研究終了報告書

「解釈可能なインタラクティブ深層学習」

研究期間：2019年10月～2022年3月
研究者：谷 林

１． 研究のねらい

This project aims to propose a novel framework to automatically generate annotation, contextual labelling and semantically reasoning via invasively collecting user' gaze and texture input. In the research plan, there are three goals: (1) Automatically generate accurate pixel-wise label from user' gaze and texture input. (2) Interpret deep learning of its reasoning. (3) A large scale system to generate the pixel-wise label for medical image.

２． 研究成果

（１）概要

Thanks to the support from ACT-X, this project has resulted in the fruitful outputs. These results are from joint international collaborations across the world. In the original plan, there are three goals: (1) Automatically generate accurate pixel-wise label from user' gaze and texture input. (2) Interpret deep learning of its reasoning. (3) A large scale system to generate the pixel-wise label for medical image.
I have completed first two goals. However, the goal 3 is not completed due to the COVID-19.

（２）詳細

Thanks to the support from ACT-X, this project has resulted in the fruitful outputs including 6 international journals (IEEE JHBI, IEEE TMM, PR, etc.) and 8 international conference proceeding papers (CVPR, ICCV, ECCV, BMVC, etc.).

This project aims to propose a novel framework to automatically generate annotation, contextual labelling and semantically reasoning via invasively collecting user' gaze and texture input.

Goal 1: Research on generate accurate pixel-wise label from user' gaze and texture input.
This goal is completed. Our CVPR 2021 work achieves discriminative attribute localization guided by the actual human gaze and the attribute descriptions on

general image. To further extend it to medical image, our BMVC 2021 work (1) generates the pixel-wise segmentation on both 2D and 3D medical data based on the actual human gaze. However, due to the physical contact restriction of Covid-19, I have only collected limited number of gaze data. Therefore, I have proposed a novel self-supervised method that allows the model training under limited and low quality data. This work (2) is published in a top AI conference: ICCV 2021.

Goal 2: Research on deep learning interpretability.
To achieve the goal 2, several strategies are proposed to explain the reasoning of deep learning. The representative work (3) is published in a top medical journal IEEE JHBI that explains the deep neural network from the perspective of Koch's Postulates, the foundation in evidence-based medicine. The proposed method could automatically find the location and types of symptoms that the DR detector identifies as evidence to make prediction. To incorporate more external knowledge, a knowledge graph based method has been proposed to automatically reason the semantic dependencies among objects. This method is published in ICIP and could be used to explain the high-level semantic contexts.

Goal 3: Research on a large scale system to generate the pixel-wise label for medical image.
However, the goal 3 of a system for large-scale medical data is not completed due to the COVID-19. My research is particularly disturbed by the restricted physical contact for two reasons: 1. The medical relevant data is highly sensitive that could only be assessed locally. 2. The collection of gaze requires recording the doctors' action with a special device, gaze tracker.

３． 今後の展開

Now this research has attracted global attention from both academic and industrial fields. As far as I know, several groups across the world are starting collecting and utlising the gaze data following my proposed methodology.

The proposed system has potential to become the standard system for clinics to conduct routine medical imaging examine. This would also be used to for the annotation of general 2D image and even 3D data. With this system, large scale data with labelling could be collected at almost zero cost.

４． 自己評価

Thanks to the support from ACT-X, I have the unique opportunity to collaborate with global researchers on the breakthrough research on utlising human gaze to enhance the artificial intelligence. During the research, I have published several papers on top AI conferences and journals.

In the original plan, I have proposed three goals. Though the first two goals are completed, the third goal is postponed due to Covid-19. My research is particularly disturbed by the restricted physical contact for two reasons: 1. The medical relevant data is highly sensitive that could only be assessed locally. 2. The collection of gaze requires recording the doctors' action with a special device, gaze tracker.

Apart from the grant itself, the support from ACT-X advisors and other peer ACT-X researchers is equally important for my research. Without these advice and support during 2 years, it is impossible to achieve the existing progress. ACT-X also provides a valuable chance to communicate and collaborate with other peer ACT-X researchers.

The proposed research supported by ACT-X has attracted global interesting. Now, multiple academic and industrial institutes are following this work and investing on collecting large scale gaze to enhance the attention mechanism of AI. It would a keystone technique to relieve the data-hungry bottleneck of AI by providing large-scale pixel-wise label at almost zero-cost.

５． 主な研究成果リスト

（1）代表的な論文（原著論文）発表

研究期間累積件数：3件

Yifei Huang, Xiaoxiao Li, Lijin Yang, **Lin Gu***, Yingying Zhu, Hirofumi Seo, Qiuming Meng, Tatsuya Harada, Yoichi Sato. Leveraging Human Selective Attention for Medical Image Analysis with Limited Training Data. Proceeding of The British Machine Vision

Conference (BMVC). 2021.　*Lin Gu is the corresponding author.

