



Conference Book



Sponsors

Hosted and Organized by :



Korea University Institute for Artificial Intelligence

**KOREA
UNIVERSITY**



Korea Brain Education Society

Supported by :



Jeju Convention & Visitors Bureau



International Association of Pattern Recognition

Welcome Message from the General Chairs

It is again a great pleasure to welcome you to the 6th Asian Conference on Pattern Recognition. Due to the COVID-19 situation, ACPR 2021 will be held in a mixture of virtual and on-site methods. As a successful tradition we are having its sixth event of the biennial pattern recognition conference. We aim to provide an international forum for researchers to present their latest on-going research achievements and to share ideas and discussions in a wide range of pattern recognition and its applications. This casual scientific gathering will spark new technical and scientific ideas. We would like to express our thanks and gratitude to reviewers, authors, speakers, lecturers, and participants in ACPR 2021. Sincere appreciation goes to our invited speakers and lecturers from all around the world for sharing their valuable time and insights with us: Lei Zhang from The Hong Kong Polytechnic University, Andreas Dengel from DFKI & University of Kaiserslautern, Jure Leskovec from Stanford University, Shang-Hong Lai from Microsoft, Che-Tsung Lin from Chalmers University of Technology, Chiou-Ting Hsu from National Tsing Hua University, Taiji Suzuki from the University of Tokyo, Junyeong Kim from KAIST, Zhijian Ou from Tsinghua University. Through having fruitful time with prospective colleagues, we wish you to share innovative thoughts with others. In addition, we cannot thank enough to our local committee members who worked hard so that this could happen. ACPR 2021 could not happen without their dedication and devoted inputs from committee and staff members of ACPR 2021. Finally, this conference could not run well without our sponsors and we would like to give our sincere thanks to our hosts, organizing institution, and supporters. Sincerely we hope that you will have a great time at ACPR 2021.

November, 2021

Seong-Whan Lee, Cheng-Lin Liu, and Yasushi Yagi

General Co-Chairs for ACPR 2021

Committee

Steering Committee

Seong-Whan Lee (Korea Univ., Korea)

Cheng-Lin Liu (CASIA, China)

Umapada Pal (Indian Statistical Institute, India)

Tieniu Tan (Chinese Academy of Sciences, China)

Yasushi Yagi (Osaka Univ., Japan)

General Chairs

Seong-Whan Lee (Korea Univ., Korea)

Cheng-Lin Liu (CASIA, China)

Yasushi Yagi (Osaka Univ., Japan)

Program Chairs

Qingshan Liu (Nanjing Univ. of Sci.& Tech., China)

Hajime Nagahara (Osaka Univ., Japan)

Christian Wallraven (Korea Univ., Korea)

Publication Chairs

Unsang Park (Sogang Univ., Korea)

Wei Xiong (Institute for Infocomm Research, Singapore)

Publicity Chairs

Jean-Marc Ogier (University of Larochele, France)

Umapada Pal (Indian Statistical Institute, India)

Richard Zanibbi (Rochester Institute of Technology, USA)

Workshop Chairs

Soo-Hyung Kim (Chonnam National Univ., Korea)

Byoungchul Ko (Keimyung Univ., Korea)

Zhaoxiang Zhang (CASIA, China)

Tutorial Chairs

Chang D. Yoo (KAIST, Korea)

Jingdong Wang (Microsoft Research Asia, China)

Exhibition/Demo Chairs

Sung Chan Jun (Gwangju Institute of Science and Technology, Korea)

Dong Gyu Lee (Kyungpook National Univ., Korea)

Sponsorship Chair

Jaesik Choi (KAIST, Korea)

Finance Chair

Wonzoo Chung (Korea Univ., Korea)

Local Organizing Chairs

Tae-Eui Kam (Korea Univ., Korea)

Sungjoon Choi (Korea Univ., Korea)

Webmaster

Hong-Gyu Jung (Korea Univ., Korea)

Ji-Hyun Lee (Korea Univ., Korea)

Technical Program

Nov 9th, DAY 1

Type	Authors	Title	Session	Start	End
Tutorials	Junyeong Kim	Deep Learning for Vision and Language Reasoning	1	9:00	12:00
	Zhijian Ou	State-of-the-Art of End-to-End Speech Recognition	2	9:00	12:00
	Shang-Hong Lai, Che-Tsung Lin, Chiou-Ting Hsu	Deep Domain Adaptation for Computer Vision	1	14:00	17:00
	Taiji Suzuki	Deep Learning Theory from Statistics to Optimization	2	14:00	17:00

