AIS-THB

Artificial Intelligence Thursday Poster Discussions

Thursday, Dec. 5 12:45PM - 1:15PM Room: AI Community, Learning Center

AI IN

AMA PRA Category 1 Credit ™: .50

Participants
Ian Pan, MA , Providence, RI (*Moderator*) Consultant, MD.ai

Sub-Events

AI200-SD-THB1 A Federated Convolutional Denoising Autoencoder for MRI Applications

Station #1

Participants
Sebastian Niehaus, MSc, Berlin, Germany (*Abstract Co-Author*) Data Scientist, AICURA Medical GmbH
Alberto Merola, PhD, Berlin, Germany (*Presenter*) Data Scientist, AICURA Medical GmbH

For information about this presentation, contact:

sebastian.niehaus@aicura-medical.com

CONCLUSION

The proposed federated approach improves the performance of independent CDAEs without direct access to the data, enabling us to effectively address the critical issues of data availability (in terms of both sample size and restricted access) and model generalization for clinical applications.

Background

Denoising is an essential part of medical images preprocessing and autoencoders became the state-of-the-art method for it. Their performance depends on training images with noise characteristics ideally as heterogeneous as that found in the real word. This is best achieved by pooling images from different hospitals with varied noise sources, which crucially entails data privacy limitations.

Evaluation

240 MPRAGE MRI volumes from the publicly available ADNI dataset (adni.loni.usc.edu) were equally divided into 4 virtual hospitals (VHs). Data were first interpolated to a common matrix size (96x128x128) and rescaled. Noisy volumes were then created from these by simulating, five noise sources commonly found in clinical MRI: thermal noise, Gibbs ringing, k-space spikes, ghosting and zipper artifacts. For each VH a unique combination of the noise artifacts was used. For the denoising we trained 3D convolutional denoising autoencoders (CDAEs): 1. on the dataset of a single VH; 2. on the pooled dataset; 3. with a federated learning approach. The federated CDAE (3.) is obtained by training single models for each VH independently and combining them after random intervals. We quantified the performances of the three models by calculating the structural similarity index (SSIM) comparing the original images and the calculated test denoised images using the single-VH CDAE (SSIM=0.69), the pooled CDAE (SSIM=0.79) and the federated CDAE (SSIM=0.73).

Discussion

Image denoising with the federated CDAE outperforms results obtained with the single-hospital CDAE, showing the effectiveness of the federated learning approach. This indicates that it is possible to train CDAE models with relatively small local datasets and high noise heterogeneity and then combine them with federated learning into a more accurate and transferable model.

AI250-SD-THB2 Medical Federated Deep Learning (MedFDL) for Automatic Body Part Labeling of CT Scout Images

Station #2

Participants
Ross W. Filice, MD, Washington, DC (*Presenter*) Co-founder, DexNote LLC; Research Grant, NVIDIA Corporation; Advisor, BunkerHill Health, Inc
Ian Pan, MA , Providence, RI (*Abstract Co-Author*) Consultant, MD.ai
Anouk Stein, MD, Paradise Valley, AZ (*Abstract Co-Author*) Consultant, MD.ai, Inc; Stockholder, MD.ai, Inc
Laura P. Coombs, PhD, Reston, VA (*Abstract Co-Author*) Nothing to Disclose
George L. Shih, MD, New York, NY (*Abstract Co-Author*) Consultant, MD.ai, Inc; Stockholder, MD.ai, Inc;

For information about this presentation, contact:

ross.w.filice@gunet.georgetown.edu

PURPOSE

Body part information in radiology exam metadata is often unreliable. Deep learning models that can determine body parts from CT scout images could help estimate radiation dose and improve clinical workflow through hanging protocols and relevant priors. Patient privacy concerns limit data sharing across institutions which otherwise might facilitate a more generalizable model. We hypothesize that a multi-label, multi-class deep learning model can reliably identify body parts in CT scout images and that sequential federated training on two different datasets can attain comparable performance to pooled data.

