



深圳市大数据研究院

Shenzhen Research Institute of Big Data

2019 MICCAI Overview

Changmiao Wang

CUHKSZ & SRIBD

Nov. 28. 2019



主要内容

- *MICCAI的介绍*
- *2019 MICCAI Shenzhen*
- *MICCAI的研究趋势*
- *MICCAI青年科学家奖*
- *MICCAI 海报*



MICCAI

- MICCAI会议创建于1998年，由早期的计算机视觉与虚拟现实及医用机器人-CVRMed (Computer Vision, Virtual Reality and Robotics in Medicine)，医学机器人和计算机辅助手术- MRCAS (Medical Robotics and Computer Assisted Surgery) 和计算生物医学可视化-VBC (Visualization in Biomedical Computing) 三大会议合并而成。目前已被公认为是医学图像计算、医疗机器人、人工智能、辅助介入、计算生物医学等领域顶级的国际会议，论文接受率仅为30%。
- MICCAI是计算机影像处理计算 (MIC) 以及计算机辅助介入 (CAI) 两个领域的综合性会议。2004年时，MICCAI Society正式成立。



MICCAI

- MIC中包含的课题包括配准、机器学习、图像分割、传统CAD（计算机辅助检测）以及临床和生物学应用。CAI集中在介入部分，包括追踪和导航、介入式影像、医用机器人等。
- MICCAI理事会全程负责MICCAI申办的过程。MICCAI每年在美洲、欧洲、亚洲三地轮流举办。MICCAI的理事会成员里，通常有投票权的13-15人左右，由他们来推选会议地点。



2019 MICCAI Shenzhen

- 13-17 Oct 2019, Shenzhen, China
- 参会人员：
 - 1998年 400人左右；
 - 2017年 > 1000人；
 - 2019年 >2300人；



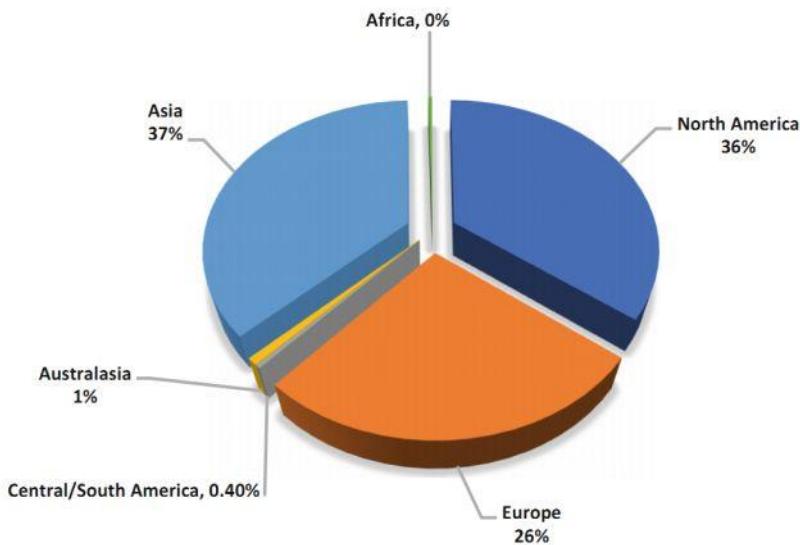
2019 MICCAI Shenzhen Paper Overview

- 投稿数量相比2018年增加63%
- 共录取538篇paper, 录取率31%
- 审稿人: 1300 (原700左右)
- 双盲审 ($>=3$)
- 入选文章强调临床价值, 深度学习占主导 (~80%)
- 华人领域主席: 13.8% → 43.5%
- 2018年, 亚洲录取比例18% (美洲33%,欧洲49%)
- 2019年, 亚洲录取比例37% (美洲36%,欧洲26%)
- 45篇MRI, 30篇CT, 18篇Ultrasound, 7篇OCT, 5篇PET,
- 21篇Surgery
- 12篇CNN, 19篇GAN, 28篇Generative, 93篇Deep, 150篇Network

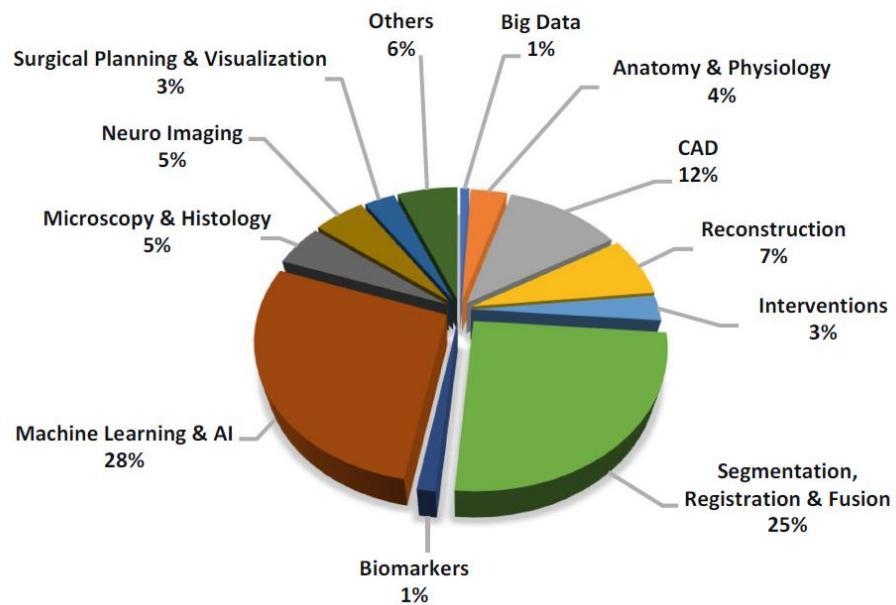


Accepted MICCAI 2019 Papers

By Region of First Author



By Technical Keyword



2019 MICCAI Shenzhen Paper Overview

Tasks

- Segmentation & Registration (174)
- Reconstruction, Synthesis and Image Quality (69)
- Detection and Classification (100)
- Computer Assisted Interventions (45)



2019 MICCAI Shenzhen Paper Overview

■ Data Modality

- MRI
 - Pathological Image
 - CT
 - Microscopy and Photo-optical Imaging
 - OCT
 - Fundus
 - X-ray

The image shows a presentation slide titled "Full-Spectrum" featuring a grid of medical imaging icons. The icons include PET-CT, SPECT, PET, MRI, Ultrasound, X-ray, and other unlabeled circular icons. Below the icons, the text "Full-stack Full-Spectrum AI / Empowerment and Efficiency" is displayed. In the bottom right corner of the slide, there is a small green "Next" button.

Organ

Brain Tumor

Lung Cancer

Liver Cancer

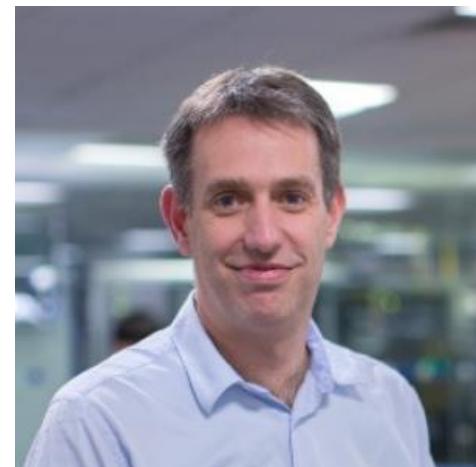
Breast Cancer

Kidney and Tumor



MICCAI研究趋势

- 2014年当选为MICCAI Society fellow的英国帝国理工学院教授Daniel Rueckert
- “Deep learning is now dominating everything in this area. There are virtually no talks without deep learning.”（深度学习“统治”了医学影像分析领域，现在几乎无人不谈深度学习）



MICCAI研究趋势

- MICCAI创始主席James Duncan
- 1998年，第一届MICCAI在麻省理工学院举办时，没有人预想到它会发展成为如今的规模，因为那时候的参会人数只有400人左右。但是近几年来，AI技术渗透到了每个角落，也随之点燃了医学影像分析。我们看到，现在图像重建、增强、分类等方向都在大范围应用深度学习和其他有意思的方法。”



MICCAI研究趋势

- 现在的研究更在于**如何将深度学习结合自己问题的领域知识**，才能达到原创性。
- 每一种方法都各有利弊，**深度学习和知识模型相结合**的方法，学者们一直在探索。“模型意味着把问题简化，用数学的方式来表示问题，但是里面有很多问题无法用这个模型来描述，而深度学习的参数非常多，可以解决很多具体的问题。怎么样把两者结合起来，非常重要。我永远认为，研究不能只偏向其中一个方向。”



MICCAI研究趋势

- 第一个方面是深度学习的自动化。周少华表示，深度学习受很多人工的东西影响：人工标注的数据、人工设计的网络架构、loss函数也是要人工设置。“所以，我觉得，一个比较大的趋势是去自动化完成这些工作。目前，这些方面的研究也比较多。”
- 第二个方面是成像与分析的紧密结合。“成像与分析是整个影像链中两个有机组成部分。我们现在是有了图像之后再进行分析，但是如果在成像端与分析直接结合，也可以做很多有意思的事情，也符合端到端学习的思路。”
- 第三个方面就是联邦学习（Federated Learning）。联邦学习是一种新兴的人工智能基础技术，能够让开发者与各企业机构利用分散在多个位置的训练数据对中心深度神经网络（DNN）进行训练的学习范式，这个方法可以支持各企业机构针对共享模型开展协作，而无需共享任何临床数据。
- 沈定刚教授也认为，研究方法都是来自于实际应用场景。联邦学习对于医疗数据的隐私性来说是一个很好的方法，既可以保证数据“不出院”，又能够利用不用医院的数据训练同一个算法。
- 第四个方面，也是周少华博士研究的侧重点，在于通用表征学习（Universal Representation Learning, URL），尝试用一种通用的学习方法来同时处理多个任务，每个任务可以有不同的领域。
- 通用表征学习的好处在于，单个任务的数据量不大，难以训练一个好的模型。如果将所有任务放在一起，就会有更多的数据，进而更好地提升模型的性能。“我们希望可以学到一个通用性的表达，对所有的任务都能适用。”
- 随着AI应用的不断推广与落地，可以预见的是，“古今结合”和从临床需求衍生出的新方法将不断涌现。



MICCAI青年科学家奖 (Young Scientist Award, YSA)

- Models Genesis: Generic Autodidactic Models for 3D Medical Image Analysis

作者：Zongwei Zhou, Vatsal Sodha, Md Mahfuzur Rahman Siddiquee, Ruibin Feng, Nima Tajbakhsh, Michael Gotway, Jianming Liang

点评：这篇文章的贡献是设计了一个针对三维医学图像分析的预训练模型，这样解决了以前大家只能用ImageNet里的二维数据训练出来的预模型，并且得到更好的效果；在5个医学图像的分割和分类问题上取得领先的效果；在作者的口头发言中也给出了开源代码<https://github.com/MrGiovanni/ModelsGenesis>
<https://zhuanlan.zhihu.com/p/86366534>

- Deep Multi Label Classification in Affine Subspaces

作者：Thomas Kurmann, Pablo Márquez Neila, Sebastian Wolf, Raphael Sznitman

点评：主要的贡献是在多类分类任务中，将不同类型的样本映射到预先定义好的具有依附关系、并且均匀分布在整个特征空间中的相应子空间，而不是像传统方法中只是简单将不同类型的样本映射到距离较远的不同子空间。



MICCAI青年科学家奖 (Young Scientist Award, YSA)

- Fully convolutional boundary regression for retina OCT segmentation

作者：Yufan He, Aaron Carass, Yihao Liu, Bruno Jedynak, Sharon Solomon, Shiv Saidha, Peter Calabresi, Jerry Prince

点评：眼底图像分割一般是先做像素分割，然后估计出不同层的边界。不过，这样的分割方法没法端到端地优化整个分割任务，而且每一层边界的平滑性和层与层之间的前后关系都没有在一个网络中综合优化。这篇文章的主要想法是将所有的分割任务（像素分割、边界估计和相邻边界的关系）都放在一个网络来优化，这样可以达到整体优化的目的，并且得到好的结果。

- Clustering of longitudinal shape data sets using mixture of separate or branching trajectories

作者：Vianney Debavelaere, Alexandre Bône, Stanley Duralleman, Stéphanie Allassonnière

点评：这篇文章主要想解决随时间变换的形状数据的聚类问题；作者提出了一个混合模型来解决生长过程中一类变成多类（例如两类）的问题。在脑老化的应用中，该方法可以发掘出海马形状随年龄老化分成两个子类



MICCAI青年科学家奖 (Young Scientist Award, YSA)

- Diagnosis-guided multi-modal feature selection for prognosis prediction of lung squamous cell carcinoma

作者：Wei Shao, Tongxin Wang, Zhi Huang, Jun Cheng, Zhi Han, Daoqiang Zhang, Kun Huang

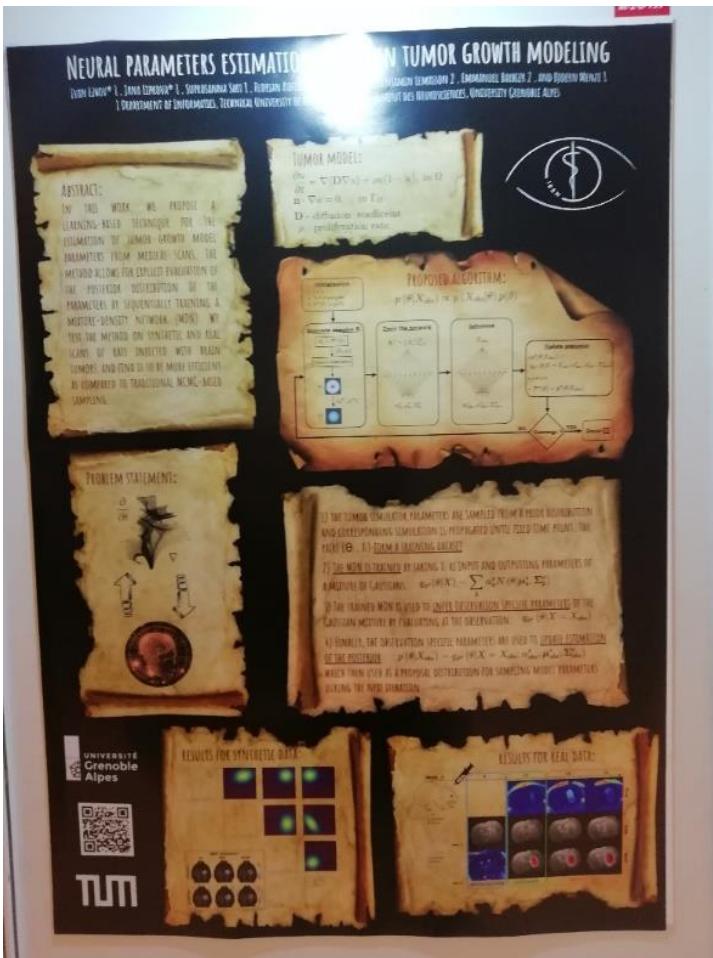
点评：这篇文章要解决的问题是用病理图像和基因数据来预测生存期。通常的特征选择方法是用单任务的方法完成特征选择。这篇文章的基本思想是把生存期预测和临床诊断信息预测作为多目标任务来进行特征选择（虽然临床诊断信息的预测不是这篇文章的目标）。在公开数据集上，跟其他方法相比，有3%左右的精度提高。



 Airway Segmentation	 Area Bound_Segmentation	 Attention	 classification	 conclusion
 Context Memory_Segmentation_Ultrasound	 data augmentation	 denoising	 GAN	 GAN Data augmentation
 GAN_3D	 GAN_CTMRI	 GAN_Data_Enhance	 GAN_Segmentation_CT	 GAN_Unsupervised
 GAN-CGN	 Graph	 Interpretable	 Local_Grobal	 Multi-instance Learning survive prediction
 Multi-view	 Projective Adversarial Network_Segmentation	 reconstruction	 Reinforcement Learning	 Segmentation
 Segmentation 3D	 Segmentation and Registration	 Segmentation_2D+3D	 Segmentation_Capsule Network	 Segmentation_Mixed Supervised
 Segmentation_Vessel	 Semi-supervised	 Transfer Learning	 Transformer	 U-net
 unsupervised anomaly location	 Weakly supervised	 X_ray	 肺结节	 可参考的
 皮肤病	 乳腺	 细胞	 眼底	 支气管分割



Tumor Growth



Neural parameters estimation for brain tumor growth modeling

Deep Probabilistic Modeling of Glioma Growth

<https://github.com/jenspetersen/probabilistic-unet>



Classification

Multi-Instance Multi-Scale CNN for Medical Image Classification

Shaohua Li¹ Yong Liu¹ Xiuchao Sui¹ Cheng Chen¹ Gabriel Tjio² Daniel Ting² Rick Goh¹

¹Institute of High Performance Computing, A*STAR
²Singapore National Eye Centre Singapore

Common Challenges

- Medical image data are small
- Labels are at whole-image level, not ROI-level (weak labels)
- ROI patterns in images are often scale-invariant

Contributions

- A multi-scale convolutional layer extracts ROI patterns in multiple scales
- A top-k pooling layer aggregates ROI features (multi-occurrence & multi-scale) across whole image, so weak labels can be used for training

Advantages

- A versatile framework for classification of various medical image data (both 3D & 2D)
- Small RAM footprint: Train on 3D Images with a single GPU
- Performed consistently better than 11 baselines

Details

MSCore:

- Extract primary features with a pretrained CNN
- Reuse primary feature into m scales
- Filter each scale with a set of secondary conv Filters
- Same set of filters are shared by different scales
- Scale-invariant ROI patterns are captured
- Number of parameters are reduced, good for small data

Top-k pooling:

- Get highest k activations from each channel of activations
- Features along Slice, Scale, Height & Width axes are pooled
- Compute the weighted sum of the top-k activations
- The k weights are optimized with backprop

Top-k pooling

MIMs on a 3D input image (J-th channel only)

Methods

	Cirrus	Spectralis	MDD (2D)	CRC-MSI (2D)	Avg
MN-Pre	0.574	0.506	0.856	0.880	0.829
Resnet + top-k					
Pyramid + top-k	0.638	0.371	0.965	0.855	0.707
Resnet + image pyramid + top-k					
MI-Pre-Conv	0.972	0.990	0.961	0.870	0.948
Resnet + conv3D + top-k					
MIMs-Nonlocal	0.956	0.975	0.961	0.879	0.942
Resnet + image pyramid + top-k					
Pyramid+MIMs	0.848	0.881	0.966	0.673	0.842
Resnet + image pyramid + conv(2,2) + top-k					
MI-Pre-Trident	0.930	1.000	0.966	0.897	0.948
Resnet + dilated conv2D + top-k					
3D-CNN	0.983	1.000	0.972	0.880	0.959
MIMs-score 6k, scale multistage	0.959	0.981	0.970	0.888	0.952
FeaturePy-4k					
ResNet34+scatnet	0.699	0.734	0.824	0.667	0.731
ResNet34+non-sharing					
MIMs	0.986	1.000	0.972	0.901	0.985
Resnet + image pyramid + top-k					
MIMs-patchcls	0.874	0.722	/	/	/
MIMs-topk+arg prediction					
MI-Pre-Conv+patchcls	0.754	0.227	/	/	/

12 methods on four image datasets (in AUROC)

Methods	mean	max	min	k = 2	k = 3	k = 4	k = 5
AUROC on Cirrus	0.829	0.960	0.375	0.986	0.990	0.986	0.986

Top-k pooling (J-th channel only) on m scales

mashash@gmail.com

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7000000/>

Spatial-Frequency Non-Local Convolutional LSTM Network for pRCC classification

Yi Zhou^{1*}, Yina Lin^{2*}, Xiangfeng Mai³, Jürgen Gallenmüller¹, Dianna Waldmann-Greuter¹, Bangwei Li¹, Xiaohui Hu¹, Xiaofei Zhou², Kangning Sun², Jürgen Münch¹

¹ Institute of Computer Science, Technische Universität München, Munich, Germany
² The Affiliated Ningpu Dong-Tower Hospital, Ningpu University Medical School, Ningpu, Jiangsu, China
³ Department of Nuclear Medicine, University of Bern, Bern, Switzerland
<https://doi.org/10.1109/CVPR45500.2021.9530010>

Abstract

This work develops a spatial-frequency non-local convolutional LSTM network for 3D image classification. (1) Compared to traditional networks, the proposed model has the ability to extract features from both the spatial and frequency domains, which allows the two feature-domain features to interact to the classification. (2) The non-local block in our architecture enable it to capture the long-range dependencies directly in the feature space. (3) To simplify the classification task and improve the performance, we utilize a two-stage framework that localizes lesions in the first step, and classifies them in the second. (4) We evaluate our method on a challenging and important clinical task, i.e., the differentiation of papillary renal cell carcinoma (pRCC) into subtype 1 and subtype 2.

2. Methodology

- The proposed framework consists of a localization and a classification stage. We leverage the idea of transfer learning in the localization stage.
- The proposed network consists of a feature-extraction component and a fusion component. The feature-extraction component includes three feature-extraction paths, where the top one is designed for extracting features from spatial domain and the other two are for frequency domain.
- The non-local block can capture long-range dependencies in the feature space, which is a more general extension of the self-attention mechanism.
- We develop our network with bidirectional Convolutional LSTM to learn the 3D context of slices sequentially and alleviate the demand of the memory.

3. Network Architecture

Fig 1. The overall framework of the proposed approach.

Fig 2. Proposed spatial-frequency non-local convolutional LSTM network.

5. Conclusion

- Experimental results demonstrate the network benefits from the proposed strategies including extracting features from both the spatial and frequency domains and employing the non-local block to capture the long-range dependencies.
- Our network shows competitive and often superior performance compared to state-of-the-art classification networks and even clinical experts.
- Currently, more data is being collected and annotated. In the future, increasing the amount of training data may further enhance the performance of the method.

6. References

- Novikov, A.A., et al.: Deep sequential segmentation of organs in volumetric medical scans. IEEE transactions on medical imaging (2016)
- Wang, X., et al.: Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7794–7803 (2016)

Tab 1. Ablation study and comparison between the proposed network and state-of-the-art methods

Method	Accuracy(%)	Sensitivity(%)	Specificity(%)
Ours	84.2	80.3	85.7
Spatial-Frequency CLSTM	80.7	71.2	81.6
Spatial-uniform CLSTM	77.2	66.7	78.6
Spatial CLSTM	70.2	63.6	71.4
3D ResNet	62.1	46.5	62.5



Data augmentation

The poster is titled "Hydranet: Data Augmentation for Regression Neural Networks" and is presented by Marleen de Brujne^{1,2}, Florien Dubost¹, Hieah Adams¹, M. Arfan Ikram¹, Melike W. Vermeij¹, and Marleen de Brujne^{1,2}. It is from Erasmus University Medical Center Rotterdam, Delft University of Technology, and University of Copenhagen.