Nov 10th, DAY 2

Type	Paper ID	Authors	Title	Session Chair	Start	End
Opening				Christian Wallraven	8:30	9:00
KEYNOTE	-	Lei Zhang	Gradient centralization and feature gradient decent for deep neural network optimization	Cheng-Lin Liu	9:00	10:00
BREAK					10:00	10:20
MEDICAL	196	Kun Yan, Zhizhe Liu, Shuai Zheng, Zhenyu Guo, Zhenfeng Zhu and Yao Zhao	Coarse-to-fine learning framework for semi-supervised multimodal MRI synthesis	Han-Jeong Hwang	10:20	10:40
	44	Zhan Zhang, Yun-Tian Wang, Bang Li, An Guo and Cheng-Lin Liu	Deep Rejoining Model for Oracle Bone Fragment Images		10:40	11:00
	140	Guodong Sun, Zejin Wang, Guoqing Li and Hua Han	Robust Frequency-Aware Instance Segmentation for Serial Tissue Sections		11:00	11:20
	72	Gun-Hee Lee, Han-Bin Ko and Seong-Whan Lee	Joint Dermatological Lesion Classification and Confidence Modeling with Uncertainty Estimation		11:20	11:40
LUNCH BREAK					11:40	14:00
HUMAN SIGNALS	195	Luhui Yue, Junxia Li and Qingshan Liu	Causal Intervention Learning For Multi-Person Pose Estimation	Sungjoon Choi	14:00	14:20
	20	Pengfei Zhang, Xiang Li, Jianjun Qian, Zhong Jin and Jian Yang	Hierarchical Attentive Upsampling on Input Signals for Remote Heart Rate Estimation		14:20	14:40

HUMAN IN THE LOOP	104	Doo Yon Kim and Christian Wallraven	Label quality in AffectNet: results of crowd-based re-annotation	Jin Xie	14:40	15:00
	102	Ashish Menon, Piyush Singh, P. K. Vinod and C. V. Jawahar	Interactive Learning for Assisting Whole Slide Image Annotation		15:00	15:20
	89	Hoe Sung Ryu, Uijong Ju and Christian Wallraven	Predicting decision-making in the future: Human versus Machine		15:20	15:40
BREAK					15:40	16:00
KEYNOTE	-	Andreas Dengel	Combining Bird Eye View and Grass Root View for Earth Observation	Christian Wallraven	16:00	17:00
POSTERS #1				Serin Park	17:00	18:30

Poster List

Paper ID	Authors	Title
13	Motoshi Abe and Takio Kurita	q-Softplus Function: Extensions of Activation Function and Loss Function by using q-Space
18	Shungo Fujii, Naoki Okamoto, Toshiki Seo, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi	Super-class Mixup for Adjusting Training Data
42	Jiacan Zheng, Zhihui Lai, Jianglin Lu and Jie Zhou	Low-Rank Correlation Analysis for Discriminative Subspace Learning
43	Huanyu Li, Cuicao Zhang, Chunlei Li, Zhoufeng Liu, Yan Dong and Shuili Tang	Rapid and high-purity seed grading based on pruned deep convolutional neural network
49	Wei Tian, Dongjun Xin and Chuzheng Wang	The Comparison of Multiple Attention Mechanisms Fused with Residual Neural Network for the Classification of Focal Liver Lesions
73	Ji-Seon Bang and Seong-Whan Lee	Motor Imagery Classification based on CNN-GRU Network with Spatio-Temporal Feature Representation
76	Young Eun Kim and Seong-Whan Lee	Domain Generalization with Pseudo-Domain Label for Face Anti-Spoofing
80	Hyun Seung Lee and Christian Wallraven	Visualizing the embedding space to explain the effect of knowledge distillation
81	Kaiheng Dang, Jianhuang Lai, Junhao Dong and Xie Xiaohua	Adversarial Training Inspired Self-attention Flow for Universal Image Style Transfer
88	Haram Joo, Inhyeok Jeong and Sang Wan Lee	Estimating the level of inference using an order-mimic agent
93	Tian Li, Kun Huang, Yuhan Zhang, Mingchao Li, Weiwei Zhang and Qiang Chen	Multi-stage domain adaptation for subretinal uid classication in cross-device OCT images
94	Yuichi Kamata, Moyuru Yamada, Keizo Kato, Akira Nakagawa and Takayuki Okatani	Symbolizing Visual Features for Pre-training with Unlabeled Images
125	He Sun, Mingkun Li and Chun-Guang Li	Hybrid Contrastive Learning with Cluster Ensemble for Unsupervised Person Re-identification