METHOD AND MATERIALS

565 CT scout images from institution 1 and 534 from institution 2 were anonymized, cropped to remove blank space, resized to 256 x 256, and then labeled with any number of 7 body parts (head, neck, chest, abdomen, pelvis, upper extremity, lower extremity). Labeled data was randomly split into 80% train, 10% validation, and 10% test sets. A multi-label, multi-class model based on MobileNetV2 architecture using the Keras deep learning library was trained first on data solely from institution 1, then institution 2, and then on pooled and shuffled data. Accuracy was compared for single institution models, naive cross-testing, single institution models retrained sequentially, and pooled data.

RESULTS

Frequency of body part labels ranged from 7.9-41.1% for institution 1 and 6.1-43.0% for institution 2. For institution 1, 81.5% of test predictions were completely correct for all labels while 98.1% were at least partially correct (92.6-100% across all body parts); results were 84.9% and 96.2% for institution 2 (90.6-100%). Many incorrect labels were defensible as scout images often partially included additional body parts. Results for the model from institution 1 were 58.5% and 75.5% when tested naively on institution 2; the reverse was 55.6% and 96.3%. If model 1 is re-tuned on data from institution 2, results are 79.2% and 98.1% with the reverse 77.8% and 98.1%. When all data is pooled to generate a single model results are 79.4% and 96.3%.

CONCLUSION

A deep learning model can reliably detect most body parts from CT scout images. Sequential federated training on two different datasets across institutions produces comparable results to isolated training on pooled data.

CLINICAL RELEVANCE/APPLICATION

Reliably detecting body part from CT scout images can help with radiation dose estimation as well as clinical workflow.

AI202-SD-THB3 Vulnerability of Deep Learning based Computer-Aided Diagnosis: Experimental Adversarial Attack Against CT Lung Nodule Detection Model

Station #3

Participants
Chang Yong Heo, BS, Seoul, Korea, Republic Of (*Presenter*) Nothing to Disclose
Jaewon Lee, Suwon , Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Sihwan Kim, MSc, Seoul, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Do Il Lee, Suwon, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Chul Kyun Ahn, Seoul, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Changwon Kim, Seoul, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Jong H. Kim, PhD, Seoul, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose

CONCLUSION

Deep learning models possessed vulnerability to perturbed input, and showed varying degree of performance degradation to different attacks. Deep learning models for CAD system needs to be verified with respect to their vulnerability to perturbation and adversarial attacks.

Background

Recent developments of deep learning technique have drawn attention from medical imaging community with outstanding performance and appear to give promise for future applications in computer-aided diagnosis. However, there still exist concerns about inherent uncertainty of the behavior of deep learning models, which needs to be thoroughly investigated before clinical translation. Adversarial attack is a useful technique for testing deep learning models by exposing them to a set of intentionally perturbed examples and evaluating the performance degradation. This study investigates the vulnerability of deep learning models for basic object classification and CT nodule detection tasks.

Evaluation

We evaluated the vulnerability of three deep learning models each trained with MNIST, CIFAR-10, and LIDC-IDRI dataset. Four latest adversarial attack algorithms were employed to generate adversarial examples for perturbing the first two deep learning models, and selected an appropriate attack algorithm for use in the test of the deep learning model for CT lung nodule detection.

Discussion

The classification performance of MNIST-trained deep learning model degraded from 0.98 before attack to 0.70, 0.78, 0.01, and 0.02 after attack by four different algorithms. The performances of CIFAR-10-trained model also degraded from 0.73 before attack to 0.11, 0.16, 0.13, and 0.02 after attacks. Performance of the CT lung nodule detection model showed gradual degradation according to the increasing degree of perturbation: AUROC was 0.95 before attack, and decreased to 0.915, 0.903, 0.890 after attack; sensitivity was 0.877 before attack, and decreased to 0.854, 0.807, 0.717 after attack.

