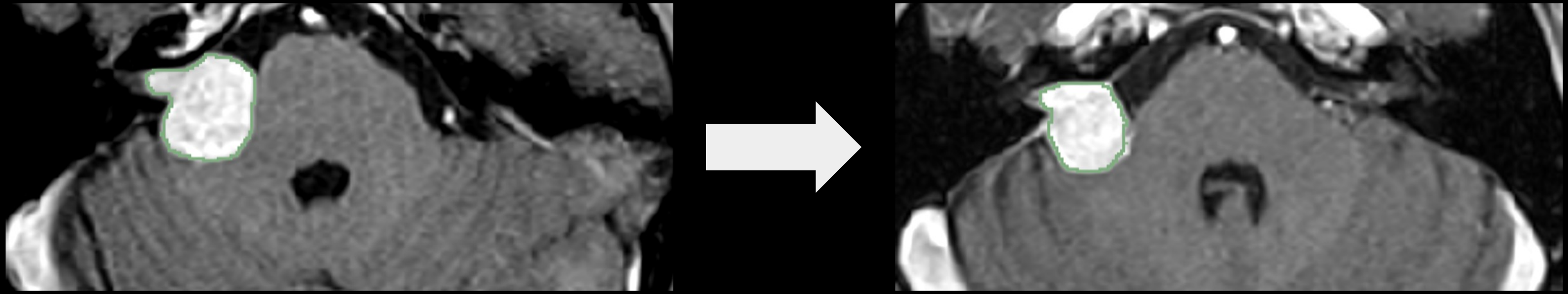
Supervised learning

Underlying assumption of **supervised training** on data distributions:

Source (Training) = Target (Test)

Source

Target



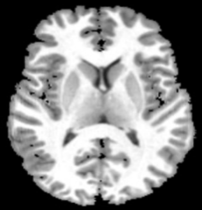
Domain shift in medical applications

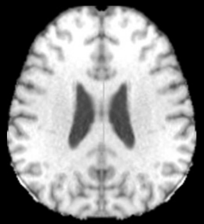
In practice:

Source (Training) ≠ Target (Test)

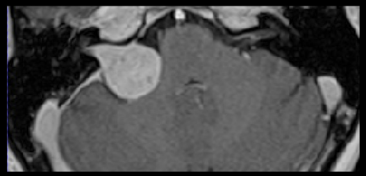
Source | Target

1. **Different acquisition protocols:**
- Scanner characteristic (manufacturer, strength)
- Sequence parameters
- Type of acquisition (axial, coronal, sagittal, isotropic - slice thickness)





2. **Different imaging modalities:**
CT vs MR
Contrast-enhanced T1 vs T2

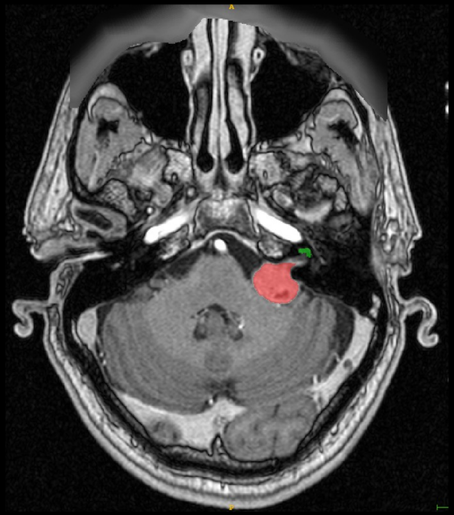


CNNs have been shown to have poor generalization capability

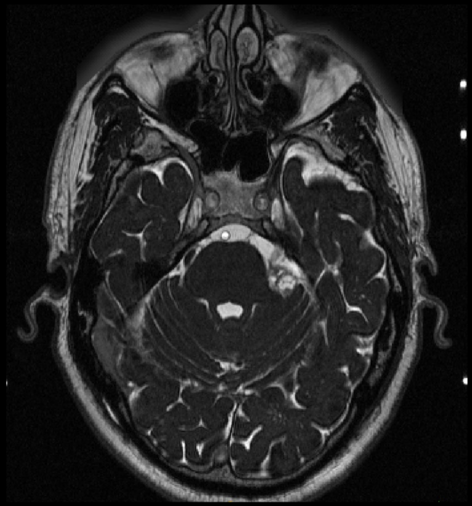
Unsupervised Domain Adaptation (UDA)

Goal: Bridging the domain distribution discrepancy between the source domain and the target domain without any target **labelled** data.

Source



Target



Various UDA approaches...

Transforming the source data in target-like data:

→ data augmentation
→ generative models (e.g., CycleGAN) [4,6]

Minimizing the discrepancy between the feature distributions:

→ distribution discrepancy loss
→ discriminative adversarial loss [1,2,3,4,6]

Self-training:

→ self-supervision via pretext tasks [5]

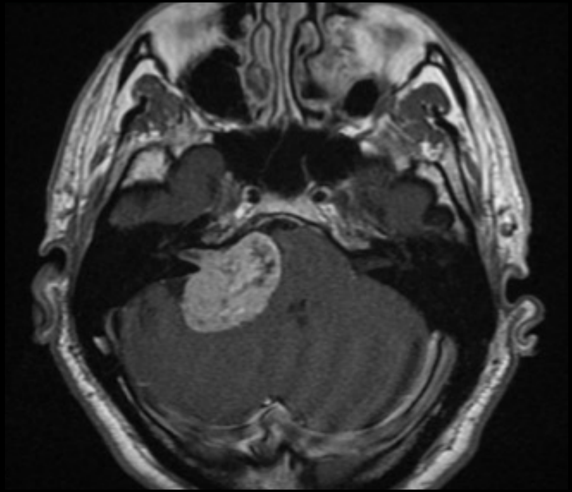
Large range of techniques can be used

… tested on different problems

	Public	Large testing set (>20)	Multi-Class Problem	Cross-modality
Traumatic brain injuries [1]		✓		
Liver Segmentation [2]		✓		✓
White Matter Lesions [5]	✓			
Cardiac structure segmentation [3,4,6]	✓		✓	✓

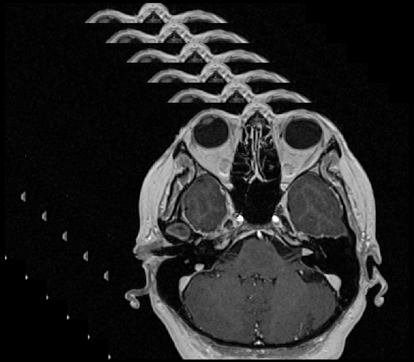
Need for a benchmark on a large, publicly available, multi-class dataset

Vestibular Schwannoma

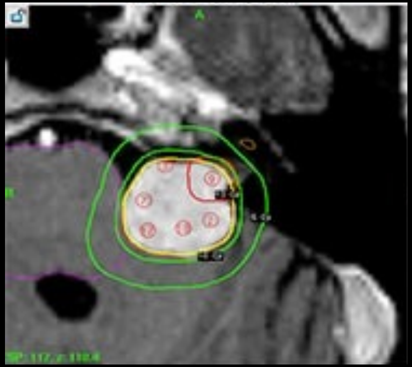


- A **benign** (non-cancerous) slow growing tumour.
- Arises from one of the **balance nerves.**
- Tumours may be found by accident or because patients present **symptoms** (e.g. hearing loss, balance disturbance).
- **1 in 1,000 people** will be diagnosed with a VS in their lifetime.

Current management



Surveillance



Stereotactic Radiosurgery

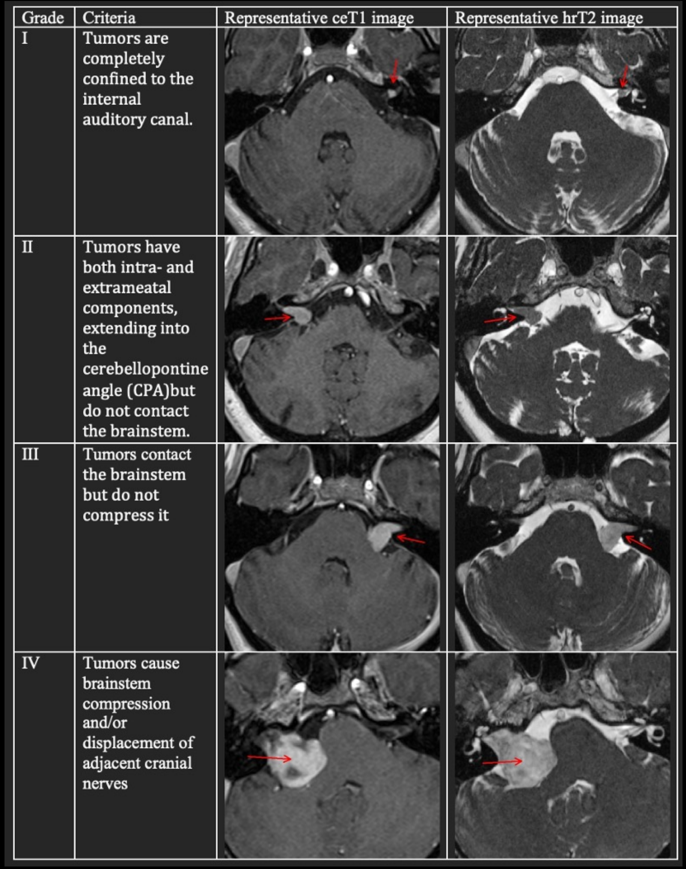
Surgery

Choice based on:
- Tumour's growth
- Symptomatic vs asymptomatic
- Koos grade: quantify the impact of the tumour on surrounding brain structures (e.g. brainstem)

Koos grading system

Classification system for VS that captures many of the characteristics that treatment decisions are typically based on.

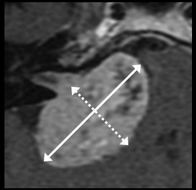
Used in clinical routine for decision-making.



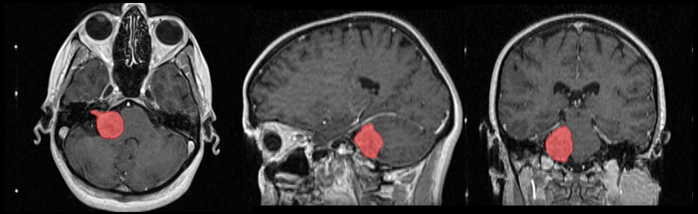
Grade	Criteria	Representative ceT1 image	Representative hrT2 image
I	Tumors are completely confined to the internal auditory canal.		
II	Tumors have both intra- and extrameatal components, extending into the cerebellopontine angle (CPA)but do not contact the brainstem.		
III	Tumors contact the brainstem but do not compress it		
IV	Tumors cause brainstem compression and/or displacement of adjacent cranial nerves		

Need for automated segmentation tools

Measuring tumour's growth:
- **Linear measurement** (maximal diameter)

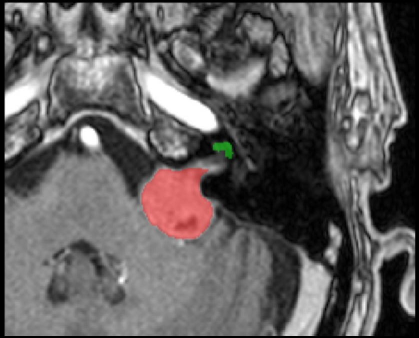


- **Volumetric assessment**



→ more accurate and sensitive method
→ superior at detecting subtle growth

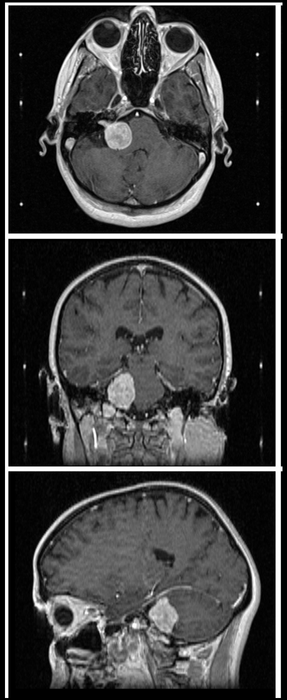
Stereotactic Radiosurgery:
requires accurate, individualised contouring of:
- clinical target volume (VS **tumour**)
- "organs" at risk (**cochleas**)



Challenge task: automatic segmentation **tumour** and **cochleas**

Imaging protocol

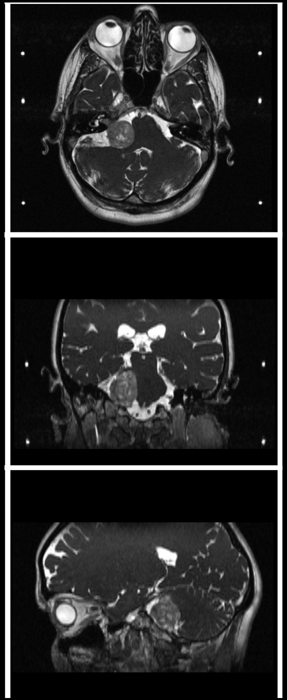
Contrast-Enhanced T1 (ceT1)



Gold standard for VS

Risks associated with gadolinium-containing contrast agents

High-Resolution T2 (hrT2)



Gold standard for cochleas

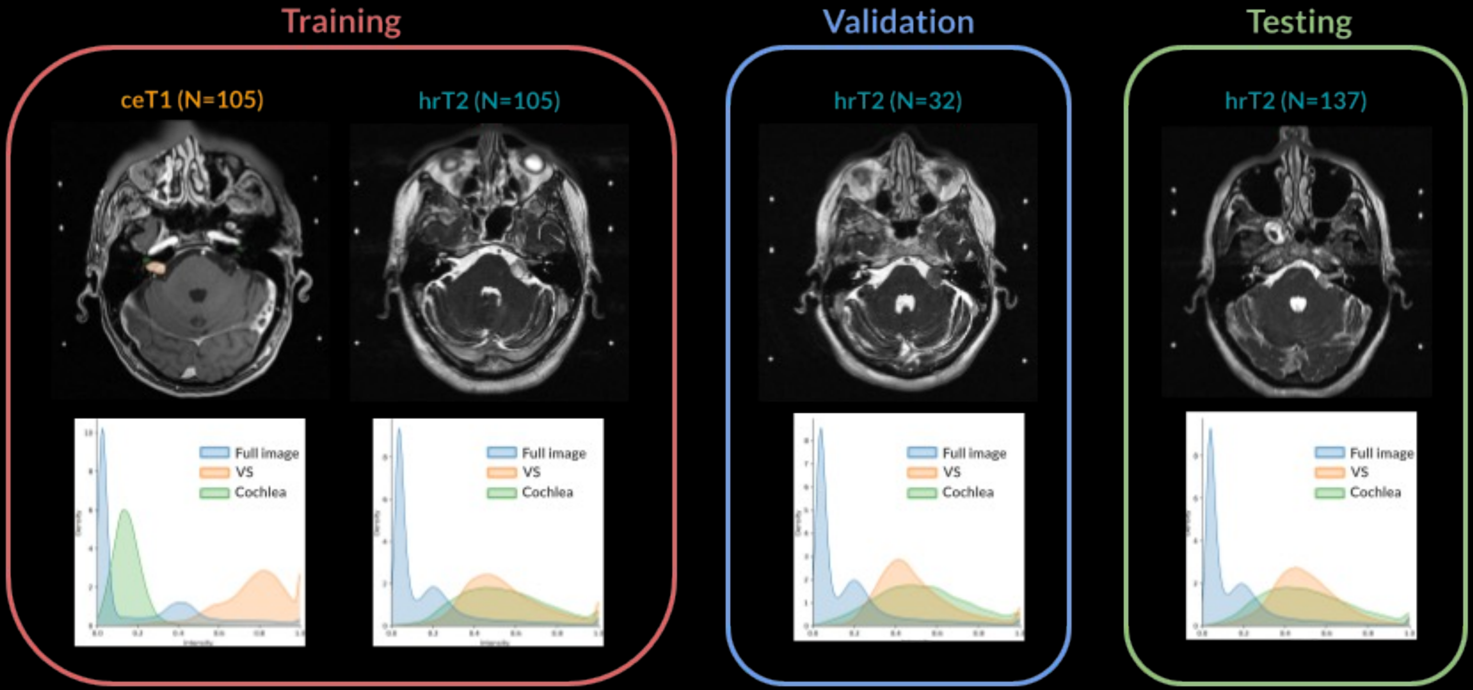
Growing interest in using non-contrast imaging sequences for VS

10 times more cost-efficient than ceT1 imaging

crossMoDA 2021: Challenge task and dataset

Dataset:

- All images were obtained on a 32-channel Siemens Avanto 1.5T scanner
- Image resolution: 0.5×0.5×1.0mm or 0.5×0.5×1.5mm
- Consecutive patients

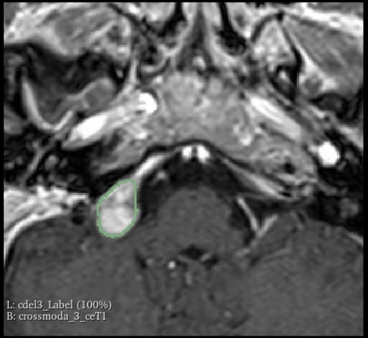


A challenging task

Vestibular Schwannoma

- Uniform on ceT1
- Borders may not be clear on hrT2

ceT1

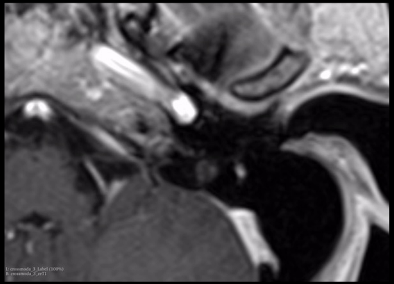


hrT2

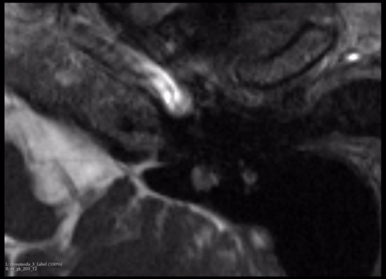
Cochlea

- Two sides
- Very small structure (92 ±14 mm^3 - 0.002% voxels)
- Unclear borders on ceT1

ceT1

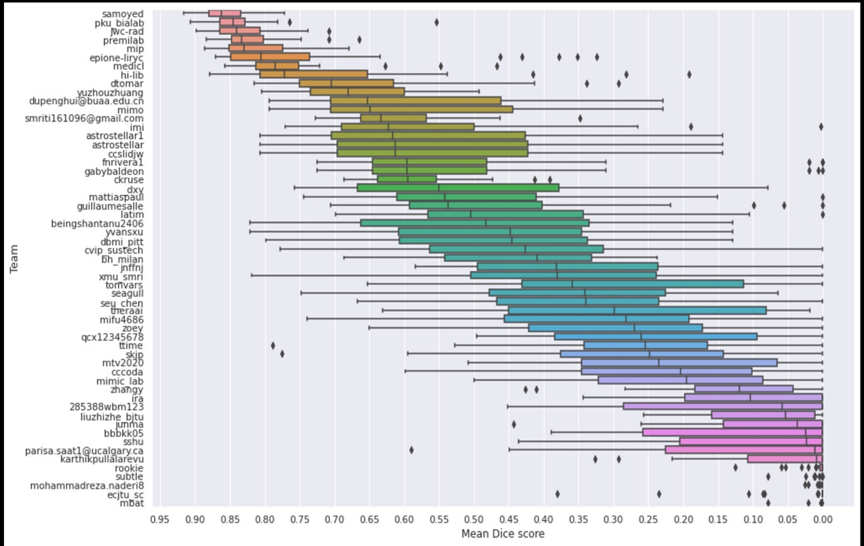


hrT2



crossMoDA 2021: Main insights

- Large **variability of techniques**
- Cross-modality domain adaptation is a **challenging task**. On the validation leaderboard:
  - 47 teams (85%) underperformed (<60% mean Dice Score).
  - Only 5 teams (10%) reached a high performance (>80% mean Dice Score).
- The top performing teams used a similar unsupervised approach (**CycleGAN + nnUnet + self-supervision**).