128	Yujia Zhai, Fulin Tang and Yihong Wu	3OFRR-SLAM: Visual SLAM with 3D-assisting Optical Flow and Refined-RANSAC
144	Lifang Wu, Heng Zhang, Ge Shi and Sinuo Deng	Weakly Supervised Interaction Discovery Network for Image Sentiment Analysis
147	Jian Pang, Weifeng Liu, Baodi Liu, Kai Zhang, Dapeng Tao and Xiaoping Lu	Interference Distillation for Underwater Fish Recognition
150	Jinhyo Shin and Wonzoo Chung	Motor Imagery Classification based on local log Riemannian distance matrices selected by confusion area score
178	Zhoufeng Liu, Chengli Gao, Chunlei Li, Ning Huang and Zijing Guo	Unsupervised Fabric Defect Detection Based on DCGAN with Component-encoder
190	Sreyasee Das Bhattacharjee and Junsong Yuan	Proactive Student Persistence Prediction in MOOCs via Multi-Domain Adversarial Learning
191	Jinhong Yu and Kun Sun	A Multi-feature Embedding Method for Robust Image Matching
55	Ayan Banerjee, Shivakumara Palaiahnakote, Soumyajit Pal, Umapada Pal and Cheng-Lin Liu	DCT-DFT-FFT based Method for Text Detection in Underwater Images
10	Sangwon Kim, In-Su Jang and Byoungchul Ko	Image Registration Between Real Image and Virtual Image Based on Self-Supervised Keypoint Learning
28	Xuran Xu, Tong Zhang, Chunyan Xu and Zhen Cui	Circulant Tensor Graph Convolutional Network for Text Classification
50	Abhishek Srivastava, Sukalpa Chanda and Umapada Pal	Exploiting Multi-Scale Fusion, Spatial Attention and Patch Interaction Techniques for Text-Independent Writer Identification
51	Tamal Chowdhury, Sukalpa Chanda, Saumik Bhattacharya, Soma Biswas and Umapada Pal	Contact-Less Heart Rate Detection in Low Light Videos
65	Kento Terauchi and Keiji Yanai	Pose Sequence Generation with a GCN and an Initial Pose Generator
95	Miseon Shim, Seung-Hwan Lee and Han-Jeong Hwang	Development of a resting-state electroencephalography (EEG)-based computer-aided diagnosis (CAD) system for first-episodic drug-naïve female major depressive disorder
111	Daehyun Cho and Christian Wallraven	Paperswithtopic: Topic Identification from PaperTitle Only
118	Juho Lee, Jin Woo Choi and Sungho Jo	Subject-Independent Motor Imagery EEG Classification Based on Graph Convolutional Network
157	Yuhang Chen, Bolin Ni and Gaofeng Meng	Joint Semantic Segmentation and Edge Detection for Waterline Extraction
170	Pilhyeon Lee, Sunhee Hwang, Seogkyu Jeon and Hyeran Byun	Subject Adaptive EEG-based Visual Recognition
172	Wonjun Ko, Eunjin Jeon and Heung-Il Suk	Spectro-Spatio-Temporal EEG Representation Learning for Imagined Speech Recognition
175	Jinyoung Choi, Sehyeon Jang and Sung Chan Jun	Is notable EEG feature extracted over time-dependent cognitive load variation during intelligence tests?

Nov 11th, DAY 3						
Type	Paper ID	Authors	Title	Session Chair	Start	End
KEYNOTE	-	Jure Leskovec	Graph Neural Networks and Beyond	Seong-Whan Lee	9:00	10:00
BREAK					10:00	10:20
MATCHING	131	Jin Zhang, Mingyang Zhao, Xin Jiang and Dong-Ming Yan	Robust Multi-view Registration of Point Sets with Laplacian Mixture Model	Christian Wallraven	10:20	10:40
	168	Jegoon Ryu and Sei-Ichiro Kamata	A Fast and Accurate Point Pattern Matching Algorithm based on Multi-Hilbert Scans		10:40	11:00
	17	Zhixiong Pi, Changxin Gao and Nong Sang	Siamese Tracking with Bilinear Features		11:00	11:20
	91	Xinlei Qi, Yaqing Ding, Jin Xie and Jian Yang	Planar Motion Estimation for Multi-camera System		11:20	11:40
LUNCH BREAK					11:40	14:00
DEPTH	48	Jian Ren, Jin Xie and Zhong Jin	Hierarchical Multi-Scale Architecture Search for Self-Supervised Monocular Depth Estimation	Qingshan Liu	14:00	14:20
	14	Takahiro Suzuki, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi	1D Self-Attention Network for Point Cloud Semantic Segmentation using Omnidirectional LiDAR		14:20	14:40
	45	Yiming Tu and Jin Xie	UnDeepLIO: Unsupervised Deep Lidar-Inertial Odometry		14:40	15:00
BREAK					15:00	15:20
SECURITY	100	Glen Brown, Jesus Martinez-Del-Rincon and Paul Miller	Least Privilege Learning	Byoung Chul Ko	15:20	15:40
	120	Sreyasee Das Bhattacharjee and Junsong Yuan	Multi-modal Co-training for Fake News Identification using Attention-aware Fusion		15:40	16:00
	84	Pei-Kai Huang, Ming-Chieh Chin and Chiou-Ting Hsu	Face Anti-Spoofing via Robust Auxiliary Estimation and Discriminative Feature Learning		16:00	16:20
TEXT AND DOCUMENTS	23	Yu-Ting Yang, Yan-Ming Zhang, Xiao-Long Yun, Fei Yin and Cheng-Lin Liu	CASIA-onDo: A New Database for Online Handwritten Document Analysis	Tae-Eui Kam	16:20	16:40

	179	Nam Tuan Ly, Trung Tan Ngo and Masaki Nakagawa	A Self-Attention based Model for Offline Handwritten Text Recognition		16:40	17:00
POSTERS #2				Daehyun Cho	17:00	18:30

