IETR
INSTITUT D'ÉLECTRONIQUE ET DE TÉLÉCOMMUNICATIONS DE RENNES

INSA
INSTITUT NATIONAL
DES SCIENCES
APPLIQUÉES
RENNES

Tittle :
Deep-learning-based model for saliency prediction

Advisors :

Supervisor :
Ass. Pr. Lu Zhang, ☎ (+33)2 23 23 88 12 ✉lu.ge@insa-rennes.fr
Co-supervisor :
Ass. Pr. Le Meur, ☎ (+33)2 99 84 74 25 ✉ olemeur@irisa.fr

Laboratory : VAADER Team, IETR Laboratory, UMR CNRS 6164, Rennes, France

Advisors short bio :

Dr **Lu ZHANG** is an associate professor at the National Institute of Applied Sciences (INSA) of Rennes, France. She is also a member of the Institute of Electronics and Telecommunications of Rennes (IETR), UMR CNRS 6164. She received the M.S. degree from Shanghai Jiaotong University in 2007. Then she participated in the Engineering Leadership Program (ELP) in National Instruments (NI) at Shanghai for two years. From October 2009 to November 2012, she was a phD student at the University of Angers, and at laboratories LISA (renamed as LARIS now) and IRCCyN (renamed as LS2N now) in France. Her thesis topic was “Numerical observers for the objective quality assessment of medical images”. Then she worked on the Quality of Experience (QoE) in Telemedicine as a research engineer before she joined INSA and IETR in September 2013. Her PhD thesis was awarded (in french, “prix de thèse”) by IEEE France Section, SFGBM, AGBM and GdR CNRS-Inserm Stic-Santé.
Since 2010, Dr. Lu ZHANG is the co-author of 14 international journal papers, 27 conference papers and 5 french conference papers. She co-supervised 6 PhD students, 4 of them defended or will defend their theses before 2020. Dr. Lu ZHANG became a member of the Video Quality Experts Group (VQEG) in 2013. She was an invited speaker at the 6th Qualinet General Meeting. She was invited to give seminars by several chinese universities or research institutes several times. She co-chaired the special session on “Quality Assessment for Medical Imaging Applications” in QoMEX 2018. She is the project leader of an ANR (France National Agency for Research) ASTRID (Specific Support for Defence Research Projects and Innovation) project from 2018 to 2020.
Personal Website : http://luzhang.perso.insa-rennes.fr/

Dr **Olivier Le Meur** obtained his PhD degree from the University of Nantes in 2005. From 1999 to 2009, he has worked in the media and broadcasting industry. In 2003 he joined the research center of Thomson-Technicolor at Rennes where he supervised a research project concerning the modelling of the human visual attention. Since 2009 he has been an associate professor for image processing at the University of Rennes 1. In the IRISA/SIROCCO team his research interests are dealing with the understanding of the human visual attention. It includes computational modelling of the visual attention and saliency-based applications (video compression, objective assessment of video quality, retargeting).

Thesis topic :

Visual attention is the mechanism allowing to focus our visual processing resources on behaviorally relevant visual information. Eye movements, revealing where and how observers look within a scene, are the key factor of visual attention [1]. Eye movements are mainly composed by fixations and saccades. Fixations aim to bring objects of interest onto the fovea, where the visual acuity is maximum. Saccades are ballistic changes in eye position, allowing to jump from one position to another [2].

The research of the model for predicting eye fixations provides a strong theoretical and application support for understanding the nature of our visual system. As an important technic of artificial intelligence (AI), the model can be expand to a lot of applications such as judging the level of fatigue of a driver [3], compressing an image according to human attention [4] , patient diagnosis, human–computer interfaces and helping robot to develop spontaneous view movement for inspection [5].

Deep learning models are loosely related to information processing and communication patterns in a biological nervous system, such as neural coding that attempts to define a relationship between various stimuli and associated neuronal responses in the brain [6]. Until now, it is the most important and efficient model in the field of AI [7-8]. Since the visual system of human is biologically composed of stimuli and neuronal responses between eyes and brain, the deep learning model turns out to be the most natural way to model our visual system. Traditional modeling methods often construct a probabilistic distribution [9], like Hidden Markov Models (HMM), to predict the saccade amplitudes and orientations according to some features extracted from image. owever, this representation is far to be able to grasp the complexity of viewing behavior, and many properties of the visual system have been left aside. The deep learning method can automatically extract high-level features from the data and model the distribution of eye fixations in a more accurate way.