More details in our *Medical Image Analysis* paper.

crossMoDA 2021: Main limitations

Domain gap between the source and target images is large, as it corresponds to different modalities
The intra-domain data was **homogeneous**:

→ Lack of robustness may occur when the same modalities are acquired with **different settings**

Tilburg study (Cornelissen et al):
- Fully supervised model trained on London data
- London (testing data): mean dice score of **92.0±5.1%**
- **Tilburg** (testing data): mean dice score of **64.5±32.%**

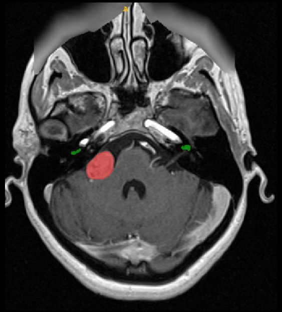
crossMoDA 2022: multi-institutional dataset

		London	Tilburg
Scanner		Siemens Avanto 1.5T	Philips Ingenia 1.5T
ceT1	Sequence	MPRAGE	3D-FFE
	In-plane res	0.4x0.4mm	0.8x0.8mm
	Slice thickness	1.0 to 1.5mm	1.5mm
	In-plane matrix	512x512	256x256
hrT2	Sequence	3D CISS or FIESTA	3D-TSE
	In-plane res	0.4x0.4mm	0.5x0.5mm
	Slice thickness	1.0 to 1.5 mm	1.0 mm
	In-plane matrix	384x384 or 448x448	512×512

crossMoDA 2022: doubling dataset size

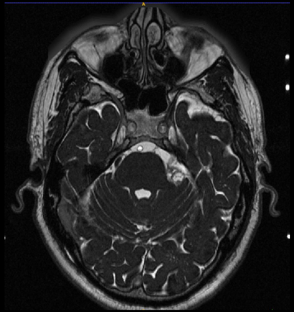
Training

Contrast-enhanced T1



105+105=210

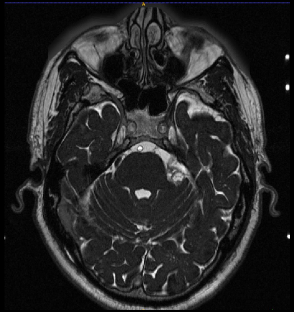
T2



105+105=210

Validation

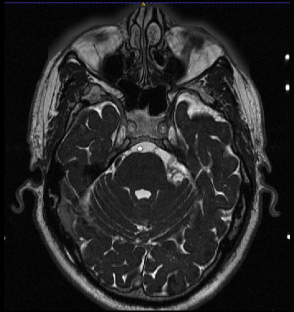
T2



32+32=64

Testing

T2



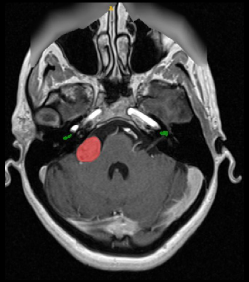
137+134

crossMoDA 2022: new task

Classification task: Koos grade classification

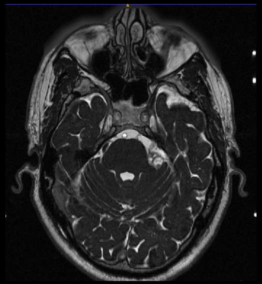
Training

Contrast-enhanced T1



105+105=210

T2

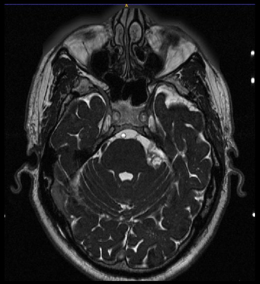


105+105=210

+Koos grade
+GIF parcellation

Validation

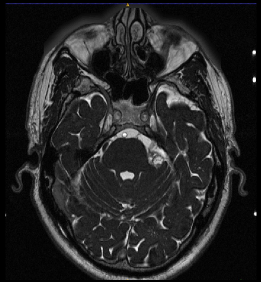
T2



32+32=64

Testing

T2



137+134

Oral presentations
Task 1: segmentation

Sun, September 18
Reuben Dorent

Multi-view Cross-Modality MR Image Translation for Vestibular Schwannoma and Cochlea Segmentation

Bogyeong Kang

Department of Artificial Intelligence, Korea University

kangbk@korea.ac.kr

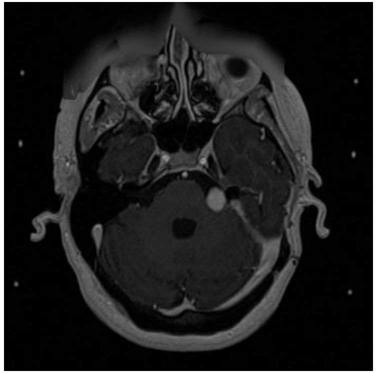
MICCAI2022 Singapore

Goal

- Segmentation of VS and Cochlea in hrT_2 scans

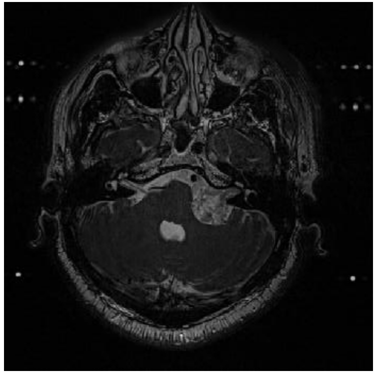
Training

contrast-enhanced T_1 (ceT_1)



Annotation label O

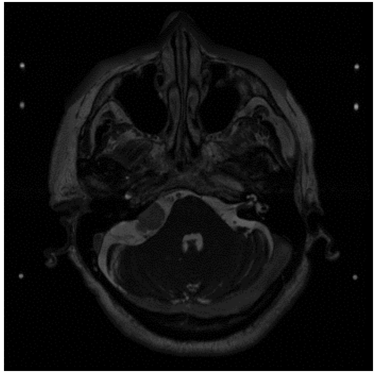
high-resolution T_2 (hrT_2)



Annotation label X

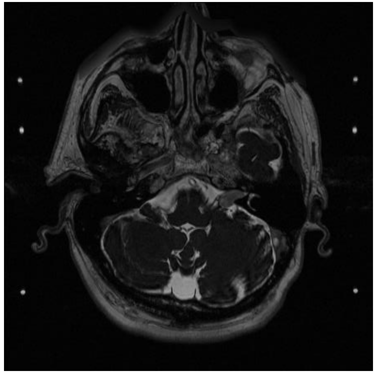
Validation

high-resolution T_2 (hrT_2)



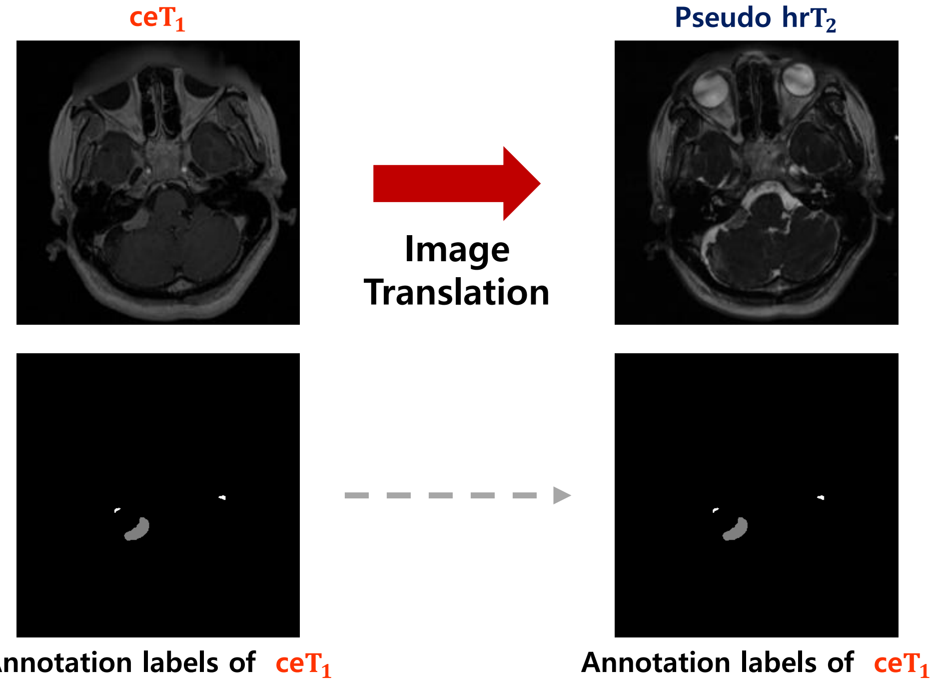
Testing

high-resolution T_2 (hrT_2)

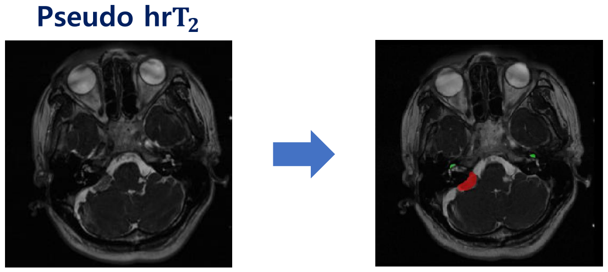


Approach

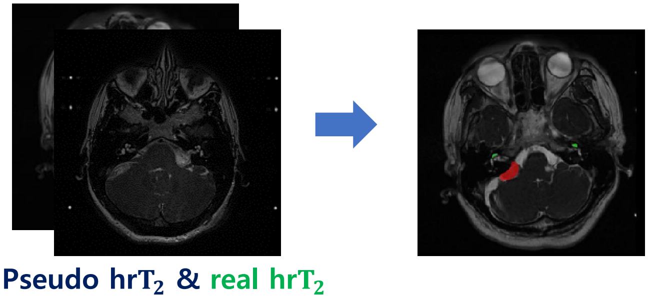
Step1: Image translation



Step2: Segmentation

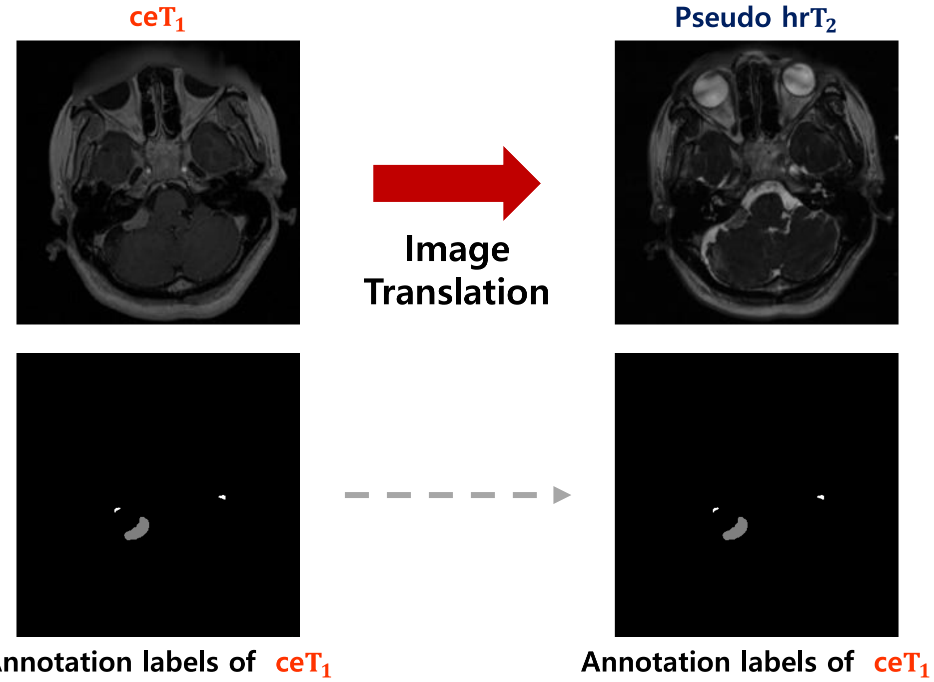


Step3: Self-training



Motivation

Step1: Image translation



- Preserve the structures in the ceT1
- Reflect the characteristics of the hrT2

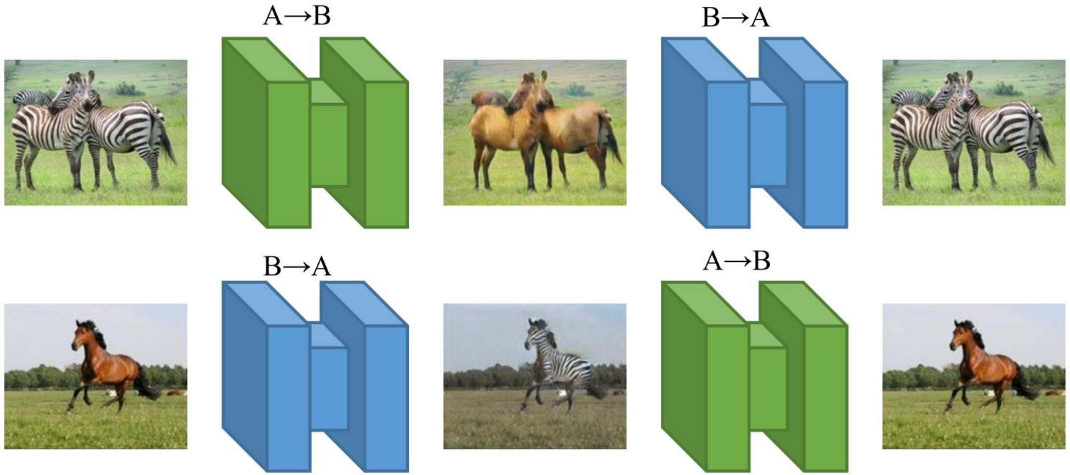
Image translation performance ↑

Segmentation performance ↑

MAI Lab EDICAL

Image translation

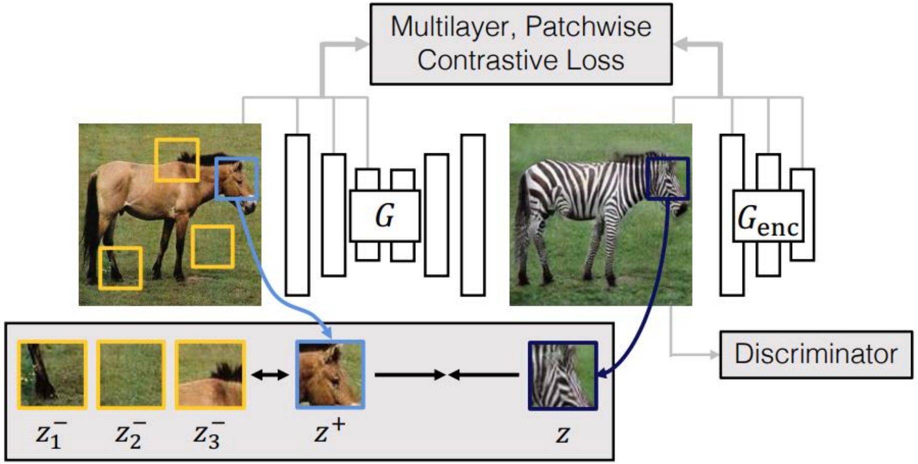
- **CycleGAN** (Zhu et al., 2017)
  - use pixel-level cycle-consistent constraint
  - use cycle-consistency loss: pixel-level reconstruction loss
  - learn the mapping from the output domain to the input domain



$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1]$$

Image translation

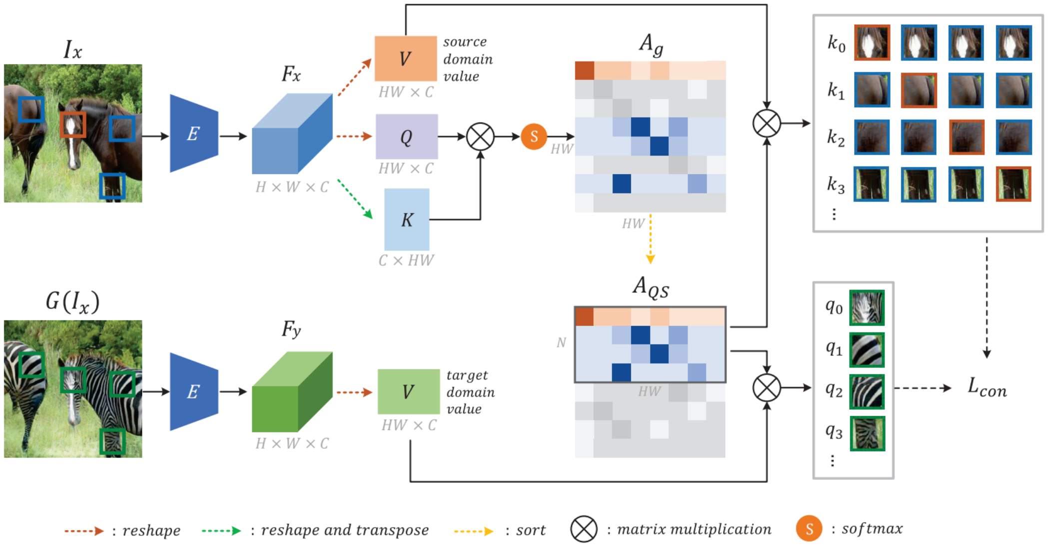
- **CUT** (Park et al., 2020)
  - use **patch-level contrastive** constraint
  - constrain the features from the same location to be close
  - calculate contrastive loss between **randomly** selected patches
  - contain some patches less information of the source domain



Patchwise Contrastive Learning

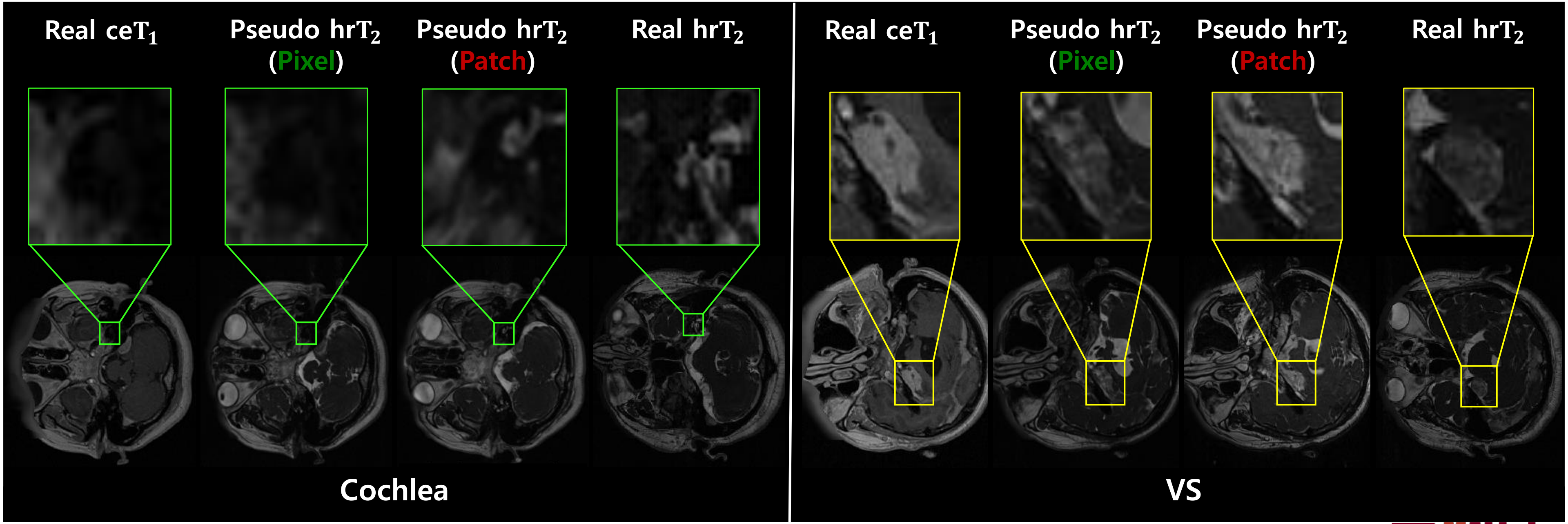
Image translation

- **QS-Attn** (Hu et al., 2022)
  - use patch-level contrastive constraint
  - select the domain-relevant patches
  - better preserve the structures of VS & cochlea