Poster List

Paper ID	Authors	Title
6	Kai Li, Yongsheng Zhang, Weiming Yang, Lei Zhang and Zhenchao Zhang	Intelligent recognition of point source target image control points with simulation datasets
7	Keiichi Yamada and Susumu Matsumi	One-Shot Image Learning Using Test-Time Augmentation
12	Jiang Zhentao, Chen Qiang, Jiang Bo, Leng Cong and Cheng Jian	Weakly Supervised Salient Object Detection with Box Annotation
25	Fengming Sun, Lufei Huang, Xia Yuan and Chunxia Zhao	Foreground-Background Collaboration Network for Salient Object Detection
38	Fan Gao, Ying Hu and Yan Yan	Deep Siamese Network with Co-Channel and Cr-Spatial Attention For Object Tracking
41	Serin Park and Christian Wallraven	Comparing Facial Expression Recognition in Humans and Machines: Using CAM, GradCAM, and Extremal Perturbation
53	Yikai Bian, Jin Xie and Jianjun Qian	Unsupervised Domain Adaptive Point Cloud Semantic Segmentation
59	Tingting Wang, Yang Xu, Zebin Wu and Zhihui Wei	Spatial-Spectral Joint Correction Network for Hyperspectral and Multispectral Image Fusion
63	Jui Chang, Yu-Ting Pang and Chiou-Ting Hsu	Towards the Target: Self-Regularized Progressive Learning for Unsupervised Domain Adaptation on Semantic Segmentation
108	Jin Liu and Yazhou Liu	Spatial Pyramid-based Wavelet Embedding Deep Convolutional Neural Network for Semantic Segmentation
112	Qifeng Luo, Ting-Bing Xu and Zhenzhong Wei	CFFNet: Cross-scale Feature Fusion Network for Real-time Semantic Segmentation
132	Farzeen Munir, Shoaib Azam, Unse Fatima and Moongu Jeon	ARTSeg: Employing Attention for Thermal images Semantic Segmentation
143	Qiuyan Tang, Jialin Ye, Fukang Chen and Xia Yuan	Attention Guided Multi-level Feedback Network for Camouflage Object Detection
145	Sha Xie, Yuhan Zhang, Mingchao Li, Zexuan Ji, Songtao Yuan and Qiang Chen	Gaussian Distribution Prior Based Multi-view Self-supervised Learning for Serous Retinal Detachment Segmentation
151	Sangpil Kim and Jihyun Bae	Hand Type Estimation for Hand Action Recognition
152	Yulan Deng, Lunke Fei, Genping Zhao, Wei Jia, Jie Wen and Chunwei Tian	Joint Multi-feature Learning for Facial Age Estimation
163	Haetsal Lee, Junghyun Cho, Ig-Jae Kim and Unsang Park	Distance-GCN for Action Recognition
167	Hiroya Kawai, Takashi Kozu, Koichi Ito, Hwann-Tzong Chen and Takafumi Aoki	Is CNN Really Looking at Your Face?
184	Tsung Hsuan Wu, Chun Lung Yang, Li Ling Chiu, Ting Wei Wang, Gueter Josmy Faure and Shang Hong Lai	CONFIDENCE-AWARE ANOMALY DETECTION IN HUMAN ACTIONS
188	Yufei Sun, Yong Li and Zhen Cui	NFW: Towards National and Individual Fairness in Face Recognition
11	Chuangquan Lin, Zhihui Lai and Jianglin Lu	Fast Deep Asymmetric Hashing for Image Retrieval

22	Jiaman Tang, Jinwen Wu, Fei Yin and Linlin Huang	Offline Handwritten Mathematical Expression Recognition via Graph Reasoning Network
60	Dong-Gyu Lee and Dong-Ok Won	Spatio-temporal Weight of Active Region for Human Activity Recognition
87	Bing Liu, Zheng Wang and Jufu Feng	PalmNet: A Robust Palmprint Minutiae Extraction Network
99	Xu Liang, Zhaoqun Li, Dandan Fan, Jinyang Yang, Guangming Lu and David Zhang	SaME: Sharpness-Aware Matching Ensemble for Robust Palmprint Recognition
117	Oskar Natan and Jun Miura	Semantic Segmentation and Depth Estimation with RGB and DVS Sensor Fusion for Multi-view Driving Perception
148	Truong Thanh-Nghia, Hung Tuan Nguyen, Cuong Tuan Nguyen and Masaki Nakagawa	Learning Symbol Relation Tree for Online Handwritten Mathematical Expression Recognition