The rise of deep learning methods has shown that many difficult computer vision problems can be solved by machine learning algorithms relying on Convolution Neural Networks (CNNs), Deep Belief Networks (DBN) and Recurrent Neural Networks (RNN). Take CNN as an example, when applied on images, CNNs consist of multiple layers of small neuron collections, which process portions of the input image. The convolutional layer is the core building block of a CNN. The layer's parameters consist of a set of learnable kernels; those weights are learned during the back-propagation step, which aims to reduce the predictor error, i.e. the difference between the prediction and the actual value. A number of deep learning-based saliency model has been very recently proposed [10-13]. Kruthiventi et al. [10] proposed a first-of-its-kind fully convolutional neural network for accurate saliency prediction. Their model can automatically learn features in a hierarchical fashion and predict saliency map in an end-to-end manner. As the first end-to-end CNNs trained and tested for the purpose of saliency prediction, Pan .et al. [12] proposed a shallow convnet trained from scratch, and a another deeper solution whose first three layers are adapted from another network trained for classification. Although these existing models have made a large promotion in this area, their works are mostly based on the context of image and do not consider the cognitive features such as gender, age and emotion of observers. Thus, it still has a lot of room for promotion in this area.

The objective of the proposed Phd thesis is to design a new deep-learning-based model for predicting eye fixations. Specifically, the objective is separated as follow:

1. The host foreign team have already proposed two saccadic models [1][9]. The first objective of the Phd theses is to revisit the proposed saccadic models in order to improve their ability to predict visual scanpaths and saliency maps.

2. The PhD candidate are supposed to get deep in the theoretical foundation in deep learning method and build effective network boost in a significant manner the performance.
3. The PhD candidate will then extend the model to the application of new media, e.g. omnidirectional images, HDR images, and to other potential field such as driver monitoring and human–computer interfaces.

References:

[1]. Le, Meur O, and A. Coutrot. "Introducing context-dependent and spatially-variant viewing biases in saccadic models. " Vision Research 121(2016):72-84.

[2]. Privitera, Claudio M. "The scanpath theory: its definition and later developments." Electronic Imaging International Society for Optics and Photonics, 2006:60570A-60570A-5.

[3]. Mandal, Bappaditya, et al. "Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State." IEEE Transactions on Intelligent Transportation Systems PP.99(2017):1-13.

[4]. Guo, C., and L. Zhang. "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. " Oncogene 19.1(2009):185-198.

[5]. Li, Songpo, and X. Zhang. "Implicit Intention Communication in Human–Robot Interaction Through Visual Behavior Studies." IEEE Transactions on Human-Machine Systems PP.99(2017):1-12.

[6]. Olshausen, B. A., and D. J. Field. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images." Nature 381.6583(1996):607-9.

[7]. Schmidhuber, J. "Deep Learning in neural networks: An overview. " Neural Networks the Official Journal of the International Neural Network Society 61(2015):85-117.

[8]. Lecun, Y., Y. Bengio, and G. Hinton. "Deep learning." Nature 521.7553(2015):436.

[9]. Le, Meur O, and Z. Liu. "Saccadic model of eye movements for free-viewing condition." Vision Research. (2015):152-164.

[10]. Sss, Kruthiventi, K. Ayush, and R. V. Babu. "DeepFix: A Fully Convolutional Neural Network for Predicting Human Eye Fixations. " IEEE Transactions on Image Processing PP.99(2015):1-1.

[11]. Jetley, Saumya, N. Murray, and E. Vig. "End-to-End Saliency Mapping via Probability Distribution Prediction." Computer Vision and Pattern Recognition IEEE, 2016:5753-5761.

[12]. Pan, Junting, et al. "Shallow and Deep Convolutional Networks for Saliency Prediction." Computer Vision and Pattern Recognition IEEE, 2016:598-606.

[13]. Jiang, Ming, et al. "SALICON: Saliency in Context." Computer Vision and Pattern Recognition IEEE, 2015:1072-1080.