Abstract: Human gaze is a cost-efficient physiological data that reveals human underlying attentional patterns. The selective attention mechanism helps the cognition system focus on task-relevant visual clues by ignoring the presence of distractors. Thanks to this ability, human beings can efficiently learn from a very limited number of training samples. Inspired by this mechanism, we aim to leverage gaze for medical image analysis tasks with small training data. Our proposed framework includes a backbone encoder and a Selective Attention Network (SAN) that simulates the underlying attention. The SAN implicitly encodes information such as suspicious regions that is relevant to the medical diagnose tasks by estimating the actual human gaze. Then we design a novel Auxiliary Attention Block (AAB) to allow information from SAN to be utilized by the backbone encoder to focus on selective areas. Specifically, this block uses a modified version of a multi-head attention layer to simulate the human visual search procedure. Note that the SAN and AAB can be plugged into different backbones, and the framework can be used for multiple medical image analysis tasks when equipped with task-specific heads. Our method is demonstrated to achieve superior performance on both 3D tumor segmentation and 2D chest X-ray classification tasks. We also show that the estimated gaze probability map of the SAN is consistent with an actual gaze fixation map obtained by board-certified doctors.

2. Multitask AET with Orthogonal Tangent Regularity for Dark Object Detection. Ziteng Cui, Guo-Jun Qi, **Lin Gu***, Shaodi You, Zenghui Zhang, Tatsuya Harada. Proceeding of International Conference on Computer Vision (ICCV). 2021. *Lin Gu is the corresponding author

Abstract: Dark environment becomes a challenge for computer vision algorithms owing to insufficient photons and undesirable noises. Most of the existing studies tackle this by either targeting human vision for better visual perception or improving the machine vision for specific high-level tasks. In addition, these methods rely on data argumentation and directly train their models based on real-world or over-simplified synthetic datasets without exploring the intrinsic pattern behind illumination translation. Here, we propose a novel multitask auto encoding transformation (MAET) model that combines human vision and machine vision tasks to enhance object detection in a dark environment. With a self-supervision learning, the MAET learns an intrinsic visual structure by encoding and decoding the realistic illumination-degrading transformation considering the physical noise model and image signal processing (ISP). Based on this representation, we achieve object detection task by decoding the bounding box coordinates and classes. To avoid the over-entanglement of two tasks, our MAET disentangles the object and degrading features by imposing an orthogonal tangent regularity. This forms a parametric manifold along which multitask predictions can be geometrically formulated by maximizing the orthogonality between the tangents along the outputs of respective tasks. Our framework can be implemented based on the mainstream object detection architecture and directly trained end-to-end using the normal target detection datasets, such as COCO and VOC. We have achieved the state-of-the-art performance using synthetic and real-world datasets.

3. Yuhao Niu, **Lin Gu**, Yitian Zhao, Feng Lu. Explainable Diabetic Retinopathy Detection and Retinal Image Generation. IEEE Journal of Biomedical and Health Informatics. 2021.

Though deep learning has shown successful performance in classifying the label and severity stage of certain diseases, most of them give few explanations on how to make predictions. Inspired by Koch's Postulates, the foundation in evidence-based medicine (EBM) to identify the pathogen, we propose to exploit the interpretability of deep learning application in medical diagnosis. By isolating neuron activation patterns from a diabetic retinopathy (DR) detector and visualizing them, we can determine the symptoms that the DR detector identifies as evidence to make prediction. To be specific, we first define novel pathological descriptors using activated neurons of the DR detector to encode both spatial and appearance information of lesions. Then, to visualize the symptom encoded in the descriptor, we propose Patho-GAN, a new network to synthesize medically plausible retinal images. By manipulating these descriptors, we could even arbitrarily control the position, quantity, and categories of generated lesions.

We also show that our synthesized images carry the symptoms directly related to diabetic retinopathy diagnosis. Our generated images are both qualitatively and quantitatively superior to the ones by previous methods. Besides, compared to existing methods that take hours to generate an image, our second level speed endows the potential to be an effective solution for data augmentation.

（2）特許出願

研究期間全出願件数:0件（特許公開前のものも含む）

1	発明者	
	発明の名称	
	出願人	
	出願日	
	出願番号	
	概要	
2	発明者	
	発明の名称	
	出願人	
	出願日	
	出願番号	
	概要	

（3）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

Cardiology and Cardiac Surgery chapter in the clinical textbook Artificial Intelligence in Clinical Medicine