Nov 12th, DAY 4

Type	Paper ID	Authors	Title	Session Chair	Start	End
ALGORITHMS	103	Yonghui Liang, Lu Zhang, Yuqing He, Na Xu and Mingqi Liu	SARNN: A Spatiotemporal Prediction Model for Reducing Error Transmissions	Takio Kurita	9:00	9:20
	32	WeiQi Zhao, Haobo Jiang and Jin Xie	Fast Proximal Policy Optimization		9:20	9:40
	27	Xiaoling Zhou and Ou Wu	Drop "Noise" Edge: An Approximation of the Bayesian GNNs		9:40	10:00
BREAK					10:00	10:20
TRANSFORMERS	24	Lufei Huang, Fengming Sun and Xia Yuan	COMatchNet: Co-Attention Matching Network for Video Object Segmentation	Christian Wallraven	10:20	10:40
	106	Tao Wang, Zhihui Lai and Heng Kong	TFNet: Transformer Fusion Network for Ultrasonic Image Segmentation		10:40	11:00
	61	Bolin Ni, Gaofeng Meng, Shiming Xiang and Chunhong Pan	NASformer: Neural Architecture Search for Vision Transformer		11:00	11:20
AWARDS & CLOSING				Christian Wallraven	11:20	12:00
LUNCH BREAK					12:00	13:00
WORKSHOP	-	-	Future Impact of AI in Real-World Applications	Jong-Ha Lee, Byoung Chul Ko	13:00	17:00

Keynote Speakers



[Professor Lei Zhang](#) (Dept. of Computing, The Hong Kong Polytechnic University)

Title of Talk : Gradient centralization and feature gradient decent for deep neural network optimization

Abstract: The normalization methods are very important for the effective and efficient training of deep neural networks (DNNs). Many popular normalization methods operate on weights, such as weight normalization and weight standardization. We propose a very simple yet effective DNN optimization technique, namely gradient centralization (GC), which operates on the gradients of weights directly. GC simply centralizes the gradient vectors to have zero mean. It can be easily embedded into the current gradient based optimization algorithms with just one line of code. GC demonstrates various desired properties, such as accelerating the training process, improving the generalization performance, and the compatibility for fine-tuning pre-trained models. On the other hand, existing DNN optimizers such as stochastic gradient descent (SGD) mostly perform gradient descent on weight to minimize the loss, while the final goal of DNN model learning is to obtain a good feature space for data representation. Instead of performing gradient descent on weight, we propose a method, namely feature SGD (FSGD), to approximate the output feature with one-step gradient descent for linear layers. FSGD only needs to store an additional second-order statistic matrix of input features and use its inverse to adjust the gradient descent of weight. FSGD demonstrates much better generalization performance than SGD in classification tasks.



[Professor Andreas Dengel](#) (DFKI & University of Kaiserslautern, Germany)

Title of Talk : Combining Bird Eye View and Grass Root View for Earth Observation

Abstract: This talk will address the multiple opportunities presented by the use of AI approaches to combine the analysis of satellite imagery and data collected on the ground. In particular, three exemplary areas, namely disaster management, air pollution monitoring, and agricultural yield management, will be discussed to demonstrate the added value of both data sources. For the first domain, I will show how multimedia data from real-time social media monitoring and spectral data from Earth observation can be tied together to provide deep insight and foresight into disasters. In the second case, I will explain how mapping human settlements in combination with Sentinel 5 data can reveal the impact of air pollution. In the third case, I will discuss some insights into monitoring agricultural processes using sensor data from agricultural machinery. In doing so, I will highlight the potential to combine the relevant findings with approaches to classifying land cover and land use using remote sensing.



[Professor Jure Leskovec](#) (Dept. of Computer Science at Stanford University)

Title of Talk : Graph Neural Networks and Beyond

Abstract: Machine learning on graphs is an important and ubiquitous task with applications ranging from drug design to friendship recommendation in social networks. The primary challenge in this domain is finding a way to represent, or encode, graph structure so that it can be easily exploited by machine learning models. In this talk I will discuss recent advancements in the field of Graph Neural Networks that automatically learn to encode graph structure into low-dimensional embeddings, using techniques based on deep learning. I will provide a conceptual overview of key advancements in this area of representation learning on graphs, including graph convolutional networks and their representational power. We will also discuss development of graph-learning benchmarks as well as open research problems.

Tutorials

Title: Deep Learning for Vision and Language Reasoning

Abstract

Understanding visual information along with the natural language appears to be desiderata in recent research communities. Notable efforts have made toward bridging the fields of computer vision and natural language processing, and have opened the door to methods from visual question answering to video-grounded dialogue. However, it is widely accepted that in order to develop truly intelligent AI systems, we need to bridge the gap between perception and cognition. The purpose of this tutorial is to present the history and recent approaches of various vision-and-language reasoning tasks including visual/video question answering, and visual/video dialogue. In this tutorial, we will provide an intuitive explanation of these topics, from the basic building blocks including attention, transformers, to recent trends such as causal learning in detail. Further, this tutorial will also cover recent advances in vision-language pre-training methods based on Transformer architectures, which show state-of-the-art performance in various downstream tasks. The limitations of current approaches are also discussed

Program Schedule

First Part (1h30m)

- Visual Question Answering
 - Attention-based Approaches
 - Debiasing Approaches
 - Causal Inference Approaches
- 2. Video Question Answering
 - Single-modal Video Question Answering
 - Multi-modal Video Question Answering

Second Part (1h30m)

- Visual Dialogue
 - Attention-based Approaches
 - Co-reference Approaches
 - Causal Inference Approach
- Video-grounded Dialogue
 - RNN seq-2-seq Approaches
 - Transformer Approaches
- Vision-language Pre-training

