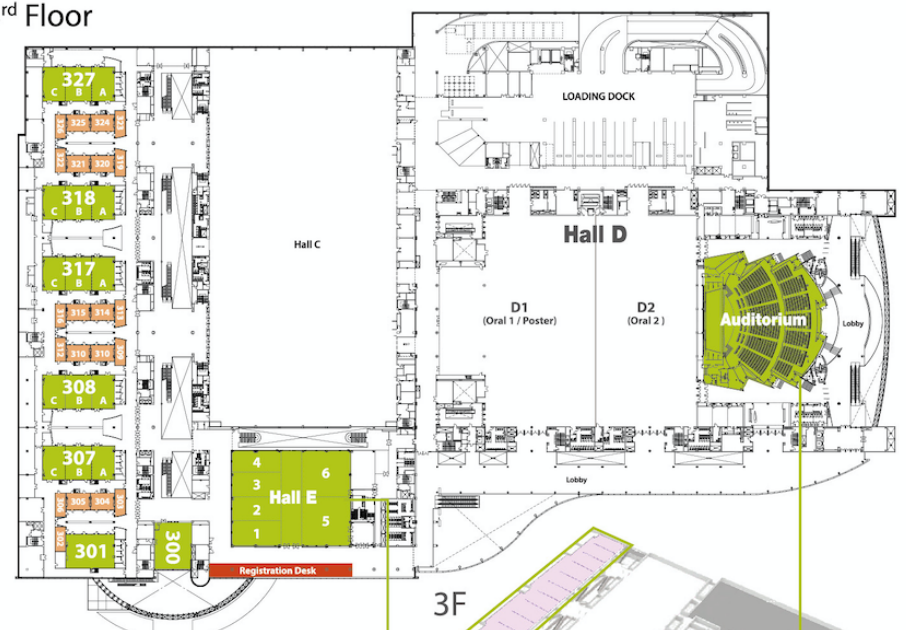

ICCV 2019
Seoul, Korea
IEEE / CVF
International Conference on
Computer Vision 2019
Oct. 27 - Nov. 2, 2019
Pocket Guide
(Workshops & Tutorials)

ICCV 2019 Seoul, Korea

# COEX CONVENTION CENTER

## 3rd Floor





## 4th Floor


# Message from the General & Program Chairs

Welcome to Seoul and to the 17th International Conference on Computer Vision, jointly sponsored by the IEEE and the Computer Vision Foundation. The first ICCV was held 32 years ago, in 1987. Very quickly, the conference became a must-attend event for all those working in the field. ICCV has grown spectacularly, as have all vision conferences. When this meeting was planned, the General Chairs envisaged a conference of about 2500 attendees. The evidence suggests there will be about 7,000 of you reading this document in Seoul when the conference is held.

The conference received 4323 valid submissions -- an increase of 100% over the previous ICCV, held in 2017. After a careful selection process coordinated by the Program Chairs, 1075 papers were accepted for publication and presentation in the main program. The resulting acceptance rate of 25% reflects the high standard of ICCV and is consistent with the rates of past ICCV conferences. 172 area chairs and 2506 reviewers (including 383 emergency reviewers) worked diligently over a period of almost nine months to make these decisions. Each paper received at least three full reviews, and the acceptance decisions were made within AC pairs in consultation with additional expert AC's as necessary. Following the best practice of our community, the Program Chairs did not place any restrictions on acceptance. Per PAMI-TC policy, Program Chairs did not submit papers, which allowed them to be free of conflict in the review process.

Out of all accepted papers, 200 were selected for oral presentations based on AC recommendations. This year, following the example set by CVPR 2019, the oral presentations are short — 6 minutes each including transition/questions — so that more papers may receive exposure. All papers have poster presentations. Award papers were selected from a pool of 12 papers nominate by ACs; final recommendations were made by an external award committee.

We would like to thank everyone involved in making ICCV 2019 a success. This includes the organizing committee, the area chairs, the reviewers, authors, demo session participants, donors, exhibitors, and everyone else without whom this meeting would not be possible. The General Chairs and Program Chairs particularly thank a few unsung heroes that helped us tremendously: Eric Mortensen for mentoring the publication chairs and managing camera-ready and program efforts; the PCO team who organized space and registrations; Gérard Medioni and Ramin Zabih for helpful support and advice on various occasions; and the Microsoft CMT support team for the tremendous help with prompt responses.

Finally, we thank all of you for attending ICCV and making it one of the top venues for computer vision research in the world. We hope that you also have some time to explore Seoul before or after the conference. Enjoy ICCV 2019!!

General Chairs: **Kyoung Mu Lee**
**David Forsyth**
**Marc Pollefeys**
**Xiaoou Tang**

Program Chairs: **In So Kweon**
**Nikos Paragios**
**Ming-Hsuan Yang**
**Svetlana Lazebnik**

# ICCV 2019 Organizing Committee

| | |
|---|---|
| **General Chairs:** | Kyoung Mu Lee<br>David Forsyth<br>Marc Pollefeys<br>Xiaoou Tang |
| **Program Chairs:** | In So Kweon<br>Nikos Paragios<br>Ming-Hsuan Yang<br>Svetlana Lazebnik |
| **Workshop Chairs:** | Yoichi Sato<br>Jingyi Yu |
| **Tutorial Chairs:** | Bohyung Han<br>Juan Carlos Niebles |
| **Finance Chairs:** | Gérard Medioni<br>Ramin Zabih |
| **Publication Chairs:** | Eric Mortensen<br>Jongwoo Lim |
| **Advisory Committee:** | Katsushi Ikeuchi<br>Richard Hartley |
| **Corporate Relation Chairs:** | Chen Change Loy<br>Yunchao Gong |
| **Local Arrangements Chairs:** | In Kyu Park<br>Seon Joo Kim |
| **Doctoral Consortium Chairs:** | Federica Bogo<br>Jonas Wulff |
| **Student Volunteers Chairs:** | Iro Armeni<br>Kuk-Jin Yoon |
| **Demo & Exhibition Chairs:** | Junseok Kwon<br>Ouyang Wanli |
| **Presentation Chairs:** | Hongdong Li<br>Soochahn Lee |
| **Publicity Chairs:** | Bolei Zhou<br>Jifeng Dai |
| **Web & Social Media Chairs:** | Minsu Cho<br>Serena Yeung |
| **Web Masters:** | Juhong Min<br>Sanghyun Son |

**QR Codes for the Conference Mobile App** (by IEEE CPS)

Apple iTunes App Store (iOS)

Google Play (Android)



## Saturday, October 26

**1500–1900 Registration** (Hall E Lobby)

# Sunday, October 27

**NOTE:** Use the QR code for each tutorials's website for more information on that tutorial. Here's the QR code to the ICCV Tutorials page.

**0730–1700 Registration** (Hall E Lobby)

**1000–1100 Morning Break**

**1230–1330 Lunch** (On your own)

**1530–1630 Afternoon Break**

## Tutorial: Global Optimization for Geometric Understanding With Provable Guarantees

**Organizers:** Luca Carlone
Tat-Jun Chin
Anders Eriksson
Fredrik Kahl

**Location:** Room 402

**Time:** Half Day - AM (0845-1230)

**Description:** This tutorial aims to give an in-depth introduction to global optimization tools, including convex and semidefinite relaxations, applied to 3D vision problems. The first goal of the tutorial is to motivate the need for global solvers by providing real-world examples where the lack of robustness results from the difficulty in solving large optimization problems to optimality. The second goal is to provide the attendees with basic mathematical and algorithmic concepts, and survey important recent advances in the area. The third goal is to outline several open research avenues: global optimization has an enormous untapped potential and it is hoped that this tutorial will inspire researchers to use modern optimization tools to solve several outstanding challenges in geometric vision.

## Tutorial: From Image Restoration to Enhancement and Beyond

**Organizers:** Radu Timofte
Shuhang Gu
Martin Danelljan
Zhiwu Huang
Robby T. Tan
Dengxin Dai

**Location:** Room 401

**Time:** Half Day - AM (0800-1230)

**Description:** Recent years have shown a large interest and tremendous advances have been achieved in image and video restoration and enhancement. A large number of solutions were proposed ranging from handcrafted designs to fully learned and generative models. Gradually the focus in image restoration shifted from improved fidelity of the results to improved perceptual quality. At the same time, the studied corruptions departed from the standard synthetic/artificial corruptions in controlled environments to fully realistic and in the wild settings – a fertile soil for developing semi- and unsupervised solutions.

This tutorial on the current state-of-the-art in the fields of image restoration and enhancement with applications to autonomous driving and smartphone cameras. Moreover, this tutorial will convey the importance of the restoration and enhancement for the subsequent higher-level computer vision tasks.

**Notes:**

## Tutorial: Everything You Need to Know to Reproduce SOTA Deep Learning Models

**Organizers:** Hang Zhang
Tong He
Zhi Zhang
Zhongyue Zhang
Haibin Lin
Aston Zhang
Mu Li

**Location:** Auditorium

**Time:** Half Day - AM (0800-1215)

**Description:** In this tutorial, we will walk through the technical details of the state-of-the-art (SOTA) algorithms in major computer vision tasks, and we also provide the code implementations and hands-on tutorials to reproduce the large-scale training in this tutorial.

## Tutorial: Interpretable Machine Learning for Computer Vision

**Organizers:** Bolei Zhou
Zeynep Akata
Andrea Vedaldi
Trevor Darrell
Alan L. Yuille

**Location:** Auditorium

**Time:** Half Day - PM
(1400-1800)

**Description:** This tutorial is designed to broadly engage the computer vision community with the topic of interpretability and explainability in computer vision models. We will review the recent progress we made on visualization, interpretation, and explanation methodologies for analyzing both the data and the models in computer vision. The main theme of the tutorial is to build up consensus on the emerging topic of the machine learning interpretability, by clarifying the motivation, the typical methodologies, the prospective trends, and the potential industrial applications of the resulting interpretability.

## Tutorial: Understanding Color and the In-Camera Image Processing Pipeline for Computer Vision

**Organizers:** Michael S. Brown

**Location:** Room 401

**Time:** Half Day - PM
(1330-1815)

**Description:** Color is not a well-understood topic in computer vision. This tutorial aims to address this issue by providing a thorough background on color theory and its relationship to the in-camera processing pipeline and computer vision applications. The tutorial is organized into two parts. The first part provides a background on color theory and color representations, namely the CIE 1931 XYZ color space and its derivatives commonly found in computer vision (sRGB, L*ab, Yuv, etc.). The first part of the tutorial will also discuss routines applied onboard cameras to convert the low-level sensor raw-RGB responses to their final standard RGB (sRGB) colors. These routines include computational color constancy (auto white balance), colorimetric conversion, image demosaicing, image denoising, tone-mapping, super-resolution, and general color manipulation. The second part of this tutorial discusses recent research in the computer vision community on many of these camera pipeline components. The second part of the tutorial will conclude with a discussion on various misconceptions about color and camera images made in many areas of computer vision.