Introduction

- Data augmentation method
- For CNNs optimized with image-level labels
- Labels should represent a countable quantity
- Evaluated on brain lesion quantification
- Count of enlarged perivascular spaces
- Volume of white matter hyperintensities
- Quantification from 3D MRI images

Method

- Combines existing samples: real samples
- To construct many more new samples: virtual samples
- The CNNs are optimized to predict a single label for a random set of images
- The CNN architecture is replicated with shared weights to process simultaneously all images in the set
- The predictions of each subnetwork are summed to output a single label

Real **Virtual**

Diagram showing the combination of real and virtual samples. Real samples are shown as 3x3 grids with counts 3, 2, 6. Virtual samples are shown as 3x3 grids with counts 3+2=5, 9, 2. A total of 8 real and 6 virtual samples are shown.

Interrater agreement performance with only 25 training scans with a single image-level label per scan, and without pretraining.

Neural Network Architecture

The architecture consists of:

- Input
- Two parallel paths:
 - Path 1: Conv 32, 3x3x3; Conv 32, 3x3x3; Maxpool, 2x2x2; Conv 64, 3x3x3; Conv 64, 3x3x3; FC 1
 - Path 2: Data Augmentation → Data Regressor → Data Regressor → Data Regressor → Data Regressor
- FC 1
- Output

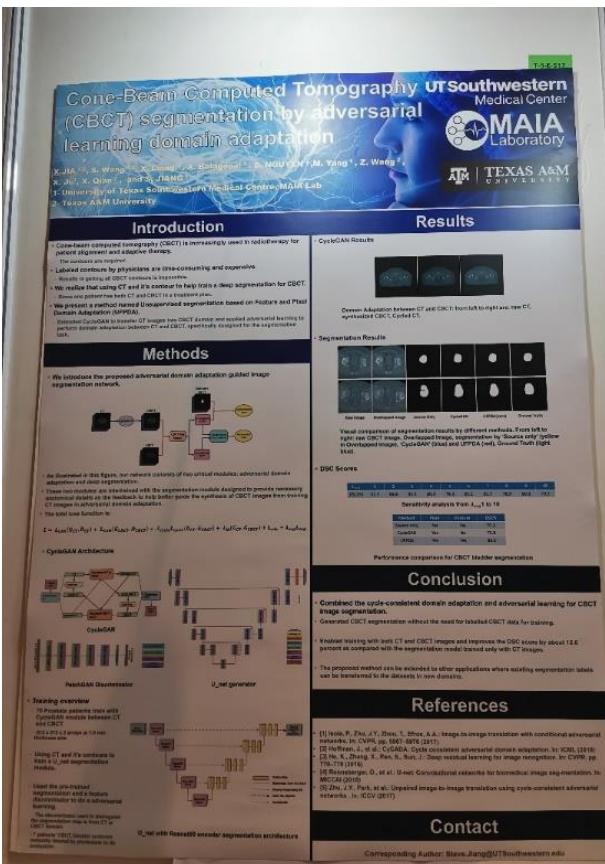
Contact Information

Marleen de Brujne
marleen.debrujne@erasmusmc.nl
Florian Dubost
florian.dubost1@gmail.com

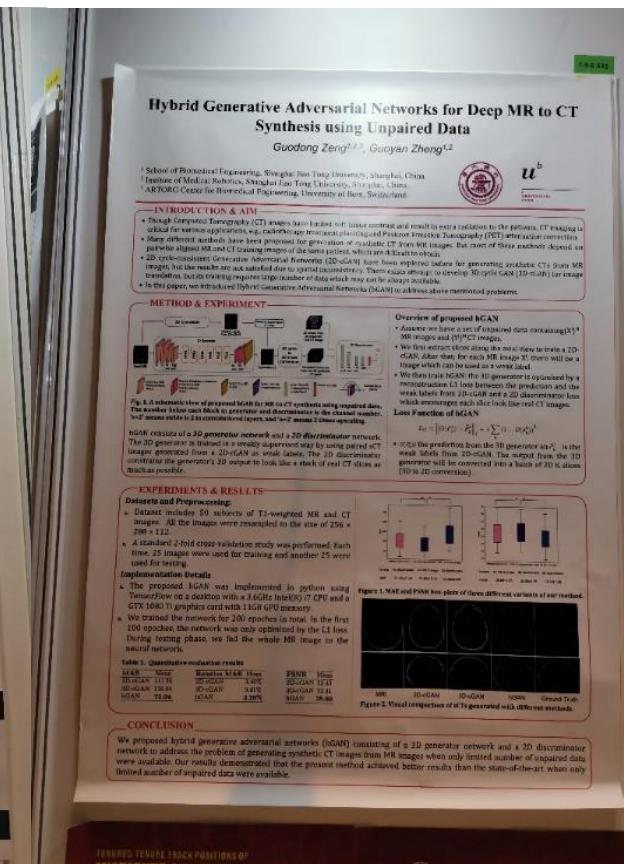


GAN

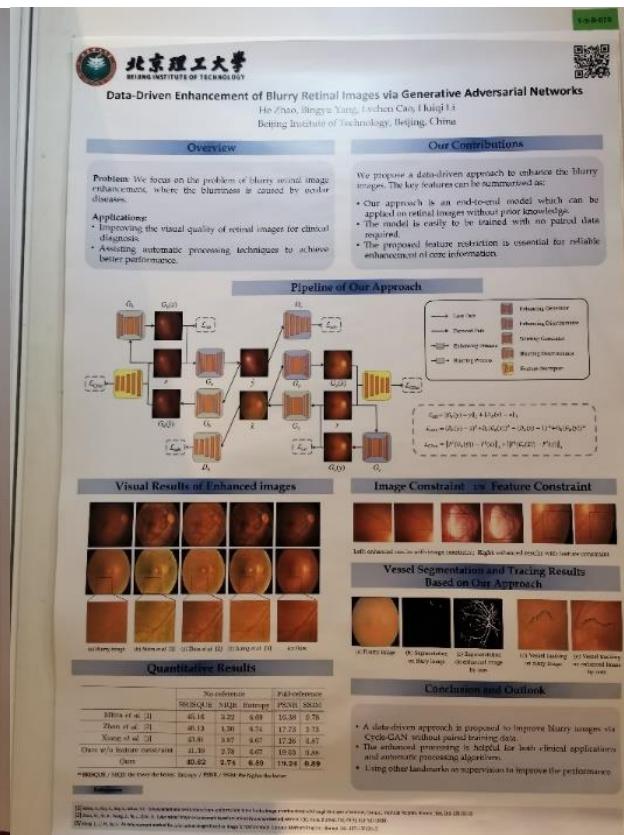
GAN + Segmentation + CT



GAN + MRI + CT



GAN + Blurry Retinal Images

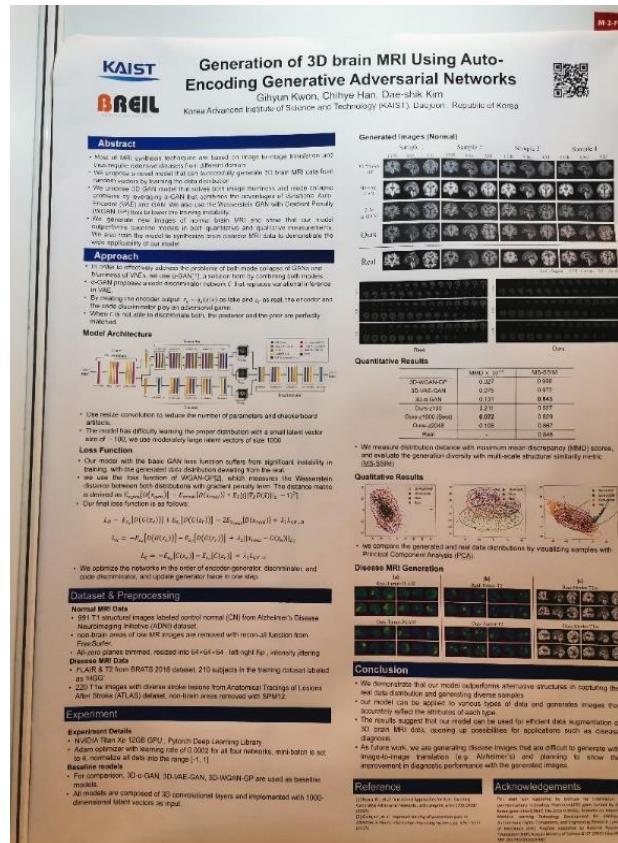


深圳市大数据研究院

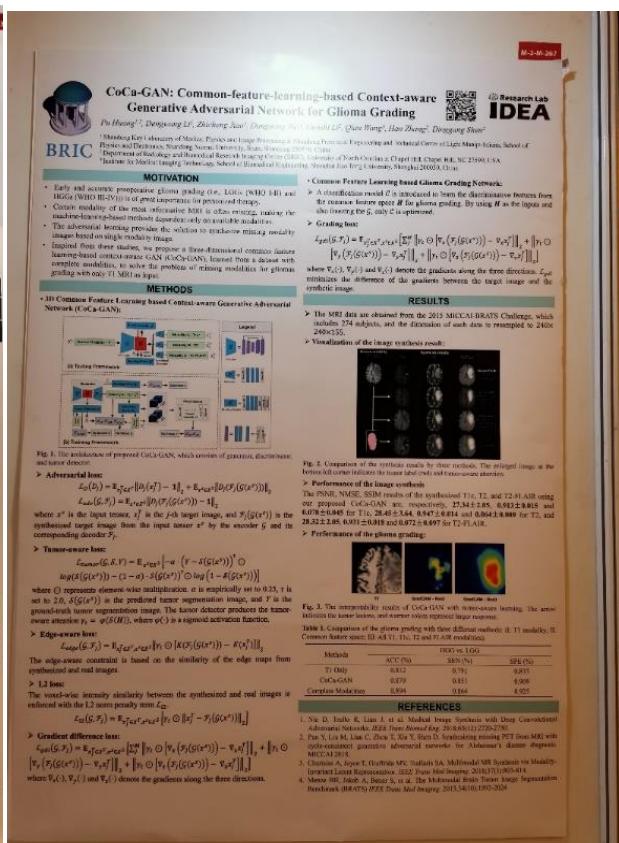
Shenzhen Research Institute of Big Data

GAN + 3D

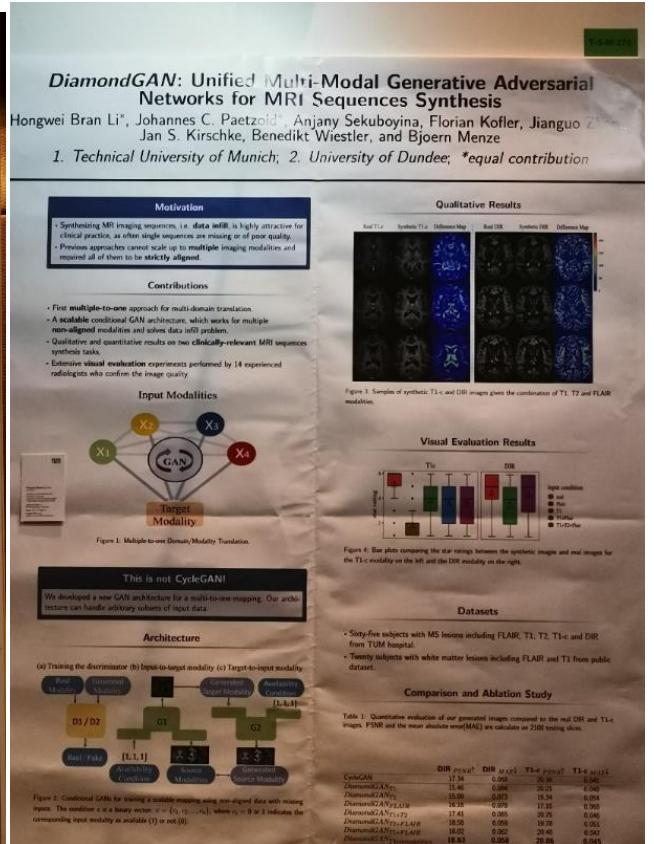
Auto-Encoding GAN

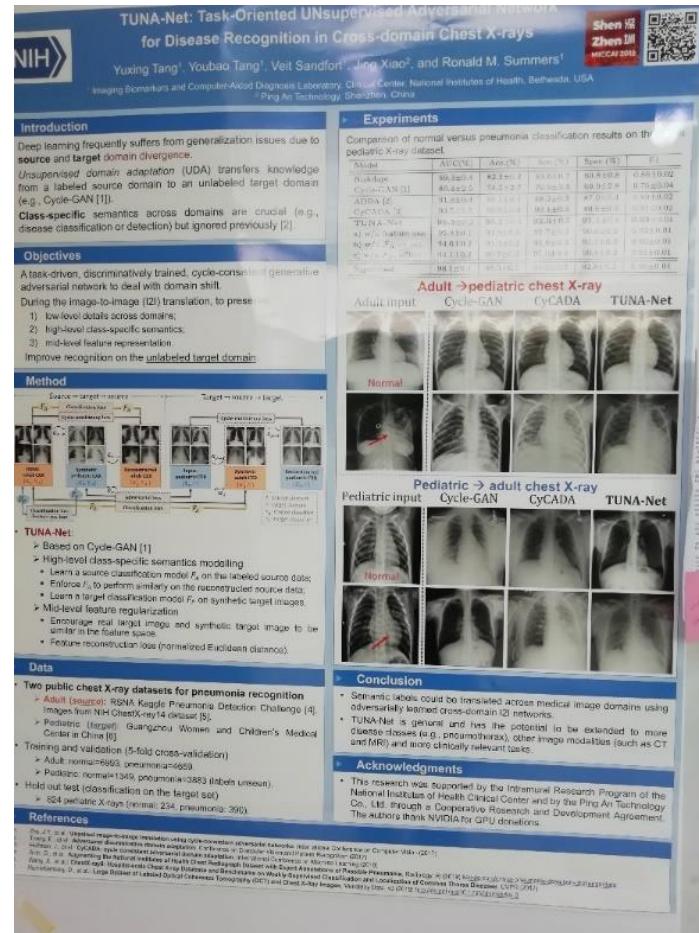
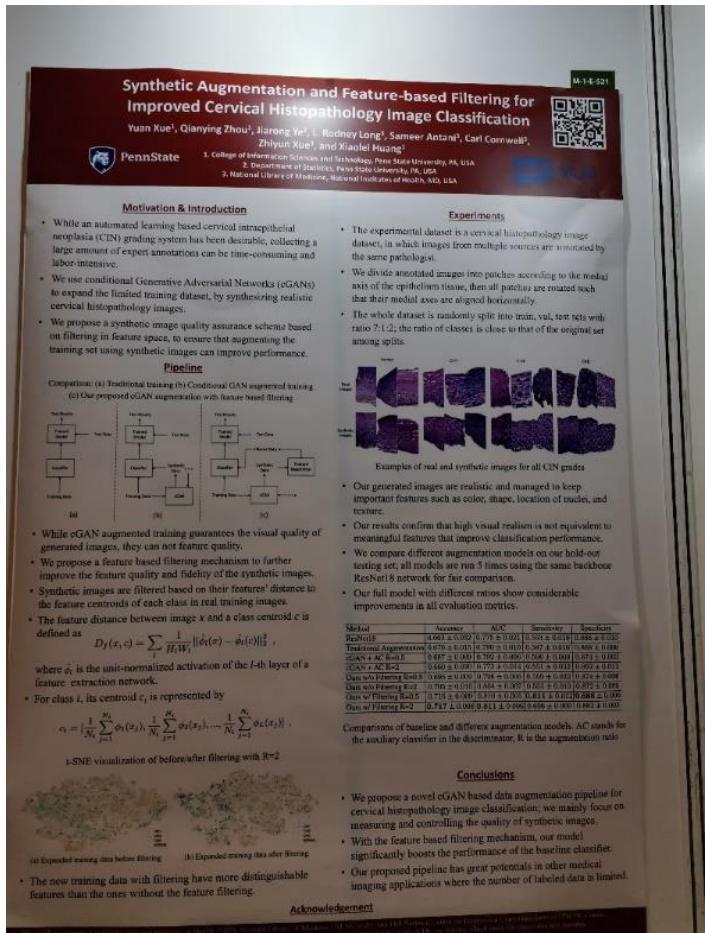


Context-aware GAN

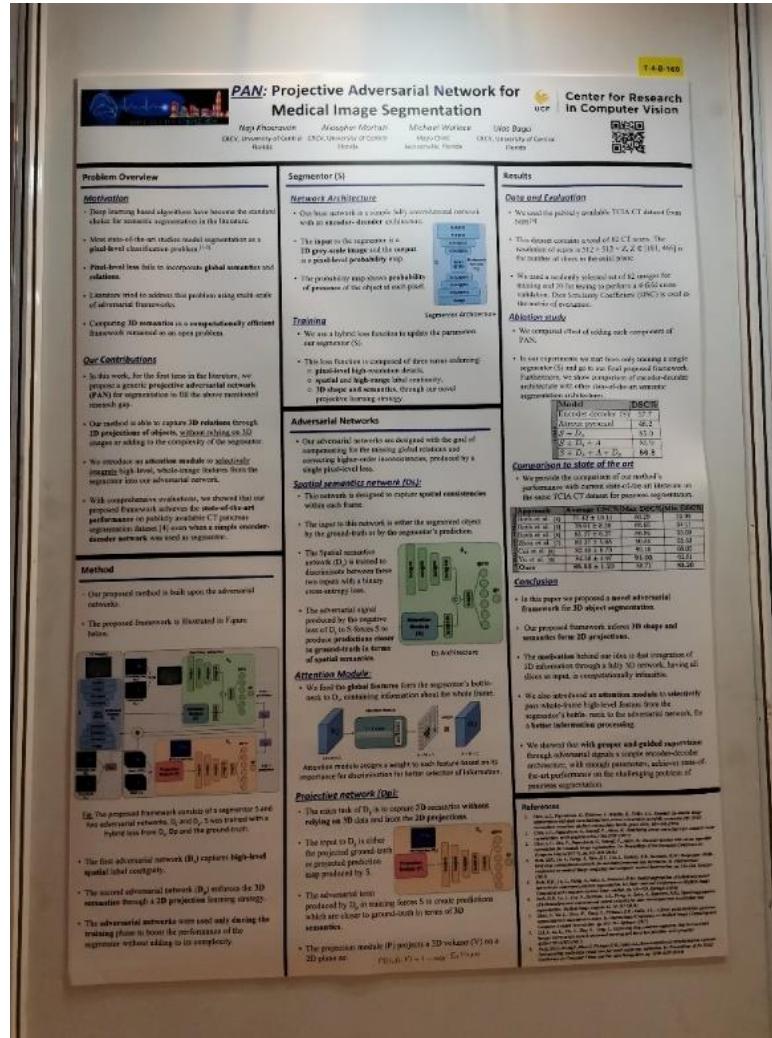


Multi-modal GAN





PAN: Projective Adversarial Network



CGAN

CISTIB Center for Computational Imaging & Simulation Technologies in Biomedicine
 The University Of Sheffield
 UNIVERSITY OF LEEDS

Missing Slice Imputation in Population CMR Imaging via Conditional Generative Adversarial Nets

Le Zhang¹, Marco Pereanez², Christopher Bowles³, Stefan K. Piechnik², Stefan Neubauer³, Steffen E. Petersen⁴, Alejandro F. Frangi²

¹Center for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB), University of Sheffield, UK
²Center for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB), University of Leeds, UK
³Oxford Center for Clinical Magnetic Resonance Research (OCMR), Division of Cardiovascular Medicine, University of Oxford, John Radcliffe Hospital, UK
⁴Cardiovascular Medicine at the William Harvey Research Institute, Queen Mary University of London and Barts Heart Center, Barts Health NHS Trust, UK

Introduction
 Accurate ventricular volume measurements depend on complete heart coverage in cardiac magnetic resonance (CMR) from where most diagnostic indicators of normal/bnormal cardiac function are available non-invasively. However, incomplete coverage, especially missing basal or apical slices in CMR sequences is insufficiently addressed in population imaging and current clinical research studies yet has important impact on volume calculation accuracy. In this work, we propose a new deep architecture, called Missing Slice Imputation Generative Adversarial Network (MSGAN), to learn key features of cardiac short axis (SAX) slices across different positions, and use them as conditional variables to effectively infer missing slices in the query volumes.

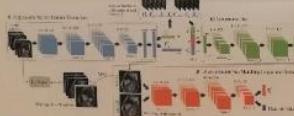
Methods


Fig. 1. Structure of the proposed MSGAN network for cardiac missing slice imputation.

The regressor net D maps each slice of the input volume to a vector containing intensity and position features. The central point feature of each position cluster in the whole training set is learned and used for generator G . Correspondingly the intensity feature and random noise to the inferred position cluster center feature, the new latent vector PC , is fed to G .

$$L_d = D_{PC}(f_{PC}(T_{PC}(D(x))) - \sum_{i=1}^n \|f_i - D(f_i)\|^2)$$

$$L_g = D_{PC}(f_{PC} - D(f_{PC}, x, f_{PC}'))$$

Both D and G are updated based on the L1 loss between the original and synthetic volumes.

$$\lambda_{GAN} = 100 \cdot \text{loss}(G)$$

The discriminative net D forces the output slice to be realistic and plausible for a given position label. Rather than maximizing the output of the discriminative net generated data, the objective of feature matching [6] is employed to optimize G to match the statistics of features in an intermediate layer of D . The objective function is defined in the following equation:

$$\lambda_{FM} = \min_{\theta_D} \max_{\theta_G} \mathbb{E}_{x \sim p_{data}} [\log(1 + D(f_{PC}(T_{PC}(D(x)))))] + \sum_{i=1}^n \mathbb{E}_{f_i \sim p_{data}} [\log(D(f_i))]$$

$$+ \mathbb{E}_{f_i \sim p_{data}, f_{PC}' \sim p_{GAN}} [\log(1 - D(f_i, f_{PC}'))]$$

From Fig. 2, we can observe that the synthetic slices produced by MSGAN have better image quality against the Mean, GMM and SGAN methods. Table 2 shows that MBS reduces ED and ES volumes by an average of 12% and 20%, respectively.

Conclusions
 In this paper, we proposed a novel deep MSGAN to implement missing slice generation in cardiac cine MRI and contribute to the missing data imputation problem neglected by the medical imaging community. Using these synthetic slice data could further augment the training samples for improvement of the segmentation module, which will be our future work.