Lecturer



Junyeong Kim

Post-doc Researcher, Korea Advanced Institute of Science and Technology, South Korea

Junyeong Kim is a Post-doc researcher in the Artificial Intelligence and Machine Learning Lab., in the School of Electrical Engineering at Korea Advanced Institute of Science and Technology. He received a B.S. and M.S. and Ph.D. degrees in Electrical Engineering at KAIST, in 2015, 2017, and 2021, respectively. His research interest lies in video-language inferences including video question answering and video-grounded dialogue and vision-language reasoning. He focuses on developing AI agents that can 'observe' and 'converse' as a human does. He has written 5 top-tier conference papers including CVPR, AAAI, ECCV. He received the Outstanding Ph.D. Thesis Award in 2021 from the School of Electrical Engineering, KAIST.

Title: State-of-the-Art of End-to-End Speech Recognition

Abstract

Over the last several years, there has been growing interests in developing end-to-end (E2E) automatic speech recognition (ASR) systems. E2E ASR is characterized by eliminating the construction of GMM-HMMs and phonetic decision-trees, training deep neural networks directly and, even further, removing the need for pronunciation lexicons and training acoustic and language models jointly rather than separately. Examples of such models include Connectionist Temporal Classification (CTC), attention based encoder-decoder (AED), and RNN transducer (RNNT) respectively.

The purpose of this tutorial is to present a systematic introduction to state-of-the-art of E2E ASR. First, we will introduce basic E2E ASR methods within the probabilistic graphical modeling (PGM) framework. We will separate neural network architectures and probabilistic model definitions, with emphasis on comparing and connecting different E2E ASR models that have been considered in the literature. Then, we will present a number of advanced topics for improving E2E ASR, including data-efficiency, low-latency streaming recognition, neural architecture search, multilingual and crosslingual ASR, and contextual biasing. Finally, the tutorial will point out some open questions about existing E2E ASR methods and discuss future directions to address these challenges. In addition, we will introduce open-source toolkits such as <https://github.com/thu-spmi/CAT> to help the audience to get familiar with the state-of-the-art techniques of E2E ASR.

Program Schedule

First Part

- Basics for end-to-end speech recognition
 - Probabilistic graphical modeling (PGM) framework
 - Classic hybrid DNN-HMM models
 - Connectionist Temporal Classification (CTC)
 - Attention based encoder-decoder (AED)
 - RNN transducer (RNNT)
 - Conditional random fields and sequence discriminative training

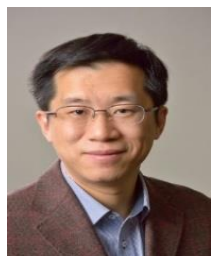
Second Part

- Improving end-to-end speech recognition
 - Data-efficiency
 - Low-latency streaming recognition
 - Neural architecture search
 - Multilingual and crosslingual ASR
 - Contextual biasing
- Open questions and future directions

Lecturer

Zhijian Ou

Associate Professor, Tsinghua University, China



Zhijian Ou received his Ph.D. from Tsinghua University in 2003. Since 2003, he has been with the Department of Electronic Engineering in Tsinghua University and is currently an associate professor. From August 2014 to July 2015, he was a visiting scholar at Beckman Institute, University of Illinois at Urbana-Champaign, USA. He has actively led national research projects as well as research funding from Intel, Panasonic, IBM, Toshiba and Apple. He currently serves as associate editor of IEEE/ACM Transactions on Audio, Speech and Language Processing, member of IEEE Speech and Language Processing Technical Committee, and was General Chair of SLT 2021, Tutorial Chair of INTERSPEECH 2020. His research interests are speech and language processing (particularly speech recognition and dialogs) and machine intelligence (particularly with probabilistic graphical models and deep learning).

Title: Deep Domain Adaptation for Computer Vision

Abstract

Deep learning methods suffer from domain shifts between training data source domain and target application domain. In recent years, several domain adaptation methods have been proposed to reduce the domain shift for practical applications of deep neural network models. In this first part of the tutorial, we will first provide an overview of different domain adaptation problems and survey some representative domain adaptation techniques, including distribution alignment, adversarial domain adaptation, and data augmentation with image domain transfer. Some application examples of applying domain adaptation techniques to face recognition and face anti-spoofing problems are discussed.

For the second part of the tutorial, we focus on GAN-based image-translation models to achieve domain adaptation with large and complex domain shifts, such as day-to-night. AugGAN could transform on-road driving images to a desired domain while image-objects would be well preserved. Later, AugGAN is further extended to a multimodal structure-consistent GAN, called Multimodal AugGAN, which can transform daytime on-road driving images to their nighttime counterparts with different ambient light levels. We will discuss some results of these image-translation models across different weathers, times-of-the-day, and datasets and their applications to train object detectors on a target domain.