**Notes:**

# Sunday, October 27

**NOTE:** Use the QR code for each workshop's website to find the workshop's schedule. Here's the QR code to the ICCV Workshops page.

**0730–1700 Registration** (Hall E Lobby)

**1000–1100 Morning Break**

**1230–1330 Lunch** (On your own)

**1530–1630 Afternoon Break**

## Vision Meets Drones: A Challenge

**Organizers:** Pengfei Zhu
Longyin Wen
Dawei Du
Xiao Bian
Qinghua Hu
Haibin Ling

**Location:** Room E1

**Time:** Full Day (0830-1700)

**Description:** Drone, or general UAVs, equipped with cameras have been fast deployed in our daily life with range of applications, including agricultural, aerial photography, fast delivery, and surveillance. Consequently, automatic understanding of visual data collected from drones becomes highly demanding, making computer vision and drones more and more closely. Based on the proposed large-scale drone-based object detection and tracking datasets with fully manual annotations, we present the second VisDrone challenge to advance state-of-the-art methods in object detection and tracking for drone based scenes. During our workshop, the keynote speakers and winning teams will share us with new ideas in applications for drone based scenes.

## Computer Vision for Wildlife Conservation

**Organizers:** Jianguo Li
Weiyao Lin
Hanlin Tang
Greg Mori
Joachim Denzler

**Location:** Room 327 A

**Time:** Full Day (0850-1735)

**Description:** This workshop aims to enhance the social responsibility of the CV community, and bring together researchers from both the CV community and wildlife conservation community together to advance wildlife conservation using CV techniques from 3 aspects:

1. Welcome contributed papers in a broad area of CV for wildlife conservation.
2. Organize a challenge on dataset we collected for Amur tiger conservation with tasks like tiger detection, pose estimation and re-identification.
3. Foster new ideas and directions on "CV for wildlife conservation" with invited talks and panel discussions from both communities.

## Visual Recognition for Medical Images

**Organizers:** Hoo-Chang Shin
Kyunghyun Cho
Donggeun Yoo

**Location:** Room 327 B-C

**Time:** Full Day (0900-1700)

**Description:** During last few years, visual recognition based on deep learning is receiving more attention in the medical image domain, where there is still much room for compensating human ability with machine vision. This workshop is dedicated to addressing the current challenges of visual recognition model development in medical image domain. By bringing leading researchers together and let them present, discuss, and share their up-to-date research outcomes, we expect this workshop contributes to solving fundamental research problems both in the field of visual recognition and medicine.

## Joint COCO and Mapillary Recognition Challenge

**Organizers:** Tsung-Yi Lin
Holger Caesar
Peter Kontschieder
Alexander Kirillov
Piotr Dollár

**Location:** Room 301

**Time:** Full Day (0900-1800)

**Description:** Benchmark challenges provide a focal point for the community to test the accuracy of algorithms, identify the state of the art, and discover novel directions for research. This workshop will host the Joint COCO and Mapillary Recognition Challenges, plus feature a new teaser challenge on large-vocabulary, few-shot instance segmentation (LVIS). While these challenges look at the general problem of visual recognition, the underlying datasets and the specific tasks in the challenges probe different aspects of the problem. This year the challenge will features instance and panoptic segmentation tracks on COCO and Mapillary datasets as well as key-points and densepose estimation tracks on COCO only.

## Disguised Faces in the Wild

**Organizers:** Rama Chellappa
Nalini Ratha
Richa Singh
Mayank Vatsa
Maneet Singh

**Location:** Room 308 A

**Time:** Full Day (0845-1700)

**Description:** Biometrics systems including face recognition systems can be attacked by various methods including presentation attacks. Disguises in face images are a form of presentation attack on face analytics. In the entertainment industry face disguise is an accepted norm. Disguise accessories such as sunglasses, masks, scarves, or make-up modify or occlude different facial regions which makes face recognition a challenging task. Disguise as a covariate involves both intentional and unintentional changes on a face through which one can either obfuscate his/her identity or impersonate someone else's identity. The problem can be further exacerbated due to unconstrained environment or "in the wild" scenarios. The 2nd International Workshop on Disguised Faces in the Wild focuses on understanding the state-of-the-art on face recognition in the presence of disguise variations. The Disguised Faces in the Wild 2019 (DFW2019) competition was also organized in conjunction with the workshop, in order to push the state-of-the-art on disguised face recognition. The scope of this workshop extends beyond face recognition under disguised variations to recognizing partially occluded faces or faces with spoofing variations. We believe that research in the field of disguised face recognition would facilitate the development of robust algorithms, applicable in several real world applications. As part of this workshop, the keynote speakers will discuss the current challenges faced by face recognition systems, especially in the context of digital/physical attacks.

## Robust Subspace Learning and Applications in Computer Vision

**Organizers:** Thierry Bouwmans
Sajid Javed
Soon Ki Jung
Paul Rodriguez
Namrata Vaswani
René Vidal
Brendt Wohlberg
El-Hadi Zahzah

**Location:** Room E4

**Time:** Full Day (0800-1800)

**Description:** Robust subspace learning/tracking/clustering either based on robust statistics estimation on reconstruction error and on decompositions into low-rank/sparse plus additive matrices/tensors provide suitable frameworks for many computer vision applications. The goals of this workshop are thus three-fold: 1) designing robust subspace methods for computer vision applications; 2) proposing new adaptive and incremental algorithms with convergence guarantees that reach the requirements of real-time applications such as background/foreground detection, and 3) proposing robust algorithms to handle the key challenges in computer vision application. This workshop also adress how to brigde robust subpace learning and deep learning to introduce more robustness in deep learning.

## Large-Scale Video Object Segmentation Challenge

**Organizers:** Ning Xu
Linjie Yang
Yuchen Fan
Thomas S. Huang
Jianchao Yang

**Location:** Room 318 B-C

**Time:** Full Day (0900-1800)

**Description:** Video object segmentation is an important video understanding task. We present the second large-scale video object segmentation challenge which is based on a recently published dataset (YouTube-VOS) with 4000+ YouTube videos and 90+ categories. In this year's challenge, in addition to the popular semi-supervised video object segmentation task, we also propose a new video segmentation task called video instance segmentation. The new task requires automatic segmentation and tracking of all instances in a video, which is the extension of image instance segmentation in the video domain. The workshop is split into two sessions, with the morning session focusing on semi-supervised video object segmentation and the afternoon session focusing on video instance segmentation. During each session, the winning teams and invited speakers will share us with the state-of-the-art methods for video segmentation and understanding.

## Intelligent Short-Video

**Organizers:** Guodong Guo
Niculae Sebe
Ying Shan
Changhu Wang
Ying Wu

**Location:** Room E3

**Time:** Full Day (0900-1700)

**Description:** Short-video refers to video content with length ranging from a few seconds to a few minutes. They are usually played on various mobile devices during fragmented leisure time. Short-video gained popularity on the social media platforms such as Vine and Snapchat, and enjoys explosive growth outside of US especially in China. There are a handful of mega Apps with over 100 million DAU (daily active users) in China, feeding various forms of short videos to billions of users. In contrast to the red-hot short-video industry and enthusiastic users, the response from the research community of computer vision has been sporadic. The workshop will strive to gather the most up-to-date information from both industry and academia, and take a holistic view of short-video through the lens of algorithm design and research. We also hope the workshop will inspire more computer vision researchers to join the cause of defining short-video as an emerging research field, and make it more and more intelligent.

## Statistical Deep Learning in Computer Vision

**Organizers:** Ping Luo
Mete Ozay
Hongyang Li
Chaochao Lu
Lei Huang
Wenqi Shao
Xianfeng Gu
Alan L. Yuille
Xiaogang Wang
Yi Ma
Lizhong Zheng
Wenyan Wu

**Location:** Room E5-E6

**Time:** Full Day (0830-1700)

**Description:** We consider statistical approaches employed to improve our understanding of deep learning, and to develop methods to boost their properties, with applications in computer vision, such as object recognition, detection, segmentation, tracking, scene description, visual question answering, robot vision, image enhancement and recovery. The workshop will consist of invited talks, oral talks, poster sessions and a research panel. Our target audience is graduate students, researchers and practitioners who have been working on development of novel statistical deep learning algorithms and/or their application to solve practical problems in computer vision.

## Computer Vision for Road Scene Understanding and Autonomous Driving

**Organizers:** Mathieu Salzmann
Jose Alvarez
Lars Petersson
Fredrik Kahl
Bart Nabbe
Lukas Neumann
Andrea Vedaldi
Andrew Zisserman
Bernt Schiele

**Location:** Room 300

**Time:** Full Day (0830-1700)

**Description:** Analyzing road scenes using cameras could have a crucial impact in many domains, such as autonomous driving, advanced driver assistance systems (ADAS), personal navigation, mapping of large scale environments and road maintenance. For instance, vehicle infrastructure, signage, and rules of the road have been designed to be interpreted fully by visual inspection. As the field of computer vision becomes increasingly mature, practical solutions to many of these tasks are now within reach. Nevertheless, there still seems to exist a wide gap between what is needed by the automotive industry and what is currently possible using computer vision techniques. The goal of this workshop is to allow researchers in the fields of road scene understanding and autonomous driving to present their progress and discuss novel ideas that will shape the future of this area. In particular, this workshop aims to bridge the gap between the community that develops novel theoretical approaches for road scene understanding and the community that builds working real-life systems performing in real-world conditions.

**Notes:**

## Real-World Recognition From Low-Quality Images and Videos

**Organizers:** Yuqian Zhou
Yunchao Wei
Zhangyang Wang
Jiaying Liu
Ding Liu
Shaogang Gong
Jeffrey Cohn
Nicu Sebe
Honghui Shi
Thomas S. Huang

**Location:** Room 317 B-C

**Time:** Full Day (0900-1800)

**Description:** What is the current state-of-the-art for recognition and detection algorithms in non-ideal visual environments? While the visual recognition research has made tremendous progress in recent years, most models are trained, applied, and evaluated on high-quality (HQ) visual data, such as ImageNet benchmarks. However, in many emerging applications such as robotics and autonomous driving, the performances of visual sensing and analytics are largely jeopardized by low-quality (LQ) visual data acquired from complex unconstrained environments, suffering from various types of degradations such as low resolution, noise, occlusion and motion blur. Other factors, such as contrast, brightness, sharpness, and out-of-focus, all have various negative effects on visual recognition. We organize this one-day workshop to provide an integrated forum for researchers to review the recent progress of robust recognition models from LQ visual data, and the novel image restoration algorithms. RLQ 2019 consists of challenge, keynote speech, paper presentation, poster session, special session on privacy and ethics of visual recognition, and a panel discussion from the invited speakers.