REFERENCES

[1] Alomar, S., Alomar, T., Blasius, V., Cevher, S.: CGANs for Synthetic Data Generation in Medical Ultrasound. *bioRxiv* 2017, 1–20 (2017)

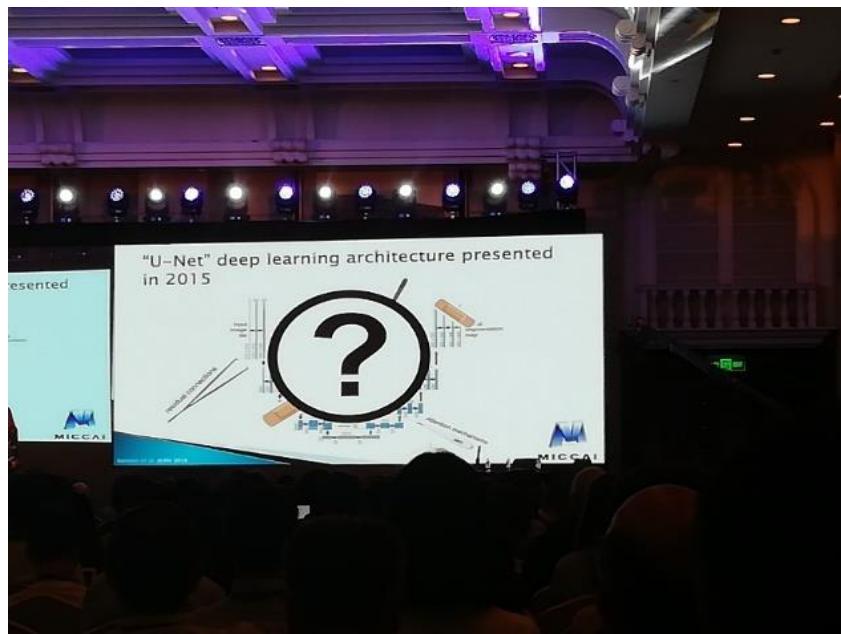


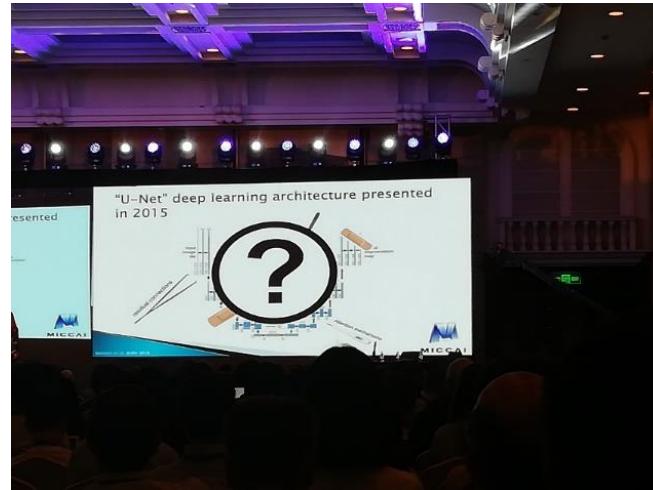
深圳市大数据研究院
 Shenzhen Research Institute of Big Data

Segmentation

高频词：Semi-supervised, unsupervised, attention, cascaded, multi-task

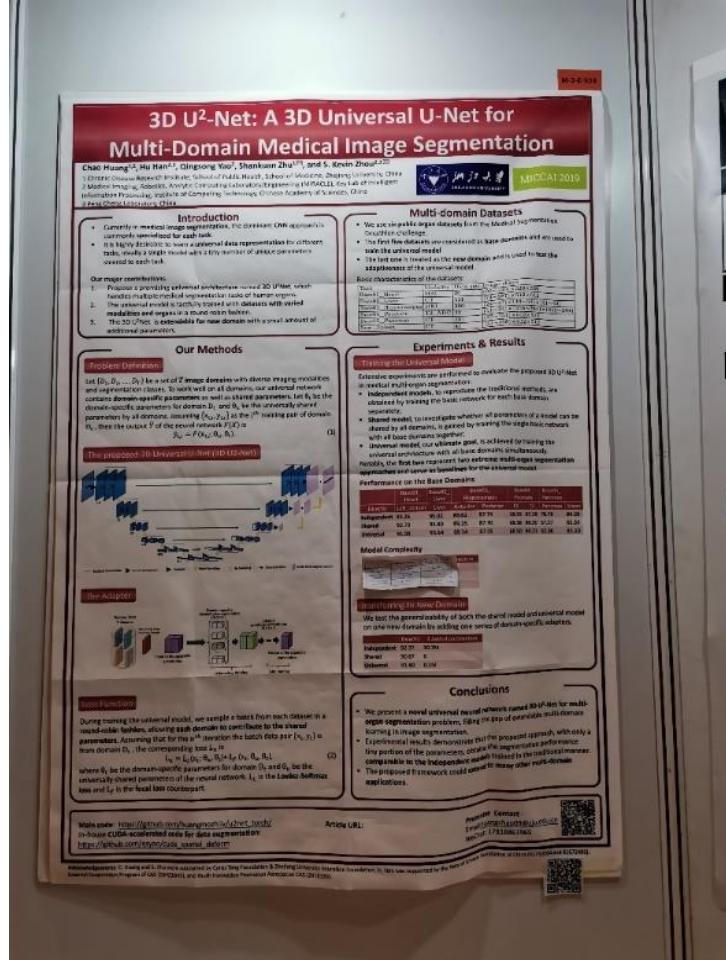
unet结构泛滥，然后各种模块改进。其中nnUnet因为广泛适用，效果好，代码开源，因而知名度比较高。





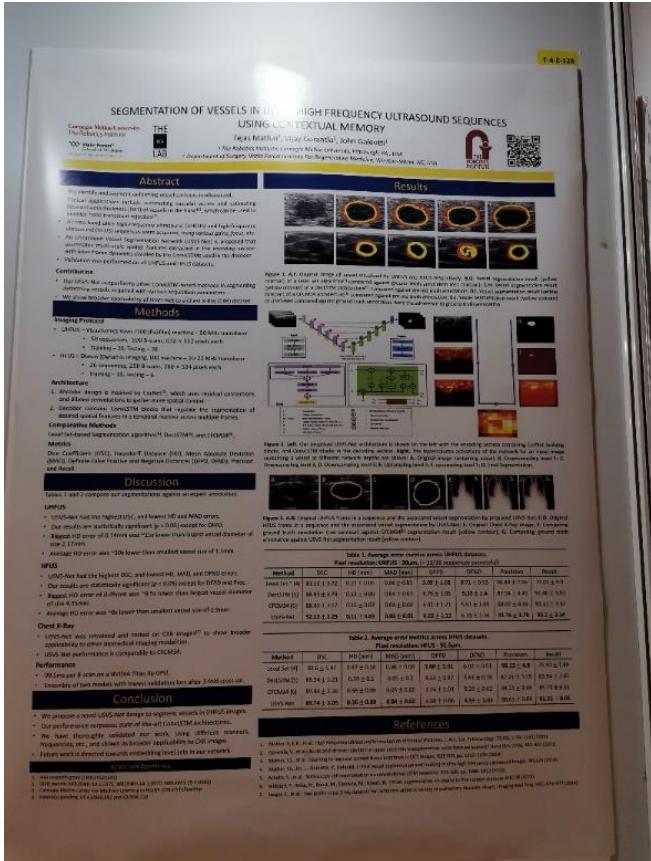
U-net

3D U²-Net: A 3D Universal U-Net for Multi-Domain Medical Image Segmentation

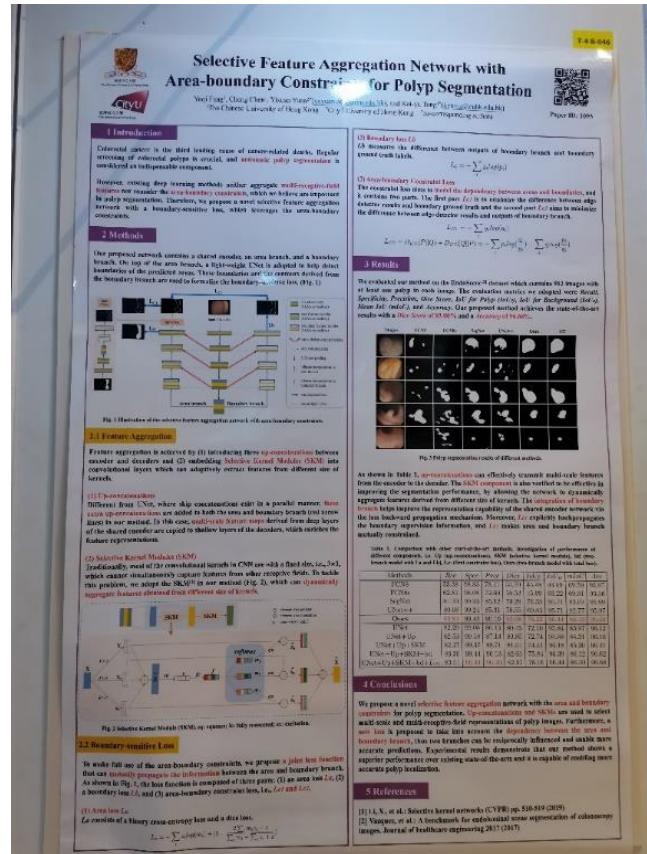


深圳市大数据研究院
Shenzhen Research Institute of Big Data

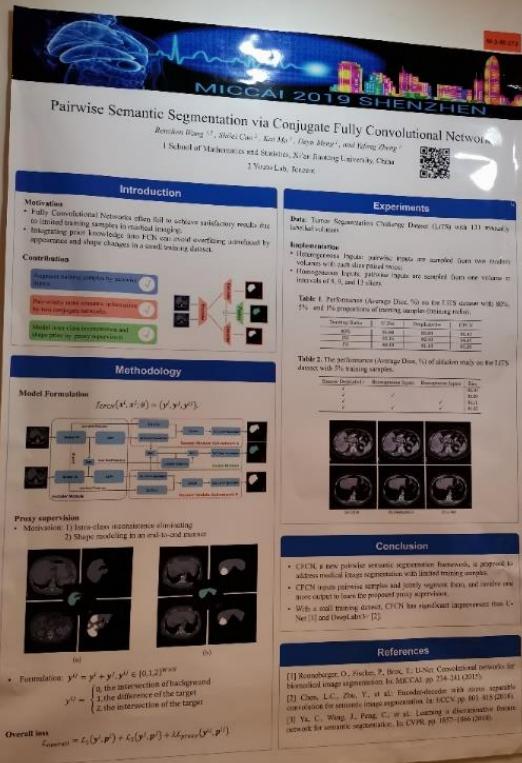
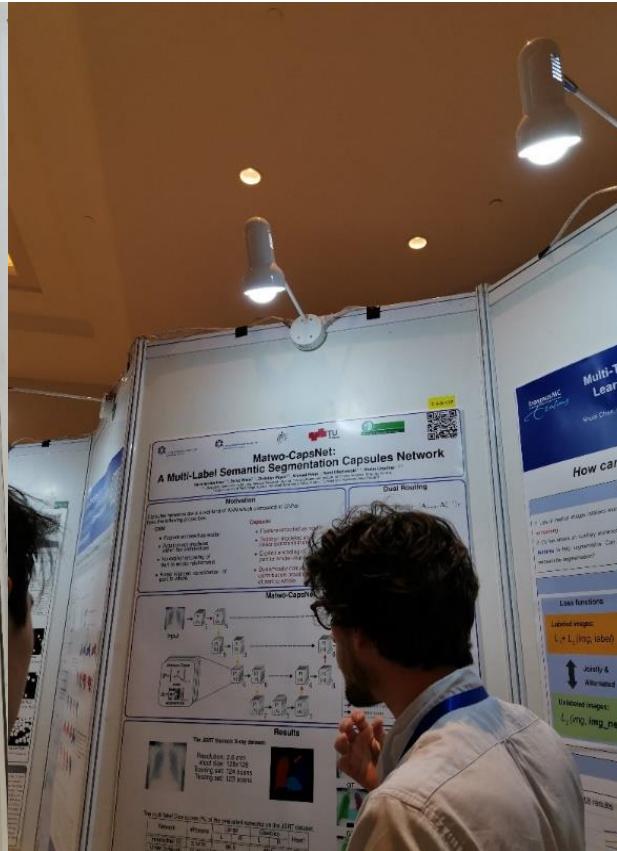
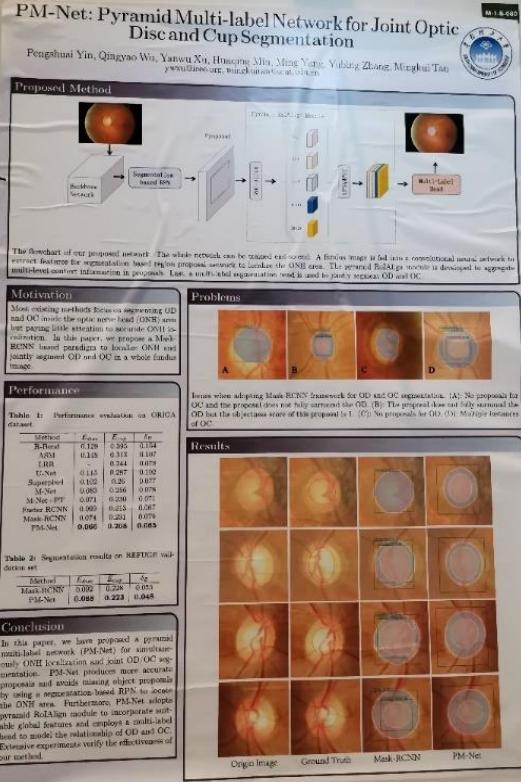
Context Memory _Segmentation_Ultrasound

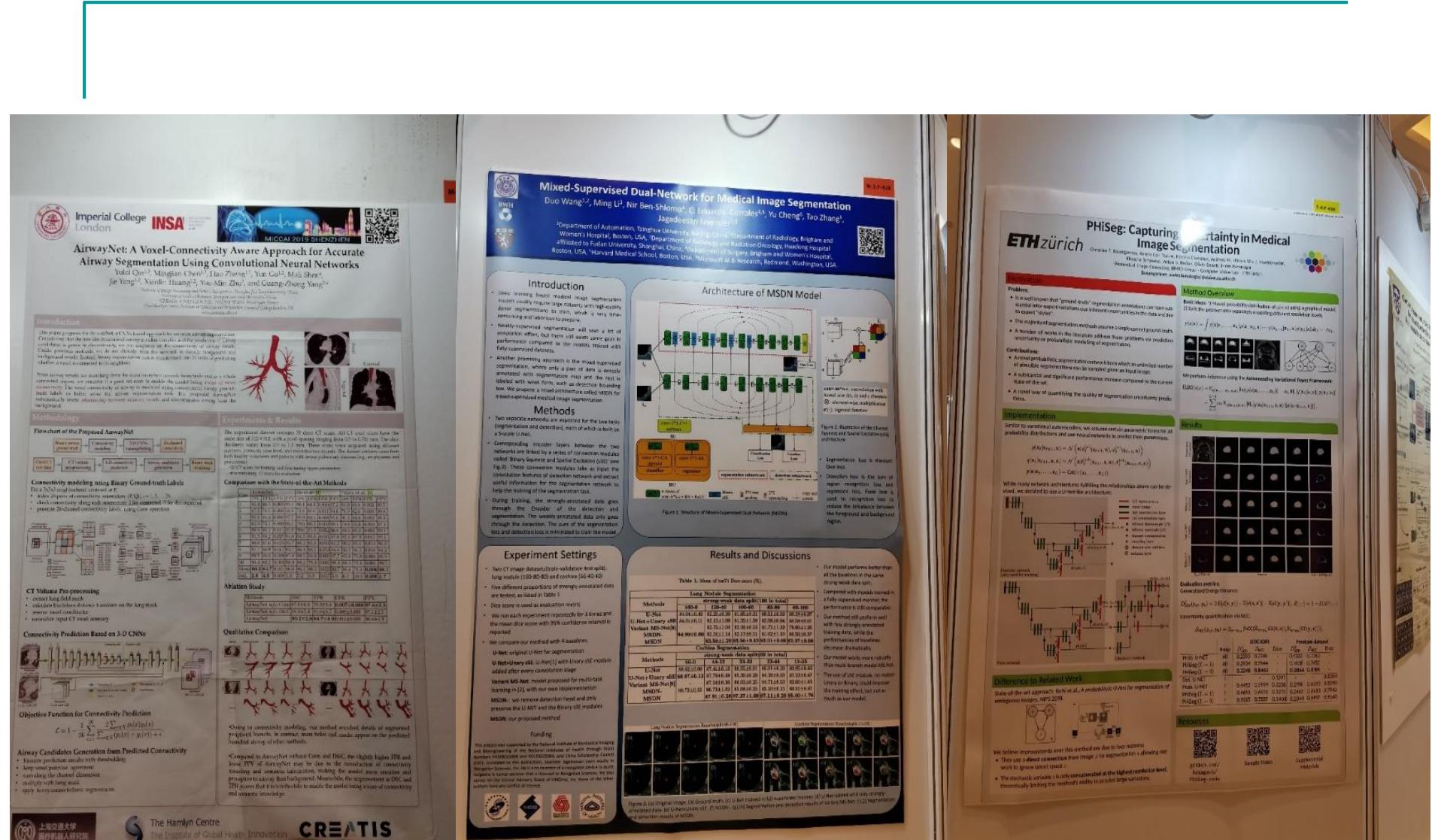


Area Bound_Segmentation



深圳市大数据研究院
Shenzhen Research Institute of Big Data





Segmentation + 3D

3D Tiled Convolution for Effective Segmentation of Volumetric Medical Images

Guodong Zeng^{1,2,3}, Guoyan Zheng^{1,2}

¹ School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China.

² Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China.

³ ARTORG Center for Biomedical Engineering, University of Bern, Switzerland.

INTRODUCTION & AIM

- Convolutional Neural Networks (CNNs) have achieved state-of-the-art performance in many different 2D medical image analysis tasks. But in clinical practice, lots of the medical imaging data available are in 3D.
- 3D CNNs benefit from more spatial context when compared with 2D CNNs, but most state-of-the-art 3D methods depend on patch processing due to GPU memory restrictions.
- The size of the input patch is usually small if no specialised hardware with large GPU memory is used, limiting the incorporation of larger context information for a better performance.
- In this paper, we proposed a 3D Tiled Convolution (3D-TC) which learns a number of separate kernels within the same layer. It can significantly reduce the required GPU memory for 3D medical image processing task, but also with improved performance.

METHOD & EXPERIMENT

Figure 1. A schematic view of 3D tiled convolutions which are implemented as a periodic down-shuffling operation with low-resolution convolutions. Here the down-shuffling factor is $(2, 2, 2)$.

Figure 2. A schematic view of how to augment existing FCNs with 3D-TC and DUC for volumetric image analysis.

Definition of 3D-TC and DUC

- 3D-TC is implemented with a periodic down-shuffling operation followed by conventional 3D convolutions.
- Dense Upampling Convolution (DUC) consists of low-resolution convolutions with a periodic up-shuffling operation to generate high-resolution results.

Augment Existing FCNs with 3D-TC and DUC

- Both DUC and 3D-TC are network agnostic and can be combined with existing FCNs.
- The advantage of such a pipeline is apparent. When a 3D-TC with down-shuffling factors of (n_1, n_2, n_3) is applied to the input data, both the computational and the storage cost for the underlying networks will be reduced by a factor of $(n_1 \times n_2 \times n_3)$, allowing one to use larger patches as the input without increasing computational time. The full resolution result is then obtained at the final output by applying a DUC with up-shuffling factors of (n_1, n_2, n_3) .

EXPERIMENTS & RESULTS

Datasets and Preprocessing

- Two datasets, i.e., an in-house dataset consisting of 25 T1 hip MRI images with limited field of view and a publicly available dataset from National Institute of Health (NIH) containing 92 abdominal contrast enhanced 3D CT scans.
- Standard 5-fold and 4-fold cross-validation study were performed on Hip and Pancreas dataset, respectively.

Table 1. Results of investigation of different patch sizes on the performance of the original 3D-U-Net. The acronyms Fmean, F1, Sensitivity, Specificity, and Dice are defined in Table 2.

Method	3D-U-Net	3D-U-Net (30, 30)	3D-U-Net (60, 60)	3D-U-Net (120, 120)	3D-U-Net (240, 240)
Fmean	0.85	0.85	0.85	0.85	0.85
F1	0.85	0.85	0.85	0.85	0.85
Sensitivity	0.85	0.85	0.85	0.85	0.85
Specificity	0.85	0.85	0.85	0.85	0.85
Dice	0.85	0.85	0.85	0.85	0.85

Table 2. Results when different shuffling factors were used for the 3D LP-U-Net. The size of the input patch is fixed to $400 \times 400 \times 80$.

Method	3D LP-U-Net	3D LP-U-Net (30, 30)	3D LP-U-Net (60, 60)	3D LP-U-Net (120, 120)	3D LP-U-Net (240, 240)
Fmean	0.85	0.85	0.85	0.85	0.85
F1	0.85	0.85	0.85	0.85	0.85
Sensitivity	0.85	0.85	0.85	0.85	0.85
Specificity	0.85	0.85	0.85	0.85	0.85
Dice	0.85	0.85	0.85	0.85	0.85

Table 3. Segmentation accuracy of 3D LP-U-Net and state-of-the-art methods.

Method	3D LP-U-Net	3D-U-Net	3D-U-Net (30, 30)	3D-U-Net (60, 60)	3D-U-Net (120, 120)	3D-U-Net (240, 240)
Fmean	0.85	0.85	0.85	0.85	0.85	0.85
F1	0.85	0.85	0.85	0.85	0.85	0.85
Sensitivity	0.85	0.85	0.85	0.85	0.85	0.85
Specificity	0.85	0.85	0.85	0.85	0.85	0.85
Dice	0.85	0.85	0.85	0.85	0.85	0.85

CONCLUSION

We proposed a simple yet effective 3D tiled convolution for 3D medical image analysis tasks. The 3D-TC consists of a periodic down-shuffling operation followed by low-resolution convolutions. It can be directly applied to the input data and has the advantage of significantly reducing GPU memory. Experimental results demonstrated the effectiveness of our framework on different semantic segmentation tasks.

Table 5. Accuracy (Dice, %) comparison between 3D LP-U-Net and the state-of-the-art on the MIL pancreas segmentation dataset.

Approach	Accuracy (%)
Zhou et al. (2015) [1]	87.71 ± 10.11 ± 25.95
Zhou et al. (2017) [1]	89.71 ± 8.68 ± 90.85
Gorelick et al. (2017) [2]	90.00 ± 1.00 ± 90.00
Wang et al. (2018) [3]	91.47 ± 2.77 ± 88.96 ± 10.00
Yu et al. (2018) [4]	84.20 ± 4.27 ± 91.02 ± 8.81
Ours approach	88.00 ± 3.84 ± 90.81 ± 8.81

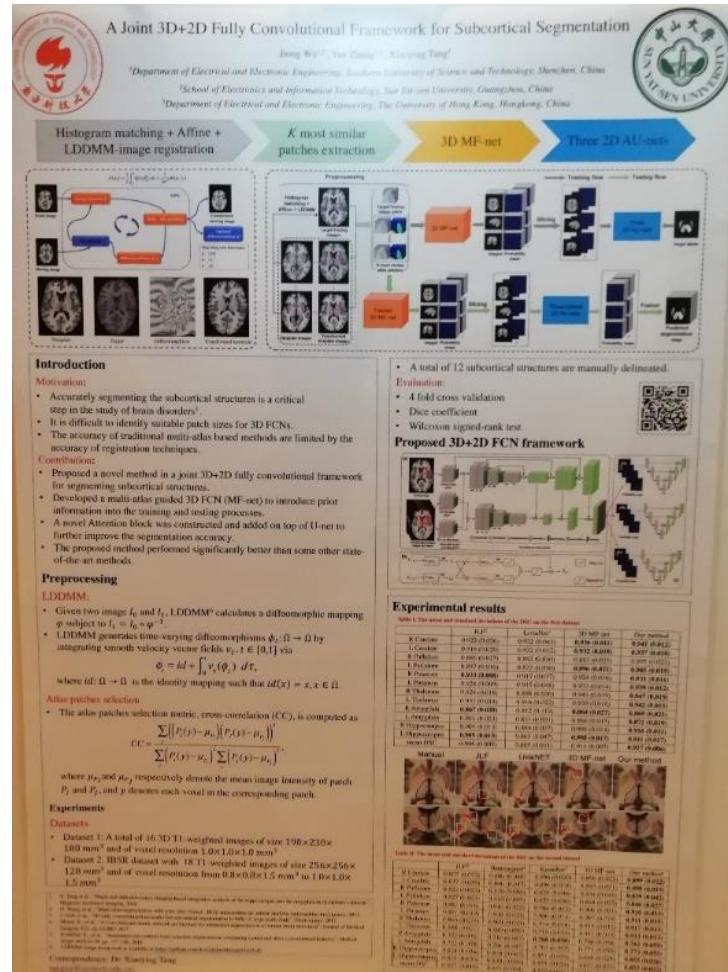
Fig. 1: Qualitative comparison of the segmentation results of 3D LP-U-Net with different shuffling factors and the 3D U-net with the largest patch size.