AI002-EC-THB Generative Adversarial Network Models for Prediction of Survival in Patients with Interstitial Lung Diseases

Custom Application Computer Demonstration

Participants
Tomoki Uemura, MS,BA, Boston, MA (*Presenter*) Nothing to Disclose
Chinatsu Watari, MD, Boston, MA (*Abstract Co-Author*) Nothing to Disclose
Janne J. Nappi, PhD, Boston, MA (*Abstract Co-Author*) Royalties, Hologic, Inc Royalties, MEDIAN Technologies
Toru Hironaka, Boston, MA (*Abstract Co-Author*) Nothing to Disclose
Hyoungseop Kim, PhD, Kitakyushu, Japan (*Abstract Co-Author*) Nothing to Disclose
Hiroyuki Yoshida, PhD, Boston, MA (*Abstract Co-Author*) Patent holder, Hologic, Inc; Patent holder, MEDIAN Technologies;

For information about this presentation, contact:

yoshida.hiro@mgh.harvard.edu

TEACHING POINTS

Generative adversarial networks (GANs) can be used for predicting survival of patients from their chest CT images. The teaching points of this exhibit are (1) to explain the role of GAN in survival analysis, (2) to introduce state-of-the-art GAN models for survival estimation of patients with interstitial lung disease (ILD), and (3) to provide recommendations for GAN-based survival estimation in chest CT images.

TABLE OF CONTENTS/OUTLINE

1. Introduction 2. Generative adversarial networks (GANs) 3. GAN-based survival analysis methods in chest CT images 4. Interstitial lung diseases (ILDs) 5. GAN-based survival estimation of patients with ILDs 6. Recommendations

AI009-EC-THB Common-Space-Learning from Multi-Modality for Missing MRI Synthesis and Glioma Grading

Custom Application Computer Demonstration

Participants
Pu Huang, PhD, Jinan , China (*Abstract Co-Author*) Nothing to Disclose
Han Zhang, Chapel Hill, NC (*Presenter*) Nothing to Disclose
Zhicheng Jiao, Chapel Hill, NC (*Abstract Co-Author*) Nothing to Disclose
Dongming Wei, Shanghai, China (*Abstract Co-Author*) Nothing to Disclose
Feng Shi, Shanghai, China (*Abstract Co-Author*) Employee, Shanghai United Imaging Healthcare Co, Ltd
Dengwang Li, Jinan, China (*Abstract Co-Author*) Nothing to Disclose
Dinggang Shen, Chapel Hill, NC (*Abstract Co-Author*) Nothing to Disclose

For information about this presentation, contact:

hanzhang@med.unc.edu

Conclusion

Our model achieved more accurate glioma grading than simply using single modality and eventually led to a comparable performance to that with complete modalities.This provides a clinically feasible grading solution with limited imaging modality

Background

Gliomas is one of the most lethal cancers.Early and accurate preoperative glioma grading based on MRI is important for personized therapy.Despite multimodality data provide complementary information,a complete set of high-resolution multimodality MRI is costly and usually impossible to acquire in clinical settings,where T1 MRI is more commonly acquired.To leverage more comprehensive multimodality information for better glioma grading instead of doing so with T1 MRI only,we introduce a novel common-space-learning-based deep-learning model for missing MRI image synthesis based on T1 MRI and use the comprehensive feature from the common space to achieve a clinically feasible glioma grading

Evaluation

In the training set,common space is estimated by utilizing four MRI modalities(T1,T1c,T2,FLAIR) with adversarial context-aware learning.In the testing set where only T1 MRI is available,we synthesize other three modalities,compute their features in the common space,and perform the glioma grading.Results show that our method achieves greatly improved accuracy,sensitivity,and specificity of 0.879,0.851,and 0.909 against the conventional T1 MRI-based grading of 0.812,0.791,and 0.833.In particular,the tumor core is correctly estimated in the synthesized T1c,which is better than the state-of-the-art methods(i.e., c-GAN).Gradient-weighted Class Activation Mapping(GradCAM)further verifies that the tumor lesions contribute to the grading

Discussion

Our model has three main innovations:1)It is trained with four modalities,both the inherent inter-modality relationship and the lesion-specific representation are explicitly encoded in common space to jointly guarantee the robustness;2)The common space can comprehensively represent multimodality information,making the grading more accurate than traditional single modality-based grading;3)The edge-ware and tumor-aware learning strategy enable to capture the context information for improvement of the synthesized tumor