## Extreme Vision Modeling

**Organizers:** Vignesh Ramanathan
Dhruv Mahajan
Laurens van der Maaten
Alex Berg
Ishan Misra
Rahul Sukthankar

**Location:** Room E2

**Time:** Full Day (0915-1730)

**Description:** This workshop provides a forum for researchers from industry and academia to discuss extreme paradigms in training computer vision models. We will focus on extremes in the scale of data, either a small handful or billions, and extremes of training labels from fully-labeled and structured to self-supervised.

Over the past few years, pre-training on extremely large-scale datasets has emerged as a clear winner in most computer vision challenges. However, most of the techniques applied to a billion images have been straightforward extensions of techniques used for million scale (ImageNet) pre-training. There are fundamental research questions that need to be revisited at the billion scale - how to model noise in weakly labeled data, train on millions of classes, address the long-tail distribution of labels etc. Even at the billion scale, there are still categories with a handful of samples which means that large-scale data alone cannot solve this problem. Thus, we focus on the second extreme - low-shot learning of visual concepts. Finally, we also wish to spark a discussion on whether restricting training/evaluation of models to either a completely "self", "weakly" or "strongly" labelled setting is practical in this age of large-scale noisy datasets.

**Notes:**

## Gaze Estimation and Prediction in the Wild

Organizers: Hyung Jin Chang
Seonwook Park
Xucong Zhang
Otmar Hilliges
Aleš Leonardis

**Location:** Room 318 A

**Time:** Half Day - AM (0830-1245)

**Description:** Knowing what a user is looking at and understanding the eye movement patterns of the user can lead to a wide variety of novel applications in various applications, and many such applications are expected to be performed in environments beyond the laboratory. Unlike many other areas in computer vision, deep learning has only recently been introduced to address these challenges in the tasks of gaze estimation and gaze prediction. We aim to encourage and highlight novel strategies with a focus on robustness and accuracy in real-world settings. This is expected to be achieved via novel neural network architectures, incorporating anatomical insights and constraints, introducing new and challenging datasets, and exploiting multi-modal training among other directions.

## Human Behavior Understanding

**Organizers:** Xavier Alameda-Pineda
Xiaoming Liu
Elisa Ricci
Albert Ali Salah
Nicu Sebe
Sergey Tulyakov

**Location:** Room 308 B-C

**Time:** Half Day - AM (0900-1250)

**Description:** As in many other computer vision tasks, deep learning has brought revolutionary advances in human behaviour understanding from visual data. Deep models are now extremely effective not only in detecting and analyzing human faces, bodies and collective activities but also in generating realistic human-like behavioral data. From full-

body deepfakes to AI-based translation dubbing, deep networks can now synthesize images and videos of humans such as they are virtually indistinguishable from real ones. The workshop will focus on recent advances and novel methodologies for generating human behaviour data, with special emphasis on approaches for forging images and videos depicting real-looking human faces and/or full bodies and on algorithms for detecting fake human-like visual data.

## 360° Perception and Interaction

**Organizers:** Min Sun
Yu-Chuan Su
Wei-Sheng Lai
Hou-Ning Hu
Cheng Sun
Liwei Chan
Silvio Savarese
Kristen Grauman
Ming-Hsuan Yang

**Location:** Room 317 A

**Time:** Half Day - AM (0830-1210)

**Description:** 360° camera is a core building block of the Virtual Reality (VR) and Augmented Reality (AR) technology that bridges the real and digital world. With the rapid growth of VR/AR technology, the availability and popularity of 360° camera are also growing faster than ever. Content sharing sites like YouTube and Facebook enable their support for 360° images and videos. People now create, share, and watch 360° content in our everyday life, just like any other media.

We believe that a workshop for research centering around 360° content can greatly boost the research in the field and that this is the right time for the workshop. A major goal of this workshop will be to bring together researchers that are working on or interested in 360° related topics. We invite researchers from communities, including computer vision, HCI, multimedia, computer graphics, and machine learning to join the workshop. This will provide a forum to discuss current progress in the field and foster collaboration. It will also provide a good introduction for researchers that are interested and want to start their research in the field.

## Multi-Discipline Approach for Learning Concepts - Zero-Shot, One-Shot, Few-Shot and Beyond

**Organizers:** Nakamasa Inoue
Hirokatsu Kataoka
Yoshitaka Ushiku
Yusuke Matsui
Koichi Shinoda
Shin'ichi Satoh
Benoit Huet
Chong-Wah Ngo

**Location:** Room 307 A-C

**Time:** Half Day - AM (0900-1220)

**Description:** From what data can we learn concepts such as objects, actions, and scenes? Recent studies on zero-shot, one-shot, and few-shot learning have shown the effectiveness of collaboration between computer vision and natural language processing. This workshop promotes deeper and wider collaboration across many research fields to scale-up these studies. With the common theme, Learning Concepts, we hope to provide a platform for researchers to exchange knowledge from their respective backgrounds. The scope of this workshop includes Few-/Low-/k-Shot Learning, One-Shot Learning, Zero-Shot Learning, Cross-Domain Learning, and Meta Learning.

**Notes:**

## E-Heritage and Dunhuang Challenge

**Organizers:** Katsushi Ikeuchi
Xudong Wang
Takeshi Masuda
Takeshi Oishi
Guillaume Caron
Rei Kawakami
Shaodi You
Tianxiu Yu
Jiawan Zhang

**Location:** Room 317 A

**Time:** Half Day - PM (1330-1750)

**Description:** CV research plays an important role in cultural heritage preservation efforts. The e-Heritage workshop aims to bring together CV researchers as well as interdisciplinary researchers in Computer Graphics, Virtual Reality, Archaeology, and Art History, etc. Moreover, this year, we have organized the first Dunhuang World Heritage Challenge, which is an open challenge on data-driven e-heritage restoration using 1000 paintings of Dunhuang Grottoes. We introduce the winners of Dunhuan Challenge in this workshop.

## Large Scale Holistic Video Understanding

**Organizers:** Vivek Sharma
Mohsen Fayyaz
Ali Diba
Luc Van Gool
Juergen Gall
Rainer Stiefelhagen
Manohar Paluri

**Location:** Room 308 B-C

**Time:** Half Day - PM (1300-1730)

**Description:** In the last years, we have seen tremendous progress in the capabilities of computer systems to classify video clips. There are lots of works in video recognition field focusing on specific video understanding tasks, such as action recognition, scene understanding, etc. Current systems are expert in some specific fields of the general video understanding problem. However, for real-world applications, such as, analyzing multiple concepts of a video for video search engines and media monitoring systems or providing an appropriate definition of the surrounding environment of a humanoid robot, a combination of current state-of-the-art methods should be used. Therefore, in this workshop, we intend to introduce the holistic video understanding as a new challenge for the video understanding efforts. This challenge focuses on the recognition of scenes, objects, actions, attributes, and events in the real world videos. To be able to address such tasks, we also introduce our new dataset named Holistic Video Understanding~(HVU dataset) that is organized hierarchically in a semantic taxonomy of holistic video understanding. Almost all of the real-world conditioned video datasets are targeting human action or sport recognition. So our new dataset can help the vision community and bring more attention to bring more interesting solutions for holistic video understanding. The workshop is tailored to bringing together ideas around multi-label and multi-task recognition of different semantic concepts in the real world videos. And the research efforts can be tried on our new dataset.

## Open Images Challenge

**Organizers:** Vittorio Ferrari
Alina Kuznetsova
Rodrigo Benenson
Victor Gomes
Matteo Malloci

**Location:** Room 402

**Time:** Half Day - PM (1300-1730)

**Description:** The Open Images Challenge follows in the tradition of PASCAL VOC, ImageNet, and COCO; but at an unprecedented scale. It features 500 object classes and the training set contains 12M object bounding-boxes, 2.1M segmentation masks, and 375k relationship triplets. The images are very varied and often contain complex scenes with several objects. The Challenge includes three tracks: (1) object class detection; (2) visual relationship detection; (3) instance segmentation. We hope that the very large and diverse training set will stimulate research into more advanced computer vision models that will exceed current state-of-the-art performance. Also, having a single dataset with unified annotations for image classification, object

detection, visual relationship detection, and instance segmentation will hopefully promote studying these tasks jointly and stimulate progress towards genuine scene understanding.

## Comprehensive Video Understanding in the Wild

**Organizers:** Kuk-Jin Yoon
Kwanghoon Sohn
Ming-Hsuan Yang
Karteek Alahari
Yale Song

**Location:** Room 307 A-C

**Time:** Half Day - PM (1300-1700)

**Description:** Comprehensive video understanding has recently received increasing attention from the computer vision and multimedia communities with the goal of building machines that can understand the video like humans. Currently, most works for untrimmed video recognition mainly focus on isolated and independent problems such as action recognition or scene recognition. While they address different aspects of video understanding, there exist strong mutual relationships and correlations among action and scene. To achieve the very accurate human level understanding of untrimmed videos, the comprehensive understanding of various aspects such as what the actors are doing and where they are doing so is of great importance.

This workshop aims at providing a forum to exchange ideas in comprehensive video understanding with a particular emphasis on the untrimmed video summarization with temporal action and scene recognition in untrimmed videos. Papers presented in this workshop have to address one of independent video understanding problems including but not limited to

1. Video summarization
2. Scene recognition in untrimmed video
3. Action recognition in untrimmed video (or their joint problem)
4. Video summarization with temporal action and/or scene recognition

This workshop consists of two tracks: Regular and Challenge.

**Notes:**

# Monday, October 28

**NOTE:** Use the QR code for each tutorials's website for more information on that tutorial. Here's the QR code to the ICCV Tutorials page.

**0730–1700 Registration** (Hall E Lobby)

**1000–1100 Morning Break**

**1230–1330 Lunch** (On your own)

**1530–1630 Afternoon Break**

## Tutorial: Visual Recognition for Images, Video, and 3D

**Organizers:** Alexander Kirillov
Ross Girshick
Kaiming He
Justin Johnson
Georgia Gkioxari
Christoph Feichtenhofer
Saining Xie
Haoqi Fan
Yuxin Wu
Nikhila Ravi
Wan-Yen Lo
Piotr Dollár

**Location:** Auditorium

**Time:** Full Day (0900-1715)

**Description:** The purpose of this tutorial is to discuss popular approaches and recent advancements in the family of visual recognition tasks for different input modalities. We will cover in detail the most recent work on object recognition and scene understanding. Going beyond single images we will show current progress in video (detection and classification in video) and 3D visual recognition (multi-object mesh prediction). Our goal is to show existing connections between the techniques specialized for different input modalities and provide some insights about diverse challenges that each modality presents.