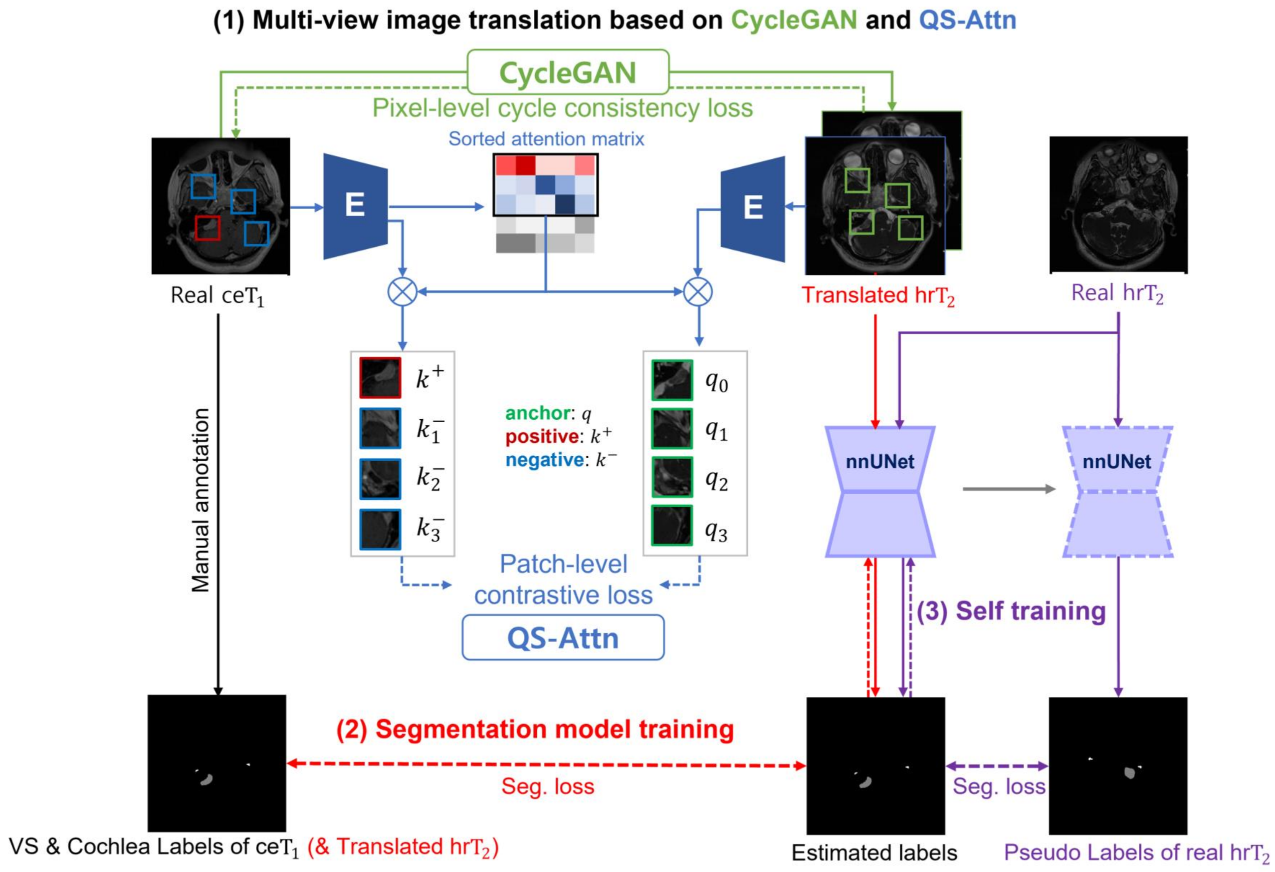


Importance of multi-view image translation

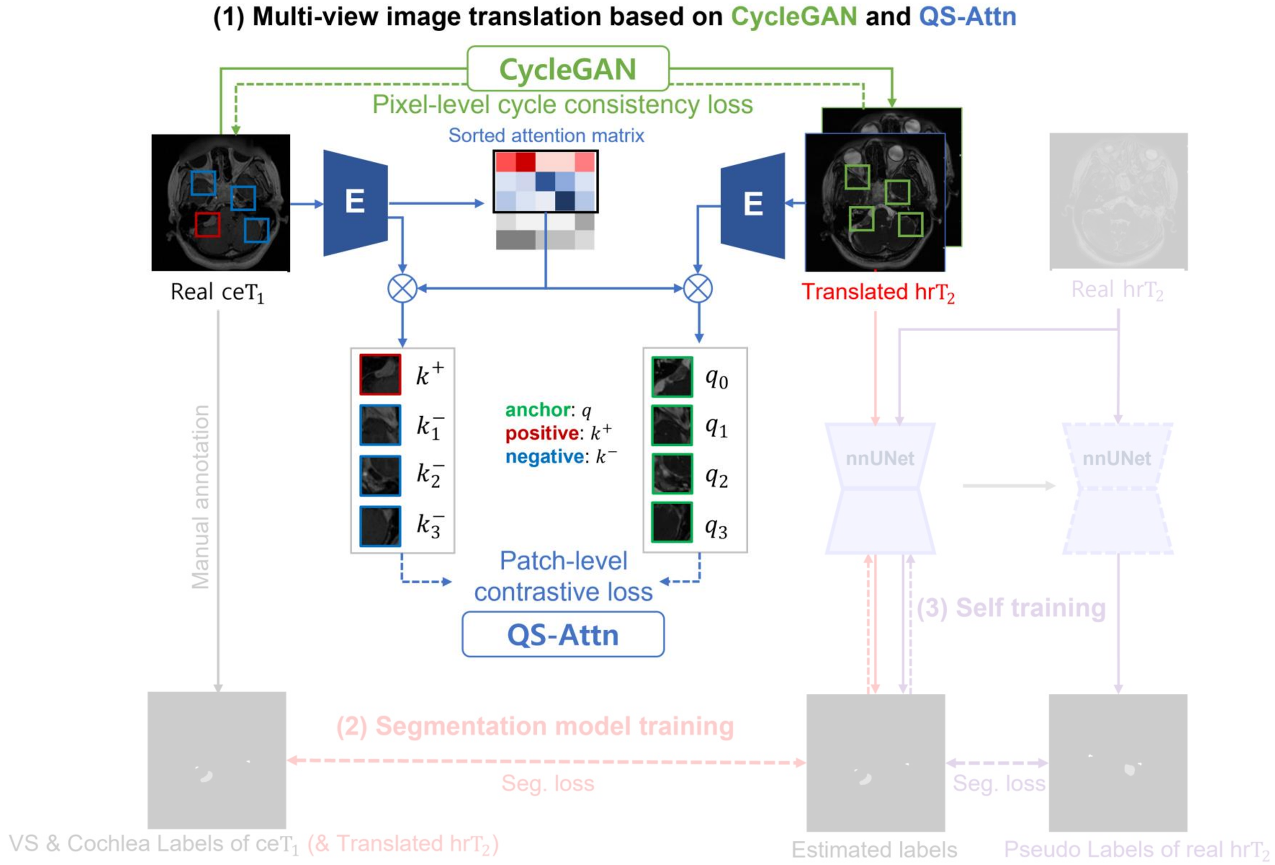
- **Pixel-level** cycle consistent constraint: better reflect **intensity**
- **Patch-level** contrastive constraint: better preserve **structures**



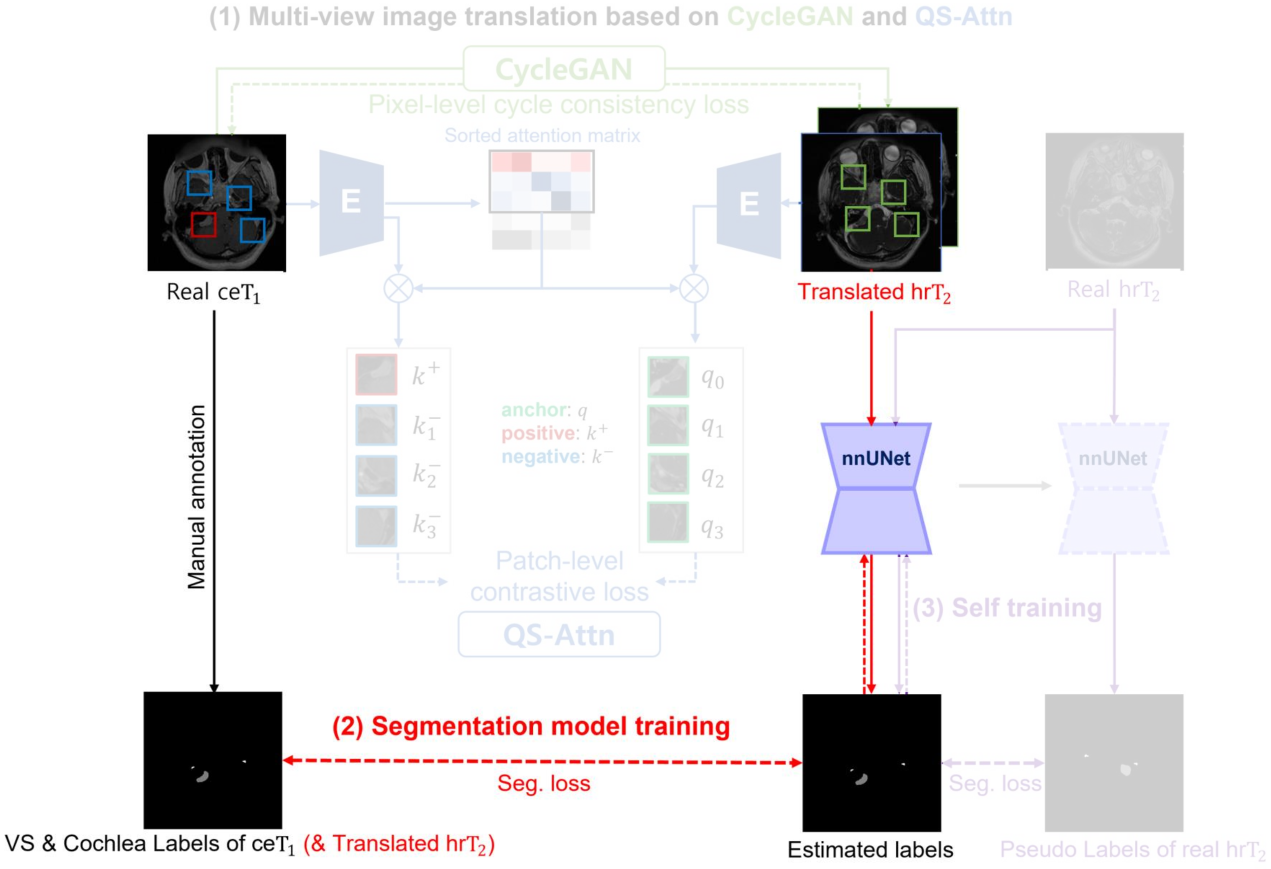
Proposed framework



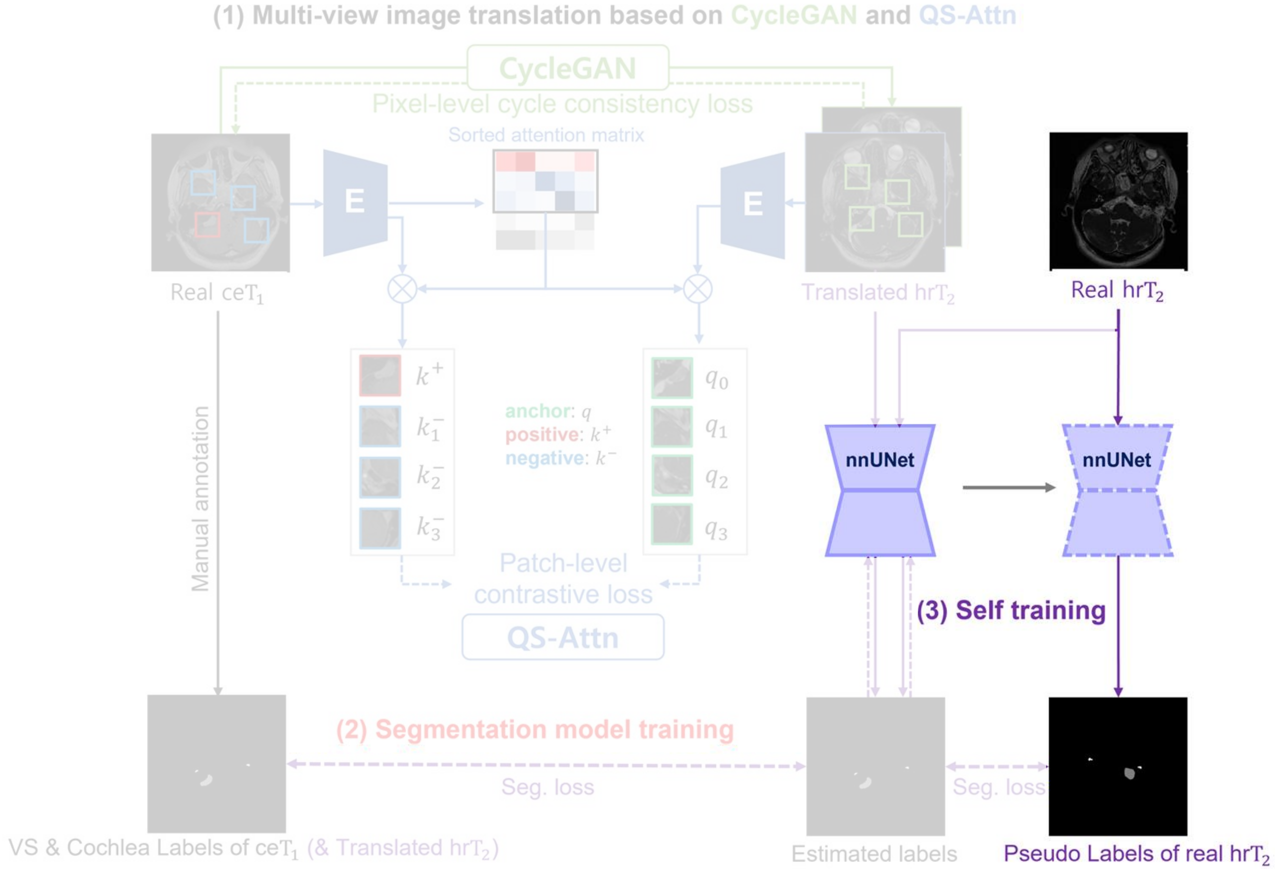
Proposed framework



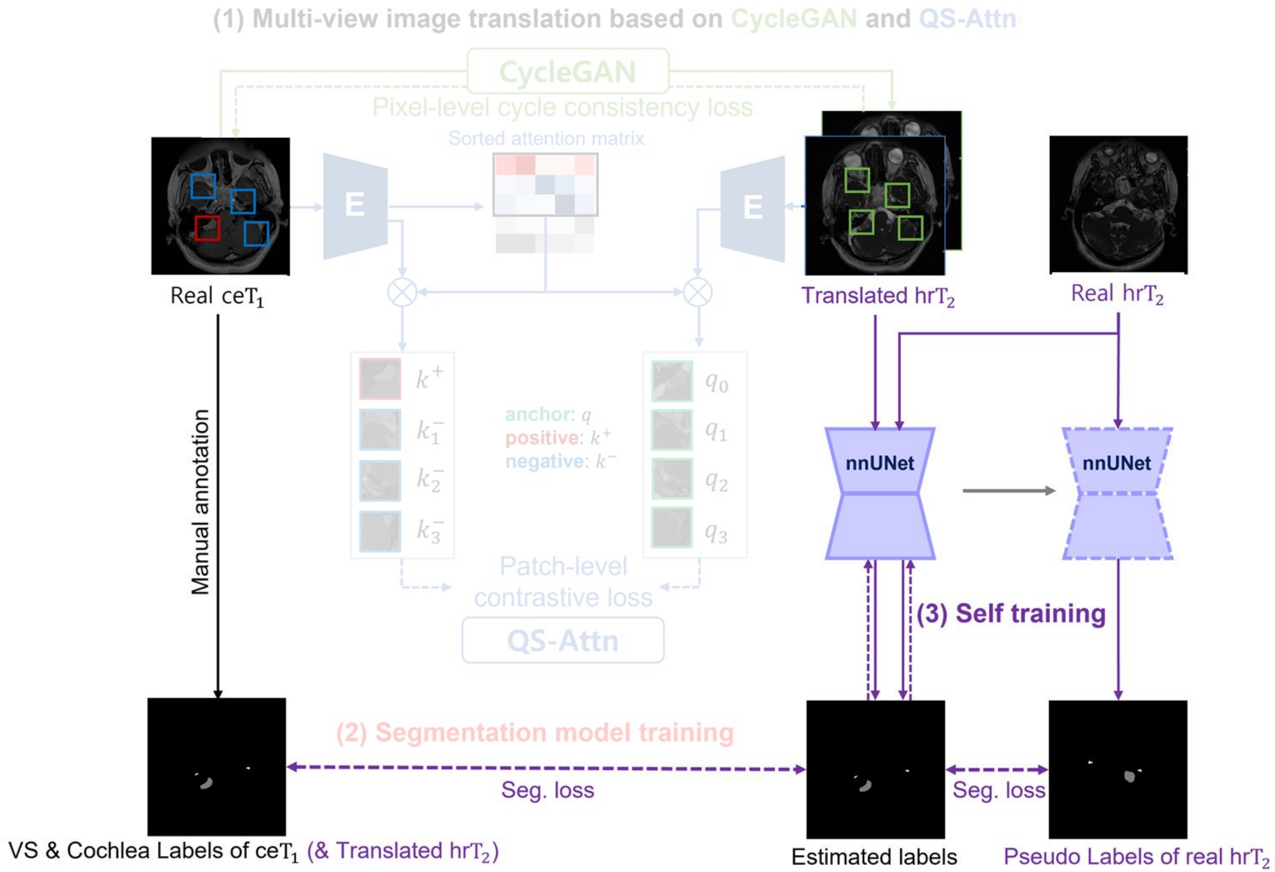
Proposed framework



Proposed framework



Proposed framework



Results: validation phase

Translation model	Dice score (↑)			ASSD (↓)	
	VS	Cochlea	Mean	VS	Cochlea
CycleGAN	0.7798 (±0.1901)	0.8066 (±0.0323)	0.7932 (±0.0972)	0.8750 (±0.9222)	0.2422 (±0.1608)
QS-Attn	0.7779 (±0.1825)	0.8158 (±0.0287)	0.7968 (±0.0929)	0.6667 (±0.3891)	**0.2365 (±0.1573)**
Proposed	**0.8043 (±0.1656)**	**0.8158 (±0.0289)**	**0.8101 (±0.0863)**	**0.5742 (±0.2461)**	0.2387 (±0.1581)