For the third part, we will discuss some different problem settings of domain adaptation. In semantic segmentation problems, when adapting from the synthetic source domain to the unlabeled target domain, some recent methods employed the idea of within-domain adaptation to alleviate the semantic inconsistency. Multi-source domain adaptation focuses on learning a domain-agnostic model, which needs to handle the conflicts across multiple domains as well as to narrow the domain gap between source and target domains. Another setting is source-free adaptation, which is a special setting of model adaptation when the source dataset is unavailable during the model adaptation. Finally, we discuss the problem setting when only multiple sources are available during the training, but there exists no target domain for distribution alignment, which is to achieve domain generalization.

Program Schedule

Part 1 (Shang-Hong Lai)

- Overview of domain adaptation (15 minutes)
 - Problem description
 - Categorization of domain adaptation problems
- Deep domain adaptation techniques (45 minutes)
 - Distribution alignment
 - Adversarial domain adaptation
 - Image domain transfer
 - Applications on face recognition and anti-spoofing

Part 2 (Che-Tsung Lin)

- Domain adaptation via GAN-based image-to-image translation (60 minutes)
 - Domain adaptation via image-translation
 - Unimodal structure-consistent image-to-image translation
 - Multimodal structure-consistent image-to-image translation
 - Results of image-translation across different weathers, times-of-the-day, and datasets

Part 3 (Chiou-Ting Hsu)

- Other domain adaptation settings (60 minutes)
 - Cross-domain adaptation vs. within domain adaptation
 - Multi-source domain adaptation
 - Source-free adaptation
 - Source-only domain generalization

Lecturers

Shang-Hong Lai

Principal Researcher, Microsoft AI R&D Center, Taiwan

Professor, National Tsing Hua University, Taiwan

Email: shlai@microsoft.com ; lai@cs.nthu.edu.tw



Shang-Hong Lai received the Ph.D. degree from University of Florida, Gainesville, USA in 1995. He worked at Siemens Corporate Research in Princeton, New Jersey, USA, as a member of technical staff during 1995-1999. Since 1999, he joined the Department of Computer Science, National Tsing Hua University, Taiwan, where he is now a professor there. Since the summer of 2018, Dr. Lai has been on leave from NTHU to join Microsoft AI R&D Center, Taiwan. He is currently a principal researcher at Microsoft AI R&D Center and leads a science team focusing on computer vision research for face related applications.

Dr. Lai's research interests are mainly focused on computer vision, image processing, and machine learning. He has authored more than 300 papers published in refereed international journals and conferences in these areas. In addition, he has been awarded around 30 patents on his research on computer vision. He has been involved in the organization for a number of international conferences in computer vision and related areas, including ICCV, CVPR, ECCV, ACCV, ICIP, etc. Furthermore, he has served as an associate editor for Pattern Recognition and Journal of Signal Processing Systems.

Che-Tsung Lin

Postdoctoral Researcher, Chalmers University of Technology, Sweden

Email: alexofntu@gmail.com ; chetsung@chalmers.se



Che-Tsung Lin received his B.S. degree in Mechanical Engineering from National Taiwan University of Science and Technology and MS. degree in the Institute of Applied Mechanics from National Taiwan University, and Ph.D. degree in the computer science department in National Tsing Hua University in 2003 and 2005, and 2020, respectively. He was an associate researcher from 2006 to 2014, a researcher from 2014 to 2020 and a senior researcher in 2020 in the Intelligent Mobility Division, Mechanical and Systems Lab, Industrial Technology Research Institute, Taiwan. From April to October in 2013, he was a

visiting researcher at the computer science department of University of California, Santa Barbara, USA. He is currently a postdoctoral researcher at Chalmers University of Technology, Sweden. His research is mainly about object detection, semantic segmentation, domain adaptation and their applications in ADAS and autonomous vehicles.

Chiou-Ting Hsu

Professor, National Tsing Hua University, Taiwan

Email: cthsu@gapp.nthu.edu.tw



Chiou-Ting Hsu received the Ph.D. degree in computer science and information engineering from National Taiwan University, Taipei, Taiwan, in 1997. From 1998 to 1999, she was with Philips Innovation Center, Taipei, Philips Research, as a senior research engineer. Since 1999, she joined the Department of Computer Science, National Tsing Hua University, Taiwan, and is now a professor there. She was a visitor scholar at Columbia University, New York, USA, in 2005, at University of Maryland, College Park, USA, in 2009, and at EURECOM, France, in 2019. She was an associate editor of *Advances in Multimedia* and the *IEEE Transactions on Information Forensics and Security* (2012-2015), and is currently an associate editor of *Journal of Visual Communication and Image Representation* and *EURASIP Journal on Image and Video Processing*. She was an elected member of the IEEE Information Forensics and Security Technical Committee (2013-2015) and of the APSIPA Image, Video, and Multimedia Technical Committee (2013-2016).

Title: Deep Learning Theory from Statistics to Optimization

Abstract

In this tutorial, I will explain the theories of deep learning, especially from the viewpoint of statistical learning theory and optimization theory. While deep learning is used in a variety of fields, there are still many theoretical aspects that remain unexplored, and vigorous research is currently being conducted from a variety of perspectives to clarify them. Deep learning theory can be roughly divided into three aspects: "approximation theory," "generalization theory," and "optimization theory," and those research is being conducted using various mathematical tools such as functional analysis and probability theory.