In conjunction with the tutorial, we are open-sourcing three new visual recognition systems for images, videos, and 3D respectively. These PyTorch-based systems contain multiple state-of-the-art methods in the corresponding domains. In our tutorial, we will pair each research talk with a talk that discusses these codebases sharing best engineering practices and showing details of implementation for each domain. We hope that such pairing will help researchers who are interested primarily in visual recognition to build and benchmark their systems easier. For researchers from different areas, we hope to make SOTA recognition systems easy to incorporate in their frameworks.

**Notes:**

## Tutorial: Holistic 3D Reconstruction: Learning to Reconstruct Holistic 3D Structures From Sensorial Data

**Organizers:** Zihan Zhou
Yasutaka Furukawa
Yi Ma

**Location:** Room 300

**Time:** Half Day - AM
(0830-1230)

**Description:** The perception of holistic scene structures, that is, orderly, regular, symmetric, or repetitive patterns and relationships in a scene, plays a critical role in human vision. When walking in a man-made environment, such as office buildings, a human can instantly identify parallel lines, rectangles, cuboids, rotational symmetries, repetitive patterns, and many other types of structure, and exploit them for accurate and robust 3D localization, orientation, and navigation. In computer vision, the use of such holistic structural elements has a long history in 3D modeling of physical environments, especially man-made environments, from data acquired by a variety of sensors such as monocular and binocular vision, LiDAR, and RGB-D sensors. These methods have shown great success and potential in creating high-fidelity 3D models, increasing the accuracy, robustness, and reliability of 3D vision systems, and facilitating modern 3D applications with a high-level, compact, and semantically rich scene representation.

In this context, this tutorial aims at bringing together the current research advances and discussing the state-of-the-art methods in 3D modeling of structured scenes and its applications. The tutorial will review the fundamental theory of multiview geometry of 3D structures; analyze traditional and recent geometric approaches in utilizing holistic 3D structures; present an overview of current confluence of learning-based approaches and geometry-based approaches. Finally we discuss possible future directions in combining reconstruction and recognition for 3D modeling of man-made environments.

## Tutorial: Large-Scale Visual Place Recognition and Image-Based Localization

**Organizers:** Eric Brachmann
Torsten Sattler
Giorgos Tolias

**Location:** Room 300

**Time:** Half Day - PM
(1330-1740)

**Description:** Given a database of geo-tagged images or images of known places, the goal of visual place recognition algorithms is to determine the place depicted in a new query image. Traditionally, this problem is solved by transferring the geo-tags or place identities of the most similar database images to the query image. Highly related to the visual place recognition problem is the task of visual localization: Given a scene representation computed from a database of geo-tagged images, e.g., a 3D model recovered via Structure-from-Motion, visual localization approaches aim to estimate the full 6 Degree-of-Freedom (6DOF) pose of a query image, i.e., the position and orientation from which the image was taken. Both place recognition and visual localization are fundamental steps in many Computer Vision applications, including robotics, autonomous vehicles (self-driving cars), Augmented / Mixed / Virtual Reality, loop closure detection in SLAM, and Structure-from-Motion. This tutorial covers the state-of-the-art in place recognition and visual localization, with three goals:

1. Provide a comprehensive overview over the current state-of-the-art. This is aimed at first- and second-year PhD students and engineers from industry who are getting started with or are interested in this topic.
2. Have experts teach the tricks of the trade to more experienced PhD students and engineers who want to refine their knowledge on place recognition and localization.
3. Highlight current open challenges in place recognition and localization. This outlines what current algorithms can and cannot do.

# Monday, October 28

**NOTE:** Use the QR code for each workshop's website to find the workshop's schedule. Here's the QR code to the ICCV Workshops page.

**0730–1700 Registration** (Hall E Lobby)

**1000–1100 Morning Break**

**1230–1330 Lunch** (On your own)

**1530–1630 Afternoon Break**

## Computer Vision for Physiological Measurement

**Organizers:** Wenjin Wang, Daniel McDuff, Sander Stuijk

**Location:** Room E4

**Time:** Full Day (0850-1610)

**Description:** Measuring physiological signals from the human face and body using cameras is an emerging topic that has grown rapidly in the last decade. Various human vital signs (e.g., heart rate (variability), respiration rate, blood oxygenation saturation, pulse transit time) can be measured by a remote camera without skin-contact, which is convenient and comfortable for long-term continuous vital signs assessment. The use of cameras also enables the analysis of human behaviors/activities and high-level visual semantics that can facilitate health monitoring and human understanding (e.g. affective computing). In this workshop, we will discuss recent advances and progress made by researchers in camera based physiological measurement, and its future challenges and potentials. We hope this workshop can increase the communication within our field and bring useful ideas/applications for neighboring fields in computer vision.

## Scene Graph Representation and Learning

**Organizers:** Ranjay Krishna, Jia Deng, Michael Bernstein, Fei-Fei Li

**Location:** Room 318 B-C

**Time:** Full Day (0830-1800)

**Description:** Graphs have enabled the innovation, adoption and use of numerous new spectral-based models like graph convolutions and graph-based evaluation metrics like SPICE. Modeling graphical data has historically been challenging for the machine learning community, especially when dealing with large amounts of data. Traditionally, methods have relied on Laplacian regularization through label propagation, manifold regularization or learning embeddings. Soon, operators on local neighborhoods of nodes became popular with their ability to scale to larger amounts of data and parallizable computation. Today's choice of architecture, the graph convolution, has become the de facto choice when dealing with graphical data. Graph convolutions, and similar techniques are slowly making their way into computer vision tasks and have recently been combined with RCNN to perform scene graph detection.

We hope to discuss the importance of structure in computer vision. How should we be representing scenes, videos, and 3D spaces? What connections to language and knowledge bases could aid vision tasks? How can we rethink the machine learning community's traditional relation-based representation learning? How can we both use and build upon spectral methods like random walks over graphs, message passing protocols, set-invariant neural architectures, and equivariant structured outputs? What are the shortcomings with our current representations and learning based methods and how can we remedy these problems? What tasks and directions should we be urging the community to move towards?

## YouTube-8M Large-Scale Video Understanding

**Organizers:** Joonseok Lee
Apostol (Paul) Natsev
Cordelia Schmid
Rahul Sukthankar
George Toderici
Ke Chen
Julia Elliott
Nisarg Kothari
Hanhan Li
Joe Yue-Hei Ng
Sobhan Naderi Parizi
Walter Reade
David Ross
Javier Snaider
Balakrishnan Varadarajan
Sudheendra Vijayanarasimhan
Yexin Wang
Zheng Xu

**Location:** Room 317 B-C

**Time:** Full Day (0900-1800)

**Description:** Many recent breakthroughs in machine learning and machine perception have come from the availability of large labeled datasets, such as ImageNet, which has millions of images labeled with thousands of classes, and has significantly accelerated research in image understanding. Google announced the YouTube-8M dataset in 2016, which spans millions of videos labeled with thousands of classes, with the hope that it would spur similar innovation and advancement in video understanding. YouTube-8M represents a cross-section of our society, and was designed with scale and diversity in mind so that lessons we learn on this dataset can transfer to all areas of our lives, from learning, to communication, to entertainment. The 3rd YouTube-8M Large-Scale Video Understanding Kaggle challenge and Workshop focus on temporal localization within a video. Segment/frame-level annotation or temporal localization is an important challenge in video understanding with various applications, such as searching within a video or discovering interesting action moments. In practice, segment-level annotation data is very hard and expensive to collect at large scale, making this problem very difficult. Thus, the main focus of this year's challenge is how to leverage noisy video-level labels and a small subset of segment-level calibration set jointly in order to better annotate and temporally localize concepts of interest.

## WIDER Face and Person Challenge

**Organizers:** Wanli Ouyang
Chen Change Loy
Dahua Lin
Hongsheng Li
Yuanjun Xiong
Qingqiu Huang
Dongzhan Zhou
Shuo Yang
Yantao Shen
Shuang Li
Wei Xia
Hongwei Qin
Kun Wang
Xingyu Zeng
Quanquan Li
Junjie Yan
Yuzhu Tang

**Location:** Room 327 B-C

**Time:** Full Day (0850-1730)

**Description:** Following the success of the First WIDER Challenge Workshop, we organize a new round of challenge in conjunction with ICCV 2019. The challenge centers around the problem of precise localization of human faces and bodies, and accurate association of identities. It comprises of four tracks:

- **WIDER Face Detection:** Aims at soliciting new approaches to advance the state-of- the-art in face detection.
- **WIDER Pedestrian Detection:** Has the goal of gathering effective and efficient approaches to address the problem of pedestrian detection in unconstrained environments.
- **WIDER Cast Search by Portrait:** Presents an exciting challenge of searching cast across hundreds of movies.
- **WIDER Person Search by Language:** Aims to seek new approaches to search person by natural language.

## Video Retrieval Methods and Their Limitations

**Organizers:** Ian Soboroff
Keith Curtis
Asad A. Butt
George Awad
Klaus Schoeffmann
Luca Rossetto
Werner Bailer

**Location:** Room 318 A

**Time:** Full Day (0900-1700)

**Description:** With the vastly increasing amount of video data being created, searching in video is a common task in many application areas, such as entertainment, surveillance, or education. The success of video search relies crucially on indexing video content, which is often done based on textual information, after extracting text or adding labels based on detection or classification of the visual or audio content. Video search systems are thus often built by integrating a set analysis components, many of which rely on computer vision algorithms, and fusing their results to create an efficiently searchable index. This has the consequence that the performance of video search systems is impacted by many factors, which makes the analysis of which components of the system contribute to the success or failure in a particular case difficult. The fact that many of the components have moved to deep neural networks (DNN) based approaches in recent years has not made this analysis easier. Benchmarking initiatives for video analysis and retrieval, such as TRECVID, have significantly contributed to a more systematic evaluation and have tremendously fostered the evolution of systems. However, their results show that there are usually outliers in the performance of a system on specific queries or datasets. In the existing literature, these aspects of comparative analysis and failure analysis are not sufficiently explored. This workshop will discuss contributions in video search using two types of queries: Generic search (natural language queries), and Instance Search (search by visual example).