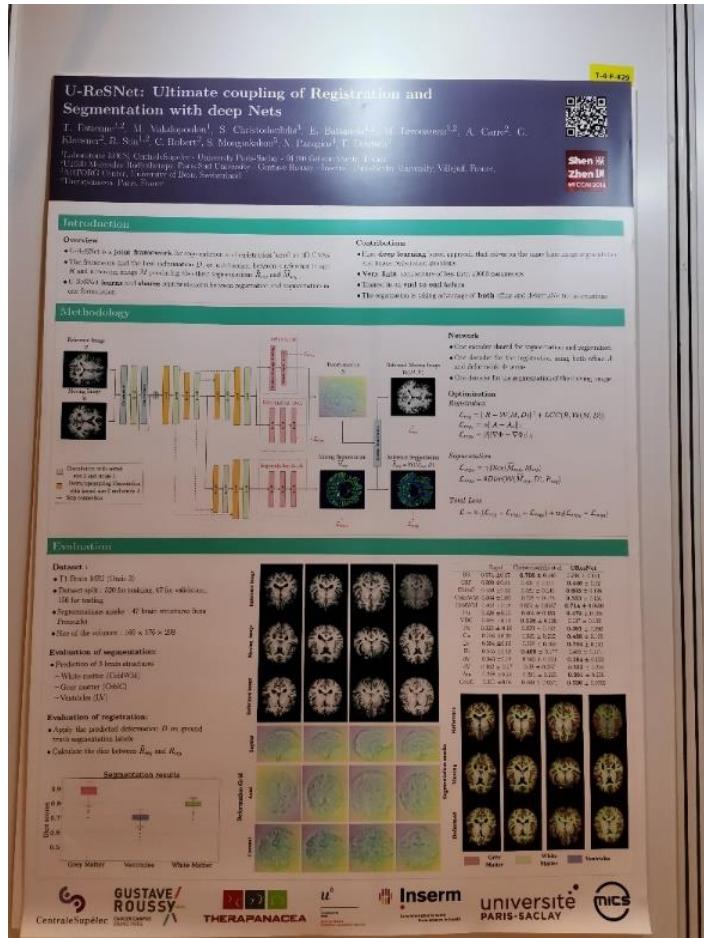
深圳市大数据研究院

Shenzhen Research Institute of Big Data

Segmentation 3D + 2D



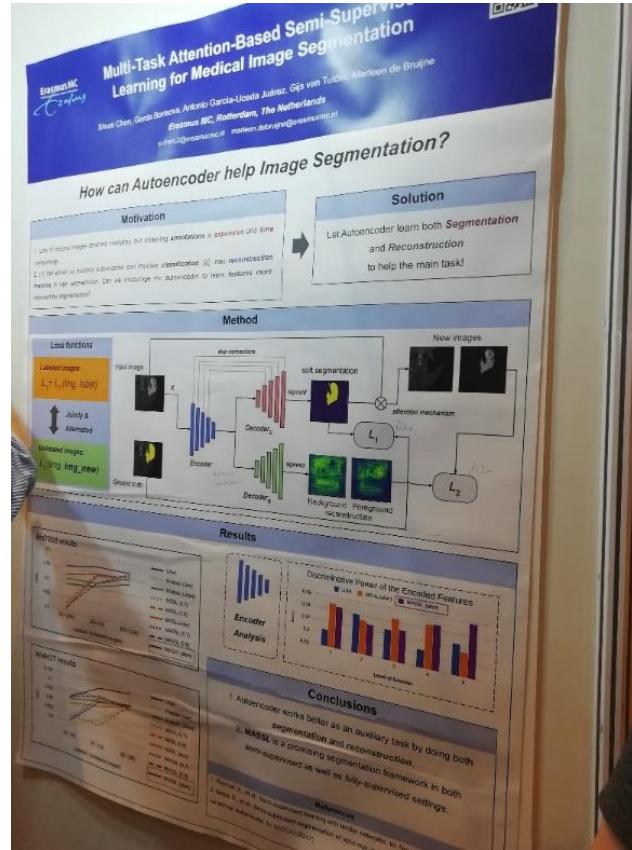
Segmentation and Registration



- 高频词：segmentation & registration
许多配准的文章都在把分割和配准协同起来做。
- 配准 8%
- winter for image registration
- 配准tutorial
 - <https://learn2reg.github.io/>
 - <http://www.cs.ucf.edu/~bagci/teaching/mic16/lec17.pdf>
 - <https://github.com/natandrade/Tutorial-Medical-Image-Registration>



Segmentation & Reconstruction



Denoising

Consensus Neural Network for Few-shot Imaging Denoising with Only Noisy Training Samples

Defan Wu^{1,2}, Kuang-Guo^{1,2}, Kyung-Sang Kim^{3,4}, Xiang Li^{1,2}, and Qunzheng Li²

¹Center for Advanced Medical Computing and Analytics, ²Garrison Center for Medical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
dewu@mit.edu, liqunzheng@mit.edu

INTRODUCTION

Deep learning based image denoising and restoration demonstrated significantly improved performance compared to conventional methods. However, most methods assume that we have several datasets.

- Acquiring high-quality training data is often hard to do for this modality, which is a complicated task.
- For many applications, high-quality training images are difficult to acquire, e.g., ultrasound images, medical denoising to deal with CT scans, CT, MRI, etc.

The proposed consensus neural network aims to learn medical image denoising without high-quality reference images.

BACKGROUND: Noise2Noise

- Conventional training: map noisy images to clean images.
- Noise2Noise training: map noisy images to another noisy modality.

Mathematically, conventional Noise2Net training is

$$w_{\text{out}} = \arg \min F_1/(x_i + n_i, 0) - (x_i + n_i, 0)$$

And Noise2Noise training is

$$w_{\text{out}} = \arg \min F_2/(x_i + n_i, 0) - (x_i + n_i, 0)$$

CONSENSUS LOSS

Motivation: Having two noisy environments $x_i + n_i, x_j + n_j$, we would like to increase the RMSE value averaging

$$\text{RMSE} = \sqrt{\frac{(f(x_i + n_i, 0) - f(x_j + n_j, 0))^2}{2}}$$

The equivalent consensus loss function

$$R_{\text{cons}} = \arg \min F_3 \frac{1}{2} \|f(x_i + n_i, 0) - (x_i + n_i)\|^2 + \frac{1}{2} \|f(x_j + n_j, 0) - (x_j + n_j)\|^2 - \frac{1}{2} \|f(x_i + n_i, 0) - f(x_j + n_j, 0)\|^2$$

Testing time output:

$$x = \frac{f(x_i + n_i, 0) + f(x_j + n_j, 0)}{2}$$

Additional implementation:

- Weight decay: λ_1 .
- Loss consistency: $F_1[\hat{x}_i - x^{true}]$, where x^{true} is a noisy estimation, such as FBP, SENSE, etc.
- The additional regularizations are helpful in further preventing over-smoothing and artifacts.

CONCLUSION

The proposed consensus network can provide promising denoising results. Despite it needs access to two different training datasets, some techniques are required instead of complete. We have already successfully applied to medical.

(1) Dual energy CT; (2) 4D Perfusion CT; (3) We are exploring stage-wise methods now.

Improving Robustness of Medical Image Diagnosis with Denoising Convolutional Neural Networks

Fei-Fei Xu¹, Jin Peng¹, RuiXuan Wang¹, Qiong Zhang¹, Wei-Shi Zheng¹

¹School of Data and Computer Science, Sun Yat-sen University, China

Introduction

Adversarial noises critically fool the CNN classifiers. This work tries to improve the robustness of medical image classification systems by bringing denoising ability to CNN classifiers with a naturally embedded auto-encoder and high-level feature invariance to general noises.

Methodology

Clean Image x_i → High-level features → Output $f(x_i)$ → Cross-entropy Loss

Noisy Image x'_i → Duplicate → $f(x'_i)$ → Cross-entropy Loss

Neighbor Loss: $L_n = \frac{1}{N} \sum_{i=1}^N \|f(x_i) - f(x'_i)\|$

Neighborhood loss emphasizes similarity between images and clean images in semantic feature space.

$L = L_c + \lambda_n L_n + \lambda_a L_a$ λ_n and λ_a are hyper-parameters

Reconstructed images \hat{x}_i, \hat{x}'_i → AE Loss

Encoder → Decoder → Output

$L_a = \frac{1}{N} \sum_{i=1}^N (\|x_i - \hat{x}_i\|_2 + \|x_i - \hat{x}'_i\|_2)$

AE loss minimizes the distances between raw images with both reconstructed clean and noisy images.

Experiments & Conclusion

It performs better when combined with existing methods

Defense	FGSM	IFGSM	C&W			
λ_n	λ_a	$\epsilon = 4$	$\epsilon = 8$	$\epsilon = 4$	$\epsilon = 8$	C&W
N/A	N/A	19.27	21.99	3.78	3.78	1.54
1	0	41.66	29.55	49.41	31.8	18.91
0	10	46.34	32.51	56.15	38.65	21.28
1	10	57.92	48.86	65.37	54.37	35.22

Notes: Performance is on SKIN4 test set. Blue model is ResNet18, FGSM, IFGSM and C&W are methods to generate noisy images.

Contributions:

1) A novel mechanism to improve robustness.

Anatomical Priors for Image Segmentation via Post-Processing with Denoising Autoencoders

Agostina Larrazabal, Cesar Martinez, Enzo Ferrante

¹Research Institute for Signals, Systems and Computational Intelligence, sinc²
CONICET, Universidad Nacional de La Plata
Santa Fe, Argentina

Motivation

In general, pixel-level predictors are not designed to account for shape and topological properties.

The lack of contextual information could lead to incorrect predictions in areas with similar intensities.

Qualitative results

Random Forest Post-DAE U-Net 5-ep U-Net 100-ep U-Net 500-ep U-Net 1000-ep

Post-processing with Denoising Autoencoders

Post-DAE is a simple post-processing method which produces anatomically plausible segmentations by improving pixel-level predictions from arbitrary classifiers incorporating shape and topological priors.

Post-DAE: Annotates the class labels in multi-class learning via iterative post-processing. We learn a few directional prior of anatomically plausible segmentations, and use it to impose shape constraints by post-processing anatomical segmentations made obtained with arbitrary methods.

Quantitative results

We evaluated Post-DAE to improve segmentation masks of different qualities produced by various U-Net models (trained with different number of epochs) and a Random Forest model. The performance of Post-DAE was compared with a standard CRF post-processing step [2].

From left to right: Input image, Ground truth, Post-DAE segmentation, Random Forest segmentation.

During training, a Segmentation Function is used to de-annotate the ground truth images after each iteration and analyze noisy images.

The number is plotted to minimize the reconstruction error measured by a loss function based on the Dice coefficient.

Conclusions

DAEs can be used as a simple and effective post-processing step to implement alternative priors than classical segmentation methods.

Post-DAE can be easily implemented, is independent of image dimensions and improves results produced with arbitrary methods.

References

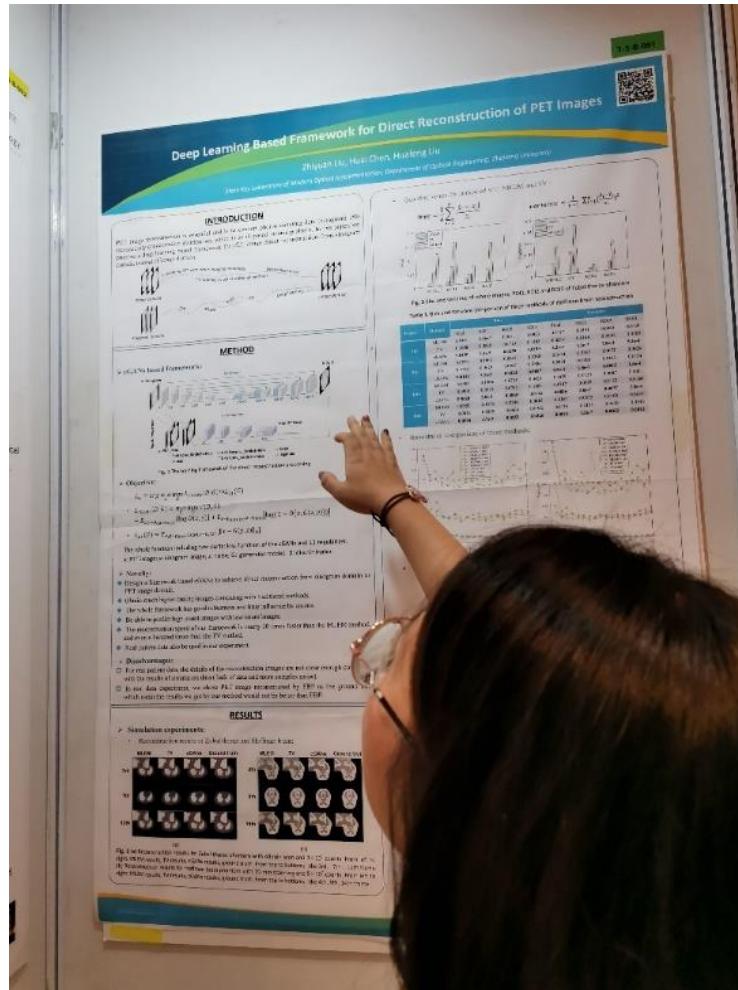
Notes: Performance is on SKIN4 test set. Blue model is ResNet18, FGSM, IFGSM and C&W are methods to generate noisy images.

Acknowledgement. This work is supported in part by Natio Research and Development Plan and by the Guangdong Provincial Key Laboratory of Intelligent Manufacturing and Machine Vision.

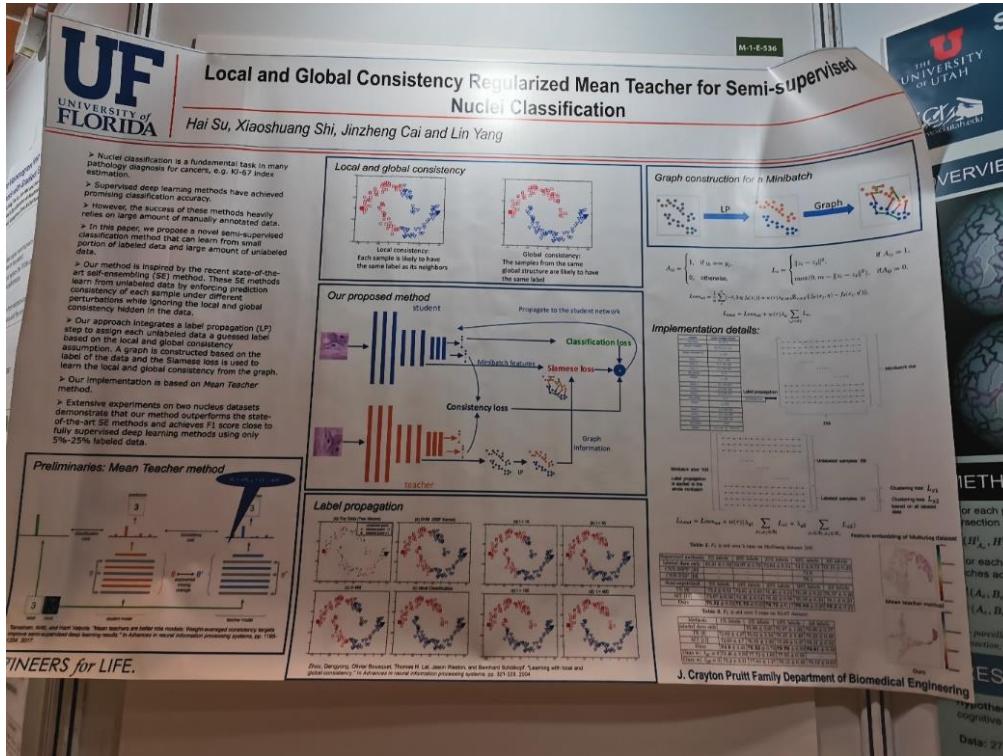
We then aggregated all the results in a single scatter plot showing the improvement made by a fully connected CRF (Krahenbuhl, 2011) and Post-DAE.



Reconstruction



Semi-supervised



MICCAI 2019, 13-17 Oct 2019, Shenzhen, China

Uncertainty-aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation

Lequan Yu¹, Zhiqiang Li¹, Xiomeng Li², Chi-Wing Fu², and Philip-Ann Heng^{1,2}
¹ Dept. of Computer Science and Engineering, Chinese University of Hong Kong
² STone Robotics Institute, The Chinese University of Hong Kong

Introduction

Background:
 Automated segmentation of left atrium (LA) in magnetic resonance (MR) images is important. It is expensive and tedious to delineate reliable annotations from 3D medical images. Unlabeled data are generally abundant.

Contributions:
 A novel semi-supervised framework for LA segmentation is leveraged the unlabeled data. An uncertainty-aware scheme to enable the student model to gradually learn from reliable teacher targets. Achieve 4.15% Jaccard and 2.85% Dice gains by utilizing the unlabeled data and outperform other state-of-the-art semi-supervised methods.

Network Architecture

Uncertainty-Aware Mean Teacher Framework

- Update the teacher's weights θ' as an exponential moving average (EMA) of the student's weights θ following: $\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_t$.
- Estimate the uncertainty with the Monte Carlo Dropout in Bayesian networks^[1].
- Perform 1 stochastic pass on the teacher model under random dropout and input Gaussian noise for each input volume.
- Choose the predictive entropy as the metric to approximate the uncertainty.

$$\mu_i = \frac{1}{T} \sum_t p_i^t \quad U_i = -\sum_t \log p_i^t$$

- Use an uncertainty-aware consistency loss to filter out unreliable predictions and preserve only reliable ones: $\mathcal{L}_c(f', f) = \sum_i I(u_{i0} < H) \frac{\|f_i - f'_i\|^2}{\sum_i I(u_{i0} < H)}$

Experiments

Datasets:
 - Atrial Segmentation Challenge (300 3D gadolinium-enhanced MR scans)
 - 80 scans for training and 20 scans for evaluation

Comparison with baselines and other semi-supervised methods :

Method	# scans used	Labeled	Unlabeled	Jaccard (%)	Dice (%)	Metric
Yuanlin X. Net ^[2]	16	64	24.13	73.58	8.76	17.95
Borodits V. Net ^[3]	16	0	86.03	78.08	9.02	4.29
Yuanlin X. Net ^[4]	16	0	86.03	78.08	9.02	4.29
Our UA-MT ^[5]	16	93	89.14	83.82	1.22	3.76
Self-training ^[6]	26	64	86.95	77.26	2.27	9.19
DAN ^[8]	26	64	87.87	78.46	4.65	9.49
ASNet ^[9]	26	64	87.87	78.46	4.65	9.49
PSNet ^[10]	16	64	88.15	79.30	2.44	9.65
UA-MT-UN (ours)	16	93	86.98	77.37	3.77	10.03
UA-MT ^[5]	16	93	86.21	72.06	7.32	-

Impact of uncertainty-aware module and different labeled/unlabeled ratios:

Method	# scans used	Labeled	Unlabeled	Jaccard (%)	Avg. Dice (%)	Metric
MT-Drop	16	64	68.23	79.29	7.76	10.50
Our UA-MT	16	64	88.88	80.51	2.58	7.22
Bojanian V. Net ^[11]	0	0	79.93	78.12	5.48	21.11
Our UA-MT	8	72	84.25	73.48	3.96	18.83
Bojanian V. Net ^[11]	16	64	88.32	79.00	3.05	19.45
Our UA-MT	21	58	90.18	82.18	2.73	8.98

Qualitative Results:

(a) Image (b) Supervised (c) Our method (d) Uncertainty

References

- [1] Bai et al., Semi-supervised learning for network-based cardiac mr image segmentation. MICCAI 2017.
- [2] Kondali et al., What uncertainties do we need in Bayesian deep learning for computer vision? NeurIPS 2018.
- [3] Li et al., Semi-supervised brain lesion segmentation with an adapted mean teacher model. IPMA 2019.
- [4] Li et al., Semi-supervised skin lesion segmentation via Transformation Consistent Self-ensembling. BMVA 2018.
- [5] Li et al., ASNet: Attention based semi-supervised deep networks for medical image segmentation. MICCAI 2018.
- [6] Zhang et al., Deep adversarial networks for biomedical image segmentation utilizing unannotated images. MICCAI 2017.



深圳市大数据研究院
 Shenzhen Research Institute of Big Data

Weakly supervised

Biomarker Localization by Combining CNN Classifier and Generative Adversarial Network

Rong Zhang¹, Shuxian Tang², Ruixuan Wang³, Jingjing Chen³, Hanjian Liu³ and Wei - Shi Zheng¹
¹Sun Yat-Sen University, ²University of Jinan, ³Zhengzhou Ophthalmic Center of Sun Yat-Sen University

Introduction

Objective: precisely localize biomarkers with image-level annotations.
 Solution: a novel architecture by combining a CNN classifier and GAN.

Methods

The proposed architecture combines a CNN classifier (C), a discriminator (D) and a generator (G). The discriminator helps the Generator output normal images and the classifier helps the encoder-decoder only change biomarker regions to generate normal outputs.

Experiments & Conclusion

G & C: G & C localized only part of biomarkers.
G & D: G & D localized most biomarker regions but altered some normal regions.
G & C & D: G & C & D precisely localized most biomarkers.

Compared with CAM (3rd column) and Grad-CAM (4th column), our approach (6th column) gave much more precise localization of biomarkers with irregular shapes and scattered distributions.

Conclusion:

- 1) A novel framework and training strategy.
- 2) Weakly supervision with image-level annotation.

Email: Rong Zhang(zhangr25@mail2.sysu.edu.cn); Ruixuan Wang (wangruix5@mail.sysu.edu.cn)

Acknowledgement : This work is supported in part by the Guangdong Key Research and Development Program, and the National Natural Science Foundation of China.