In this tutorial, we will provide an intuitive explanation of these topics, from the basics of statistical learning theory to recent trends in theoretical research such as double descent, neural tangent kernels and mean-field analysis. In particular, we will provide theoretical analyses of questions such as "Why does deep learning produce high prediction performance?" and "Why does it generalize despite the huge number of parameters?". The limitations of current theories are also discussed.

Program Schedule

First Part (1h45m)

- Introduction to statistical learning theory and deep learning theory
- Representation ability of deep neural network
 - Universal approximator
 - Depth separation
 - Approximation error of function spaces with smoothness: Holder space, Sobolev space, Besov space
 - Benefit of adaptivity of deep learning: separation between kernel method and deep learning
- Generalization error on over-parameterized models
 - Generalization gap of deep neural network models
 - Double descent and benign overfitting

Coffee Break (15m)

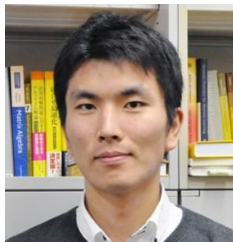
Second Part (1h15m)

- Optimization and generalization
 - Introduction to Neural tangent kernel (NTK) and mean-field regime
 - Convergence analysis of SGD on NTK
 - Optimization on mean-field regime: McKean-Vlasov process
 - Theories of gradient Langevin dynamics

Lecturer

Taiji Suzuki

Associate Professor, the University of Tokyo, Japan



Taiji Suzuki is currently an Associate Professor in the Department of Mathematical Informatics at the University of Tokyo. He also serves as the team leader of "Deep learning theory" team in AIP-RIKEN. He received his Ph.D. degree in information science and technology from the University of Tokyo in 2009. He has a broad research interest in statistical learning theory on deep learning, kernel methods and sparse estimation, and stochastic optimization for large-scale machine learning problems. He served as area chairs of premier conferences such as NeurIPS, ICML, ICLR, AISTATS and a program chair of ACML. He received the Outstanding Paper Award at ICLR in 2021, the MEXT Young Scientists' Prize, Outstanding Achievement Award in 2017 from the Japan Statistical Society, Outstanding Achievement Award in 2016 from the Japan Society for Industrial and Applied Mathematics, and Best Paper Award in 2012 from IBISML.

Workshop

Title: Future Impact of AI in Real-World Applications

Organizers

Jong-Ha Lee, Keimyung University, [http://203.247.8.200/index.php/Faculty:Jong-Ha_Lee_\(KR\)](http://203.247.8.200/index.php/Faculty:Jong-Ha_Lee_(KR))

Byoung Chul Ko, Keimyung University, <https://cvpr.kmu.ac.kr/member.htm>

Abstract

The workshop will focus on various area of future impact of AI. The aim of this workshop is to bring researchers and scientists from academia, medical area with engineers from industry together to discuss about various impact of cutting-edge technologies of AI in future society. The workshop will take a deep dive into the capabilities of Edge Insights for Academia and Industrial via a tutorial utilizing real-world AI applications. For example, breast cancer can be detected via smartphone level infrared camera for detecting lesion and target mass at home. Despite the existence of some commercial AI systems such as autonomous vehicle, we are at the beginning of a long research pathway towards a future generation of deep AI. The workshop focuses on numerical and computational aspects of future impact of AI and on these relations to various AI techniques.

Workshop Schedule

Time	Item
13:00pm - 13:05pm	Opening Remarks (Prof. Jong-Ha Lee)
13:05pm - 13:40pm	Keynote Talk: Prof. Soo Hyung Kim (Chonnam National University)
13:40pm - 14:10pm	Keynote Talk: Prof. Chang-Hee Won (Temple University)
14:10pm - 14:30pm	Coffee Break
14:30pm - 15:05pm	Keynote Talk: Prof. Suha Kwak (POSTECH)
15:05pm - 15:35pm	Keynote Talk: Prof. Chih-Chung Hsu (National Cheng Kung University)
15:35pm - 15:45pm	Coffee Break
15:45pm - 16:45pm	Oral Presentation
15:45pm - 15:55pm	G2CN: Geometric Graph Convolutional Network for Facial Expression Recognition
15:55pm - 16:05pm	Histological Image Segmentation and Classification Using Entropy-Based Convolutional Module
16:05pm - 16:15pm	Vision Trnasformer Based Dynamic Facial Emotion Recognition
16:15pm - 16:25pm	Development of a system capable of diagnosing and treating Alzheimer's disease: a technique experiment using cadaver
16:25pm - 16:35pm	Remote Bio Vision: Perfusion Imaging Based Non-Contact Biosignal Measurement Method
16:35pm - 16:45pm	Calibrating a Multiple-View Thermal Camera
16:45pm - 16:55pm	Closing Remarks (Prof. Byoung Chul Ko)

For more information, please go to <http://ai-in-real.github.io>