## Closing the Loop Between Vision and Language

**Organizers:** Mohamed Elhoseiny
Anna Rohrbach
Xin Wang
Leonid Sigal
Marcus Rohrbach

**Location:** Room 317 A

**Time:** Full Day
(0850-1800)

**Description:** This workshop features invited talks, challenges, contributed spotlights and posters at the intersection of Computer Vision and NLP. Topics include visual question answering, generating textual descriptions from images and video, learning language embeddings of images, visual dialog, referring expression comprehension, vision-and-language navigation, and embodied question answering. Throughout the day, we are excited to welcome and hear from our invited speakers including Sanja Fidler, Mohit Bansal, Yejin Choi, Gunhee Kim, and Devi Parikh. The day concludes with a panel session discussing what we have learned in the last decade of vision & language research and what are the challenges for the next decade. The workshop will also feature a new edition of the Large Scale Movie Description Challenge (LSMDC), and the first VATEX Captioning Challenge for Multilingual Video Captioning. The LSMDC presents a new challenge this year, aiming at multi-sentence movie description generation. When describing sequences of events, it is important to distinguish "who is who", thus, the challenge will have a focus on identifying movie characters. The VATEX dataset is a new large-scale multilingual video description dataset, which contains over 41,250 videos and 825,000 captions in both English and Chinese. This year's Captioning Challenge aims to benchmark progress towards models that can describe videos in both languages.

## Neural Architects

**Organizers:** Samuel Albanie
Li Shen
Jie Hu
Barret Zoph
Andrea Vedaldi
Andrew Zisserman

**Location:** Room 308 B-C

**Time:** Full Day (0915-1730)

**Description:** Deep Neural Networks (DNNs) now represent a fundamental building block of many machine perception methods. The reason is simple—these models achieve exceptional performance. DNNs represent the state-of-the-art for core competencies such as image classification, object detection and semantic segmentation as well as for integrated approaches to higher level tasks including environment mapping and video understanding. While their usefulness is clear, our understanding of how best to design these models remains far from complete.

The goal of this workshop is to bring together researchers to discuss questions and ideas relating to various aspects of the structure and design of DNNs. We hope to consider in particular the following the two questions:

1. What we have learned as a community from our experience of designing these models?
2. Which research directions are most promising for improving existing architectures?

## 3D Reconstruction in the Wild

**Organizers:** Jan-Michael Frahm
Adrian Hilton
Tomas Pajdla
Akihiro Sugimoto

**Location:** Room 301

**Time:** Full Day (0850-1800)

**Description:** Research on 3D reconstruction has long focused on recovering 3D information from multi-view images captured in ideal conditions. However, the assumption of ideal acquisition conditions severely limits the deployment possibilities for reconstruction systems, as typically several external factors need to be controlled, intrusive capturing devices have to be used or complex hardware setups need to be operated to acquire image data suitable for 3D reconstruction. In contrast, 3D reconstruction in unconstrained settings (referred to as 3D reconstruction in the wild) usually imposes only limited to no restrictions on the data acquisition procedure and/or on data capturing environments, and therefore, represents a far more challenging task. The goal of this workshop is to foster the development of 3D reconstruction techniques capable of operating in unconstrained conditions which are robust and/or real-time, and perform well on a variety of environments with different characteristics. Towards this goal, we are interested in all parts of 3D reconstruction techniques ranging from multi-camera calibration, feature extraction, matching, data fusion, depth learning, and meshing techniques to 3D modeling approaches capable of operating on image data captured in the wild.

## Visual Object Tracking Challenge

**Organizers:** Matej Kristan
Aleš Leonardis
Jiři Matas
Michael Felsberg
Roman Pflugfelder
Joni-Kristian Kämäräinen

**Location:** Room E3

**Time:** Full Day (0900-1740)

**Description:** The Visual Object Tracking (VOT) Challenges provide the tracking community with a precisely defined and repeatable way of comparing short-term trackers and long-term trackers as well as a common platform for discussing the evaluation and advancements made in the field of visual tracking. VOT2019 is the seventh in a row of highly successful VOT challenges. In addition to RGB short-term, long-term and real-time tracking challenges, two novel challenges addressing multi-spectral tracking are introduced (RGB-depth and RGB-thermal). The VOT2019 program contains presentation of challenge results, talks from authors of the winning trackers, presentations of contributed papers, a keynote talk and a panel discussion.

## Person in Context Challenge

**Organizers:** Si Liu
Chen Qian
Yue Liao
Lejian Ren
Guanghui Ren
Hongyi Xiang
Guanbin Li
Fei Wang
Yanjie Chen;

**Location:** Room 308 A

**Time:** Full Day (0900-1710)

**Description:** Cognition involves both recognizing and reasoning about our visual world. Among the tens of thousands of categories of the world, the human is MOST special one. Understanding the human, including his/her action, pose, identity, appearance, the interactions between multiple human, the interaction between person and object etc, is a very good breakthrough point to cognition of the world. In the PIC 1.0, we presented the Person in Context (PIC) dataset to enable the comprehensive understanding of human in the image by converting images into scene graphs. The PIC 2.0 aims to further investigate the potential application value of scene graph from general relations to specified relations.

## Image and Video Synthesis: How, Why and What If?

**Organizers:** Shiry Ginosar
Taesung Park
Jun-Yan Zhu
Ming-Yu Liu
Aaron Hertzmann

**Location:** Room E5-E6

**Time:** Full Day (0845-1800)

**Description:** Generative modeling approaches are now at the point where high definition images can be synthesized from noise vectors and conditional methods enable video synthesis and future prediction. These technologies are reaching the point when they work well enough to both fascinate and disturb the general public, and to provide a rich unexplored medium of expression for artists.

While the end results seem similar, the approaches taken in visual synthesis range from conditional generative adversarial networks, through variational auto encoders to traditional graphics tricks of the trade. Moreover, the goals of synthesis research vary from modeling statistical distributions in machine learning, through realistic picture-perfect recreations of the world in graphics, and all the way to providing tools of artistic expression.

Additionally, there is a disconnect between research aimed at synthesis and practitioners interested in forensics. The issue of fake content synthesis and detection has recently become relevant to the public at large as a result of current political and social trends, and we can no longer afford to operate in two parallel universes.

## Autonomous Driving

**Organizers:** Dengxin Dai
Simon Hecker
Marius Cordts
Wim Abbeloos
Daniel Olmeda Reino
Jiri Matas
Roberto Cipolla
Luc Van Gool

**Location:** Room 401

**Time:** Full Day (0850-1800)

**Description:** Autonomous driving (AD) will have a substantial impact on people's daily life, both personally and professionally. As such, developing automated vehicles is becoming the core interest of several industrial and academic players. With so much effort poured into this field, all technologies concerned with AD are enjoying great progress. While it is exciting to see rapid advances in so many sub-fields, it is becoming hard to keep an overview of topics related to Autonomous Driving. Our goal therefore is to provide a better overview of recent challenges and trends for both researchers and practitioners.

## Low Power Computer Vision

**Organizers:** Alexander C. Berg
Bo Chen
Yiran Chen
Yen-Kuang Chen
Eui-Young Chung
Jaeyoun Kim
Svetlana Lazebnik
Yung-Hsiang Lu
Sungroh Yoon

**Location:** Room 403

**Time:** Full Day (0830-1730)

**Description:** This workshop investigates how to implement computer vision on systems with limited resources, such as mobile phones. The workshop features invited speakers from past competition winners as well as leaders in industry and academia. The accepted papers present new solutions for low-power computer vision. This workshop includes a competition using recent accelerators designed specifically for artificial intelligence.

## Autonomous Driving – Beyond Single-Frame Perception

**Organizers:** Li Erran Li
Ruigang Yang
Dragomir Anguelov
Silvio Savarese

**Location:** Room 402

**Time:** Full Day (0845-1830)

**Description:** The ICCV 2019 Workshop on Autonomous Driving — beyond single frame perception builds on the CVPR 2019 Workshop on Autonomous Driving with a focus on multi-frame perception, prediction, and planning for autonomous driving. It aims to get together researchers and engineers from academia and industries to discuss computer vision applications in autonomous driving. In this one day workshop, we will have invited speakers and panel discussions to present the current state of the art, as well as the limitations and future directions for computer vision in autonomous driving, arguably the most promising application of computer vision and AI in general.

## Assistive Computer Vision and Robotics

**Organizers:** Giovanni Maria Farinella
Marco Leo
Gerard G. Medioni
Mohan Trivedi
Antonino Furnari

**Location:** Room E1

**Time:** Full Day (0830-1800)

**Description:** While the end results seem similar, the approaches taken in visual synthesis range from conditional generative adversarial networks, through variational auto encoders to traditional graphics tricks of the trade. Moreover, the goals of synthesis research vary from modeling statistical distributions in machine learning, through realistic picture-perfect recreations of the world in graphics, and all the way to providing tools of artistic expression.

**Notes:**

## Linguistics Meets Image and Video Retrieval

**Organizers:** Amrita Saha
Hui Wu
Adriana I. Kovashka
Andrei Barbu
Xiaoxiao Guo
Karthik Sankaranarayanan
Samarth Bharadwaj
Yupeng Gao
Soumen Chakrabarti
Rogerio S. Feris

**Location:** Room E2

**Time:** Half Day - AM (0830-1200)

**Description:** Image and video retrieval systems have been one of the widely studied areas in computer vision for decades. In recent years, the need for effective retrieval systems has intensified, finding its use in many application domains, such as e-commerce, surveillance and Internet search. Over the past few years, the advent of deep learning has propelled the research of visual content retrieval and the field has been evolving at a fast pace. Amongst progress on core topics in image retrieval such as efficient search, ranking algorithms, and recommender systems, there has been a burgeoning trend on exploiting natural language understanding in the context of visual media retrieval. The initial attempts at the intersection of visual content retrieval systems and natural language understanding have explored topics such as interactive search using natural language feedback, image and video retrieval based on natural language queries, and task-oriented visual dialog agents for image retrieval. These recent works are opening up new paths forward, centering around open issues such as a) how can comprehension and communication of language enhance visual search? and b) how can information retrieval (IR) tools, algorithms and infrastructure assist multimodal knowledge acquisition, interaction and interpretability? The goal of the workshop is to bring together emerging research in the areas of information retrieval, computer vision and natural language understanding to discuss open challenges and opportunities and to study the different synergistic relations in this interdisciplinary area.