Results: validation phase

Translation model	Dice score (↑)			ASSD (↓)	
	VS	Cochlea	Mean	VS	Cochlea
CycleGAN (*w/o.* ST)	0.7798 (±0.1901)	0.8066 (±0.0323)	0.7932 (±0.0972)	0.8750 (±0.9222)	0.2422 (±0.1608)
QS-Attn (*w/o.* ST)	0.7779 (±0.1825)	0.8158 (±0.0287)	0.7968 (±0.0929)	0.6667 (±0.3891)	0.2365 (±0.1573)
Proposed (*w/o.* ST)	0.8043 (±0.1656)	0.8158 (±0.0289)	0.8101 (±0.0863)	0.5742 (±0.2461)	0.2387 (±0.1581)
Proposed (*w.* ST)	**0.8520 (±0.0889)**	**0.8488 (±0.0235)**	**0.8504 (±0.0466)**	**0.4748 (±0.2072)**	**0.1992 (±0.1524)**

* ST: self-training

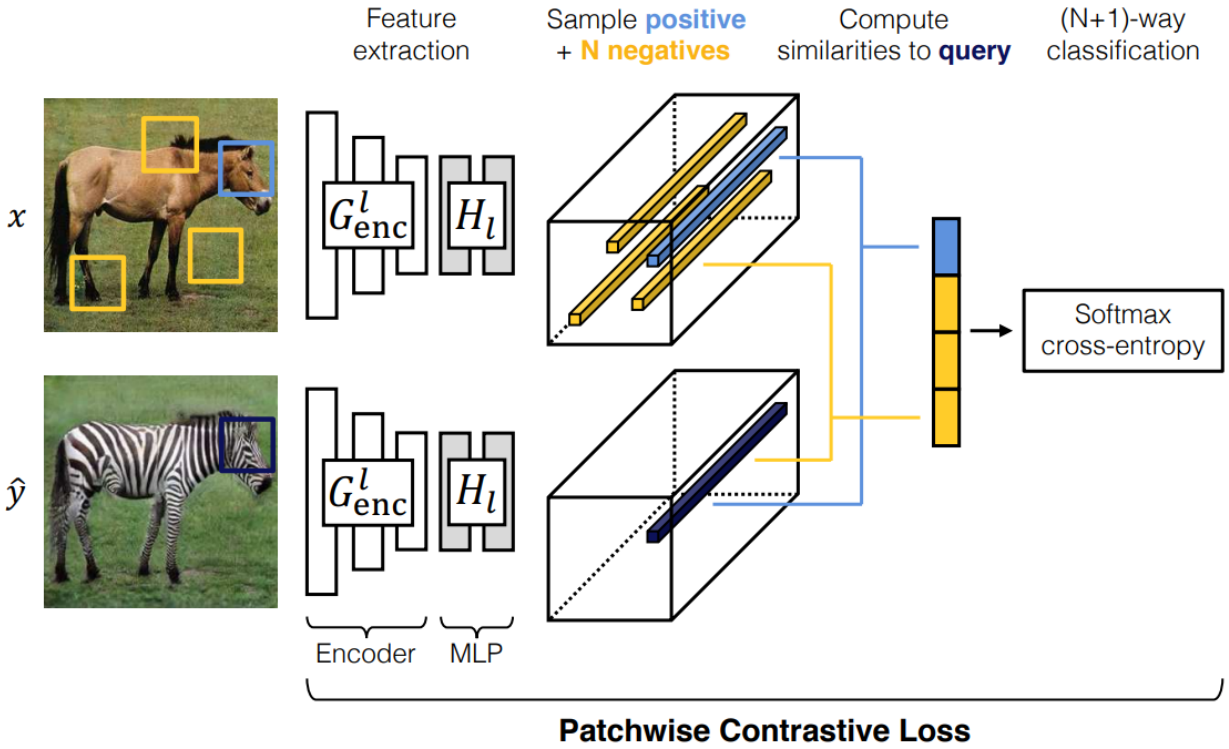
Conclusion

- Design a **multi-view image translation framework**
- Adopt **CycleGAN & QS-Attn in parallel** for image translation
- **Reflect various perspectives** (i.e., intensity & texture, structure)

Thank you

Appendix

- CUT



Appendix

- Preprocessing

1. Resample to $0.41 \times 0.41 \times 1.5$
2. Slice 2D images along the axial plane
3. Center crop & resize to 256×256

Appendix

MICCAI2022 Singapore

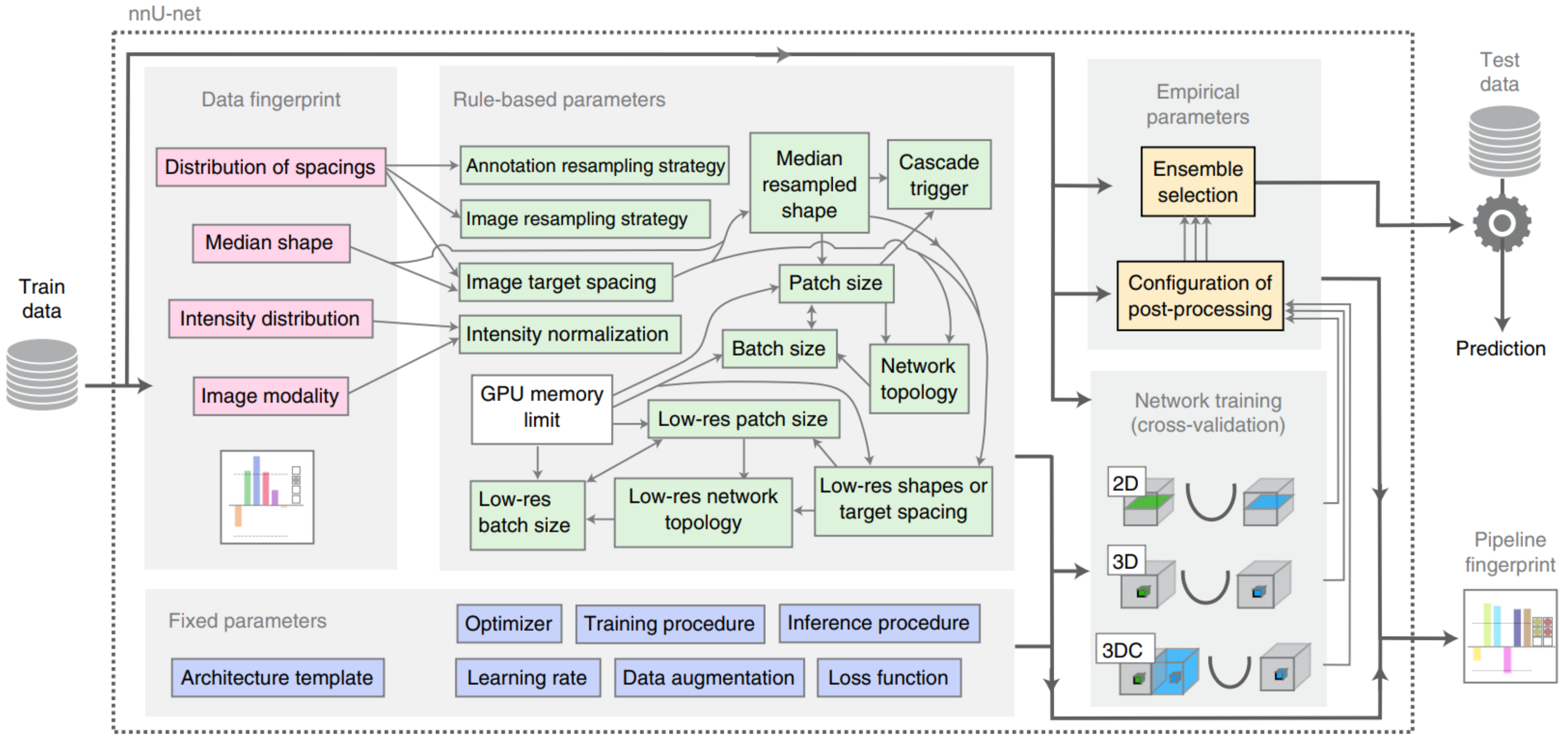
▪ Only CycleGAN *w.* Self-training

* ST: self-training

Translation model	Dice score (↑)			ASSD (↓)	
	VS	Cochlea	Mean	VS	Cochlea
CycleGAN (*w/o.* ST)	0.7798 (±0.1901)	0.8066 (±0.0323)	0.7932 (±0.0972)	0.8750 (±0.9222)	0.2422 (±0.1608)
QS-Attn (*w/o.* ST)	0.7779 (±0.1825)	0.8158 (±0.0287)	0.7968 (±0.0929)	0.6667 (±0.3891)	0.2365 (±0.1573)
Proposed (*w/o.* ST)	0.8043 (±0.1656)	0.8158 (±0.0289)	0.8101 (±0.0863)	0.5742 (±0.2461)	0.2387 (±0.1581)
CycleGAN (*w.* ST)	0.8234 (±0.1098)	0.8154 (±0.0278)	0.8194 (±0.0582)	0.8052 (±0.8004)	0.2318 (±0.1578)
Proposed (*w.* ST)	**0.8323 (±0.1017)**	**0.8265 (±0.0283)**	**0.8294 (±0.0546)**	**0.5273 (±0.2028)**	**0.2259 (±0.1570)**

Appendix

- **nnUNet** (Isensee et al., 2021)



Oral presentations
Task 1: segmentation

crossMoDA

MICCAI2022 Singapore
25th International Conference on Medical Image Computing and Computer Assisted Intervention
September 18-22, 2022
Resorts World Convention Centre Singapore

Sun, September 18
Reuben Dorent

Unsupervised Domain Adaptation in Semantic Segmentation Based on Pixel Alignment and Self-Training (PAST)

Hexin Dong[1] Fei Yu[1] Mingze Yuan[1] Jie Zhao[1,2] Bin Dong[4,3,2] Li Zhang[1,2]✉

[1]Center for Data Science, Peking University, Beijing, China

[2]National Biomedical Imaging Center, Peking University, Beijing, China

[3]Center for Machine Learning Research, Peking University, Beijing, China

[4]Beijing International Center for Mathematical Research (BICMR), Peking University, Beijing, China

✉ Corresponding author: Li Zhang (zhangli_pku@pku.edu.cn)

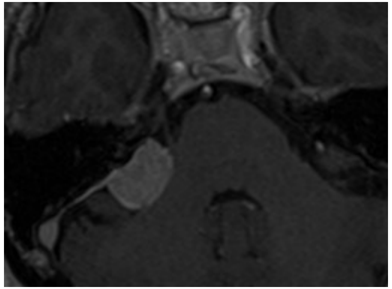
Introduction

- **Problem Setting：**
  - 3D Semantic Segmentation
  - Domain Adapataion
  - Few shots learning (210 source images & 210 target images)

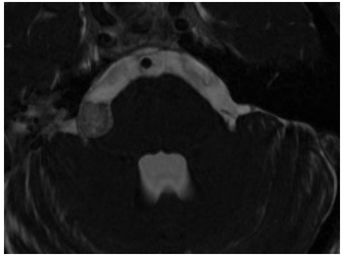
- **Domain Adaptation：**
  - Pixel alignment method
  - Feature alignment method
  - Self training method

- Preprocess :
- ☐ Center crop
- ☐ Normalization

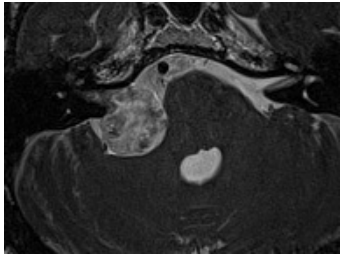
Source Domain:



Target Domain:



London data



Tilburg data

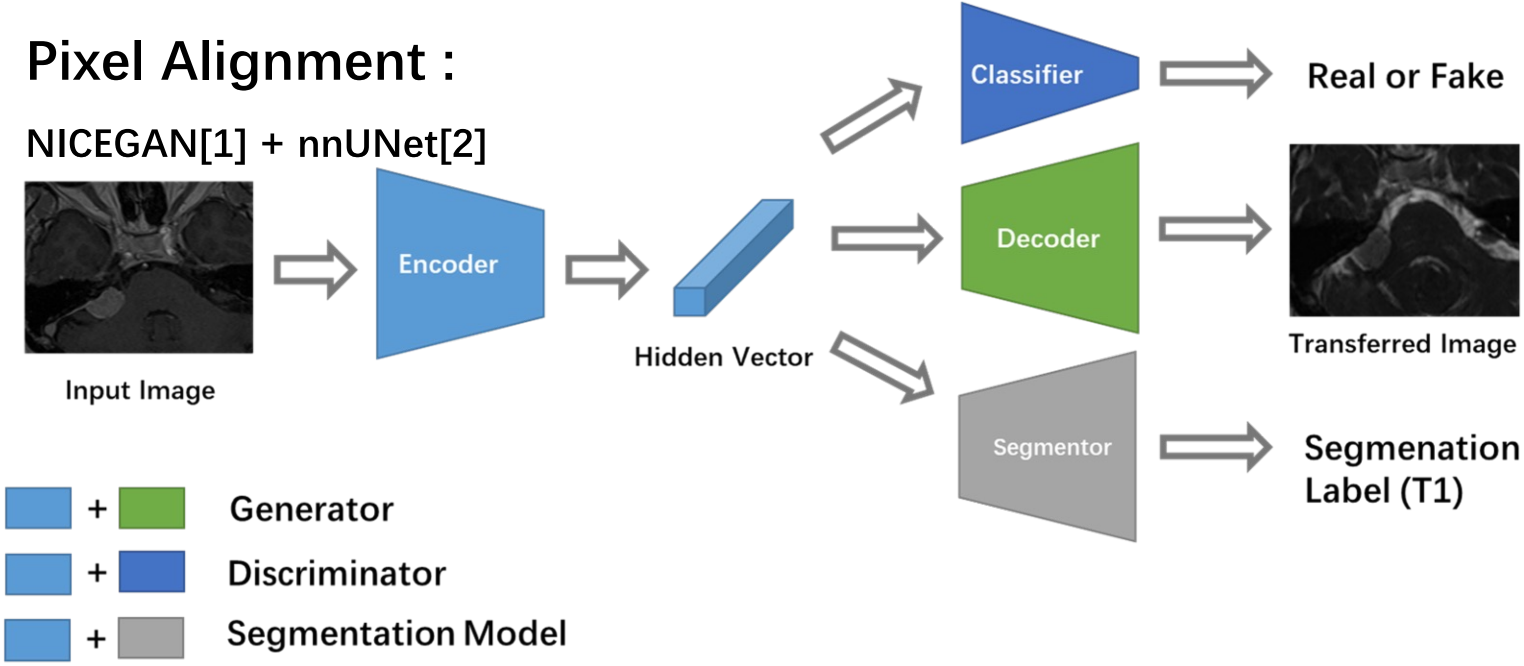
PAST1.0[3] in CrossModa2021:

- We proposes an unsupervised cross-modality domain adaptation approach based on pixel alignment and self-training（PAST）.
- Pixel alignment stage aims to transfer ceT1 scans to hrT2 scans.
- Self training stage aims to finetune the model with generated hrT2 labels and ceT1 labels.
- PAST performs well on VS while have some problems on cochlea.

[3]. Hexin Dong, Fei Yu, Jie Zhao, Bin Dong and Li Zhang, Unsupervised Domain Adaptation in Semantic Segmentation Based on Pixel Alignment and Self-Training, arXiv:2109.14219, 2021

- **Pixel Alignment :**

- NICEGAN[1] + nnUNet[2]



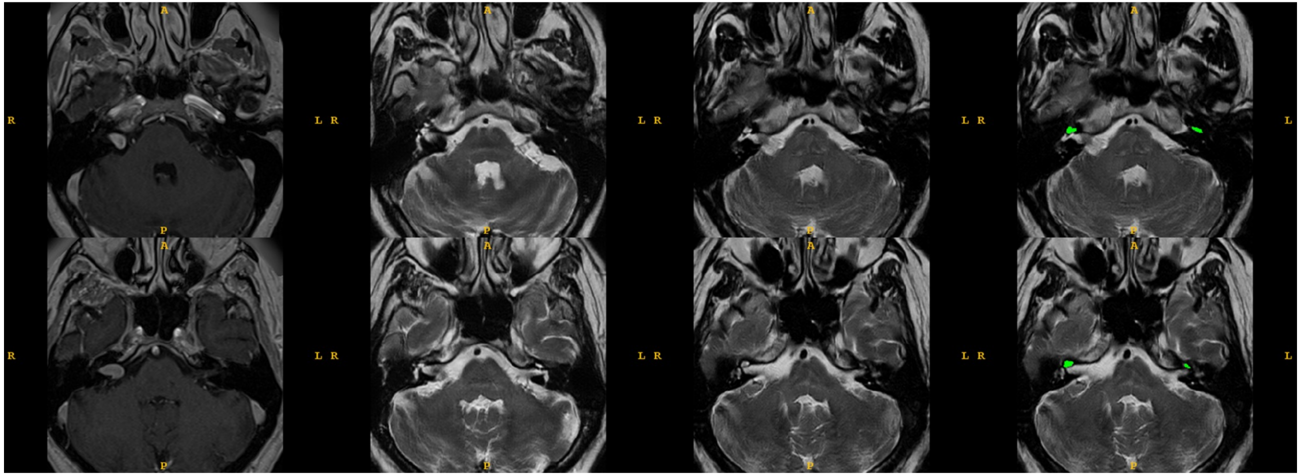
[1].Chen et.al. Reusing discriminators for encoding: Towards unsupervised image-to-image translation. CVPR 2020
[2].Isensee et.al. Automated design of deep learning methods for biomedical image segmentation. arXiv preprint arXiv:1904.08128

Pixel Alignment :

- Train two model with different architecture .
- Named as ResUnetPA and nnUnetPA.

Results:

Model Name	VS Dice	Cochlea Dice	Mean Dice
nnUnetPA	0.6716	0.8280	0.7498
ResUnetPA	0.6729	0.8246	0.7487



From left to right: (1) ceT1 scans. (2) synthesized hrT2 scans without segmentor. (3) synthesized hrT2 scans with segmentor. (4) cochlea ground truth.

Self training:

- Set nnUnetPA/ResUnetPA as $S_0, K = 2, q_k = 0.6$
- Train two model based on nnUnet and ResUnet named as nnUnetPAST2 and ResUnetPAST2.

Algorithm 1 training process of the proposed method

Initialize ceT1 scans images and label (X_s, y_s), hrT2 scans images X_t, Segmentation network S, Image translation network T
Train network T with X_s and X_t
Transfer ceT1 scans X_s to $\hat{X}_s$ using T
Train network S with $(\hat{X}_s, y_s)$
Initialize concat scans images $X_c = \{\hat{X}_s, X_t\}$, self-training segmentation network $S_0 = S$
for $k \leftarrow 1$ to K **do**
    input X_c into S_{k-1} and generate pseudo label $\hat{y_c^k}$ with a fixed portion q_k
    Initialize $S_k \leftarrow S_{k-1}$
    Train S_k with $(X_c, \hat{y_c^k})$
end for
return S_k

Results:.