WEAKLY SUPERVISED SEGMENTATION FRAMEWORK WITH UNCERTAINTY: A STUDY ON PNEUMOTHORAX SEGMENTATION IN CHEST X-RAY

Xi Chongyang^{1,2}, Zhong Xue¹, Yiqiang Zhang¹, Xiang Sean Zhou¹, Qingfeng Wang¹, Ying Zhou¹, Qian Wang¹ and Tie-Zhi Cheng¹
¹United Imaging Intelligence, ²Shanghai United Imaging Information Co., Ltd, Shanghai Jiao Tong University, Shanghai, China
³School of Biomedical Engineering, Southwest Jiaotong University, Chengdu, China
⁴School of Computer Sci and Eng, Southwest University of Sci and Tech, Mianyang, China
⁵Radiology Department, Mianyang Central Hospital, Mianyang, China

MOTIVATION & OBJECTIVES

- Pneumothorax (PTX) is a lung abnormality with air leaking into the space between the lung and chest wall [1]. Large PTX can be fatal and needs immediate treatment.
- Quantification of large PTX requires segmentation of PTX regions. However, the pixel-level annotations are difficult to obtain and can be very limited and costly.
- We propose a semi-supervised segmentation method that leverages pixel-level (well-annotated and image-level weak) annotations to address the issue of less availability of pixel-level annotations.
- The shape, size, location of PTX may vary a lot and segmentation may also be affected by anatomical variations of patients.

RESULTS

- Then, the cross-entropy loss with SLSR is:

$$L_{SLSR} = \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{\lambda_i} \log(p_{ij}) + \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{\lambda_i} \log(p_{ij}^*) + \sum_{i=1}^n \sum_{j=1}^{m_i} \lambda_i p_{ij}^* \delta(p_{ij}, p_{ij}^*)$$
- For well-annotated image samples, we set $\lambda = 0$, otherwise $\lambda = 1$ for weakly-annotated image samples.

METHOD

Figure 1. Training flowchart of our weakly-supervised segmentation method.

Method Overview

- First, an image-level classification (PTX or not) model is trained for the attention masks of weakly-annotated data.
- Second, semi-supervised image segmentation model is developed to leverage well- and weakly-annotated data, where the corresponding attention masks are used. Since the attention masks do not delineate the exact pneumothorax regions and may have some errors, we employ the Spatial Label Smoothing Regularization (SLSR) technique to reserve the uncertainty for the incorrectness of the attention masks in the training phase.
- Classification Network for Attention Mask Generation
- Binary classification (PTX or not), with Resnet(10).
- The attention mining method, Guided Attention Inference Network (GAIN) [2], is performed to generate the attention masks. The sizes of the obtained attention masks are 1/8 of the input sizes.
- Mask Uncertainty with SLSR
- Since attention masks only capture rough PTX regions, we explore the uncertainty of attention masks by numerically perturbing one-hot label distribution into the probabilistic domain as shown in Figure 2.

Figure 2. Perturbing one-hot label into the probabilistic domain.

Figure 3. Visualization results.

Table 1. Comparison of different settings.

	Only image-level annotations	Small	Medium	Large	Test
only image-level annotations	0.051	0.124	0.231	0.242	
100 well-annotated cases	0.129	0.443	0.533	0.577	
+ 200 weakly-annotated cases	0.129	0.443	0.533	0.577	
+ 400 weakly-annotated cases	0.182	0.492	0.661	0.520	
+ 800 weakly-annotated cases	0.136	0.456	0.647	0.503	
+ 1600 weakly-annotated cases	0.108	0.488	0.592	0.545 [*]	
+ 2000 weakly-annotated cases	0.220	0.658	0.788	0.612 [*]	
+ 4000 weakly-annotated cases	0.287	0.527	0.794	0.590	
+ 8000 weakly-annotated cases	0.203	0.480	0.704	0.558 [*]	
3000 well-annotated cases	0.224	0.551	0.768	0.611 [*]	
+ 2000 weakly-annotated cases	0.265	0.583	0.774	0.637 [*]	
+ 4000 weakly-annotated cases	0.265	0.583	0.774	0.627	
+ 8000 weakly-annotated cases	0.265	0.583	0.774	0.614	
100 well-annotated cases [*]	0.252	0.583	0.777	0.645 [*]	
+ 200 weakly-annotated cases	0.252	0.583	0.777	0.655	
+ 400 weakly-annotated cases	0.276	0.616	0.794	0.669	
+ 800 weakly-annotated cases	0.252	0.583	0.774	0.648	

Table 1. Comparison of different settings.

[1] Zarogoulidis, P., Klimnis, I., Petrou, G., Papageorgiou, A., Kastogianni, N., Zarik, B., Brancik, P., Sezen, N., et al. Pneumothorax: from definition to diagnosis and treatment. *Journal of Thoracic Disease* 6(Sept 4), S712 (2014).

[2] Li, K., Wu, Z., Peng, K.C., Ernst, J., Fu, Z.: Tell me where to look: Guided attention inference network. *arXiv preprint arXiv:1802.01713* (2018).

[3] Jégou, S., Doumal, M., Vazquez, D., Romero, A., Bengio, Y.: The one-dimensional convolutional layer. *CoRR* abs/1609.04836 (2016).

REFERENCES



Weakly Supervised Segmentation

PseudoEdgeNet:
Nuclei Segmentation on Cells with Point Annotations
Inwan Yoo, Donggeun Yoo, Kyunghyun Paeng, (Lunit Inc.)

Problem – Weakly Supervised Segmentation

Fine mask annotation: labor-intensive
Point annotation: cheap and fast

input mask point

Point annotations (positive samples) Voronoi diagram (negative samples)

Our method – PseudoEdgeNet

LC-FCN [Laradji et al. ECCV'18]

Input I → Segmentation net f → Result $f(I)$ → Segmentation loss → Point annotation P

Small network for **capacity gap**

Remove false positives edges

Edge net g → Edge $g(I)$ → Sobel filter s → Edge loss → $s(f(I))$

PseudoEdgeNet → Attention module h → $h(I)$

$\mathcal{L}(I, P, f, g, h) = \mathcal{L}_{ce}(f(I), P) + \lambda \cdot |s(f(I)) - g(I) \otimes h(I)|$

Results

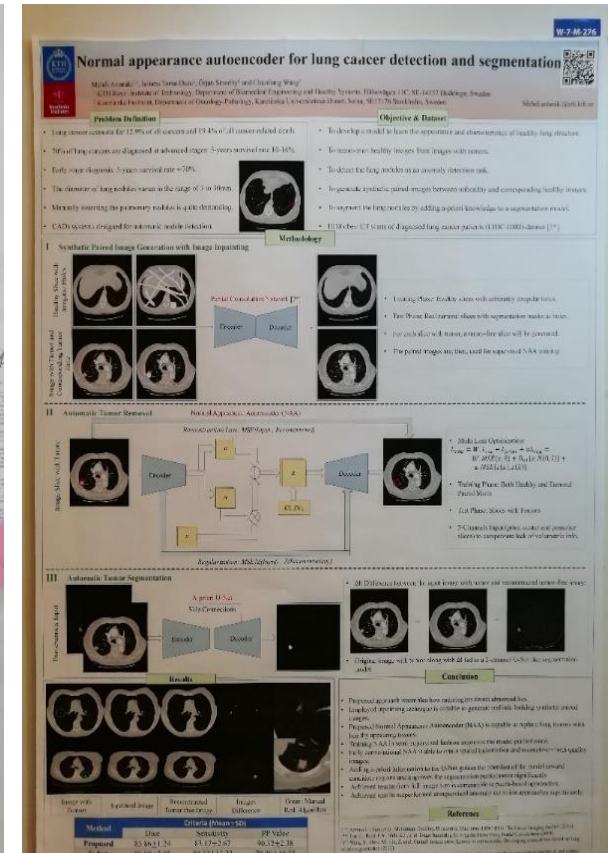
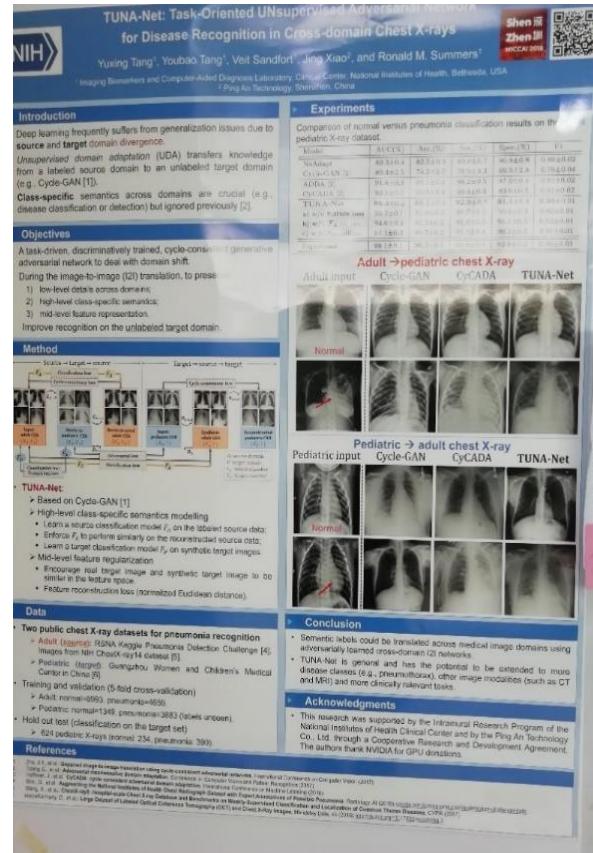
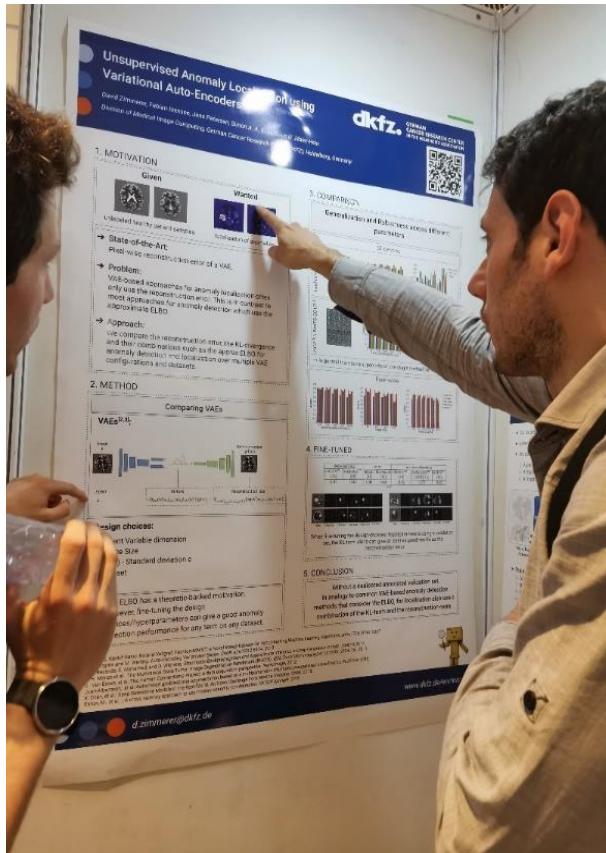
Methods	MnNuSeg	TNDC
Baseline [6]	0.5779 ± 0.0002	0.5365 ± 0.0047
DenseCRF [18]	0.5813 ± 0.0009	0.5335 ± 0.0046
PseudoEdgeNet with large g	0.5796 ± 0.0046	0.5377 ± 0.0046
PseudoEdgeNet with small g	0.6059 ± 0.0046	0.5533 ± 0.0046
PseudoEdgeNet with small g and h	0.8136 ± 0.0046	0.8058 ± 0.0046
Fully supervised / upper bound	0.8132 ± 0.0046	0.8057 ± 0.0046

*Author's open source is used. <https://github.com/wmpang/weakseg>



Unsupervised

Unsupervised anomaly location



深圳市大数据研究院
Shenzhen Research Institute of Big Data

Attention

Cross Attention

Multi-label Thoracic Disease Image Classification with Cross-Attention Networks

Congbo Ma^{1,2}, Hu Wang¹, Steven C.H. Hoi^{1,4}

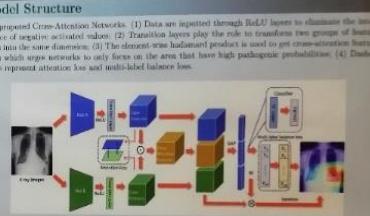
¹Singapore Management University, Singapore ²National Clinical Center of Technology, Guangzhou, China ³The University of Adelaide, Australia ⁴Machine Research Assn, Singapore, Singapore

Abstract

Automated disease classification of radiology images has been emerging as a promising technique to support clinical diagnosis and treatment planning. Unlike generic image classification tasks, a real-world radiology image classification task is more challenging, as it is far more expensive to collect the training data where the labeled data is in natural multi-label, and it is usually necessary to have very *easy* or *other domain* training data as highly class-imbalanced problem exists in practice as well. To overcome these challenges, in this paper, we propose a novel scheme of Cross-Attention Networks (CAN) for automated thoracic disease classification from X-ray images, which can effectively extract more meaningful representation from data to boost the performance through cross-attention by image-level annotation. We also design a new loss function to beyond cross entropy loss to help cross-attention process and is able to overcome the imbalance between classes and easy-dominated samples within each class. The proposed method achieves state-of-the-art results.

Cross-Attention Networks

Feature extraction networks. After raw data processing, the images are partitioned into two feature extraction networks, one for positive channels and one for negative channels. Then, two parallel multi-layer perceptrons (MLPs) are used as feature extractors, which are worked as feature extractors. Our proposed cross-attention method is flexible and the two networks can be easily substituted by other methods. Different from traditional CNNs, instead of reducing extracted features into attention layers and global average pooling layers, we keep them in their original form and then use them as cross-attention module. We propose a new cross-attention model that two networks have attention on each other, while each model generates information from the other. This is achieved by pumping images into two networks, which have different initializations or structures, the two feature extraction networks. In this way, both networks respectively to ensure that negative activation values will not interfere the cross-attention feature. Then, the output of two networks will be input to an attention layer to transform the output into the same shape. Later on, instead of concatenating two large tensors as a full attention model, we use element-wise product of the two outputs to derive a new attention loss. Through attention loss, we set a constraint to urge two networks to find mutual pathogenic areas which would smooth the cross-attention process. The formulation of attention loss is given as below:



Loss design

Attention loss. In order to further extract more meaningful features in proposed cross-attention model, we design a new attention loss. Through attention loss, we set a constraint to urge two networks to find mutual pathogenic areas which would smooth the cross-attention process. The formulation of attention loss is given as below:

$$Loss = \left| \text{norm}(\sum_{i=1}^M f_{ai}) - \text{norm}(\sum_{i=1}^M f_{bi}) \right|$$

where M and D denote the total size of feature channels within one network, n and m represent the index of each feature map, f_{ai} and f_{bi} are different feature maps from two networks respectively.

Cross-entropy loss. Cross-entropy loss is often used to measure the difference caused by imbalance between classes, which is not ideal for standard cross-entropy loss to solve. And any direct application of cross-entropy loss to multi-label setting with multi-label balanced loss, such as M-Net[2], can only handle the imbalance between positive-negative samples with each class, but also exaggerates the representation of dominated easy samples.

Footnote: $\text{norm}(f) = \sqrt{\sum_i f_i^2}$, where $f = f_1 \otimes f_2$

where $F_{1,2}$ represents the feature maps of Network A and Network B, $F_{1,2}$ represent cross-attention feature maps. Cross-entropy function computes the cross-entropy loss between two groups of feature maps. Using hadamard product, the cross-entropy loss of feature map $f_{1,2}$ would be activated on those areas that belong to two groups of feature maps $F_{1,2}$ and $F_{1,2}$ are both activated on the same areas. When the loss is too large, gradient vanishes, and the gradient of the loss function is zero. Thus, we add some other gradient blocks which also enable cross-attention feature maps to be activated on the same areas. After getting the final cross-attention feature maps, we integrate these two groups of feature maps to a global multi-label prediction. Then, we compare these two groups of feature maps to the same class. Where C and C' denote the class and the class of cross-attention feature maps. By using transformation layers, we ensure the feature maps of two groups are the same and also further reduce loss function.

Footnote: $L = \sum_{i=1}^n -w_{pi} \cdot [1 - p(Y_i)]^2 \cdot [(1 - g)(\log(p(Y_i)) + w_{pi} \cdot [1 - p(Y_i)] \cdot g)]^2 \cdot g(p(Y_i = 1 | X))$

$w_{pi} = \frac{1}{\sum_{j=1}^m w_{pj}}$

where L is the total number of analytic labels and i denotes each label of different type of diseases. w_{pi} and w_{pj} denote the number of positive and negative samples of a certain disease. $p(Y_i)$ denotes the probability of the occurrence and share of each label balance samples, which may have better performance to some "easy" examples to contribute more to the model training in some settings, especially when "easy" examples dominate the dataset greatly.

Cross-attention loss function. The cross-attention loss function is a combined loss function which is:

$$L = \alpha L_{att} + (1-\alpha)L_{ce}$$

where α represents the trade-off factor between attention loss and multi-label balanced loss. The attention loss could force the model to focus on pathogenic areas more accurately. Multi-label balanced loss helps the class balance.

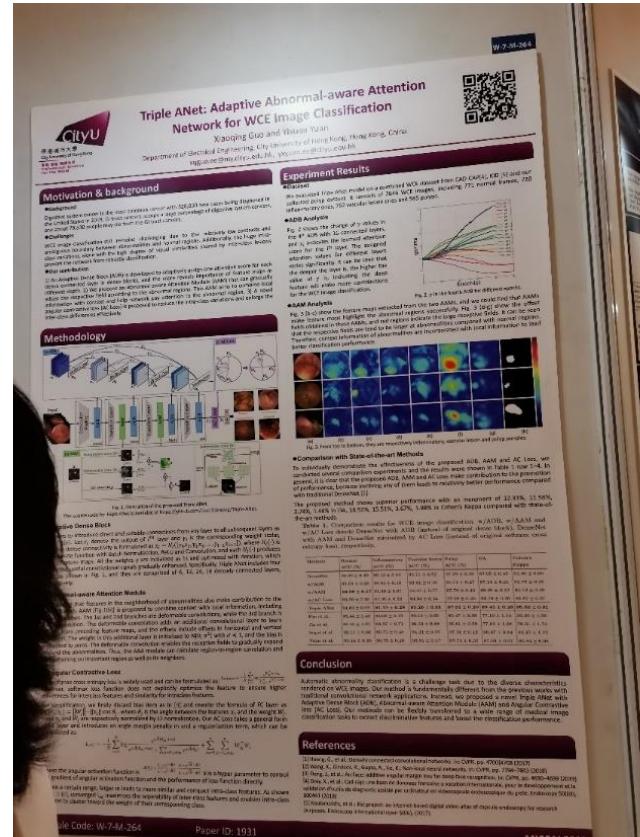
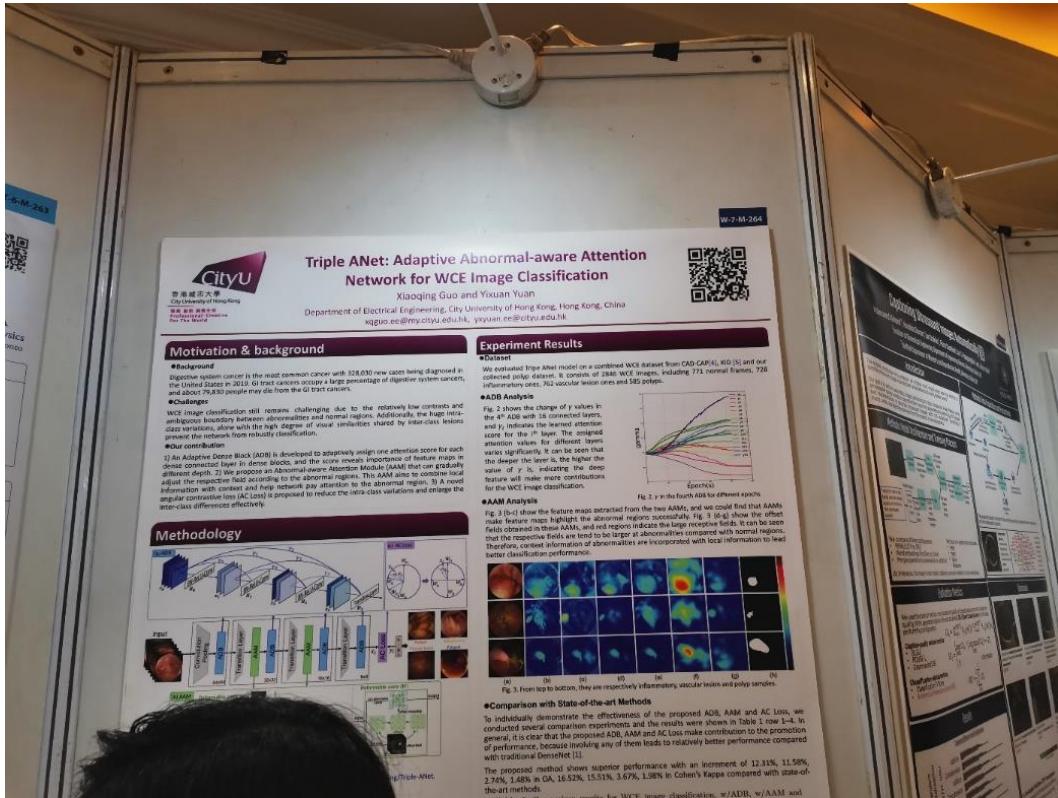
Conclusion

This paper proposed an end-to-end trainable Cross Attention Network (CAN) for multi-label classification of