## Visual Perception for Robot Navigation in Human Environment: The JackRabbot Dataset

**Organizers:** Hamid Rezatofighi
Roberto Martín-Martín
Ian Reid
Silvio Savarese

**Location:** Room 307 A

**Time:** Half Day - AM (0830-1230)

**Description:** In the recent past, the computer vision community has proposed several centralized benchmarks to evaluate and compare different machine visual perception solutions. However, existing benchmarks mainly focus on the one or few visual perception tasks defined on single RGB images or RGB video sequences. With the rise of popularity of 3D sensory data systems based on LiDAR, some benchmarks have begun to provide both 2D and 3D sensor data, and to define new scene understanding tasks on this geometric information. Nonetheless, their targeted domain application is autonomous driving.

In this workshop, we target a unique visual domain, captured from a human size robot platform using 2D and 3D sensors, tailored to the perceptual tasks related to navigation in human environments, both indoors and outdoors. We hope that this new domain provide the community an opportunity to develop visual perception frameworks for various types of autonomous navigation agents, not only self-driving cars but also other types of agents like social mobile robots. These agents require understanding both indoor and outdoor scenes in order to interact successfully with humans, predict their behaviour in these environments, and incorporate this behaviour in agent's planning and decision processes.

## Lightweight Face Recognition Challenge

**Organizers:** Jiankang Deng
Jia Guo
Debing Zhang
Yafeng Deng
Shi Song
Xiangju Lu

**Location:** Room 307 B-C

**Time:** Half Day - AM (0830-1250)

**Description:** Face recognition in static images and video sequences captured in unconstrained recording conditions is one of the most widely studied topics in computer vision due to its extensive applications in surveillance, law enforcement, bio-metrics, marketing, and so forth. Recently, methodologies that achieve good performance have been presented in top-tier computer vision conferences (e.g., ICCV, CVPR, ECCV etc.) and great progress has been achieved in face recognition with deep learning-based methods. Even though comprehensive benchmarks and extensive efforts exist for deep face recognition, very limited effort has been made towards benchmarking lightweight deep face recognition, which aims at model compactness and energy efficiency to enable efficient system deployment. In ICCV 2019, we make a significant step further and propose a new comprehensive benchmark, as well as organise the first challenge & workshop for lightweight deep face recognition.

**Notes:**

## Sensing, Understanding and Synthesizing Humans

**Organizers:** Ziwei Liu
Tinghui Zhou
Angjoo Kanazawa
Yu Rong

**Location:** Room 307 B-C

**Time:** Half Day - PM (1355-1735)

**Description:** Great progress has been achieved in human sensing, understanding and synthesis. We further identify three key issues of the future directions:

- We should take a holistic view on the integral pipeline of human sensing/understanding/synthesis, and explore the frontier problems emerged from their intersections.
- The lessons, practices and foresights from both academia and industry should be shared and discussed together.
- These topics lay the foundation for human-centric AI and will play a greater role in the age of intelligent well-being.

We hope this workshop fosters inter-disciplinary research in these fields that could profoundly advance our society.

## Moving Cameras

**Organizers:** Chung-Ching Lin
Lisa M. Brown
Rogerio S. Feris
Mubarak Shah

**Location:** Room E2

**Time:** Half Day - PM (Time TBA)

**Description:** Traditional research on intelligent video analytics has primarily focused on video analysis from fixed overhead cameras where techniques such as background modeling are commonly used for moving object detection. More recently, wearable visual sensors and cameras mounted on aerial and ground vehicles are becoming increasingly accessible in terms of cost and availability, leading to new forms of visual sensing based on moving cameras. For example, dash cams are being mounted in police vehicles for license plate recognition; police officers are starting to use body-worn cameras in patrol operations; and drones are gaining significant popularity in a variety of applications, including law enforcement. These mobile devices are significantly expand-

ing the scope of video analytics beyond traditional static cameras by providing quicker and more effective means of crime fighting, such as wide area monitoring for civil security and crowd analytics for large gathering and sports events. Combining stationary cameras with moving cameras enables new capabilities in video analytics, at the intersection of Wearables, Internet of Things, Smart Cities, and sensing. The goal of this workshop is to bring together researchers from the area of intelligent video analytics from moving cameras (body cams, dash cams, drones and other UAVs), in order to discuss emerging technology in the intersection of these areas, as well as their societal implications.

## Recovering 6D Object Pose

**Organizers:** Tomáš Hodaň
Rigas Kouskouridas
Tae-Kyun Kim
Jiří Matas
Carsten Rother
Vincent Lepetit
Ales Leonardis
Krzysztof Walas
Carsten Steger
Eric Brachmann
Bertram Drost
Juil Sock

**Location:** Room 307 A

**Time:** Half Day - PM (1330-1830)

**Description:** Object pose estimation is of great importance to many higher-level tasks such as robotic manipulation, augmented reality and autonomous driving. The introduction of consumer and industrial grade RGB-D sensors and the advent of deep learning have allowed for substantial improvement in the field. However, there still remain challenges to be addressed such as robustness against occlusion and clutter, scalability to multiple objects, effective synthesis of training data, and fast and reliable object modeling, including capturing of reflectance properties. Extending contemporary methods to work reliably and with sufficient execution speed in an industrial setting is still an open problem. Many recent methods focus on specific rigid objects but pose estimation of deformable or articulated objects and of object categories is also an important research direction. In this workshop, people working on relevant topics in both academia and industry will share up-to-date advances and identify open problems. The workshop will feature several invited talks, presentation of accepted workshop papers and presentation of the BOP Challenge 2019 awards.

## Observing and Understanding Hands in Action

**Organizers:** Tae-Kyun Kim
Guillermo Garcia-Hernando
Antonis Argyros
Vincent Lepetit
Anil Armagan
Iason Oikonomidis
Angela Yao

**Location:** Room 327 A

**Time:** Half Day - PM (1330-1730)

**Description:** The fifth edition of the HANDS workshop aims at gathering researchers with interested in computer vision problems involving hands such as 2D/3D hand detection, hand segmentation, hand pose estimation, hand tracking, and their applications. This year we emphasize hand-object interaction and RGB-based hand pose estimation via invited speakers and a public challenge competition (HANDS 2019 Challenge). Development of RGB-D sensors and camera miniaturization have opened the door to a whole new range of technologies and applications which require detecting hands and recognizing hand poses in a variety of scenarios, including AR/VR, assistive car driving, robot grasping, and health care. A majority of hand tracking data sets and papers have been focused on near-range front-on scenarios, where a single hand or multiple hands appear visible or under moderate occlusions. Most existing methods fail to address severe occlusions under hand-object or hand-hand interaction scenarios. In parallel, RGB-based (cf. depth-based) hand pose estimation has been increasingly important in recent literature with new benchmarks and methods, yet many challenges remain. The goal of this workshop is to push the boundaries of 3D hand pose estimation and its relevant problems under hand-object interaction scenarios using depth images and/or RGB images.

# Saturday, November 2

**NOTE:** Use the QR code for each tutorials's website for more information on that tutorial. Here's the QR code to the ICCV Tutorials page.

**0730–1700 Registration** (Hall E Lobby)

**1000–1100 Morning Break**

**1230–1330 Lunch** (On your own)

**1530–1630 Afternoon Break**

## Tutorial: Accelerating Computer Vision With Mixed Precision

**Organizers:** Arun Mallya
Carl Case
Paulius Micikevicius
Pavlo Molchanov
Karan Sapra
Guilin Liu
Ting-Chun Wang
Ming-Yu Liu

**Location:** Room E5

**Time:** Half Day - AM (0800-1100)

**Description:** New levels of accuracy in computer vision, from image recognition and detection, to generating images with GANs, have been achieved by increasing the size of trained models. Fast turn-around times while iterating on the design of such models would greatly improve the rate of progress in this new era of computer vision. Our tutorial will describe techniques to utilize half-precision floating point representation that allow deep learning practitioners to accelerate the training of large deep networks while also reducing memory requirements. We will demonstrate how to benefit from mixed precision training for several computer vision tasks, including image classification, detection, segmentation, and synthesis.

This tutorial will provide a deep-dive into available software packages that enable easy conversion of models to mixed precision training, practical application examples and tricks of the trade (mixed precision arithmetic, loss scaling, etc.), as well as considerations relevant to training many popular models in commonly used deep learning frameworks including PyTorch and TensorFlow.

## Tutorial: 3D Deep Learning and Applications in Autonomous Driving

**Organizers:** Li Erran Li
Hao Su

**Location:** Room E5

**Time:** Half Day - PM (1330-1830)

**Description:** 3D understanding is crucial for many applications such as self-driving cars, autonomous robots, virtual reality, and augmented reality. Different from 2D images that have a dominant representation as regular pixel arrays, 3D data can come as irregular 3D point cloud such as from LiDAR sensors. This poses challenges to deep architecture design.

Tremendous progresses have been made in recent several years. This tutorial will cover deep learning algorithms that analyze 3D data for 3D understanding such as 3D semantics segmentation, 3D object detection and tracking. Despite these advances, fundamental challenges remain for problems such as activity recognition, behavior prediction, and inferring spatial relationship of objects in 3D scene in both static and dynamic environment. Furthermore, as our world is inherently 3D, 3D deep learning could be essential to make representation learning robust to input perturbation and generalize to real-world variations with high sample efficiency (e.g. transformation invariance). This tutorial presents a timely opportunity to engage the computer vision community with the unique challenges and opportunities presented in 3D deep learning.

## Tutorial: Second- and Higher-Order Representations in Computer Vision

**Organizers:** Piotr Koniusz
Mehrtash Harandi
Lei Wang
Ruiping Wang

**Location:** Room E4

**Time:** Half Day - PM (1330-1820)

**Description:** This tutorial will include three parts intertwined by three talks from renowned speakers. Firstly, we will outline the foundations of second- and higher-order representations such as covariance and auto-correlation matrices, their derivations, connection to statistical moments, derivations of power normalising functions such as Eigenvalue Power Normalisation and the aggregations of kernel feature maps in the context of image classification, fine-grained recognition, action recognition, domain adaptation and few-shot learning. We will talk about discriminative Stein kernel which utilises label information, the sparsity structure among features to compute sparse inverse covariance matrix as representation, and kernel-based covariance matrices that go beyond standard covariances in end-to-end learning. We will present techniques for improving the robustness of second-order representations based on iterative techniques for spectral scaling (e.g., matrix square root or logarithm) and feature reweighing (e.g., democratic pooling) which can be performed by efficient forward and backward operations: unrolling the iterations as layers. Finally, we will introduce the recent advances of Riemannian metric learning. By using second-order statistics based visual representations e.g., SPD covariance matrix, linear subspace model, and Gaussian distribution, typical visual classification task can be formulated as metric learning on some specific Riemannian manifolds (e.g., SPD Riemannian manifold, Grassmann manifold, and Gaussian statistical manifold). We will discuss metric learning algorithms and their mathematical formulations, derivations, as well as their connections with the traditional linear metric learning paradigm.