Model Name	London data VS Dice	Tilburg data VS Dice	Mean Dice
nnUNetPAST2	0.8231	0.7959	0.8095
ResUNetPAST2	0.8281	0.7949	0.8115
PAST1.0	**0.8705**	0.7170	0.7935
IResUNetPAST2	0.8519	**0.8243**	0.8381
PAST2.0	**0.8705**	**0.8243**	**0.8474**

Model Name	VS Dice	Cochlea Dice	Mean Dice
nnUnetPA	0.6716	0.8280	0.7498
ResUnetPA	0.6729	0.8246	0.7487
nnUNetPAST2	0.8095	0.8547	0.8320
ResUNetPAST2	0.8115	0.8515	0.8315
PAST1.0	0.7935	0.7677	0.7806
IResUNetPAST2	0.8381	0.8412	0.8386
PAST2.0	**0.8474**	**0.8547**	**0.8511**

- nnUnetPAST2/ResUnetPAST2 fails on Tilburg scans.
- Set PAST1.0 as $S_0, K = 2\ , q_k = 0.6$ and named it as IResUnetPAST2.

Using nnUnetPAST2 to segment cochlea, IResUnetPAST2 to segment Tilburg data VS and PAST1.0 to segment London data VS achieves a better result. We named this combined version as PAST2.0.

- We proposes an unsupervised cross-modality domain adaptation approach based on pixel alignment and self-training.
- PAST2.0 improves the cochlea results with the extra segmentor in pixel alignment stage.
- Experiment results show that PAST2.0 has outperformed the non-UDA baseline significantly.
- It received rank-2 on CrossMoDA2022 validation phase Leaderboard with a mean Dice score of 0.8511.

THE END

Thank you for your listening

For any question, Please contact donghexin@pku.edu.cn.

Oral presentations
Task 1: segmentation

crossMoDA

MICCAI2022 Singapore
25th International Conference on Medical Image Computing and Computer Assisted Intervention
September 18-22, 2022
Resorts World Convention Centre Singapore

Sun, September 18
Reuben Dorent

MICCAI 2022 Singapore

25th International Conference on Medical Image Computing and Computer Assisted Intervention

September 18–22, 2022

Resorts World Convention Centre Singapore

crossMoDA 2022

LaTIM

MICCAI

Tumor blending augmentation using one-shot generative learning for crossmodal MRI segmentation

Guillaume Sallé,

Pierre-Henri Conze, Julien Bert, Nicolas Boussion, Ulrike Schick, Dimitris Visvikis, Vincent Jaouen

Laboratoire de traitement de l'information médicale (LaTIM)
INSERM, UBO, IMT Atlantique

Vestibular schwanomma (VS) treatment planning

Current clinical routine :

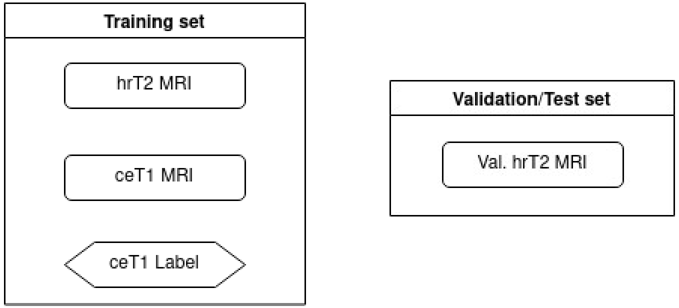
segmentation of VS : contrast-enhanced T1 MRI (**ceT1**)

segmentation of cochlea : high-resolution T2 MRI (**hrT2**)

Objective :

develop **unsupervised domain adaptation** methods to **use hrT2 only**

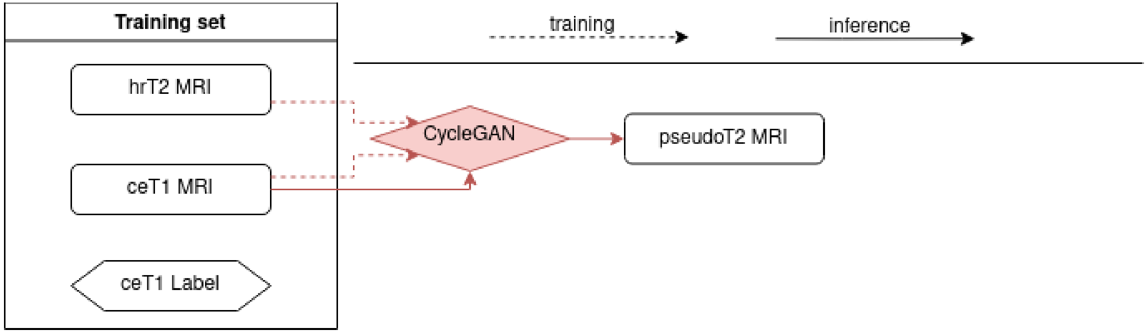
→ **cheaper** and **safer** [1]



CrossMoDA 2022 challenge (task 1) [2],[3] :

- **210** ceT1 w/ labels for **training** (105 LDN, 105 ETZ)
- **210** hrT2 w/o labels for **training** (105 LDN, 105 ETZ)
- **64** hrT2 w/o labels for **validation** (32 LDN, 32 ETZ)

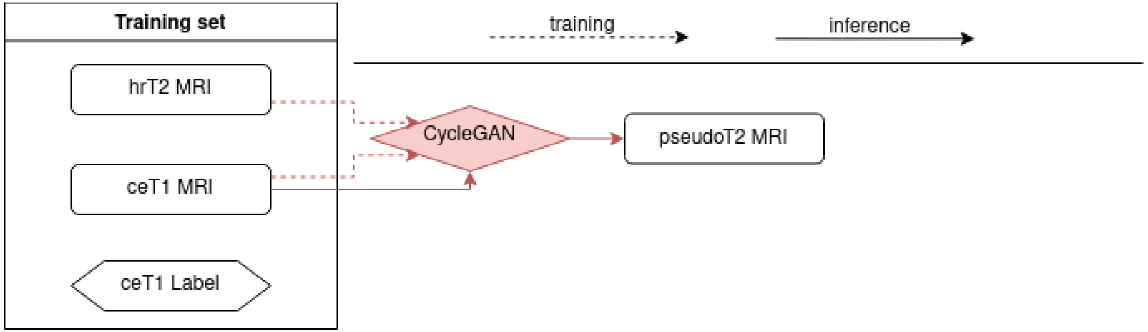
[1] Daniel H Coelho et al., “MRI surveillance of vestibular schwannomas without contrast enhancement: clinical and economic evaluation,” 2018
[2] Jonathan Shapey et al., “Segmentation of Vestibular Schwannoma from Magnetic Resonance Imaging: An Open Annotated Dataset and Baseline Algorithm,” 2021
[3] Reuben Dorent et al., “CrossMoDA 2021 challenge: Benchmark of Cross-Modality Domain Adaptation techniques for Vestibular Schwannoma and Cochlea Segmentation,” 2022



Proposed workflow

1) Image-to-image (i2i) translation using CycleGAN [4]. However :

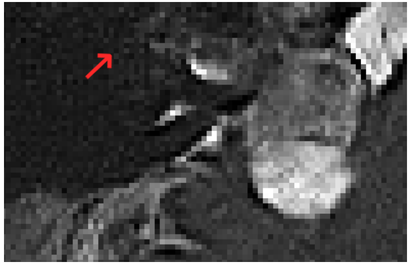
[4] Jun-Yan Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," in ICCV, 2017.



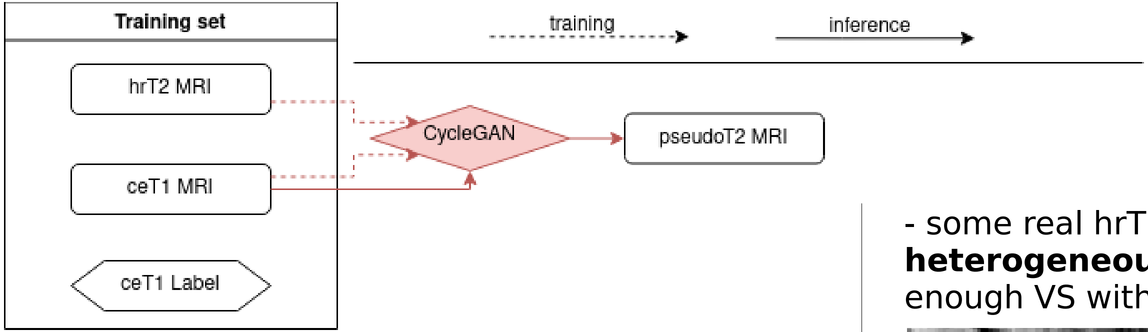
Proposed workflow

1) Image-to-image (i2i) translation using CycleGAN [4]. However :

- **small scale features** (e.g. cochlea) may **be lost** [5]



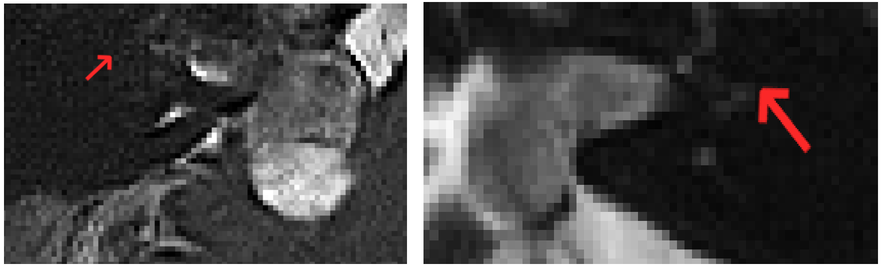
[4] Jun-Yan Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," in ICCV, 2017.
[5] Joseph P Cohen et al., "How to Cure Cancer (in images) with Unpaired Image Translation," in MIDL 2018.



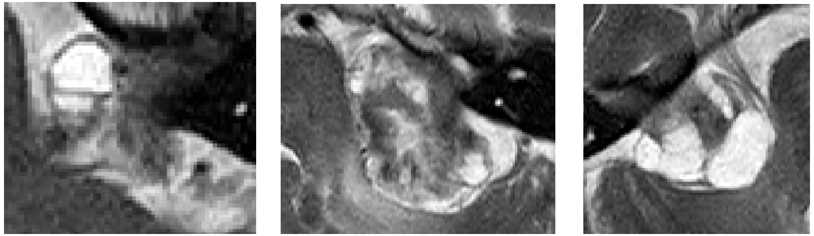
Proposed workflow

1) Image-to-image (i2i) translation using CycleGAN [4]. However :

- **small scale features** (e.g. cochlea) may **be lost** [5]

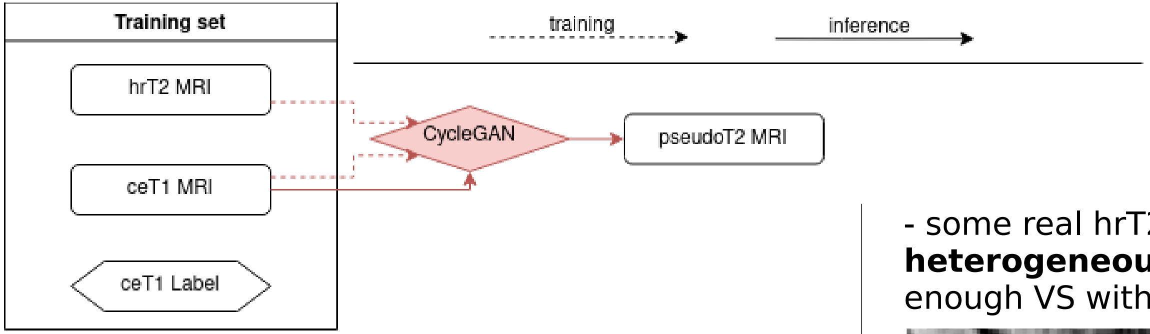


- some real hrT2 **VS** are **large**, **hypersignal** and/or **heterogeneous**. CycleGAN does not generate enough VS with these features.



Real **hrT2** from validation set

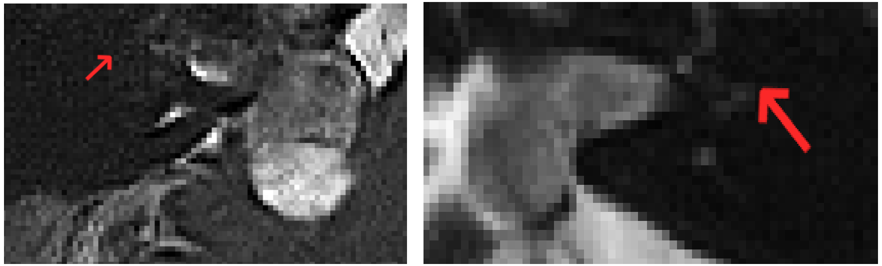
[4] Jun-Yan Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," in ICCV, 2017.
[5] Joseph P Cohen et al., "How to Cure Cancer (in images) with Unpaired Image Translation," in MIDL 2018.



Proposed workflow

1) Image-to-image (i2i) translation using CycleGAN [4]. However :

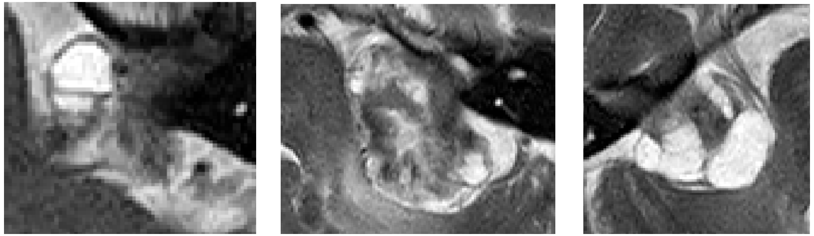
- **small scale features** (e.g. cochlea) may **be lost** [5]



→ **Objective 1 : feature preservation**

preserve cochlea before segmentation

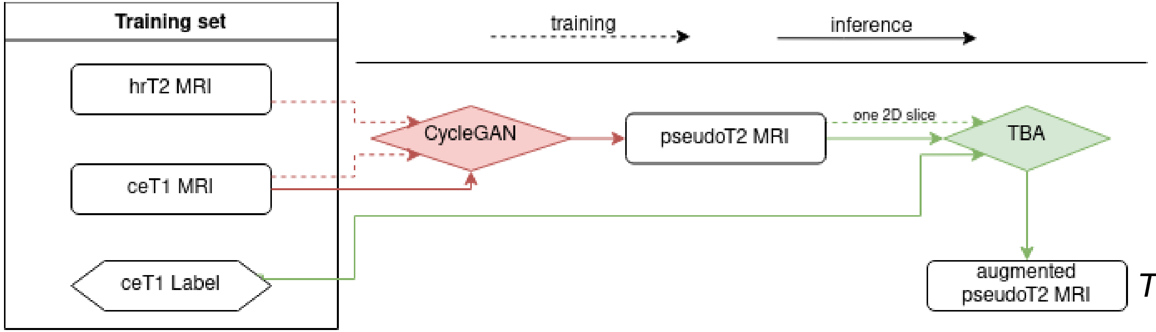
- some real hrT2 **VS** are **large**, **hypersignal** and/or **heterogeneous**. CycleGAN does not generate enough VS with these features.



Real **hrT2** from validation set

→ **Objective 2 : data augmentation**

increase VS variability (and therefore improve segmentation robustness)



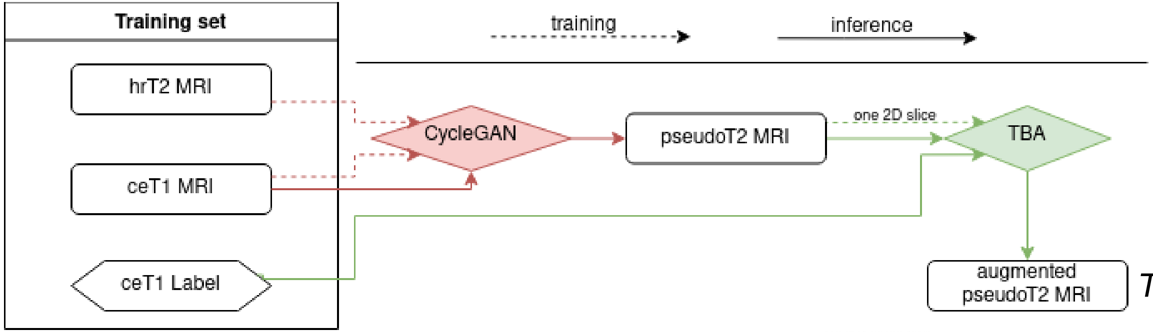
Tumor diversity ++

Proposed workflow

1) Image-to-image (i2i) translation

2) **Tumor blending augmentation** (TBA) using SinGAN [6],[7]

[6] Tamar Rott Shaham et al., “Singan: Learning a generative model from a single natural image,” in ICCV 2019
[7] Guillaume Sallé et al., “Fake tumor insertion using one-shot generative learning for a cross-modal image segmentation,” in IEEE MIC 2021.



Proposed workflow

1) Image-to-image (i2i) translation

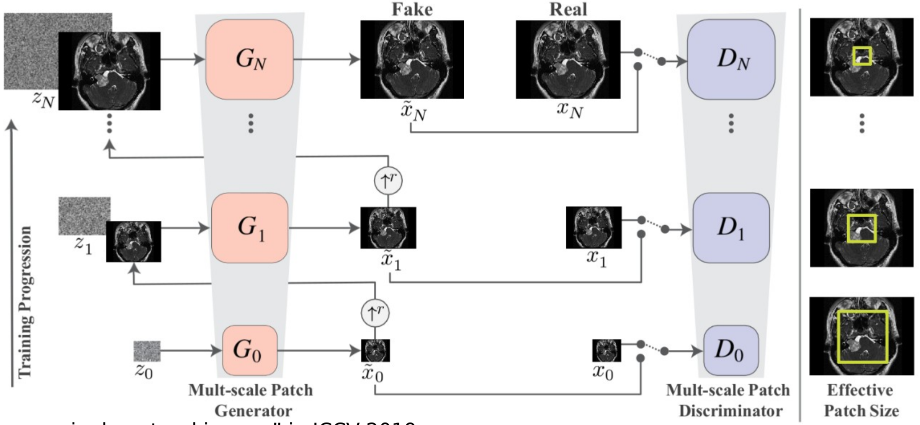
2) **Tumor blending augmentation** (TBA) using SinGAN [6],[7]

Tumor diversity ++

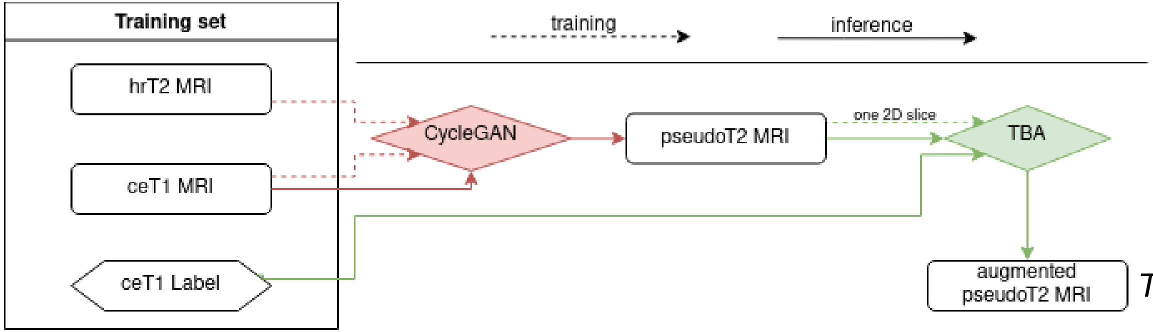
SinGAN = **multi-cascaded GAN** at different scales in a coarse-to-fine fashion. **Trained on one 2D image**

Learning process :
- first GAN learns the composition
- all others learn details at increasingly finer scales

SinGAN for harmonization :
- select a scale level
- use all above generators on a pasted object



[6] Tamar Rott Shaham et al., “Singan: Learning a generative model from a single natural image,” in ICCV 2019
[7] Guillaume Sallé et al., “Fake tumor insertion using one-shot generative learning for a cross-modal image segmentation,” in IEEE MIC 2021.