深圳市大数据研究院
Shenzhen Research Institute of Big Data

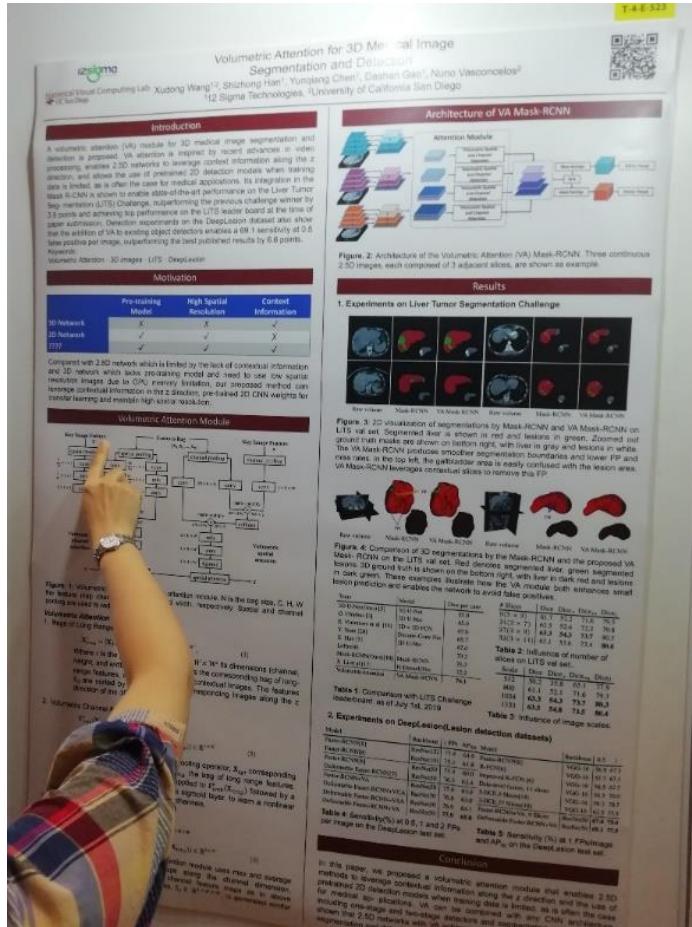
Adaptive abnormal-aware Attention



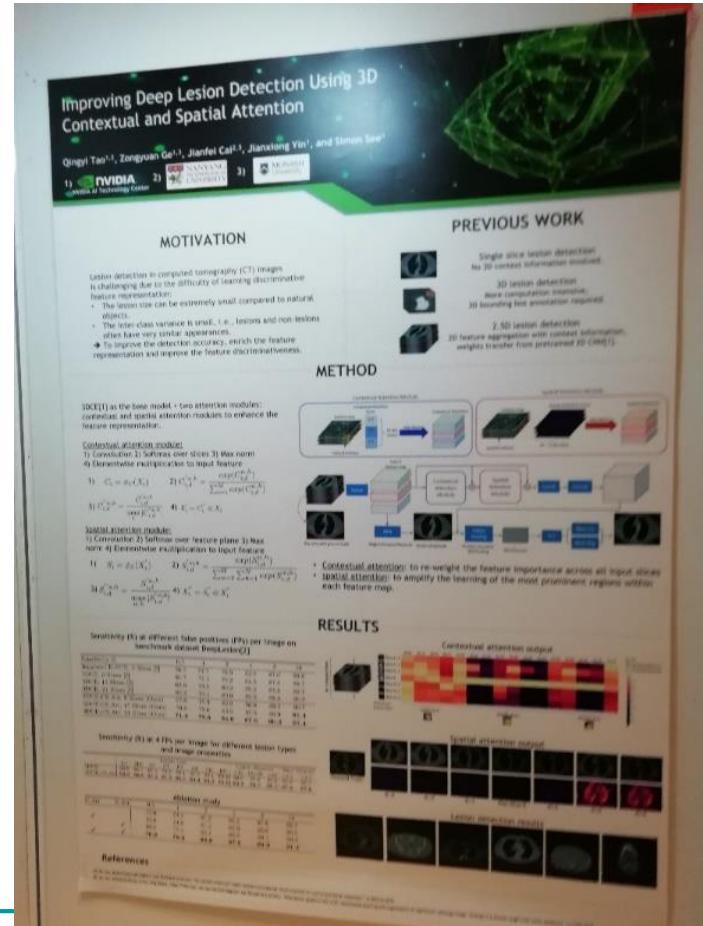
深圳市大数据研究院
Shenzhen Research Institute of Big Data

Attention + 3D

Volumetric Attention for 3D Segmentation and Detection



Improving Deep Lesion Detection Using 3D Contextual and Spatial Attention



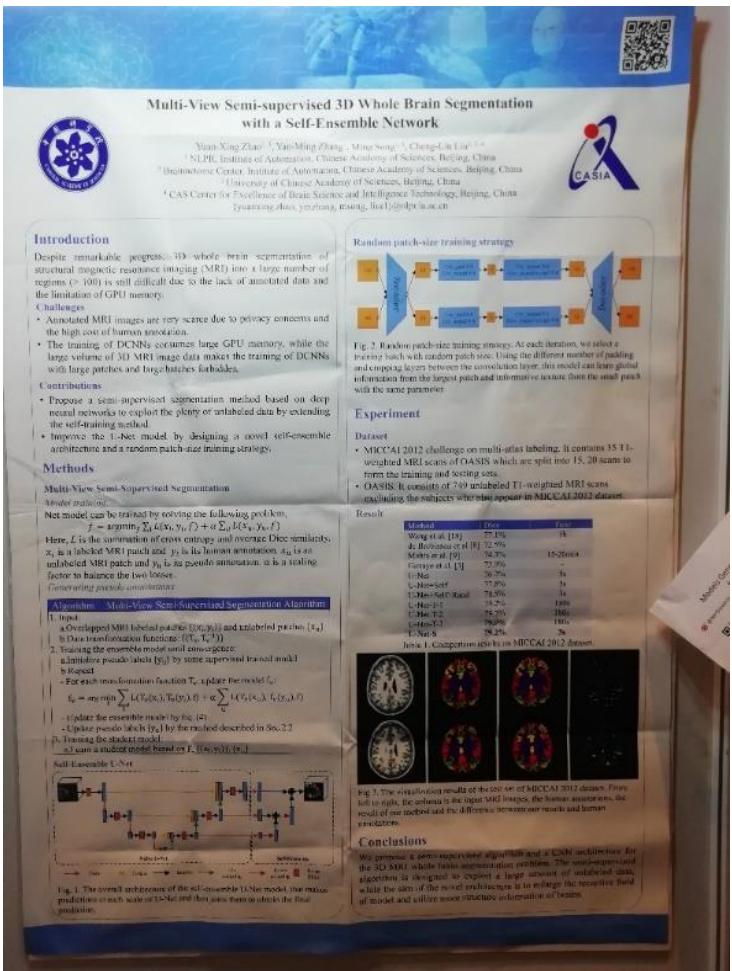
深圳市大数据研究院
Shenzhen Research Institute of Big Data

Multi-view



深圳市大数据研究院
Shenzhen Research Institute of Big Data

Multi-view



MVP-Net: Multi-view FPN with Position-aware Attention for Deep Universal Lesion Detection

Zhen Li^{1,2*}, Shu Zhang¹, Xiang Zhang¹, Junjie Huang¹, Yufan Wang^{1,2}, Yunhai Yu^{1,2}
¹Institute of Automation, Chinese Academy of Sciences / Beijing Advanced Research Institute
²Computer Science Dept., Peking University / Peking University Health Science Center

Welcome to our booth A1! Send your CV to zhengliy@imprc.ac.cn and 3nmr@163.com

Introduction

Background

It's common knowledge modeling that plays a essential role in algorithm designing. In this paper, we formulate two kinds of clinical experience into a unified deep learning framework to develop a Universal Lesion Detection, which can identify and localize different types of lesions across the whole body.

How to model clinical experience?

Motivation: Radiologists would combine information from multiple reconstructed slices under different window widths and window levels in locate positive lesions.

Proposed method: We initiate this process and propose to extract discriminating features from several frequently examined windows with limited lesion location information.

Methodology

Three-pathway Convolution Backbone with FPN:

A three pathway architecture is applied to extract the most prominent features from each representative view. The backbone network takes three slices $\times 3$ 3D context, and is shared by all paths.

Position-aware modelling: The position branch gives continuous and discrete position prediction. The discrete position supervised by classification loss is more robust to noise and can make convergence easier.

Channel-wise attention module: A channel-wise attention module is proposed to aggregate feature maps of different views.

RPN & RCNN: Standard RPN and RCNN

Experiments

Table 1. Quantitative PSC of various FPNs for lesions on the testing set of three tasks. We don't have the result with 27 slices due to memory limitation. * indicates concatenation of 3DCE with FPN as backbone.

FPN type	0.3	1	2	3	4
3DCE [1]	52.2	59.0	71.0	-	54.2
3DCE+3DCE [2]	53.9	60.7	72.7	-	56.7
3DCE+3DCE+3DCE [3]	58.7	72.0	81.0	-	58.5
3DCE+27 slices [4]	58.0	73.7	82.3	-	57.3
3DCE+3DCE+27 slices [5]	58.2	73.8	82.3	-	57.3
3DCE+RPN [6]	60.2	74.0	83.0	-	60.0
3DCE+RCNN [7]	60.2	74.0	83.0	-	60.0
3DCE+3DCE+RPN [8]	60.4	74.2	83.2	-	60.2
3DCE+3DCE+RCNN [9]	60.4	74.2	83.2	-	60.2
3DCE+3DCE+3DCE+RPN [10]	62.0	75.7	84.0	-	61.8
3DCE+3DCE+3DCE+RCNN [11]	62.0	75.7	84.0	-	61.8
Our method	62.8	76.4	84.8	86.27	63.8
Best over 27 slices [12]	71.0	76.7	85.0	-	71.0

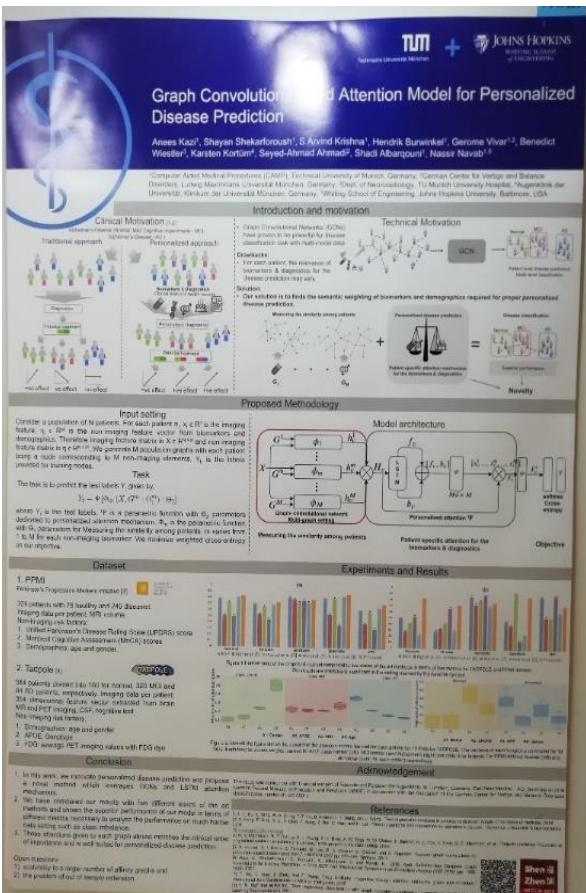
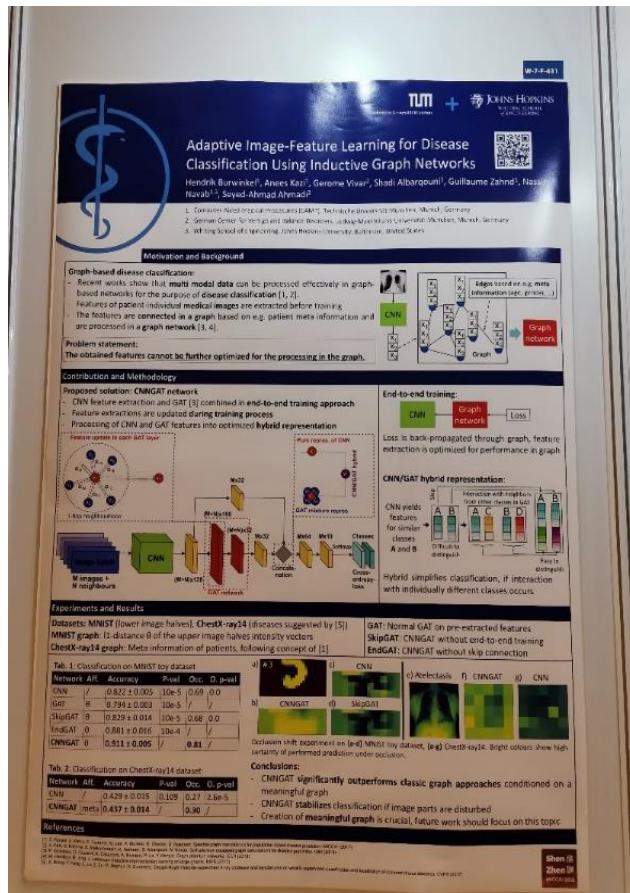
Table 2. Ablation study of our approach on the DeepLesion dataset.

Prop.	Model	Baseline	Feature	FCNN	EP+FCN	EP+FCN+RPN
✓	✓	-	-	41.20	46.14	47.20
✓	✓	✓	-	44.40	47.70	48.40
✓	✓	✓	✓	47.00	49.40	50.40
✓	✓	✓	✓	47.00	49.40	50.40

Fig.3. Case study for ADCE (leftmost column) and attention based multi-view modeling (the other three columns). Green and red boxes correspond to ground-truths and predictions respectively.



Graph Network



Graph-kernel-based Multi-task Structured Feature Selection on Multi-level Functional Connectivity Networks for Brain Disease Classification

Zhengdong Wang¹, Biao Jie¹, Mi Wang¹, Chorong Liang¹, Wen Zhou¹, Mingxia Liu², Dinggang Shen¹

¹ Department of Computer Science and Technology, Anhui Normal University, Anhui 241003, China;

² Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA;



EXPERIMENTS & RESULTS

We extensively perform experiments to evaluate the performance of our proposed g-MTSFS method. Specifically, we perform three tasks: 1) IMCI vs. SMC1, 2) ADHD vs. HC and 3) ADHD vs. HC classifications.

Table 1. Performance of all methods in three classification tasks

Methods	IMCI vs. SMC1	ADHD vs. HC	ADHD vs. HC
Baseline	56.4	62.0	63.4
NMT	60.2	66.6	67.4
Lasso	62.7	63.9	62.3
glasso	70.6	69.2	66.0
g-MTSFS(our)	76.8	76.3	66.9



Fig. 2. The accuracy of the proposed g-MTSFS method with the combination of (a) and (b) versus three classification tasks of (a) IMCI vs. SMC1, (b) ADHD vs. HC and (c) ADHD vs. HC.

METHOD

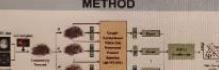


Fig. 1. Architecture of the proposed g-MTSFS framework for brain network analysis using MRI data.

► Proposed Graph-kernel-based Multi-Task Structured Feature Selection

The objective function of our proposed g-MTSFS model is defined as following

$$min_{W} \sum_{i=1}^n \|X_i^T X_i - X_i^T W\|^2 + \beta \sum_{i=1}^n \|X_i^T W\|^2 + \gamma \|W\|_F^2$$

where $W = (\frac{1}{\sqrt{n}} I_n^T, \frac{1}{\sqrt{n}} M_n^T)^T$ is the Laplacian matrix, β and γ are two positive constants that balance the contributions of three items. In practice we use inner cross-validation on the training data to determine its optimal values.

According to this definition, the objective function can be effectively solved via using accelerated proximal gradient algorithm.

► g-MTSFS Based Learning Framework

1. Network Thresholding

$$E_j^T(p, q) = \begin{cases} 0, & U_j^T(p, q) < T \\ 1, & \text{otherwise} \end{cases}$$

where $U_j^T(p, q)$ is the element of matrix U_j , corresponding to the weight (i, j , i.e., correlation coefficient) of edge between ROI p and q .

2. Feature Selection

We further perform our proposed g-MTSFS method to select the most discriminative features for improving the classification performance.

3. Classification

we use the following method to integrate these kernels:

$$A(S_i, E_j) = \sum_{p=1}^P A^p(S_i^p, E_j^p)$$

MATERIALS

In this study, we use two datasets with resting-state fMRI (rs-fMRI) data. The first dataset is from the ADNI-2 dataset (www.cnl.ucla.edu/ADNI). It contains 43 healthy subjects (21 men) and 43 patients (21 men and 22 women) with Alzheimer's disease (AD). The second dataset is from New York University site, including 118 ADHD (52 males) aged 11.00 ± 2.7 years and 98 NCs (51 males) aged 12.2 ± 2.1 years.

REFERENCES

- [1] He, R.; Liu, M.; Zhang, D.; Sun, D.; Sub-network kernel for measuring consistency of brain connectivity networks in disease diagnosis. *IEEE Transactions on Image Processing* 2019; 28(10): 3346-3354.
- [2] Li, Y.; Wang, Y.; Chen, Z.; Li, H.; Wang, Y.; Huang, Z.; Peng, J.; Tong, L.; Olf, C.; Cheng, Q.; Zhang, B.; Multivariate classification of visual sensory disorder using whole-brain functional connectivity. *Brain Structure and Function* 2019; 221(1): 101-115.



深圳市大数据研究院
Shenzhen Research Institute of Big Data

Local Global

Global and Local Interpretability for Cardiac MRI Classification

Local and Global Interpretability For Cardiac MRI Classification
J.R. Clough, I. Oksuz, E. Puyol-Antón, B. van Ginneken, A.P. King, J. Schnabel
School of Biomedical Engineering & Imaging Sciences, King's College London

Introduction

- Disease classification with deep learning models can be accurate but hard for clinicians to interpret
- How can we describe the way a model processes images using language familiar to doctors?
- Our approach combines:
 - Variational Autoencoder** which encodes input image to a low dimensional latent space
 - Concept Activation Vectors** which associate directions in that space with clinical biomarkers which are familiar to clinicians.

This allows an interrogation of the classification process and better understanding of the model's reasoning.

Data & Task

- UK Biobank dataset
- MR images: short-axis view
- Network inputs are segmentations
- Predict whether subject had previously been diagnosed with a serious heart condition e.g. coronary artery disease, myocardial infarction from a sequence of 50 segmentations covering the whole cardiac cycle.
- Biomarkers such as the ejection fraction, peak ejection rate and peak filling rate were calculated and used in the Concept Activation Vector framework.

Encoder

Decoder

Classifier Layers

- Fully Connected
- Batchnorm
- ReLU

Methods and Results

- Encoder finds **latent representation** of the input cardiac segmentations.
- Classifier is trained on this **sequence** of latent vectors - one per frame.
- In the space of activations at intermediate layer of this classifier we find the **Concept Activation Vectors**.
- This involves training a secondary linear model to predict these biomarkers on a held-out validation set of subjects.
- For each biomarker we obtain its CAV by finding the normal vector to this classification boundary.
- We trained the model with a set of 5316 subjects, used 5000 for the CAV validation set and then tested on 500.
- The final classifier had an AUC of 0.78 and the autoencoder a Dice score of 0.93.

Discussion

- Once the classifier is trained we can analyse with respect to the biomarkers.
- If gradient of the disease classification logit has a large dot-product with a CAV then moving the latent vector in that direction would increase the model's estimate of disease
- This global interpretability helps us to understand which biomarkers are most important for disease classification by this model.
- Secondly, we use the decoder to visually inspect these CAVs meaning
 - By interpolating in the latent space in the direction of one of the CAVs, and then decoding that new latent vector we see the effect of perturbations in the latent space.

First Principal Component

Local and Global Consistency Regularized Mean Teacher for Semi-supervised Nuclei Classification
Hai Su, Xiaoshuang Shi, Jinzheng Cai and Lin Yang

Nuclei classification

- Nuclei classification is a fundamental task in many pathology diagnosis for cancers, e.g. Ki-67 index estimation.
- Supervised deep learning methods have achieved promising classification accuracy.
- However, the success of these methods heavily relies on large amount of labeled annotated data.
- In this paper, we propose a novel semi-supervised classification method that can learn from small portion of labeled data and large amount of unlabeled data.
- Our method is inspired by the recent state-of-the-art self-supervised learning methods that can learn from unlabeled data by enforcing prediction consistency of each sample under different perturbations. It integrates both local and global consistency hidden in the data.
- Our approach integrates a label propagation (LP) step to assign each unlabeled data a guessed label based on the learned global consistency assumption. A graph is constructed based on the label of the data and the Siamese loss is used to learn the local and global consistency from the graph.
- > Our implementation is based on Mean Teacher method.
- > Extensive experiments on two nucleus datasets demonstrate that our method outperforms the state-of-the-art SE methods and achieves F1 score close to fully supervised deep learning methods using only 5%-25% labeled data.

Local and global consistency

Graph construction for a Minibatch

Our proposed method



深圳市大数据研究院
Shenzhen Research Institute of Big Data

X Ray

Code

Adversarial Regression Training for Visualizing the Progression of Chronic Obstructive Pulmonary Disease with Chest X-Rays

Motivation: Chronic obstructive pulmonary disease (COPD), third leading cause of death[1] and not reported by radiologists
- We present regression CNN: COPD screening with chest x-ray
- Explainability wanted for adoption, communication and debugging. Challenge: model explainability for regression

Our approach: explainability for regression by adversarial training

Method

$\pi = \text{original image linked to value } y$
 $\pi' = \text{modified image linked to } y$
 $\Delta\pi = \text{the generated disease effect}$
 $y = \text{FEV}_1/\text{FVC}$

Loss terms:
 1. R performs the original regression task: $L_{R\pi} = |R(\pi) - y|_\ell_1$
 2. G modifies R 's output to desired value: $L_{G\pi} = |R(\pi') - y|_\ell_1$
 3. Adversarial: R ignores changes from G : $L_{RG} = |R(\pi) - R(\pi')|_\ell_1$

Experiments and Results

Baseline: VA-GAN [2]:
 - Critic instead of regressor (generated vs. real with disease)

Toy Dataset: square side length ≈ 11

Chest x-ray dataset:
 - Patients with chest x-ray and a PFT
 - Longitudinal data registered for calculating scores
 - Dose: $y < 0.7$

Conclusion

- Regression target inclusion in formulation improves results.
 - Learned changes agree with radiologists on COPD effects.
 - Realistic images with no real-life discriminator.
 - A step toward building explainable models for regression.

Code available! github.com/hcbil/vrgan

Endotracheal Tube Detection and Segmentation in Chest Radiographs using Synthetic Data

Maayan Frid-Adar¹, Rula Amer¹, and Hayit Greenspan²

¹HabEdge Ltd., Tel Aviv, Israel

²Tel Aviv University, Department of Biomedical Engineering, Medical Image Processing Lab, Tel Aviv, Israel

Introduction

- Chest X-ray acquisition is recommended following intubation to ensure proper positioning of inserted tubes for patients in the ICU.
- Malposition of the ET tube can cause serious complications if not detected.
- Collecting and labeling chest X-ray for the presence of ET tubes in order to develop deep learning systems is expensive and time-consuming.

Objective

Design a deep learning system for combined ET tube detection and segmentation in chest radiographs, in the scenario of limited expert labeled data - using a public dataset and novel synthetic data generation.

Generating Synthetic Data

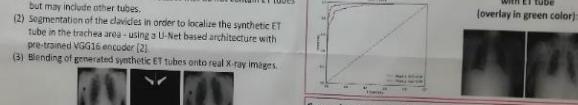
In this study, we apply a technique to insert synthetic ET tubes as an overlay to the original X-ray images taken from a publicly available dataset of chest radiographs [1].

- Steps for generating synthetic ET tubes over real X-ray images:
- (1) Selection of cases from the NIH dataset that do not contain ET tubes but may include other tubes.
 - (2) Segmentation of the clavicles in order to localize the synthetic ET tube in the trachea area - using a U-Net based architecture with pre-trained VGG16 encoder [2].
 - (3) Blending of generated synthetic ET tubes onto real X-ray images.

Results

- * Test set: 479 real chest radiographs that were collected manually one time during the development and entirely independent from all training data.