## Tutorial: Visual Learning With Limited Labeled Data

**Organizers:** Rogerio Feris
Kate Saenko
Leonid Karlinsky
Ross Girshick
Judy Hoffman
Kevin Swersky
Lorenzo Torresani

**Location:** Room E6

**Time:** Half Day - PM (1330-1800)

**Description:** Deep neural networks have shown remarkable success in many computer vision tasks, but current methods typically rely on massive amounts of labeled training data to achieve high performance. Collecting and annotating such large training datasets is costly, time-consuming, and in many cases infeasible, as for certain tasks only a few or no examples at all may be available. In this tutorial, we address the problem of visual learning with limited labeled data. We plan to focus on state-of-the-art techniques to tackle this problem, including meta-learning and metric learning approaches for few-shot classification, hallucination-based techniques for sample synthesis, recent methods for domain adaptation, and applications in computer vision tasks such as video understanding, object detection, and instance segmentation. The organizers will share their extensive experience on this topic, and provide links to resources such as relevant datasets and source code.

**Notes:**

# Saturday, November 2

**NOTE:** Use the QR code for each workshop's website to find the workshop's schedule. Here's the QR code to the ICCV Workshops page.

**0730–1700 Registration** (Hall E Lobby)

**1000–1100 Morning Break**

**1230–1330 Lunch** (On your own)

**1530–1630 Afternoon Break**

## Compact and Efficient Feature Representation and Learning in Computer Vision

**Organizers:** Li Liu
Yu Liu
Wanli Ouyang
Jiwen Lu
Matti Pietikäinen
Luc Van Gool

**Location:** Room E2

**Time:** Full Day (0855-1715)

**Description:** Feature representation is at the core of many computer vision and pattern recognition applications such as image classification, object detection, image and video retrieval, image matching and many others. In the past few years we have witnessed significant progress in feature representation and learning. The popularity of traditional handcrafted features seems to be overtaken by DeepCNNs, which can learn powerful features automatically from data and have brought about breakthroughs in various problems in computer vision. However, these advances rely on deep networks with millions or even billions of parameters, and the availability of GPUs with very high computation capability and large scale labeled datasets plays a key role in their success. In other words, powerful DeepCNNs are data hungry and energy hungry.

With the prevalence of social media networks and the portable / mobile / wearable devices to access them, comes the current concern of the limited resources these offer. Therefore, there is a growing need for feature descriptors that are fast to compute, memory efficient, and that yet exhibit good discriminability and robustness.

Given sufficient annotated data, existing features - especially those produced by deep CNNs - have yielded good performance. Nonetheless, there are many applications where only limited amounts of annotated training data can be gathered (such as with many visual inspection or medical diagnostics tasks). Such applications are challenging for many existing feature representations, and require sample-efficient techniques to learn good representations. The workshop aims at stimulating computer vision researchers to discuss the next steps in this important research area.

## 3D Face Alignment in the Wild Challenge

**Organizers:** Laszlo A. Jeni
Jeffrey F. Cohn
Lijun Yin

**Location:** Room 317 A

**Time:** Full Day (Time TBA)

**Description:** Over the past few years a number of research groups have made rapid advances in dense 3D alignment from 2D video and obtained impressive results. How these various methods compare is relatively unknown. Previous benchmarks addressed sparse 3D alignment and single image 3D reconstruction. No commonly accepted evaluation protocol exists for dense 3D face reconstruction from video with which to compare them. The 2nd 3D Face Alignment in the Wild Challenge presents a new large corpora of profile-to-profile face videos annotated with corresponding high-resolution 3D ground truth meshes to enable comparisons among alternative methods.

## Computer Vision for Fashion, Art and Design

**Organizers:** Negar Rostamzadeh
Hui Wu
Ping Luo
Julia Lasserre
Xavier Snelgrove
David Vazquez
Thomas Boquet
Yuying Ge
Wayne Zhang
Leonidas Lefakis
Ruimao Zhang
Wei Zhang
Reza Shirvany
Luba Elliott
Chris Pal
Tao Mei
Rogerio S. Feris
Kristen Grauman

**Location:** Room 317 B-C

**Time:** Full Day (0830-1800)

**Description:** Creative domains render a big part of modern society, having a strong influence on the economy and cultural life. Much effort within creative domains, such as fashion, art and design, center around the creation, consumption and analytic of creative visual content. In recent years, there has been an explosion of research in applying machine learning and computer vision algorithms to various aspects of the creative domains including generating, analyzing and processing visual content. This ever-increasing interest is most evident in two important research trends: (1) Computer Vision for Fashion and (2) Visual Content Generation for Creative Applications.

This workshop aims to bring together researchers, practitioners and artists in computer vision, machine learning and creative domains to discuss open problems in the two above mentioned areas. This involves addressing interdisciplinary problems, with all of the challenges it entails. We hope to continue the success of the first workshop on starting and cultivating conversations between artists, professionals in creative industries and computer vision scientists, and create a new space for collaboration between these communities and begin to tackle these deep problems. To provide rich opportunities to share opinions and experience in such an emerging field, we will accept paper submission on established and novel ideas, as well as workshop challenges. Following the success from last year, we will continue hosting the art gallery competition and Fashion-Gen challenge.

## Transferring and Adapting Source Knowledge in Computer Vision and VisDA Challenge

**Organizers:** Tatiana Tommasi
David Vázquez
Kate Saenko
Xingchao Peng
Ben Usman
Kuniaki Saito
Ping Hu
Judy Hoffman
Antonio M. López
Wen Li

**Location:** Room 308 A

**Time:** Full Day (0830-1715)

**Description:** This is the 6th annual TASK-CV workshop that brings together computer vision researchers interested in domain adaptation and knowledge transfer techniques. A key ingredient of the recent successes in computer vision has been the availability of visual data with annotations, both for training and testing, and well-established protocols for evaluating the results. However, this traditional supervised learning framework is limited when it comes to deployment on new tasks and/or operating in new domains. In order to scale to such situations, we must find mechanisms to reuse the available annotations or the models learned from them. TASK-CV aims to bring together research in transfer learning and domain adaptation for computer vision. The workshop is also held jointly with the VisDA Domain Adaptation Challenge, which this year focuses on multi-source and semi-supervised domain adaptation.

## Advances in Image Manipulation

**Organizers:** Radu Timofte
Shuhang Gu
Martin Danelljan
Ming-Hsuan Yang
Luc Van Gool
Kyoung Mu Lee
Eli Shechtman
Ming-Yu Liu
Zhiwu Huang
Seungjun Nah
Richard Zhang
Andrey Ignatov

**Location:** Room 318 B-C

**Time:** Full Day (0730-1840)

**Description:** Image manipulation is a key computer vision problem, encompassing multiple different tasks, including restoration and completion of image information, enhancement of visual quality, and manipulation of image content to achieve a desired effect. Recent years have witnessed an increased interest from the vision and graphics communities in these fundamental topics of research, which has lead to a substantial progress in many areas. While image manipulation directly relate to image quality enhancement and editing applications, it also forms an important step in a growing range of applications, including surveillance, automotive, electronics, remote sensing, and medical image analysis. The emergence and ubiquitous use of mobile and wearable devices offer another fertile ground for additional applications and faster methods. This workshop aims to provide an overview of the new trends and advances in areas concerning image manipulation. This workshop builds upon the success of the Perceptual Image Restoration and Manipulation (PIRM) workshop at ECCV 2018, the workshop and Challenge on Learned Image Compression (CLIC) editions at CVPR 2018 and CVPR 2019 and the New Trends in Image Restoration and Enhancement (NTIRE) editions at CVPR 2017, 2018 and 2019 and at ACCV 2016. This workshop features papers addressing topics related to image and video manipulation, restoration and enhancement and hosts several challenges covering different tasks within the aforementioned topics.

## Eye Tracking for VR and AR

**Organizers:** Robert Cavin
Jixu Chen
Ilke Demir
Stephan Garbin
Oleg Komogortsev
Immo Schuetz
Abhishek Sharma
Yiru Shen
Sachin S. Talathi

**Location:** Room 318 A

**Time:** Full Day (0900-1715)

**Description:** Due to recent advances in specialized hardware and on-device processing, virtual and augmented reality technologies are projected to receive mainstream adoption. In order to create 3D immersive experience in those virtual worlds, tracking the user behavior plays an important role for interaction and efficiency. In particular, tracking eyes and gazes of users unlocks novel display and rendering architectures that can substantially enable intuitive and adaptive user experiences, and alleviate the computational requirements to render 3D environments. As a vital condition, such eye tracking approaches should reliably work all the time, for all individuals, under all environmental conditions.

The goal of this workshop is to raise awareness for new eye tracking challenges in VR and AR, to engage the broader computer vision and machine learning communities in those discussions, and to create a benchmark for current eye tracking approaches. We release some datasets and host two competitions for that purpose: (1) semantic segmentation challenge, and (2) synthetic eye generation challenge. Successful approaches will address some outstanding questions relevant to eye tracking for VR and AR platforms that potentially solve the aforementioned generalizability problem through deep learning. The workshop will involve enlightening keynotes, selected oral and poster presentations including winners of the challenges, and a panel discussion for potential collaborations.

## Multi-Modal Video Analysis and Moments in Time Challenge

**Organizers:** Dhiraj Joshi
Mathew Monfort
Kandan Ramakrishnan
Rogerio S. Feris
David Harwath
Dan Gutfreund
Carl Vondrick
Bolei Zhou
Hang Zhou
Zhicheng Yan
Aude Oliva

**Location:** Room 307 B-C

**Time:** Full Day (Time TBA)

**Description:** Video understanding/analysis is a very active research area in the computer vision community. This workshop aims to particularly focus on modeling, understanding, and leveraging the multi-modal nature of video. Recent research has amply demonstrated that in many scenarios multimodal video analysis is much richer than analysis based on any single modality. At the same time, multimodal analysis poses many challenges not encountered in modeling single modalities for understanding of videos (for e.g. building complex models that can fuse spatial, temporal, and auditory information). The workshop will be focused on video analysis/understanding related, but not limited, to the following topics:

- Deep network architectures for multimodal learning.
- Multimodal unsupervised or weakly supervised learning from video.
- Multimodal emotion/affect modeling in video.
- Multimodal action/scene recognition in video.
- Multimodal video analysis applications including but not limited to sports video understanding, entertainment video understanding, healthcare etc.
- Multimodal embodied perception for vision (e.g. modeling touch and video).
- Multimodal video understanding datasets and benchmarks.