Proposed workflow

1) Image-to-image (i2i) translation

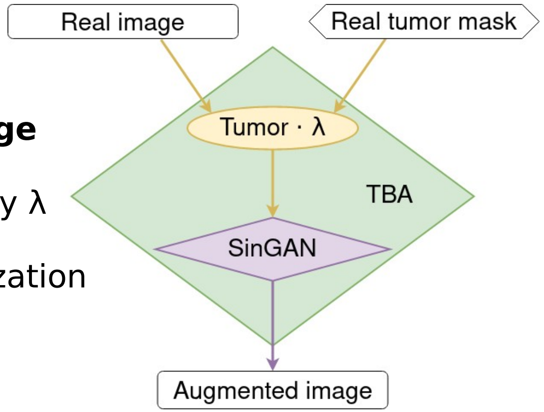
2) **Tumor blending augmentation** (TBA) using SinGAN [6],[7]

Tumor diversity ++

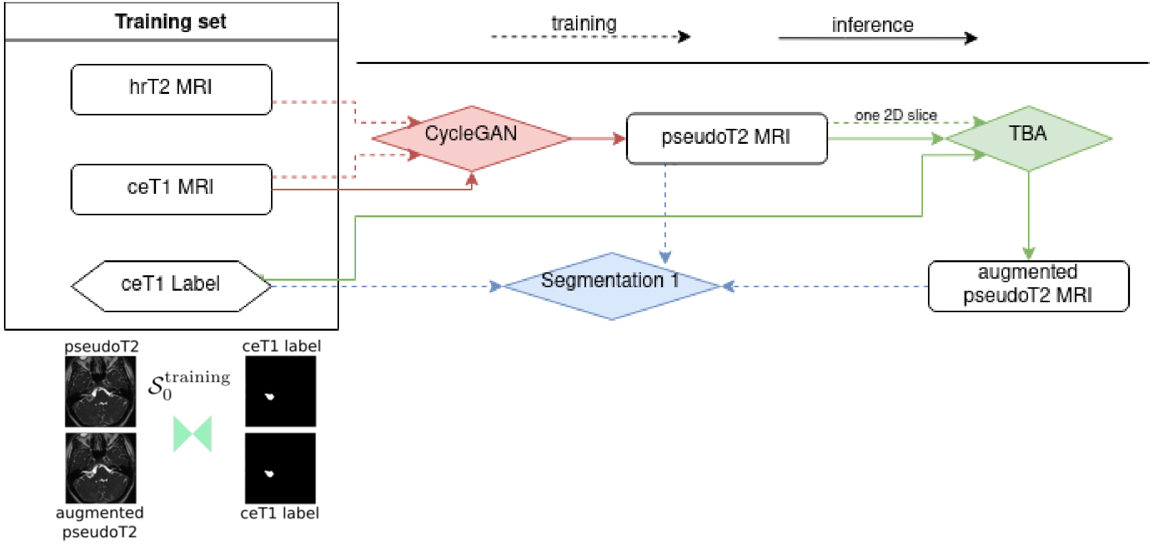
Original SinGAN object harmonization [6]

Proposed TBA to change tumor appearance :

- Scale tumor intensity by λ (intensity scaling factor)
- Apply SinGAN harmonization



[6] Tamar Rott Shaham et al., “Singan: Learning a generative model from a single natural image,” in ICCV 2019
[7] Guillaume Sallé et al., “Fake tumor insertion using one-shot generative learning for a cross-modal image segmentation,” in IEEE MIC 2021.



Proposed workflow

1) Image-to-image (i2i) translation

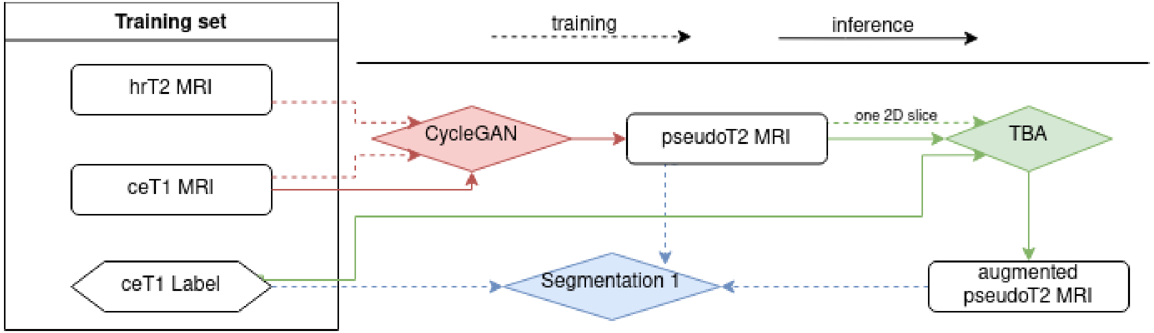
2) Tumor blending augmentation

3) Segmentation using i2i outputs and augmented data

nnU-Net [8] :
- 5-fold ensembling
- 3D full-res
- 500 epochs

iterative training [9]

[8] Fabian Isensee et al., “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” Nature methods, 2021.
[9] Hyungseob Shin et al., “COSMOS: Cross-Modality Unsupervised Domain Adaptation for 3D Medical Image Segmentation based on Target-aware Domain Translation and Iterative Self-Training,” 2022



Proposed workflow

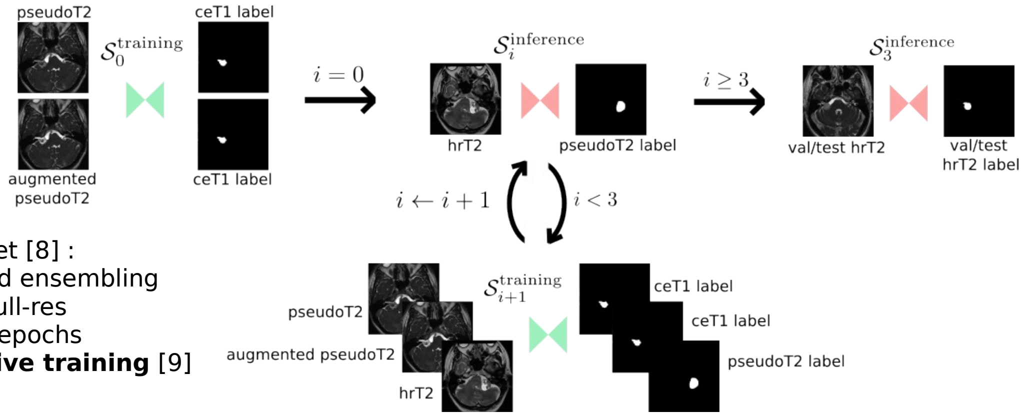
1) Image-to-image (i2i) translation

2) Tumor blending augmentation

3) Segmentation using i2i outputs and augmented data

4) Last segmentation network inferences on real hrT2

5) New segmentation model with i2i outputs, augmented data and real hrT2



nnU-Net [8] :
- 5-fold ensembling
- 3D full-res
- 500 epochs

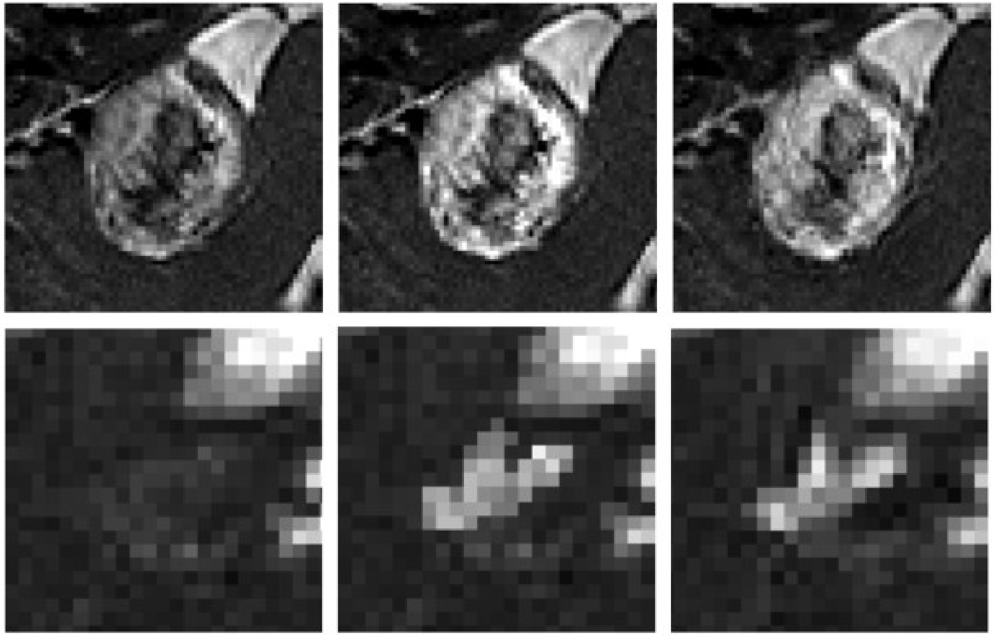
iterative training [9]

We repeat step 4&5 three times

[8] Fabian Isensee et al., “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” Nature methods, 2021.
[9] Hyungseob Shin et al., “COSMOS: Cross-Modality Unsupervised Domain Adaptation for 3D Medical Image Segmentation based on Target-aware Domain Translation and Iterative Self-Training,” 2022

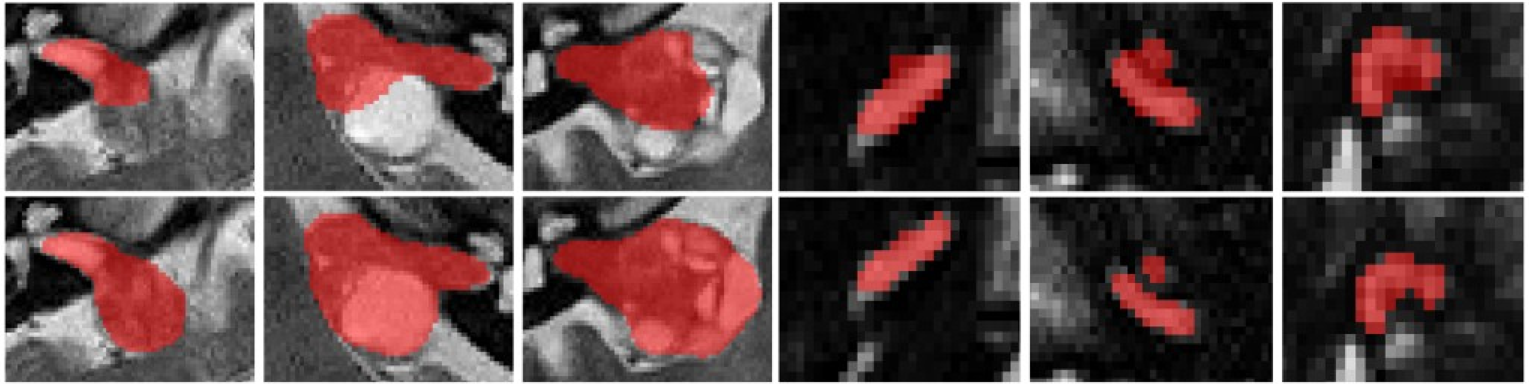
TBA diversifies
VS appearance

TBA to recover cochlea



(a) (b) (c)

Augmentation results on training pseudoT2 images. (a) original pseudoT2, (b) multiplied VS or cochlea (*mask* x 1.5 for VS, *mask* x 4.0 for cochlea), (c) augmented pseudoT2



Segmentation results on validation set. First row w/o TBA, second row w/ TBA (after 1st seg)

	DICE score	ASSD
VS	0.8682 ± 0.0601	0.4302 ± 0.1780
Cochlea	0.8506 ± 0.0294	0.1892 ± 0.1457

Quantitative scores on validation set (best submission)

Conclusion

- New **tumor blending data augmentation** technique to **diversify segmentation training sets**

- Generative model based on a **single 2D image** applied to 3D volumes

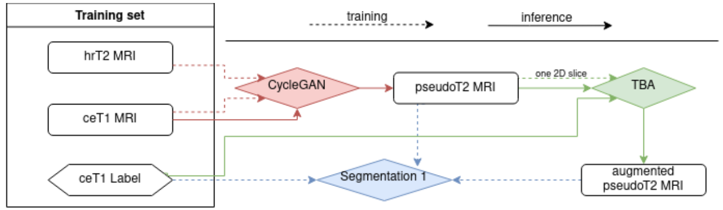
- CrossMoDa 2022 challenge :

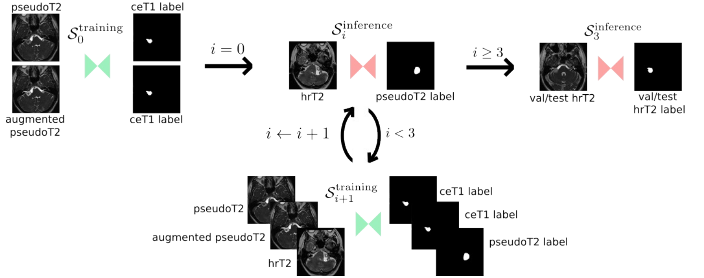
    → diversify VS appearance & enforce cochlea preservation

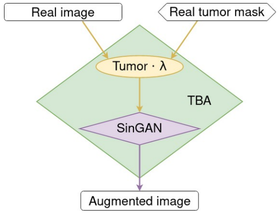
    → 1st place on the validation leaderboard

	DICE score	ASSD
VS	0.8682 ± 0.0601	0.4302 ± 0.1780
Cochlea	0.8506 ± 0.0294	0.1892 ± 0.1457

Quantitative scores on validation set (best submission)

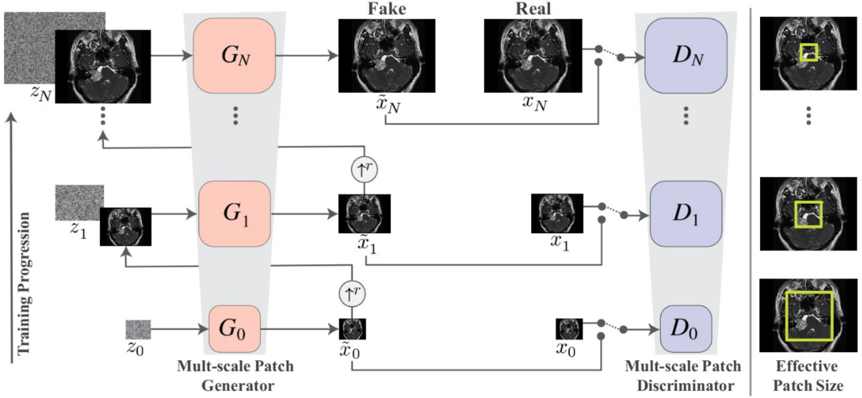






Thank you ! Questions ?

Contact : guillaume.salle@univ-brest.fr



	DICE score	ASSD
VS	0.8682 ± 0.0601	0.4302 ± 0.1780
Cochlea	0.8506 ± 0.0294	0.1892 ± 0.1457

[1] Daniel H Coelho et al., “MRI surveillance of vestibular schwannomas without contrast enhancement: clinical and economic evaluation,” 2018
[2] Jonathan Shapey et al., “Segmentation of Vestibular Schwannoma from Magnetic Resonance Imaging: An Open Annotated Dataset and Baseline Algorithm,” 2021
[3] Reuben Dorent et al., “CrossMoDA 2021 challenge: Benchmark of Cross-Modality Domain Adaptation techniques for Vestibular Schwannoma and Cochlea Segmentation,” 2022
[4] Jun-Yan Zhu et al., “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” in ICCV, 2017.
[5] Joseph P Cohen et al., “How to Cure Cancer (in images) with Unpaired Image Translation,” in MIDL 2018.
[6] Tamar Rott Shaham et al., “Singan: Learning a generative model from a single natural image,” in ICCV 2019
[7] Guillaume Sallé et al., “Fake tumor insertion using one-shot generative learning for a cross-modal image segmentation,” in IEEE MIC 2021.
[8] Fabian Isensee et al., “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” Nature methods, 2021.
[9] Hyungseob Shin et al., “COSMOS: Cross-Modality Unsupervised Domain Adaptation for 3D Medical Image Segmentation based on Target-aware Domain Translation and Iterative Self-Training,” 2022

Main implementation details

Preprocessing (from [8]) : we resampled to 0.6x0.6x1 and extracted 256x256xZ volume by computing the x and y average location of voxels higher than the 75th percentile.

CycleGAN postprocessing : we applied Van Cittert Deconvolution algorithm (VC) for VS : 1 × 1 × 2.5 mm^3 for 15 iterations. For last cochlea segmentation inferences : we applied (VC) with parameters 0.4 × 0.4 × 1.5 mm^3 for 15 iterations.

TBA :
- VS from ETZ of volumes larger than 2340 mm^3 with standard variation higher than 0.09 (6500 voxels for 29 images in total) were augmented with TBA using intensity scaling factors λ of 0.7, 1.2 and 1.5.
- VS of volumes less than 288 mm^3 (800 voxels ; 19 images in total) were augmented with TBA by using λ of 0.6, 0.8 and 1.2 (to increase the proportion of weakly appearing tumors).
- all cochlea were augmented with TBA using λ of 2, 3 and 4.

SinGAN training : default parameters except kersize=5 and scale_factor=0.85 (17 scales in total)
Augmentation is performed twice per lambda value with scale 15 and 13.

Images are resampled to 0.4x0.4x1 spacing before last segmentation model to refine masks.