Two phase Detection results: Segmentation results of real cases with ET tube (overlap in green color):



Comparison to state-of-the-art models:

	AUC	Sensitivity	Specificity	Training size [pos, neg]
Balakrishnan et al. [3]	0.92	0.95	0.75	66 [28, 38]
Chen et al. [4]	0.95	-	-	87 [44, 43]
Abdullah et al. [5]	0.99	-	-	60 [30, 30]
Trivedi et al. [6]	0.97	89.2%	93.0%	470 [221, 247]
ETT-Net+ - phase1	0.96	86.2%	93.0%	470 [221, 247]
ETT-Net+ - phase2	0.99	95.5%	96.5%	470 [221, 247]

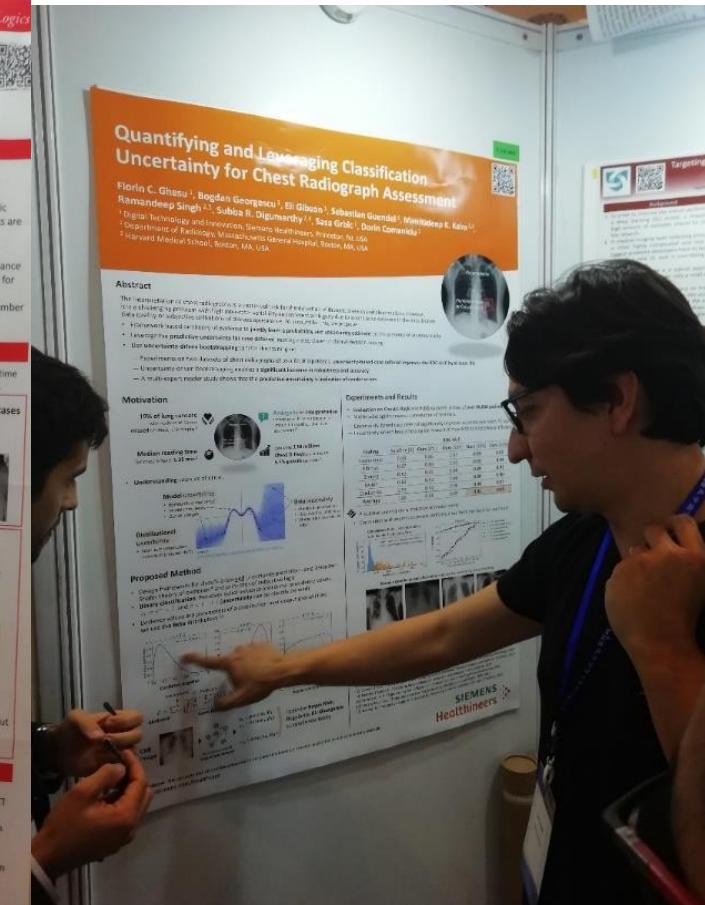
Baseline model: we trained a classification only CNN with the same real data we collected for comparison.

Comparison Conclusion:

- Our best model reached high performance with a test set size of one magnitude more than state-of-the-art methods.
- Our model was trained and tested using free public dataset without manual annotations (other methods used hand picked and annotated cases).

Conclusion

In this work, we proposed an approach for training a combined deep learning network for the tasks of detection and segmentation of the ET tube in chest radiographs without collecting and annotating data. We trained the ETT-Net in two phases using a public dataset: first with synthetic ET tubes and then with real cases retrieved during the training process for a fine-tune stage. Our model achieved a very high accuracy for the presence of ET tube in real ICU patients (0.99 AUC) and outputs high quality segmentation maps that can assist in detection of the misplacement of the tubes.



深圳市大数据研究院
Shenzhen Research Institute of Big Data

Multi-label Thoracic Disease Image Classification with Cross-Attention Networks

Congbo Ma¹, Hu Wang¹, Steven C.H. Hoi^{1,2}

¹Singapore Management University Singapore, ²South China University of Technology, Guangzhou, China, ³The University of Adelaide, Australia, ⁴Autodesk Research Asia, Singapore, Singapore

Abstract
Automatic classification of radiographic images has been emerging as a promising technique to support clinical diagnosis and treatment planning. Existing generic classification models, make a hard decision-making image classification task as significantly more challenging as it is for medical image classification. Multi-label disease detection based on the labeled data is a more multi-label, and more seriously complex from a class-wise classification perspective. The problem model needs to perceive as well. In addition, disease challenges, in this paper, we present a multi-label learning framework Cross-Attention Network (CAN) for automatic thoracic disease classification from chest X-ray images, which can effectively mitigate the cross-disease interference from data to boost the performance through cross-attention mechanism. The proposed model can learn to ignore the diseases that layout cross-energy loss to help cross-attention process and is able to overcome the imbalance ratio among different diseases.

Model Structure
The proposed Cross-Attention Networks. (1) Data are inputted through ReLU layers to eliminate the inner layer of negative activated values; (2) Transition layers play the role to transform two groups of disease stage into the same dimension; (3) The dimension-reduced product is used to get cross-attention mechanism which maps networks to only focus on the new class have high pathogenic probabilities; (4) Transfused layer performs attention loss and multi-label losses function.

Loss design
Attention loss. In order to further estimate more distinguishable features in proposed cross-attention model, we define a new attention loss. Through attention loss, we not a constraint to urge two networks to find out specific pathogenic model that can separate hard attention from each other. The formulation of attention loss is given as follows:

$$L_{att} = \text{Feature}(\sum_{i=1}^{N_{att}} f_i) - \text{Feature}(\sum_{i=1}^{N_{att}} f_i)$$

where f_i represents the feature map of N_{att} network. A larger L_{att} indicates that the cross-attention network can obtain more distinguishable features. By using normalized product, we can get a better loss function. For example, if f_1 would be normalized on class one, then only f_1 will be used to calculate the loss. If f_2 is also normalized on class one, then both f_1 and f_2 are being affected. When f_1 and f_2 are both normalized, then we can back propagate two networks, respectively. But if f_1 and f_2 are not normalized, then we can only update one network, which also receive the gradient update from the other network. That's why we can back propagate two networks, especially when "easy" examples can update the shared gravity.

Transition layers and classifier. The transition layers are composed of two groups of feature maps. D_{Att} is the number of 1 × 1 convolutional kernels for feature fusion, and D_{Att} is the number of feature maps after the transition layers. When D_{Att} is set to 1, using transition layers can only realize the cross-disease interference of one feature map. When D_{Att} is set to 2, using transition layers can realize the cross-disease interference of two feature maps. After getting the all disease information from each group of feature maps, we can then extract more information to each disease. Our model is composed of three main parts: (1) multi-label learning module, (2) cross-attention module, and (3) cross-disease loss function. Since each disease has a different probability, our model can predict the probability of images which are composed of a multi-label learning module and a separate cross-attention module.

Encoding CT Anatomy Knowledge for Unpaired Chest X-ray Image Decomposition

Zeju Li¹, Han Li¹, Hu Han^{1,3}, Gonglei Shi¹, Jiannan Wang⁴, and S. Kevin Zhou^{3,4}

¹BioMedIA group, Imperial College London, ²University of Chinese Academy of Sciences

³MIRACLE group, Institute of Computing Technology, Chinese Academy of Sciences

⁴Peng Cheng Laboratory, ⁵Southeast University, ⁶Nankai University

1. Introduction

- While clinical diagnosis using chest X-ray (CXR) is common, it is a challenging task because CXR is only a 2D projection image containing overlapped anatomies and ambiguous structure details.
- Computed tomography (CT) is a closely related medical imaging modality to CXR because CT is reconstructed from a set of X-ray projections. Leveraging the anatomy knowledge (lung, bone, etc.) embedded in CT to improve the diagnostic value of CXR is very useful for practical applications.
- We propose a decomposition generative adversarial network (DecGAN) to leverage the prior structural knowledge from the unpaired CT domain to perform CXR enhancement.

2. Methods

- DecGAN is designed upon the backbone of CycleGAN with latent space disentanglement.
- Problem formulation:

$$X_{in} = F(X; \alpha_1, \alpha_2, \alpha_3) = G_X(G_{Din}(G_E(X; \alpha_1, \alpha_2, \alpha_3)))$$

- Main contributions:

- A latent space decomposition discriminator D_{Bx} is introduced to encourage the embedding of prior CT decomposition knowledge and the separation of different components in the generated DR.

- The DRR decomposition network G_{Din} is embedded into the backbone of CycleGAN to provide the decoder enough knowledge to tackle the decomposition information in the latent space.

3. Results

- CXR Bone Suppression:** DecGAN can not only suppress bones in CXR but also preserve the most realistic results of non-bone regions.

Method	CXR	Blurred Seg.	No Adaption	CycleGAN	DecGAN	DR
GANP	2.7	4.1	4.0	6.0	2.5	2.5
PNSR	30.2	26.7	25.1	29.4	30.1	30.1

where L_{att} is the total number of multiple labels and L_{cls} is the total number of different type of disease, α_1 and α_2 are multi-label factor to indicate posterior probability of each binary class, β_1 and β_2 are the ratio of positive and negative samples of a certain disease. Furthermore, the shape and size of the training data which may have better performance to the "hard" examples to contribute more to the model training in some settings, especially when "easy" examples can update the shared gravity.

Cross-attention loss function. The cross-attention loss function is a weighted loss function which is:

$$L_{att} = \alpha_1 L_{cls} + \alpha_2 L_{att}$$

where α_1 is the responsibility factor between attention loss and multi-label balance loss. The greater α_1 is, the more attention attention loss and multi-label balance loss. The greater α_2 is, the more difficult to solve the multi-label problem, and more accurately. Multi-label balance loss helps the model to focus on the diseases which layout cross-energy loss to help cross-attention process.

Conclusion. The proposed CAN can effectively mitigate the cross-disease interference of the multi-label learning module. The CAN can not only realize the cross-disease interference of different diseases but also can realize the cross-disease interference of one disease. Furthermore, the proposed CAN can effectively increase the performance of the multi-label learning module. After getting the all disease information from each group of feature maps, we can then extract more information to each disease. Our model is composed of three main parts: (1) multi-label learning module, (2) cross-attention module, and (3) cross-disease loss function. Since each disease has a different probability, our model can predict the probability of images which are composed of a multi-label learning module and a separate cross-attention module.

Pick-and-Learn: Automatic Quality Evaluation for Noisy-Labeled Image Segmentation

Haidong Zhu¹, Jialin Shi², and Ji Wu^{1,3}

¹Department of Electronic Engineering, Tsinghua University, Beijing, China
²Institute for Precision Medicine, Tsinghua University, Beijing, China

Introduction

Background

- Deep learning methods are still struggling with noisy-labeled images during the training process.
- How to eliminate the disturbance from noisy labels for segmentation tasks without further annotations is still a significant challenge.

Motivation

- Propose a solution for network automatically evaluating the relative quality of the labels in the training set and using good ones to tune the network parameters.
- Introduce label quality evaluation strategy (QAM) for deep neural networks automatically assessing the quality of each label.
- Design an overfitting control module (OCM) to let the network maximally learn from the precise annotations during the training process.
- Experiments on the public biomedical image segmentation dataset have proved the method outperforms baseline methods and retains both high accuracy and good generalization at different noise levels.

Examples

Figure 1: Two examples of noisy labels in the segmentation problem

Network Architecture

Figure 2: End-to-end network architecture for label quality evaluation strategy

Segmentation module

- A CNN structure module for generating segmentation
- Taking the concatenation of the image and its labels as input
- Running parallelly with the separation module

Quality awareness module (QAM)

- A CNN structure module
- Taking the concatenation of the image and its labels as input
- Running parallelly with the separation module

Overfitting control module (OCM)

$f(x) = \text{tanh}(x)$

- Output quality score from $(-1, +\infty)$ to $(-1, +1)$
- Maximum possible ratio of relative scores decrease from a to a^2 .
- Relative weight can not be too small or too big

Final loss

$\text{Loss} = \sum_{i=1,3} \theta(z_i, y_i) \cdot L_i \quad z_i = \sum_{j=1,2} \Theta(x_j, y_j) = 1, 0 \leq \Theta(x_j, y_j) \leq 1$

Experiments and Results

Table 1. Results on JSRT dataset

Label percentage	Score	Knowledge	Loss	Image	Ground truth
0% noise	-	GAM	0.92	0.935	
10% noise	-	GAM	0.93	0.949	
20% noise	-	GAM	0.93	0.952	
30% noise	0.5 ~ 0.9	GAM+OCM	0.96	0.960	0.978
40% noise	0.5 ~ 0.9	GAM+OCM	0.96	0.960	0.984
50% noise	0.5 ~ 0.9	GAM+OCM	0.96	0.960	0.985
60% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
70% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
80% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
90% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
100% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
110% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
120% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
130% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
140% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
150% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
160% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
170% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
180% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
190% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
200% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
210% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
220% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
230% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
240% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
250% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
260% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
270% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
280% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
290% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
300% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
310% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
320% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
330% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
340% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
350% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
360% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
370% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
380% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
390% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
400% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
410% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
420% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
430% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
440% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
450% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
460% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
470% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
480% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
490% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985
500% noise	0.5 ~ 0.9	GAM+OCM	0.97	0.960	0.985

Figure 3: Average dice accuracy and loss plots of different noise levels on JSRT. Loss curves belong to models trained on the training set with 50% labels deleted or added from 1 to 15 pixels

Figure 4: Relative weight and variance for clean and noisy-labeled data



Multi-instance Learning survive prediction

Deep Multi-instance Learning for Survival Prediction from Whole Slide Images

Jiawen Yao, Xinlang Zhou, and Junzhou Huang
Dept. Computer Science & Engineering, University of Texas at Arlington

Problem

- Survival prediction is very crucial to healthcare for cancer patients because it would allow clinicians to make early decisions on treatments.
- It has been shown that connections exist between lung tumor morphology and prognosis. Pathological images could benefit tumor diagnosis since they can describe tumor growth with very clear details.
- Recent image-based survival models have very high risks to ignore survival-discriminative patterns if only select several tiles from very heterogeneous whole pathological image slides. State-of-the-art WSI models cannot consider the joint effects of phenotypes which could be useful for prediction.

Motivations

- Different from the existing works on learning using key patches or clusters from WSIs, we take advantages of a deep multi instance learning to encode all possible patterns from WSIs and consider the joint effects from different patterns for clinical outcomes prediction.
- The primary motivation of using multi-instance learning (MIL) is because of its ability to learn discriminative patterns based on a training set of bags, where each bag contains multiple feature vectors known as discriminative instances. This enables us to identify patients' survival outcomes with weak labels.

Pipeline

- Each patient x_i may contain multiple whole slides. At the first step, we extracted patches from all WSIs belonging to the same patient and then cluster them into different phenotypes.
- Clustering:
 - Extract VGG features for each patch and then clustering based on those features.
 - By clustering different patches from all patients into several distinguished phenotype groups, we will have different phenotype groups with various prediction powers on patients' survival.

Experiments

- Data set:
 - Lung ADC from NLST and TCGA-GBM
- Compare with seven state-of-the-arts frameworks including six baseline models using traditional imaging features and one WSI model. The following Table shows the performance comparison of the proposed methods and other existing related methods using C-index values on two datasets.

	Lung-NLST	En-Cap	NLST	WISI-Bull	LightCnn	MTLSS	WSI4SA	Proposed
Long AUC	0.552	0.547	0.490	0.531	0.542	0.549	0.612	0.676
GCR	0.499	0.561	0.538	0.521	0.539	0.577	0.585	0.637

Kaplan-Meier survival curves of different models for NLST dataset.
High risk (greater than median) groups are plotted as green lines, and low risk (less than or equal to median) groups are plotted as red lines. The log rank test is conducted to test the difference of two curves and can evaluate how well the model classify testing patients into low and high risk groups. It is shown that the proposed method can achieve the most significant log rank test outcome while some of others do not reach statistical significances.

Conclusion

- We proposed a deep multi-instance model to directly learn survival patterns from gigapixel images without annotations.
- Compared to existing image-based survival models, the developed framework can learn holistic information of the patient using bag representations and achieve much better performance compared to state-of-the-art methods.



Bone assessment

End-to-end Convolutional Neural Network for Pediatric Bone Age Assessment

Chuanbin Liu, Hongqian Xie, Yirui Liang, Zheguan Zhu, Fanchao Lin, Yongdong Zhang
University of Science and Technology of China, Hefei, Anhui 230027, China
Email: jcb592@mail.ustc.edu.cn

National Engineering Laboratory for Brain-inspired Intelligence Technology and Application

Motivations

- Pediatric bone age assessment (BAA) is a common clinical procedure to investigate endocrinology, genetic and growth disorders of children.
- The single-stage structure can extract the annotation-specific bone parts and suffers from low accuracy.
- The multi-stage structure extracts the bone parts with human prior but suffers from model generalization and resource consumption problem.

Contributions

- Propose a novel single-stage Attention-Recognition Convolutional Neural Network (AR-CNN) for bone age assessment
- Enable an end-to-end training method extracting discriminative bone parts automatically without human prior
- State-of-the-art accuracy on the public RSNA datasets with MAE of 4.38 months.

Methods

Attention Agent

- Takes the radiography as input and produces a list of rectangle regions
- Resize the input radiography X with the size of 448, and choose anchors with scales [48, 96, 192] and ratios of [1:1, 3:2, 2:3]

Recognition Agent

- Assembling age assessment

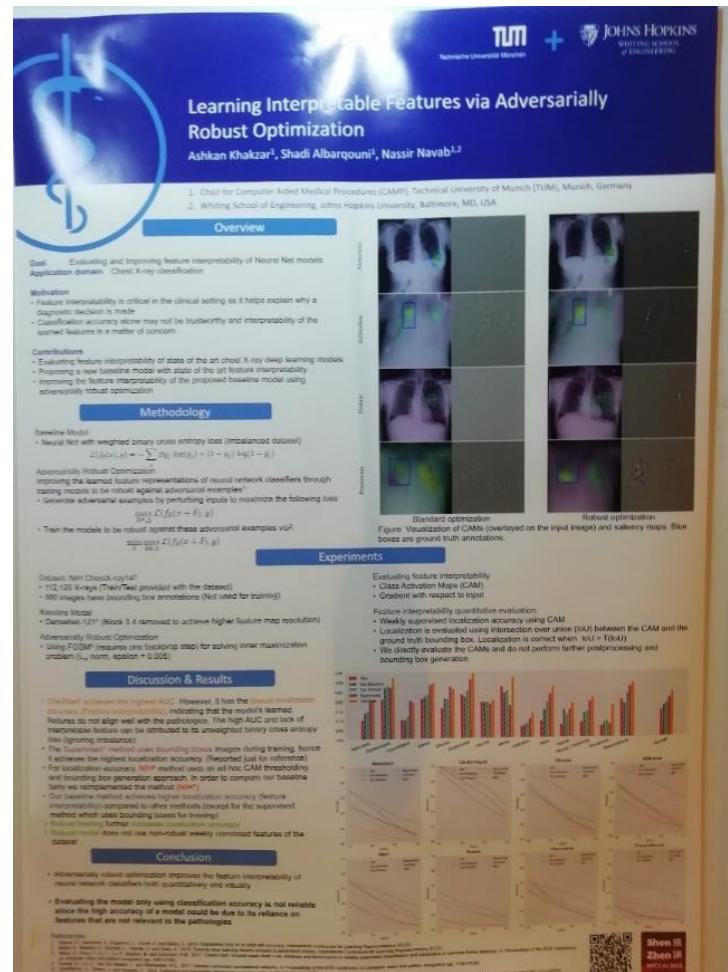
$$A_{\text{sum}} = \frac{1}{K+2} (A_C + A_S - \sum_{k=1}^K A_{R_k})$$

Results & Discussion

- Extracting three specific bone parts is enough in AR-CNN for Bone Age Assessment
- AR-CNN can extract the discriminative bone parts without human prior knowledge, and is free of segmentation or detection
- Experimental results show that AR-CNN achieves the best accuracy with MAE of 4.38 months
- Assessment result keeps strong consistency with the actual age. Meanwhile, the absolute deviation of our model is controlled within 20 months
- The human prior is not a size-invariant reference for bone age assessment task
- AR-CNN can also benefit from human knowledge and it could be a suggestion for a new clinical



Interpretable 深度学习可解释性



Transformer

Patch Transformer

Patch Transformer for Multi-tagging Whole Slide Histopathology Images

Weijian Li¹, Viet-Duy Nguyen¹, Haofu Liao¹, Matt Wilder², Ke Cheng¹, Jiebo Luo¹

¹Department of Computer Science, University of Rochester;
²Watson Inc., 703 Parkside Ave, Brooklyn NY 11226

Introduction

- Whole slide histopathology images are quiet challenging in clinical diagnosis
- Label tagging becomes a labor intensive procedure as more and more WSIs are collected nowadays
- Automatic and accurate tagging is necessary

Multi-tagging Whole Slide Images

Model	WSI	ROI	Spots	Tags
Stain	100	100	100	100
Species	100	100	100	100
Organ	100	100	100	100
Disease	100	100	100	100
Location	100	100	100	100

- Labeling each WSI with multiple attribute tags simultaneously
- Our dataset contains 4720 WSIs in total with 3 attribute tags
- Our goal is to correctly label at least 3 attribute tags for each WSI

Method

- No existing methods specifically designed for multi-tagging whole slide images
- Leverage patch information as well as estimated ROI location in general
- We designed our patch based deep model which consists of:
 - A patch transformation module
 - A multi-tag attention module
 - A modified Squeeze-and-Product Attention (SOPA) block

Patch Transformation Module

- Extracts patches from ROI level characteristics
- Only SOPA does not work well for the behavior of mining important features
- Compute attention weights following the multi-head attention algorithm
- Multiply the weighted patch representations for each tagging task

Multi-Tag Attention Module

- Aims to leverage patch level information for ROI level decision making
- Takes the output of Patch Transformation Module
- Compute attention weights following the same method as previous module
- Aggregates the weighted patch representations for each tagging task

$$W_{\text{tag}} = \text{Softmax}(W_{\text{tag}}^T \cdot \text{tanh}(\text{m} \cdot D)) \cdot W_N \in \mathbb{R}^{D \times T}, U_{\text{tag}} \in \mathbb{R}^{D \times T}$$

$$d_{\text{tag}} = \sum_{i=1}^T W_{\text{tag}} \cdot C_i = \sum_{i=1}^T W_{\text{tag}} \cdot \sum_{j=1}^N \text{softmax}(W_{\text{tag}}^T \cdot \text{tanh}(\text{m} \cdot D)) \cdot C_i$$

Experiments and Results

Quantitative Comparison

	Macro F1			Micro F1		
	Stain	Species	Organ	Stain	Species	Organ
PixelDE	0.477	0.478	0.487	0.872	0.857	0.866
PixelLM	0.487	0.794	0.875	0.875	0.884	0.871
GCR	0.601	0.632	0.602	0.917	0.857	0.891
DeepLab	0.602	0.867	0.662	0.981	0.882	0.899
TwoBranch	0.678	0.841	0.665	0.988	0.852	0.899
TwoBranch	0.692	0.889	0.671	0.982	0.892	0.944

Macro F1 performance comparing to two stage models, benefit of learning patch features in parallel.

GCN incorporates structural knowledge which helps the knowledge propagation but require detailed graph prior information.

TwoBranch model include extra knowledge: feature for stain tagging but not suitable for the other tags due to variation in patch features.