## Deep Learning for Visual SLAM

**Organizers:** Ronald Clark
Sudeep Pillai
Alex Kendall
Angela Dai

**Location:** Room 301

**Time:** Full Day
(0845-1800)

**Description:** Visual SLAM and ego-motion estimation are two of the key challenges and cornerstone requirements of machine perception. However, to enable the next generation of visual SLAM, we need to pursue better means of integrating prior knowledge and a higher-level understanding of the world. A promising means of achieving this is by harnessing deep learning. Effectively harnessing deep learning in this context has the potential to revolutionise the field and realise the long-standing goal of machine perception - robust, lifelong SLAM that can bring accurate visual localization to a wide class of consumer devices and robotic platforms - ranging from from UAVs to ground vehicles to cellphones. The purpose of this workshop will be to stimulate discussion about these topics with the goal of finding new, innovate approaches to solving the SLAM problem.

## Learning for Computational Imaging

**Organizers:** Bihan Wen
Saiprasad Ravishankar
Brendt Wohlberg
Jong Chul Ye

**Location:** Room 327 B-C

**Time:** Full Day
(0830-1800)

**Description:** The LCI workshop will feature presentations on recent research in the rapidly growing area of learning for computational imaging covering models, algorithms, theory, and diverse applications.

## Geometry Meets Deep Learning

**Organizers:** Xiaowei Zhou
Kosta Derpanis
Emanuele Rodolà
Jonathan Masci
Michael Bronstein
Kostas Daniilidis

**Location:** Room 300

**Time:** Full Day (0830-1800)

**Description:** The goal of the GMDL workshop is to encourage the interplay between geometric vision and deep learning. Deep learning has emerged as a common approach to learning data-driven representations. While deep learning approaches have obtained remarkable performance improvements in most 2D vision problems, such as image classification and object detection, they cannot be directly applied to geometric vision problems due to the fundamental differences between 2D and 3D vision problems, such as the non-Euclidean nature of geometric objects, higher dimensionality, and the lack of large-scale annotated 3D datasets. Designing geometric components or constraints to improve the performance of deep neural networks is a promising direction worth further exploration. This workshop aims to bring together experts from the areas of 3D vision, graphics, and deep learning to summarize recent advances, exchange ideas, and inspire new directions.

## Real-World Face Recognition Challenge

**Organizers:** Yandong Guo
Lei Zhang
Rama Chellappa
Erik Learned-Miller

**Location:** Room E3

**Time:** Full Day (0830-1830)

**Description:** Though almost saturated performance has been achieved on several classic face recognition tasks in academia, including LFW and Megaface, there are still many open problems for face recognition in industrial applications. For example, the training data might be quite noisy and imbalanced. Our workshop is mainly to discuss how to solve these problems. The topics we cover include but not limited to large-scale face recognition, face recognition with imbalanced training data in the low-shot learning scenario, generative model for face synthesis, how humans and face verification algorithms can work together, bias in face recognition, multimodal unsupervised or weakly supervised learning from video, etc.

## Interpreting and Explaining Visual AI Models

**Organizers:** Jaesik Choi
Seong-Whan Lee
K.-R. Müller
Seongju Hwang
Bohyung Han
David Bau
Ludwig Schubert
Yong Man Ro

**Location:** Room 308 B-C

**Time:** Full Day (0830-1700)

**Description:** Explainable and interpretable machine learning models and algorithms are important topics which have received growing attention from research, application and administration. Many advanced visual artificial intelligence systems are often perceived as black-boxes. Researchers would like to be able to interpret what the AI model has learned in order to identify biases and failure models and improve models. The present workshop focusses on explainable or interpretable AI and ML, and will aim to establish new theoretical foundations of interpreting and understanding visual artificial intelligence models including deep neural networks.

This workshop has interest including, but not limited to, the following topics:

- Explaining the decision of visual deep learning models
- Interpretable deep learning models
- Machine learning/deep learning models which generates human-friendly explanations
- Bayesian model composition/decomposition methods
- Model-agnostic machine learning explainable models
- Evaluation of explainable AI models
- Causal analysis of complex AI/ML systems
- Practical applications of explainable AI

## AutoNUE: Autonomous Navigation in Unconstrained Environments

**Organizers:** Manmohan Chandraker
C V Jawahar
Anoop M.Namboodiri
Srikumar Ramalingam
Anbumani Subramanian
Girish Varma

**Location:** Room E1

**Time:** Full Day (0900-1800)

**Description:** The AutoNUE workshop has a goal to percolate autonomous driving to domains where road infrastructure is sub-optimal for computer vision and machine learning, but which stand to gain immeasurably from its benefits. It adopts a broad view of what is entailed by driving in unconstrained environments - aspects such as changes in weather, time of day or imaging conditions, and topics like domain adaptation. This workshop also poses data challenges for autonomous driving in less constrained traffic, along with infrastructure that is not always dependable.

**Notes:**

## Physics Based Vision Meets Deep Learning

**Organizers:** Yu Li
Ying Fu
Shaodi You
Yinqiang Zheng
Feng Lu
Boxin Shi
Robby T. Tan

**Location:** Room E6

**Time:** Half Day - AM (0900-1240)

**Description:** Following the success of 1st ICCV Workshop on Physics Based Vision Meets Deep Learning (PBDL2017). We propose the 2nd workshop using the same title and topics with ICCV 2019. The goal is to encourage the interplay between physics based vision and deep learning. Physics based vision aims to invert the processes to recover the scene properties, such as shape, reflectance, light distribution, medium properties, etc., from images. In recent years, deep learning shows promising improvement for various vision tasks. When physics based vision meets deep learning, there must be mutual benefits.

## Egocentric Perception Interaction and Computing

**Organizers:** Antonino Furnari
Giovanni Maria Farinella
Dima Damen
David Crandall
Kristen Grauman

**Location:** Room 307 A

**Time:** Half Day - AM (0830-1245)

**Description:** The EPIC (Egocentric Perception, Interaction and Computing) series of events aims to bring together the different communities which are relevant to egocentric perception, including Computer Vision, Machine Learning, Multimedia, Augmented and Virtual Reality, Human Computer Interaction, and Visual Sciences. The main goal of the workshop is to provide a discussion forum and facilitate the interaction between researchers coming from different areas

of competence. The topics discussed in the fifth edition of the international workshop on Egocentric Perception, Interaction and computing will include egocentric vision for human behavioral understanding, assistive technologies, object and action recognition, eye movement analysis, and augmented reality. The workshop will include five orals and about twenty posters among full papers, extended abstracts and invited presentations from works accepted for publication at ICCV 2019. Two invited keynote talks by Marc Pollefeys and Oswald Lanz will also be part of the workshop.

## Video Turing Test: Toward Human-Level Video Story Understanding

**Organizers:** Seongho Choi
Kyoung-Woon On
Yu-Jung Heo
Haeyong Kang
Krishna Mohan Chalavadi
Ting Han
Chang Dong Yoo
Gunhee Kim
Byoung-Tak Zhang

**Location:** Room E4

**Time:** Half Day - AM
(0830-1230)

**Description:** A story in a video is highly-abstracted information that consists of a series of events from multiple scenes. Human can easily make up a story from the video, however the current state-of-the-art machine learning methodologies have still been struggled to learn a story in the video in this abstraction level. To achieve the human-level machine intelligence on video story understating, a breakthrough advancement in machine intelligence is very necessary (e.g., event extraction from multimodal video data, causal relationships inference among events, prospect and retrospect of unseen events from observed events, etc.)

To promote comprehensive discussion around the related hot research topics, we are inviting experts from many fields, including computer vision, graphics, language processing, multimedia, computational narratology, neuro-symbolic computing and speech/sound recognition as well as initiating discussions of future challenges in data-driven video understanding.

## CroMoL: Cross-Modal Learning in Real World

**Organizers:** Yan Huang
Amir Zadeh
Qi Wu
Li Liu
Louis-Philippe Morency
Liang Wang
Matti Pietikäinen

**Location:** Room 307 A

**Time:** Half Day - PM
(1300-1730)

**Description:** To understand the world around us more intelligently and better, it needs to be able to interpret multimodal signals together. With the rapid growth of multimodal data (e.g., image, video, audio, depth, IR, text, sketch, synthetic, etc.), cross-modal learning, which aims to develop techniques that can process and relate information across different modalities, has drawn increasing attention recently. It is a vibrant multidisciplinary field of increasing importance and with extraordinary potential. It has been widely applied to many tasks such as cross-modal retrieval, phrase localization, visual dialogue, visual captioning, visual question answering, language-based person search/action detection/semantic segmentation, etc. However, real world applications pose various challenges to cross modal learning, such as limited training data, multimodal content imbalance, large visual-semantic discrepancy, cross-dataset discrepancy, missing modalities, etc. To address these challenges, quite a lot of attempts motivated from various perspectives (including visual attributes, data generation, meta-learning, etc.) have been made. However, those mentioned challenges are far from being solved. The goal of this workshop is to encourage researchers to present high quality work and to facilitate effective discussions on the potential solutions to those challenges.

## Should We Pre-Register Experiments in Computer Vision?

**Organizers:** João F. Henriques
Samuel Albanie
Luca Bertinetto
Jack Valmadre

**Location:** Room 327 A

**Time:** Half Day - PM (1400-1800)

**Description:** Computer vision research has benefited tremendously from the widespread adoption of standardised public benchmarks, which enable a direct comparison of new research with the existing literature. Unfortunately, the current incentive system relies heavily on these benchmarks as a proxy for scientific progress. Researchers are incentivised to "beat the state-of-the-art", sometimes at the expense of deep scientific understanding and rigorous experimental design. Moreover, since only positive results are rewarded, negative results inevitably encountered during research are not reported, and may be repeated by many different researchers. Experiment pre-registration is a publishing and reviewing model that aims to address these issues by changing the incentive system. A pre-registered paper is a regular paper that is submitted for peer-review without any experimental results, describing instead a protocol to be followed after the paper is accepted. Thus it is important for the authors to make compelling arguments, from theory or past published evidence, and for reviewers to assess these arguments together with the quality of the experimental design rather than with the numbers themselves. This workshop is a pilot study in pre-registration applied to computer vision, with a real trial of paper reviewing and publishing under this model. The results of pre-registered experiments in computer vision will be presented by authors, who together with invited speakers will bootstrap a discussion forum on current and future evaluations and incentives.

**Notes:**

## Silver Donors


## Non-Profit Donors


## Overall Meeting Sponsors

## Platinum Donors


## Gold Donors