Challenge evaluation

Metrics:

- Dice Score Coefficient (DSC)
- Average Symmetric Surface Distance (ASSD)

Ranking method:

- Based on BraTS challenge methodology
- Participating teams are ranked for each testing subjects, for each evaluated region (i.e., VS and cochlea), and for each measure (i.e., DSC and ASSD)
- The final ranking score for each team is then calculated by firstly averaging across all these individual rankings for each patient, and then averaging these cumulative ranks across all patients for each participating team

Validation set submission process:

- Predictions submitted via *grand-challenge.org*
- 1 submission allowed per day

Testing set submission process:

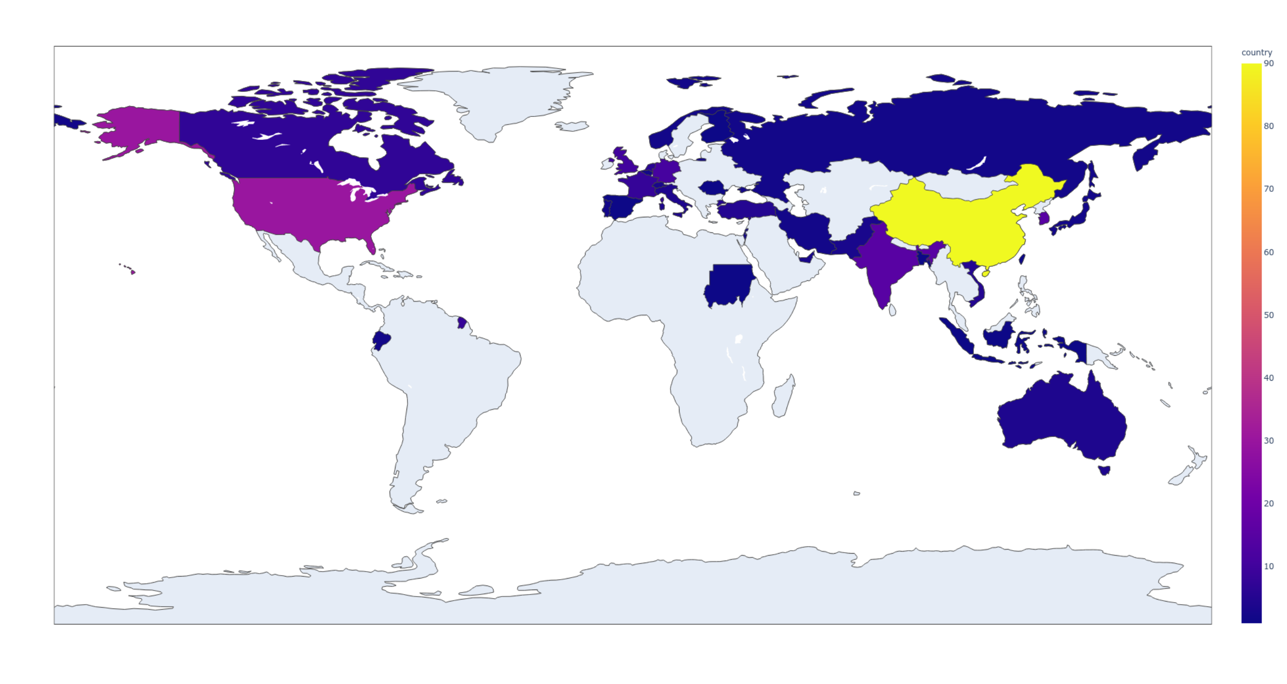
- 1 submission via a Docker container

Participation

Registration:
Number teams: **233**
Number countries: 35

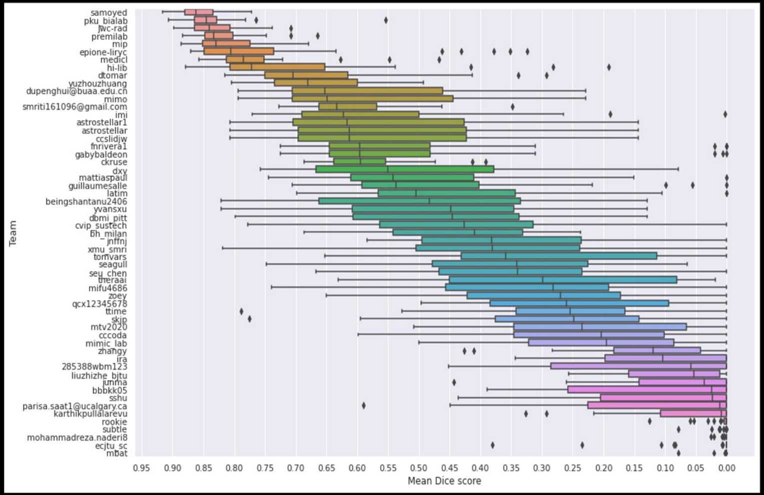
Validation:
Number teams: **27**
Number countries: 15

Testing:
Number teams: **12**
Number countries: 8

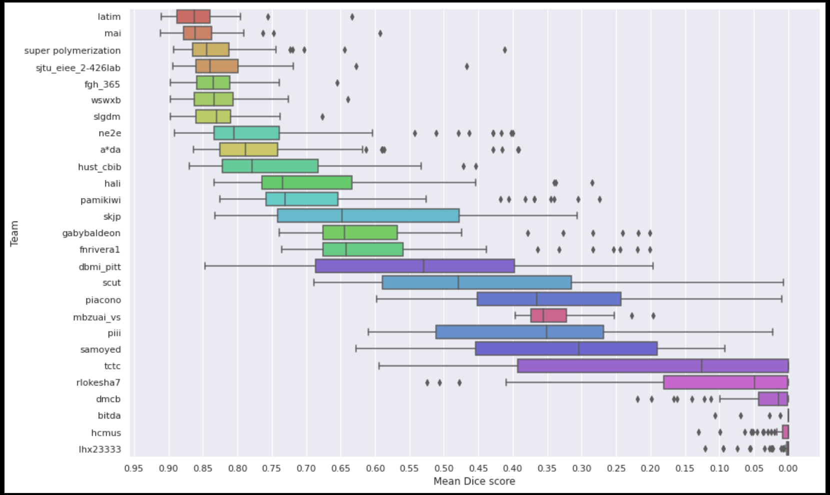


High level observations - validation

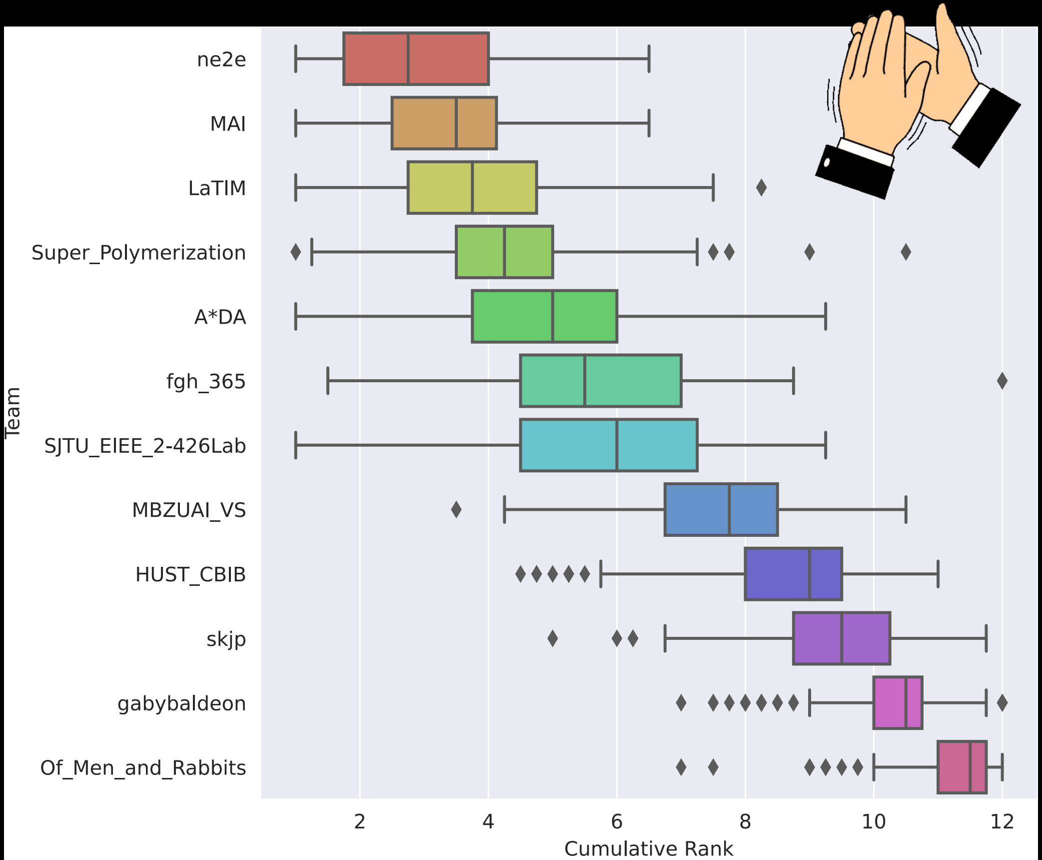
2021



2022



Results



1st – ne2e - ranking score: **3.0**

Hexin Dong, Fei Yu, Mingze Yuan, Jie Zhao, Bin Dong, Li Zhang, Luyi Han, Yunzhi Huang, Tao Tan, Ritse MannHwang

(Peking University, Beijing, China)

Prize: NVIDIA RTX 3090

2nd – MAI - ranking score: **3.4**

Bogyeong Kang, Hyeonyeong Nam), Ji-Wung, Keun-Soo Heo, Tae-Eui Kam

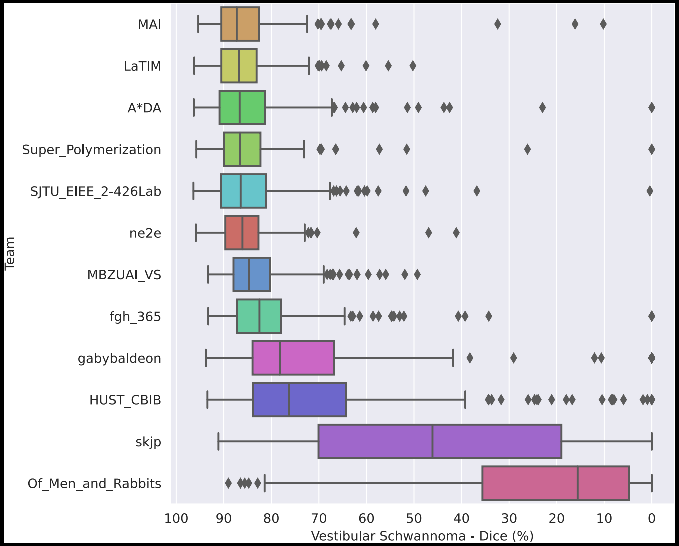
(Korea University)

3rd - LaTIM - ranking score: **3.8**

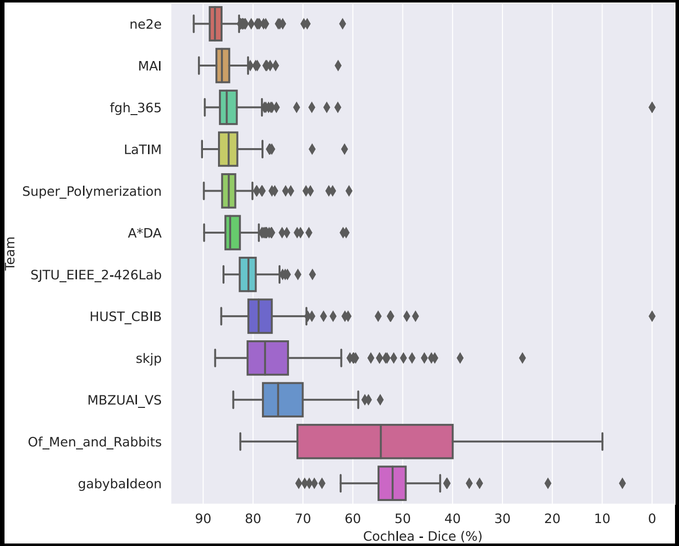
Guillaume Sallé, Pierre-Henri Conze, Julien Bert, Nicolas Boussion, Ulrike Schick, Dimitris Visvikis, Vincent Jaouen

(LaTIM, Inserm)

Overall segmentation performance



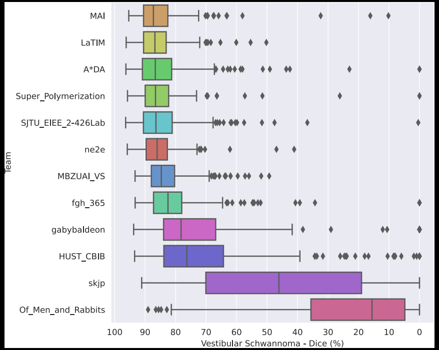
Vestibular Schwannoma

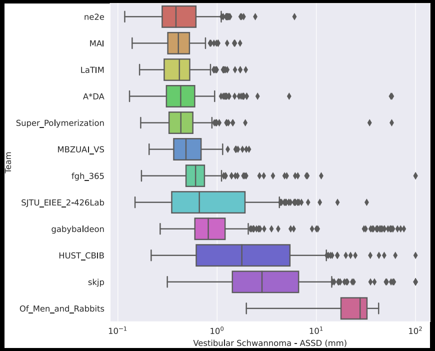


Cochleas

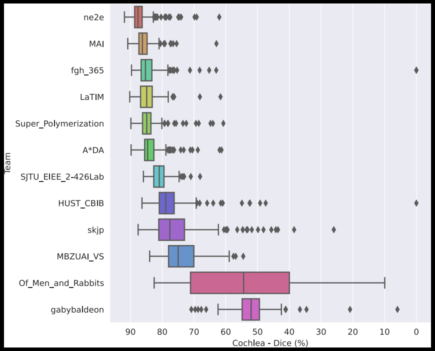
- ne2e (winner): median DSC greater than **86%** for both structures
- Top 5: median DSC greater than **84%** for both structures

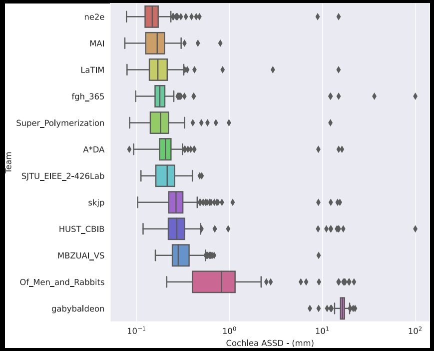
Evaluation per structure





Vestibular Schwannoma





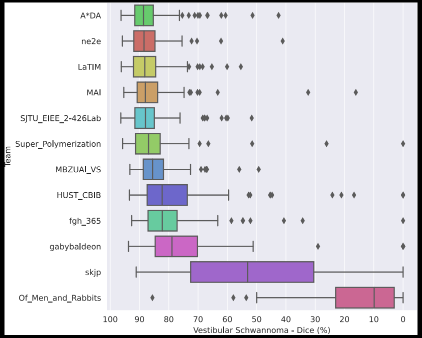
Cochleas

- More **variability** can be observed in terms of algorithm performance for the tumour than for the cochleas
- Top 10 teams: IQRs for the DSC and ASSD are respectively **2.6 and 16 times larger for VS than cochleas**
- More **outliers** for VS than for cochleas

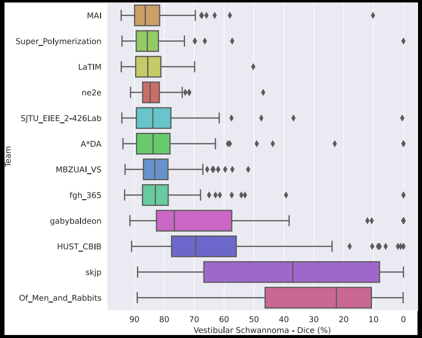
→ proposed algorithms are less robust on VS than on cochleas

cochleas are more uniform in terms of **location, volume size** and **intensity distribution** than tumours

Evaluation per center

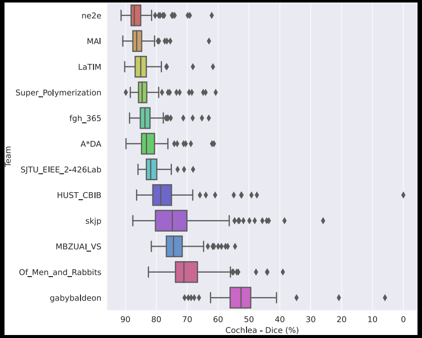


London

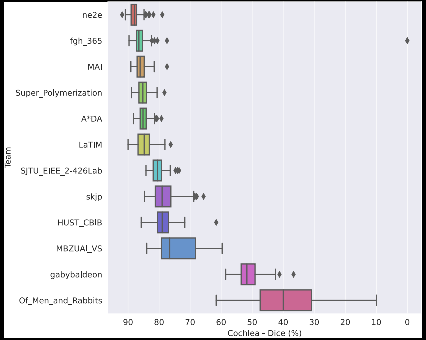


Tilburg

Vestibular Schwannoma



London



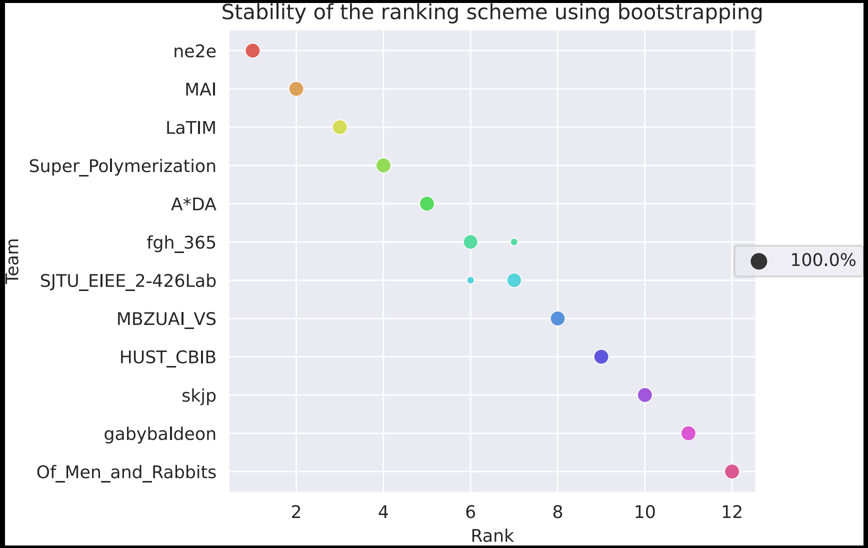
Tilburg

Cochleas

- Similar rankings for each center on cochlea
- Large changes in ranking for each center on VS
- Similar scores on cochlea (median Dice top 5 - London: 85.10%; Tilburg: 85.80%)
- Segmenting VS on Tilburg data is harder (median Dice top 5 – London: 88.10%; Tilburg: 85.4%)

Ranking stability

Bootstrapping (1,000 bootstrap samples) to investigate the ranking uncertainty and stability of the proposed ranking scheme with respect to sampling variability



The ranking stability of the proposed scheme is **excellent**

Comparison with full supervision

Team	Ranking	Vestibular Schwannoma		Cochlea	
		DSC (%)	ASSD (mm)	DSC (%)	ASSD (mm)
ne2e	1	86.1 [82.7 - 89.7]	0.38 [0.28 - 0.61]	87.6 [86.3 - 88.7]	0.15 [0.12 - 0.17]
MAI	2	87.3 [82.5 - 90.5]	0.41 [0.32 - 0.53]	86.2 [84.8 - 87.3]	0.17 [0.12 - 0.20]
LaTIM	3	86.8 [83.1- 90.5]	0.42 [0.29 – 0.43]	84.9 [83.2 - 86.8]	0.17 [0.14 - 0.21]
Super Polymerization	4	86.6 [82.3 – 90.0]	0.43 [0.33 – 0.57]	84.9 [83.6 - 86.2]	0.18 [0.14 - 0.22]
A*DA	5	86.7 [81.3 – 90.9]	0.43 [0.31 – 0.59]	84.6 [82.6 - 85.5]	0.20 [0.18 - 0.23]
Full supervision (nnUnet)		92.5 [89.2 - 94.2]	0.20 [0.14 - 0.29]	87.7 [85.8 - 89.3]	0.10 [0.09 - 0.13]

Problem almost solved

Challenge limitations

Segmentation performance depends on various parameters:

- Pre-processing step (cropping, image resampling, image normalization)
- Training strategy
- Segmentation network
- CycleGAN approach

→ **Difficult to explain** the different levels of performance reached by similar approaches

Domain gap between the source and target images is large, as it corresponds to different modalities
The intra-domain data was **homogeneous**:

→ Lack of robustness may occur when the same modalities are acquired with **different settings**

Oral presentations
Task 2: classification

crossMoDA

MICCAI2022 Singapore
25th International Conference on Medical Image Computing and Computer Assisted Intervention
September 18-22, 2022
Resorts World Convention Centre Singapore

Unsupervised Cross-Modality Domain Adaptation for Vestibular Schwannoma Segmentation and Koos Grade Prediction based on Semi-Supervised Contrastive Learning

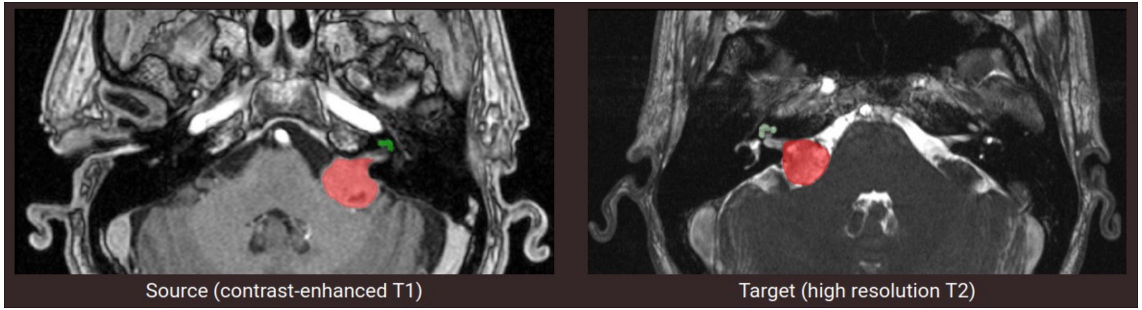
Team Members: Luyi Han, Yunzhi Huang, Tao Tan, Ritse Mann

NKI & RUMC

17-09-2022

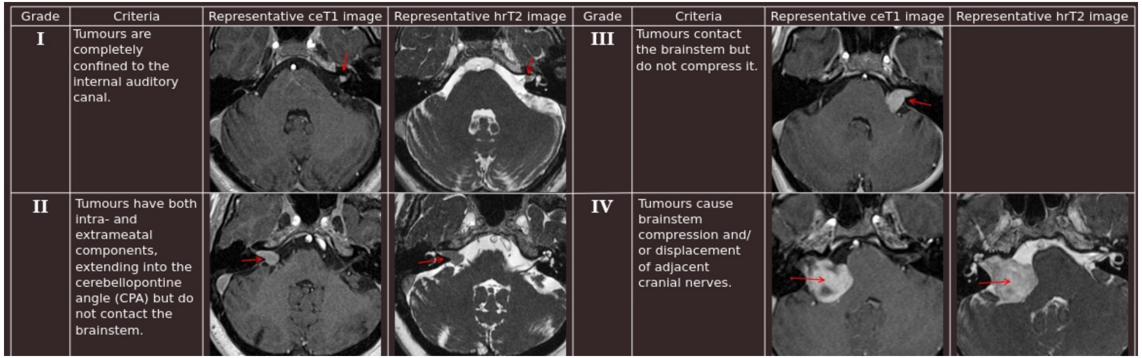
Task 1

- The goal of the segmentation task (Task 1) is to segment two key brain structures (tumor and cochlea) involved in the follow-up and treatment planning of vestibular schwannoma (VS).