Qualitative Comparison

Figure left to right: tissue, stain, species, organ

Ablation Study

	Macro F1			Micro F1		
	Stain	Species	Organ	Stain	Species	Organ
MTA	0.697	0.526	0.575	0.891	0.879	0.885
PT Head	0.697	0.596	0.665	0.891	0.893	0.894
PT Head MTA	0.697	0.606	0.665	0.892	0.893	0.895
PT Head MTA	0.697	0.602	0.666	0.893	0.894	0.895
PT Head MTA	0.698	0.606	0.666	0.893	0.895	0.895
PT Head MTA	0.698	0.606	0.666	0.893	0.895	0.895

Multi-task learning helps the model by mutual promoting each tagging task.

MTA and PT head both results are better than others. PT head with multi-task learning achieves the best performance.

Stained-Box Product Attention has a negative effect.

Feature Transformer

Feature Transformers: Privacy-preserving lifelong learners for medical imaging

Rajiv Venkateswaran, Soham Chatterjee, Prasad Sudhakar, Pavan Anandaji and Hanharan Roushankar
GE Healthcare, India

Introduction

- lifelong learning in healthcare applications is recommended by diversity in
 - 1. Data domains and subject presentation
 - Acquisition protocols
 - Pathologies
 - Diseases
- lifelong learning use cases
 - Some tasks learning
 - 1. Incremental Learning
 - 2. More data, same domain
 - 2. Domain changes
 - Distribution changes
 - 3. New task learning
 - 4. Changes in applications
- Challenges for deep learning
 - Universality of full historic data
 - If available, solution is need
 - Privacy
 - Ability to store exemption data
 - Generalization forgetting
 - Performance degrades over time

Memory and loss function

Architectural

Training with composition loss

$$\text{Loss}_{\text{FTP}}(f^{\text{old}}, f^{\text{new}}) = \left(\text{Loss}_{\text{feature}}(\text{Image}_{\text{old}}, f^{\text{old}}) + \text{Loss}_{\text{feature}}(\text{Image}_{\text{new}}, f^{\text{new}}) \right) / 2$$

Experimental results

Existing solutions

- 1. Incremental learning
 - Add more capacity, special architectures, architectures
 - Freeze weights from previous episode to retain memory
 - Store exemption, episodic memory, GANs
 - Training prior an entire component or strategy or not cost to all variance of lifelong learning
- 2. New task learning
 - Our solution - Feature Transformers
 - generic framework for all variants of lifelong learning
 - generic framework for all variants of lifelong learning
 - freeze weights from original items

Feature transformers

- 1. Cross-domain representations for adding additional data
- 2. Incremental learning
- 3. Generalization
- 4. Episodic memory of historic data
- 5. Feature transformers for different types of data
- 6. Single framework setting various ML settings
- 7. Single framework setting various ML settings

Summary and conclusion

- One single framework for all variants of lifelong learning
- incremental induction of record capacity
- privacy preserving
- no need to store images, more features



Reinforcement Learning

Searching Learning Strategy with Reinforcement Learning for 3D Medical Image Segmentation

Dong Yang · Holger Roth · Zhiye Xu · Fausto Milletri · Ling Zhang · Daguang Xu

NVIDIA

Introduction

- Training the state-of-the-art 3D deep neural network models requires careful design of the workflow, data augmentation, learning rate schedule, loss functions, optimization and so on.
- We propose a reinforcement learning based approach to search the optimal training strategy of 3D neural networks for a specific 3D medical image segmentation task.
- Indirectly, we also tune the hyper-parameters (e.g., learning rate), data augmentation strategies, data pre-processing, etc.
- The performance of the baseline model is boosted after searching, and it can achieve comparable accuracy to the manually-tuned state-of-the-art segmentation approaches.

Background

- Recent works indicate that the full potential of current state-of-the-art network models may not yet be well-explored (e.g. nnUNet, winner of the MS3D challenge [1]).
- Automatic Machine Learning (AutoML) has been recently proposed to automatically search the best learning approaches and minimize human interaction at the same time (e.g. neural architecture search).
- Since 3D segmentation model is expensive to train, efficient 3D architecture search is extremely difficult to attain. Instead, a more feasible task is represented by hyper-parameter searching which still plays a critical role on the test-time performance.

Methodology

Objectives:

- Optimal training strategy searching

Approach:

- Components
 - Agent → RL controller
 - Trained and deployed
 - Proximal Policy Optimization (PPO)
 - State → Action parameter combination
 - Action → Hyper-parameter combination
 - Reward → Job submission accuracy
- Repeated scheme
 - Agent collects reward, updates itself
 - Agent assigns new jobs with new parameters

Job scheduling

Training jobs

- Master job thread
 - New GPU instance
 - Communicate with master via job ID during submission
- Launch new training jobs

Wait for next available GPU

Not reached limit

Keep new job IDs

Send value back

Update controller RNN

Experiments

Datasets:

- We use the training datasets from the Medical Segmentation Decathlon [11]

Accuracy metric: Dice's score = $\frac{2 \times \text{TP}}{(\text{TP} + \text{FP} + \text{FN})}$

Consistent Improvement over pre-defined strategies by users

Results

MS3D challenge

Method	Volumetric Acc.	Method	Volumetric Acc.
ResNet3D	0.67	DeepLabv3	0.70
ResNet3D+CRF	0.70	DeepLabv3+CRF	0.71
ConvNeXt3D	0.70	ConvNeXt3D+CRF	0.71
Deeplabv3	0.70	ConvNeXt3D	0.72

MS3D challenge

Method	Volumetric Acc.	Method	Volumetric Acc.
MS3D	0.69	MS3D+CRF	0.70
MS3D+CRF	0.70	MS3D+ConvNeXt	0.71
ConvNeXt	0.70	ConvNeXt+CRF	0.72
ConvNeXt+CRF	0.71	ConvNeXt+MS3D	0.72
ConvNeXt+MS3D	0.72	ConvNeXt+MS3D+CRF	0.73

Conclusions

- We propose a reinforcement learning based approach to search the optimal training strategy for 3D medical image segmentation.
- The proposed method can find better training strategies than manual ones.
- The proposed approach is highly parallelizable and can be easily applied to other learning tasks.
- Our results show that the proposed method can find better training strategies than manual ones.

Reference

1. Dong Yang, Holger Roth, Zhiye Xu, Fausto Milletri, Ling Zhang, and Daguang Xu. Searching Learning Strategy with Reinforcement Learning for 3D Medical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2022*, pages 20–28. Springer, Cham, 2022.
2. J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–115, 2014.
3. S. Ioffe and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *International Conference on Machine Learning*, pages 448–456, 2015.
4. D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2014.
5. Y. Nesterov. *Introductory Lectures on Convex Optimization: Theory and Algorithms*. Springer US, Boston, MA, 2004.
6. D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2014.
7. D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2014.
8. D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2014.
9. D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2014.
10. D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2014.
11. M. Balakrishnan, S. Verma, A. Majumdar, and A. Agapiou. Convolutional neural networks for medical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017*, pages 53–60. Springer, Cham, 2017.

A Deep Reinforcement Learning Framework for Frame-by-frame Plaque Tracking on Intravascular Optical Coherence Tomography Image

Gongqiang Liu, Suyao Dong, Kunqiang Wang*, Dong Zhang, Yue Gao, Xin Chen,
Hongguo Zhang, Shua Li
School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[Kunqiang Wang@hit.edu.cn](http://www.kunqiangwang.com)

Task: Frame-by-frame Plaque Tracking on Intravascular Optical Coherence Tomography Image (IVOCT image)

The proposed framework:

Challenges:

- IVOCT images are often noisy and the low contrast of plaque edges make it difficult to identify plaques without expert's guide.
- Complex and various intravascular morphology makes it difficult to achieve comprehensive and timely tracking of plaque movement.
- It is difficult to track the movement of IVOCT images in prior, because image-to-image is impossible without an efficient analysis method during the clinical routine.

Contributions:

- For the first time, we proposed an RL-based framework to achieve accurate plaque tracking in IVOCT images.
- The proposed framework made the spatio-temporal information of adjacent frames available, combining accurate plaque detection framework and tracking framework.
- The proposed model has good robustness, because the fully-supervised and semi-supervised tracking styles are both allowed to fit the clinical practice.
- On the clinical large-scale IVOCT data, the proposed method achieves high tracking accuracy.

Spatio-temporal correlation RL module

Aggregation:

- Aggregate a right sector in a plaque object to achieve precise location and scale quantification.

Environment:

- Sequence of successive IVOCT images.

Actions:

- Action space: right sector of the current frame.
- New designed Transformer module transforms the scene flexibility: our step actions to transform the registration process between frame and next frame to achieve better tracking results.

States:

- State $S = \{I_t, H_t, H_{t-1}\}$
- I_t : IVOCT feature encoded based on current frame.
- H_t : the recorded history location and scale of detected sector region.
- H_{t-1} : the recorded history location and scale of detected sector region.

STM module:

- Spatio-temporal correlation module: the spatio-temporal correlation information based on the fact that the past location, scale, and rotation are always related to current status without extra feature of the last frame.
- Aggregation: a right sector in a plaque object to achieve precise location and scale quantification.

Localization and identification module

- Multi-task framework: multi-task loss.
- Loss function: the network generate an initial plaque location and scale for IPI.
- Oracle: the network identify whether a plaque object exists in current frame.

Plaque localization and identification module

Localization and identification:

- Multi-task framework: multi-task loss.
- Loss function: the network generate an initial plaque location and scale for IPI.
- Oracle: the network identify whether a plaque object exists in current frame.

Equation:

$$L = \frac{1}{m} \sum_{i=1}^m L_i(d_i, g_i) + \frac{1}{m} \sum_{i=1}^m L_i(e_i, c_i) \quad (2)$$

Goal design:

- Improve the tracking accuracy, avoiding over-tracking.
- Include the spatio-temporal information to the current selected action.
- Spatio-temporal information transformation happens over frame.

Goal function:

$$G = \begin{cases} d_i, & I_i = 1 \& I_{i-1} = 1 \\ d_i, & I_i = 1 \& I_{i-1} = 0, I_0 = 0, i \geq 1, \\ NUL, & I_i = 0 \end{cases} \quad (3)$$

Spatio-temporal information transformation happens over frame:

$$G = \begin{cases} d_i, & I_i = 1 \& I_{i-1} = 1 \\ d_i, & I_i = 1 \& I_{i-1} = 0, I_0 = 0, i \geq 1, \\ NUL, & I_i = 0 \end{cases} \quad (4)$$

Abstract:

- IVOCT data with 132 patients, 10100 continuous frames (excluding 154 frames with experts' label on plaque location, size and identification).
- Training set: 2000 images, testing set: 500 images.
- All images are 128 × 128 pixels and each frame has a resolution of 150 × 150 pixels.

Experiment Settings:

- Model: ConvNeXt-16, Gradient accumulation learning rate: 0.0001.
- PPO: 152-layer and variant 60M layers based on loss function in 123 training epochs.
- FAL: Feature Attention Loss, which is a loss function for tracking RF module from P-AL. The feature attention loss is defined as $L_{FAL} = \alpha \cdot L_{RF} + (1 - \alpha) \cdot L_{P-AL}$.
- Training: 172 days (18 h), model training takes 100 epochs based on the enhancement of FAL, directly comparing with the proposed method.
- Loss function: $L = L_{RF} + L_{P-AL} + L_{FAL}$.

Conclusion:

- For the first time, we proposed a novel framework for accurate and fast plaque tracking framework on IVOCT images. The proposed framework can make the spatio-temporal information of adjacent frames available and accurate plaque detection, avoidance, prediction, estimation. The proposed method has great robustness, because the fully-supervised and semi-supervised tracking styles are both allowed to fit the clinical practice.

Results:

Dataset:

- IVOCT experiments prove the superiority of the proposed framework.
- Comparing with the existing methods, the proposed method achieves better tracking accuracy and faster tracking speed.
- On the large-scale IVOCT data, the proposed method acquired the high tracking accuracy and big tracking potential in future.

Tracking performance comparison among the proposed method, visual flow field, and feature attention loss (FAL) module. The proposed method achieves better tracking accuracy and faster tracking speed than FAL. By using the FAL module, the tracking accuracy is improved by 0.02%.

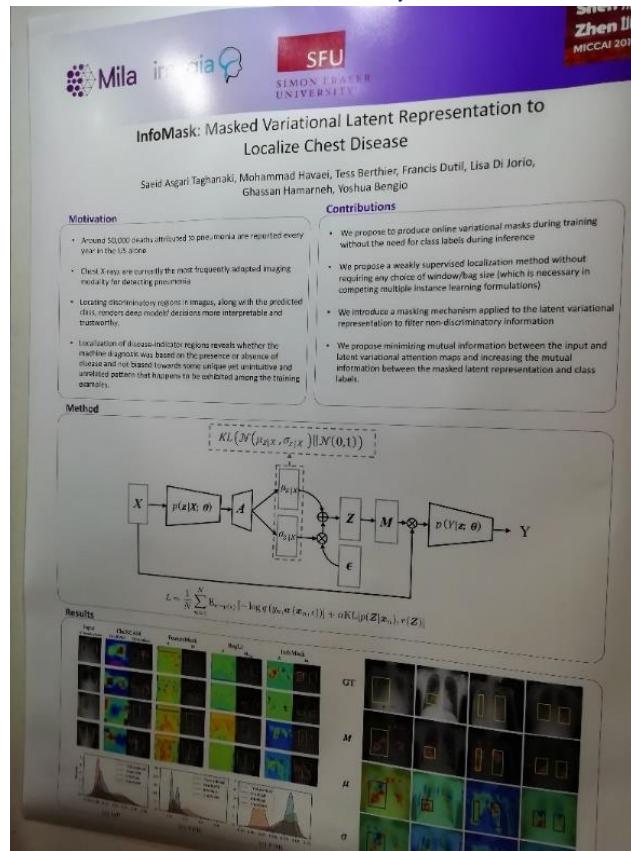
Loss:

	Proposed	Plaque local.
Mean accuracy	0.99	0.99
mean loss	0.08	0.08
mean FAL	0.92	0.93
mean visual	0.97	0.98
mean visual FAL	0.97	0.98
mean visual loss	0.02	0.02



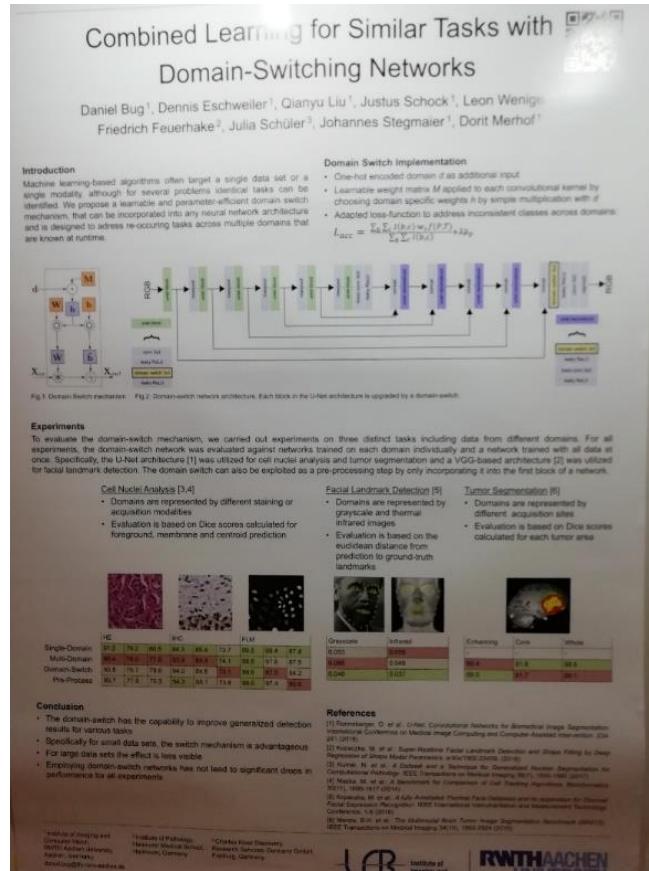
InfoMask: Masked Variational Latent Representation to Localize Chest Disease

Saeid Asgari Taghanaki, Mohammad Havaei, Tess Berthier, Francis Dutil, Lisa Di Jorio, Ghassan Hamarneh, **Yoshua Bengio**

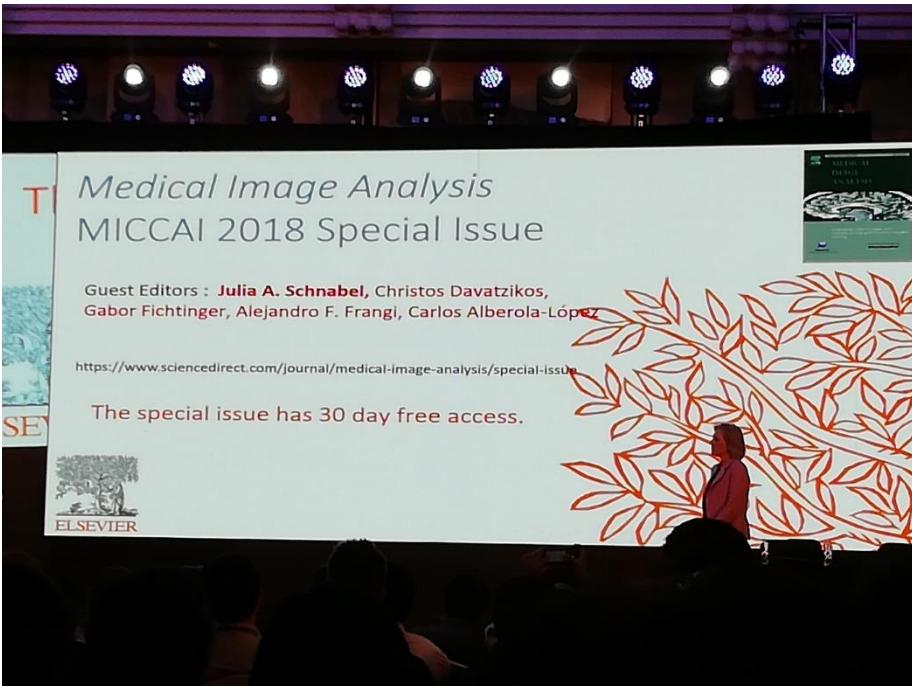


Transfer Learning

Combined Learning for Similar Tasks with Domain-Switching Networks



Resources



Resources

NOVEMBER 2019

Computer Vision News

The Magazine of the Algorithm Community

Best of MICCAI 2019
14 pages

MICCAI's Keynote Speaker
Qiyong Gong

Women in Computer Vision
Hyo Jin Kim

We Tried for You
Tesseract OCR with Python

Computer News Events



Research Paper Review:
Generating High Fidelity Images with Subscale Pixel Networks and Multidimensional Upscaling.

 RSIP VISION
Global Leader in Computer Vision and Deep Learning

<https://www.rsipvision.com/ComputerVisionNews-2019November/>
<https://www.rsipvision.com/>

 **RSIP VISION** Global Leader in Computer Vision and Deep Learning

Home Our Work Consulting Blog News / Events Magazine Careers Contact Us

Effective AI and Computer Vision Solutions for Medical Imaging

RSIP Vision provides Computer Vision and Image Processing outsourcing and services for the broadest range of medical imaging fields: cardiology, pulmonology, ophthalmology, orthopedics, radiology and more; and also for microscopy image analysis, digital pathology, pharma and all kind of machine learning projects. Our engineers are experts in artificial intelligence, deep learning and all the most advanced computer vision techniques.

We also publish [Computer Vision News](#), the online magazine of the algorithm community.

OUR FIELDS OF EXPERTISE



Resources

- [MICCAI2019-OpenSourcePapers](#)
- <https://github.com/JunMa11/MICCAI2019-OpenSourcePapers>

MICCAI 2019 Open Source Papers

Part I, LNCS Volume 11764 Optical Imaging; Endoscopy; Microscopy

Title	First Author	PDF	Code
A Deep Reinforcement Learning Framework for Frame-by-Frame Plaque Tracking on Intravascular Optical Coherence Tomography Image	Gongning Luo	LNCS	Code
Boundary and Entropy-Driven Adversarial Learning for Fundus Image Segmentation	Shujun Wang	LNCS	Code
Probabilistic Atlases to Enforce Topological Constraints	Udaranga Wickramasinghe	LNCS	Code
Synapse-Aware Skeleton Generation for Neural Circuits	Brian Matejek	LNCS	Code
Seeing Under the Cover: A Physics Guided Learning Approach for In-bed Pose Estimation	Shuangjun Liu	LNCS	Code
Ki-GAN: Knowledge Infusion Generative Adversarial Network for Photoacoustic Image Reconstruction In Vivo	Hengrong Lan	LNCS	Code
Triple ANet: Adaptive Abnormal-aware Attention Network for WCE Image Classification	Xiaoqing Guo	LNCS	Code



CALL FOR PAPERS

- You are invited to submit your full paper to MICCAI 2020 in Lima, Peru. The paper submission system will open on 15th January, 2020.
- As in previous years, you must register your intention-to-submit by Midnight, Pacific Time, 18th February, 2020. You will not be able to submit your paper if you do not register your intention-to-submit. Please see important dates on the website for further details on the submission and review the schedule.
- The full paper submission deadline will be Midnight, Pacific Time, 3rd March, 2020. There will be no extension. You will be able to edit your submission up to the deadline.
- In addition to advances in methodology, we encourage submission of papers that demonstrate clinical relevance and novel clinical applications. Topics discussed at MICCAI include, but are not limited to:
 - Applications of Big Data in Imaging; Integration of Imaging with Non-Imaging Biomarkers; Population Imaging and Imaging Genetics; Computational Anatomy and Physiology; Computational (Integrative) Pathology; Visualisation in Biomedical Imaging; Image Segmentation, Registration and Fusion; Image Reconstruction and Image Quality; Optical and Photo-acoustic Imaging; Neuro Imaging; Microscopy Image Analysis; Machine Learning, Deep Learning and Artificial Intelligence; Computer Aided Diagnosis; Interventional Imaging Systems
 - Image-Guided Interventions and Surgery; Surgical and Interventional Simulation Systems; Medical Robotics and Haptics; Surgical Data Science; Surgical Visualization and Mixed, Augmented and Virtual Reality; Surgical Planning and Simulation; Surgical Skill and Work Flow Analysis
- Submissions that specifically address accessible healthcare technologies in developing countries, through for example innovations in point-of-care imaging, are highly encouraged.
- Papers will consist of maximum 8 pages (text, figures and tables) + up to 2 pages for references only. They are submitted electronically in LNCS style, double blind review, to the CMT system.
- The papers will be evaluated by external reviewers and Area Chairs for inclusion in the scientific program of MICCAI.
- Equity, diversity and inclusion are essential to MICCAI excellence. In addition, for MICCAI 2020, we would also like to enhance the values historically cherished in LatAm: the fairness and the responsibility of the proposed models and protocols, as the accessibility and the sustainability of the suggested methods and technologies.
- An open and diverse community fosters the inclusion of voices that have been underrepresented or discouraged. We encourage submissions from members of groups that have been marginalized on any grounds.
- Summary of Important Dates for Submission to MICCAI 2020
 - Submission System Opens: 15th January, 2020
 - Intention to Submit: 18th February, 2020
 - Full Paper Submission: **3rd March, 2020**
 - Final Decision Notification: 9th June, 2020
- MICCAI 2020 Program Executive
- program-chairs@miccai2020.org



Thank You !