Task 2

- The goal of the classification task (Task 2) is to automatically classify hrT2 images with VS according to the Koos grade.



Grade	Criteria	Representative ceT1 image	Representative hrT2 image	Grade	Criteria	Representative ceT1 image	Representative hrT2 image
I	Tumours are completely confined to the internal auditory canal.			III	Tumours contact the brainstem but do not compress it.		
II	Tumours have both intra- and extrameatal components, extending into the cerebellopontine angle (CPA) but do not contact the brainstem.			IV	Tumours cause brainstem compression and/ or displacement of adjacent cranial nerves.		

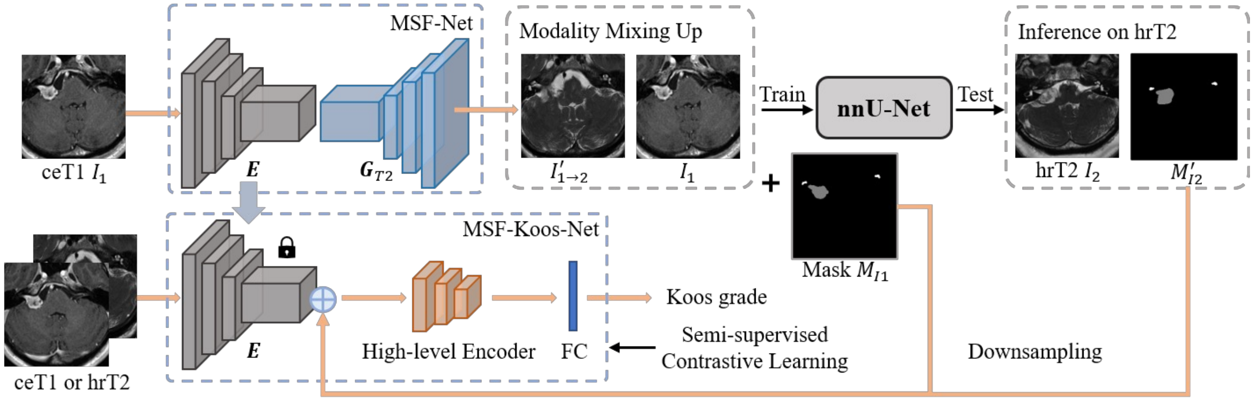
Cross-Modality Domain Adaptation Challenge 2022, *https://crossmoda2022.grand-challenge.org/*

Contributions

- We propose an unsupervised domain adaptation framework to learn the shared representation from both ceT1 and hrT2 images and recover another modality from the latent representation.
- We introduce proxy tasks of VS and GIF segmentation to restrict the consistency of image structures in domain adaptation.
- We employ a semi-supervised contrastive learning pre-train approach to improve the model performance for Koos grade prediction.

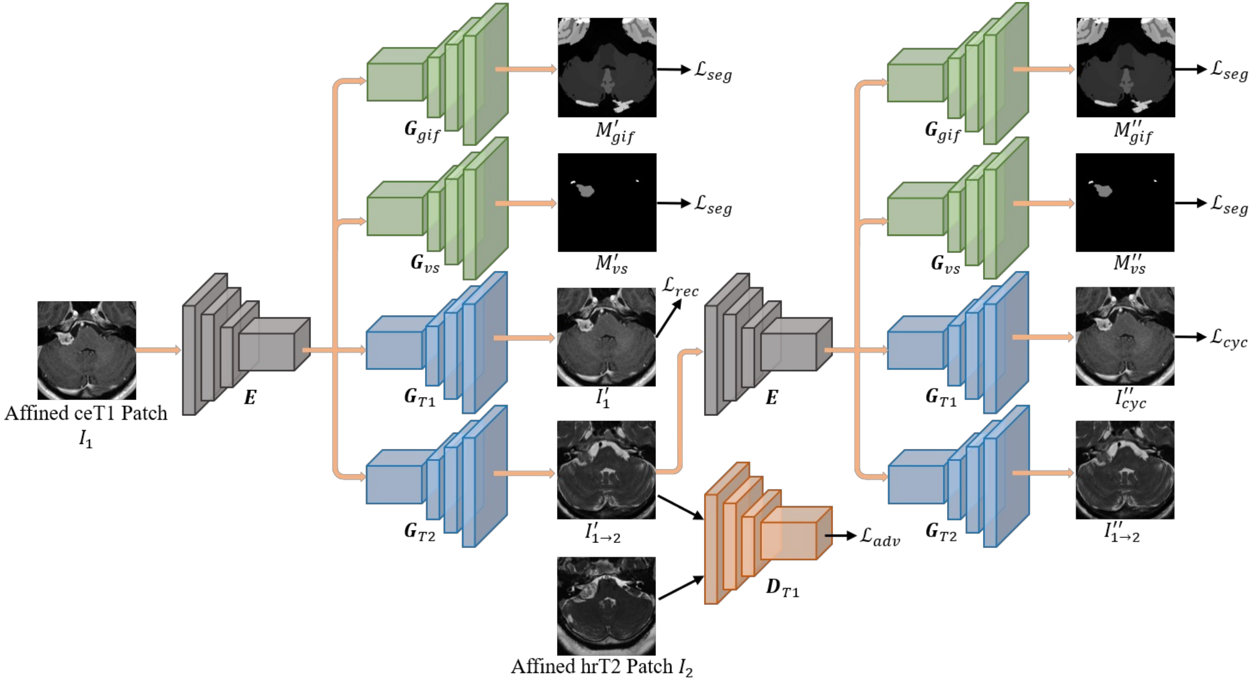
Our code is available at *https://github.com/fiy2W/cmda2022.superpolymerization*

Framework Overview



Overview of the proposed unsupervised domain adaptation segmentation and classification framework.

Our code is available at *https://github.com/fiy2W/cmda2022.superpolymerization*



The architecture of MSF-Net. The reverse transform direction (from real hrT2 to fake ceT1) is omitted for ease of illustration. Not that, both directions share weights for the model, and no proxy paths ($\boldsymbol{G}_{vs}$ and $\boldsymbol{G}_{gif}$) are involved in the reverse direction due to lack of annotations.

Our code is available at *https://github.com/fiy2W/cmda2022.superpolymerization*

Reconstruction loss

$$\mathcal{L}_{rec} = \lambda_r \cdot (\|I_1' - I_1\|_1 + \|I_2' - I_2\|_1) + \lambda_p \cdot \left(\mathcal{L}_p(I_1', I_1) + \mathcal{L}_p(I_2', I_2)\right)$$

Cycle consistency loss

$$\mathcal{L}_{cyc} = \|I''_{1\to2\to1} - I_1\|_1 + \|I''_{2\to1\to2} - I_2\|_1$$

Adversarial loss

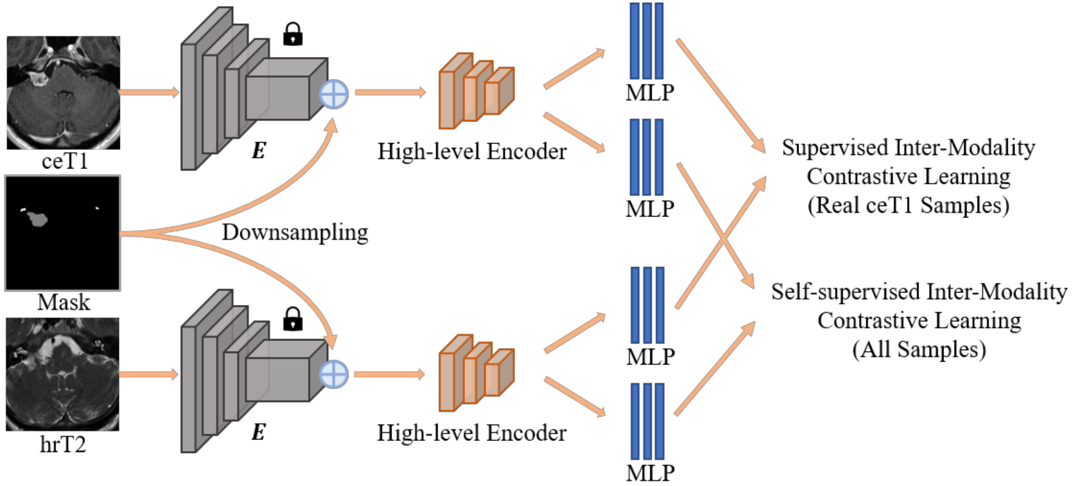
$$\min_{\mathbf{D}_{T1},\mathbf{D}_{T2}} \max_{\mathbf{G}} \mathcal{L}_{adv} = \|\mathbf{D}_{T1}(I_1) - 1\|_2 + \|\mathbf{D}_{T1}(I'_{2\to1})\|_2 + \|\mathbf{D}_{T2}(I_2) - 1\|_2 + \|\mathbf{D}_{T2}(I'_{1\to2})\|_2$$

Segmentation loss

$$\mathcal{L}_{seg} = \mathcal{L}_{ce}(M'_{vs}, M_{vs}) + \mathcal{L}_{dsc}(M'_{vs}, M_{vs}) + \mathcal{L}_{ce}\left(M'_{gif}, M_{gif}\right) + \mathcal{L}_{dsc}\left(M'_{gif}, M_{gif}\right)$$

Our code is available at *https://github.com/fiy2W/cmda2022.superpolymerization*

Part 2 Koos Prediction based on Semi-Supervised Contrastive Learning



The architecture of MSF-Koos-Net.

Our code is available at *https://github.com/fiy2W/cmda2022.superpolymerization*

Self-supervised contrastive learning

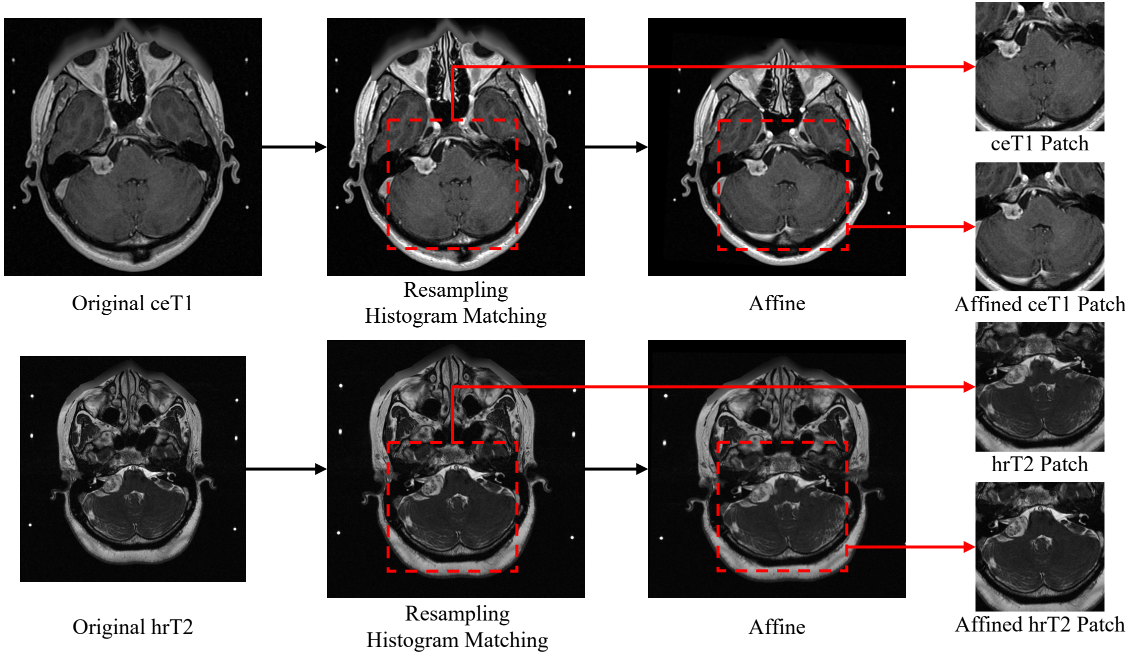
$$\mathcal{L}_{self} = -\sum_{i \in D} \log \frac{\exp\left(z_1^{(i)} \cdot z_2^{(i)} / \tau\right)}{\sum_{j \in D} \exp\left(z_1^{(i)} \cdot z_2^{(j)} / \tau\right)} \cdot \frac{\exp\left(z_1^{(i)} \cdot z_2^{(i)} / \tau\right)}{\sum_{j \in D} \exp\left(z_1^{(j)} \cdot z_2^{(i)} / \tau\right)}$$

Supervised contrastive learning

$$\mathcal{L}_{sup} = -\sum_{i \in A} \frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp\left(q_1^{(i)} \cdot q_2^{(p)} / \tau\right)}{\sum_{j \in A} \exp\left(q_1^{(i)} \cdot q_2^{(j)} / \tau\right)} \cdot \frac{\exp\left(q_1^{(p)} \cdot q_2^{(i)} / \tau\right)}{\sum_{j \in A} \exp\left(q_1^{(j)} \cdot q_2^{(i)} / \tau\right)}$$

Our code is available at *https://github.com/fiy2W/cmda2022.superpolymerization*

Data Preparing



Our code is available at *https://github.com/fiy2W/cmda2022.superpolymerization*

Segmentation results for nnU-Net utilizing generated hrT2 images with different domain adaptation methods.

Methods	VS Dice	VS ASSD	Cochlea Dice	Cochlea ASSD
CycleGAN	0.7402±0.2504	1.7556±5.3276	0.8202±0.0253	0.2325±0.1545
MSF-Net w/o VS&GIF	0.7764±0.2025	0.6905±0.6437	0.8220±0.0510	0.3097±0.2986
MSF-Net w/o GIF	0.8288±0.0838	0.7901±1.0765	0.8285±0.0354	0.2507±0.1828
MSF-Net	0.8493±0.0683	0.5202±0.2288	0.8294±0.0268	0.2454±0.2102

Koos grade prediction results for ablation study of the proposed MSF-Koos-Net.

Semi-supervised contrastive learning	Freeze pre-trained weights	MAMSE
		0.8371
√		0.6805
√	√	0.3940

Our code is available at *https://github.com/fiy2W/cmda2022.superpolymerization*

Thanks for Your Attention!

Team Members: Luyi Han, Yunzhi Huang, Tao Tan✉, Ritse Mann

NKI & RUMC

17-09-2022

Oral presentations
Task 2: classification

MICCAI2022 Singapore

25th International Conference on Medical Image Computing and Computer Assisted Intervention
September 18–22, 2022
Resorts World Convention Centre Singapore

MICCAI

Koos Classification of Vestibular Schwannoma via Image Translation-Based Unsupervised Cross-Modality Domain Adaptation

crossMoDA 2022 Challenge
Team: SJTU_EIEE_2-426Lab

Tao Yang[1], Lisheng Wang[1]
{yangtao22,lswang}@sjtu.edu.cn

Presenter: Tao Yang
September 18, 2022

SHANGHAI JIAO TONG UNIVERSITY

[1] Shanghai Jiao Tong University, Shanghai, P. R. China

Challenge evaluation

Metrics:

- Macro-Averaged Mean Absolute Error
- Takes class imbalance into account
- Depends on the difference between true and predicted label

$$MA - MAE = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{n_c}\sum_{i=1}^{n_c}|y_i - \tilde{y}_i|$$

Validation set submission process:

- Predictions submitted via *grand-challenge.org*
- 1 submission allowed per day

Testing set submission process:

- 1 submission via a Docker container

Results

1st – SJTU_EIEE_2-426Lab_class – MA-MAE: 0.26

Tao Yang, Lisheng Wang
Shanghai Jiao Tong University, China

2nd –Super Polymerization – MA-MAE 0.37

Luyi Han, Yunzhi Huang, Tao Tan, Ritse Mann
Radboud University, the Netherlands

3rd – skjp - MA-MAE: 0.84

Satoshi Kondo, Satoshi Kondo
Muroran Institute of Technology, Japan

Results

Organizing team & sponsors

Reuben Dorent

Harvard Medical School, USA
King's College London, United Kingdom

Tom Vercauteren

King's College London, United Kingdom

Jonathan Shapey

King's College Hospital NHS Foundation Trust, United Kingdom

Aaron Kujawa

King's College London, United Kingdom

Samuel Joutard

King's College London, United Kingdom

Stefan Cornelissen

Elisabeth-TweeSteden Hospital, Tilburg, Netherlands

Patrick Langenhuizen

Elisabeth-TweeSteden Hospital, Tilburg, Netherlands

Ben Glocker

Imperial College London, United Kingdom

Nicola Rieke

NVIDIA

Spyridon Bakas

University of Pennsylvania, USA

Steve Connor

King's College Hospital NHS Foundation Trust, United Kingdom

Marina Ivory

King's College London, United Kingdom

Mohamed Okasha

Ninewells Hospital NHS Tayside, Dundee United Kingdom

Anna Oviedova

Charing Cross Hospital. Imperial College Healthcare NHS Trust, London United Kingdom

NVIDIA