MACHINE LEARNING AND VISION FOR INDUSTRIAL APPLICATIONS

MALVIC

AUTUMN SCHOOL (Online) October 18-22, 2021, Kristiansand, Norway

Machine learning and computer vision have the potential to significantly improve the automation and autonomy of many industrial applications (e.g. offshore, automotive, telecommunication, gaming and multimedia) by enhancing the operational performance, decreasing cost related to manual operations, increasing benefits, minimizing losses, optimizing productivity and improving safety and security. The goal of this Autumn School MALVIC is to bring together pioneering international scientists in machine learning and computer vision with both academia and practitioners from the industrial fields on a unique setting for the discussion and demonstration of practical, hands-on machine learning and vision research and development. Offshore industrial applications and industrial process scenarios are examples for the autumn school target.

Program (Tentative)

Monday 18 October 2021

Morning

09:00 – 09:30 **Opening**

09:30 – 11:45 **Lecture 1**: Collective Intelligence "the Ant Model"
Guy Theraulaz, CNRS (France)

11:45 – 12:45 **Lunch break**

Afternoon

12:45 – 15:45 **Special program**:
To be announced

15:45 – 16:00 **Coffee break**

16:00 – 18:15 **Lecture 2**: Learning Representations
Stefano Soatto, UCLA / Amazon (USA)

MACHINE LEARNING AND VISION FOR INDUSTRIAL APPLICATIONS

MALVIC

Tuesday 19 October 2021

Morning

09:00 - 10:15 **Lecture 3:** Modern Artificial Intelligence: 1980s-2021 and Beyond
Jürgen Schmidhuber, IDSIA (Switzerland)

10:15 - 11:30 **Lecture 4**:
To be announced

11:30 - 12:30 **Lunch break**

Afternoon

12:30 - 14:00 **Lecture 5**: Deep Visual SLAM
Daniel Cremers, TU Munich (Germany)

14:00 - 14:15 **Coffee break**

14:15 - 16:30 **Lecture 6**: Mathematics of Deep learning
René Vidal, JHU / Amazon (USA), NORCE (Norway)

16:30 - 16:45 **Coffee break**

16:45 - 18:15 **Lecture 7**: Live Perception for Mobile & Web
Matthias Grundmann, Google (USA)

MACHINE LEARNING AND VISION FOR INDUSTRIAL APPLICATIONS

MALVIC

Wednesday 20 October 2021

Morning

09:00 - 11:30 **Lecture 8**: Industrial Optimization and the Search for New Algorithm
Thomas Bäck, Leiden University (The Netherlands), NORCE (Norway)

11:30 - 12:30 **Lunch break**

Afternoon

12:30 - 14:00 **Lecture 9**: Practical talk
Gal Chechik, Nvidia (Israel)

14:00 - 14:15 **Coffee break**

14:15 - 16:30 **Lecture 10**: Ensemble Approaches to Class Imbalance Learningmathias
Xin Yao, Birmingham University (USA), SUSTech (China)

16:30 - 16:45 **Coffee break**

16:45 - 18:15 **Lecture 11**: Practical talk
To be announced

MACHINE LEARNING AND VISION FOR INDUSTRIAL APPLICATIONS

MALVIC

Thursday 21 October 2021

Morning

09:00 - 11:30 **Lecture 12:** Mining for Meaning. From Vision to Language
Marius Leordeanu, Politehnica University of Bucharest (Romania)

11:30 - 12:30 **Lunch break**

Afternoon

Lecture 13 – 15: Success Stories – How did AI Shape Your Business?

12:30 - 13:00 Introduction by chair Anne Grete Ellingsen, project manager national European Digital innovation Hub candidate. Short presentation of the EUs program on investment in digital infrastructure and the benefit for SMEs and start ups.

13:00 - 13:30 **Crayon** presented by *Geir Gulliksen*

13:30 - 14:00 **Intelecy** presented by *Espen Davidsen*

14:00 - 14:15 **Coffee Break**

14:15 - 14:45 **Rocket farm** presented by *Dan Peter Rye Moen*

14:45 - 15:15 **Aquabyte** presented by *Trude Jansen Hagland*

15:15 - 15:45 **Idean** presented by *Lars Petter Aase*

15:45 - 16:00 **Closing remarks**

Friday 22 October 2021

Morning

09:00 - 11:30 **Lecture 16**: Understanding Activities in an Industrial Context
Horst Bischof, Granz University of Technology (Austria)

11:30 - 12:30 **Lunch break**

Afternoon

12:30 - 14:00 **Lecture 17**: Looking Far Ahead … Perception Challenges in the Field of Autonomous Trucking
Fridtjof Stein, Daimler (Germany)

14:00 - 14:15 **Coffee break**

14:15 - 16:30 **Lecture 18**: Practical talk
To be announced

16:30 - 17:00 **Closing remarks**

MACHINE LEARNING AND VISION FOR INDUSTRIAL APPLICATIONS

MALVIC

INVITED SPEAKERS

Guy Theraulaz: Collective Intelligence "the Ant Model"

Prof. Guy Theraulaz, CNRS Research Director, France. He is a world-leading expert in the study of collective intelligence and collective behaviors in animal and human groups. He is also a leading researcher in the field of swarm intelligence and computational biology, primarily studying social insects but also distributed algorithms, e.g. for collective robotics, directly inspired by nature.

Jürgen Schmidhuber: Modern Artificial Intelligence - 1980s-2021 and Beyond

Significant historic events appear to be occurring more frequently as time goes on. Interestingly, it seems like subsequent intervals between these events are shrinking exponentially by a factor of four. This process looks like it should converge around the year 2040.

The last of these major events can be said to have occurred around 1990 when the cold war ended, the WWW was born, mobile phones became mainstream, the first self-driving cars appeared, and modern AI with very deep neural networks came into being. In this talk, I'll focus on the latter, with emphasis on Metalearning since 1987 and what I call "the miraculous year of deep learning" which saw the birth of—among other things—(1) very deep learning through unsupervised pre-training, (2) the vanishing gradient analysis that led to the LSTMs running on your smartphones and to the really deep Highway Nets/ResNets, (3) neural fast weight programmers that are formally equivalent to what's now called linear Transformers, (4) artificial curiosity for agents that invent their own problems (familiar to many nowadays in the form of GANs), (5) the learning of sequential neural attention, (6) the distilling of teacher nets into student nets, and (7) reinforcement learning and planning with recurrent world models. I'll discuss how in the 2000s much of this has begun to impact billions of human lives, how the timeline predicts the next big event to be around 2030, what the final decade until convergence might hold, and what will happen in the subsequent 40 billion years. Take all of this with a grain of salt though.

Prof. Jürgen Schmidhuber, Scientific Director of IDSIA, Switzerland. He is a computer scientist most noted for his work in the field of artificial intelligence, deep learning and artificial neural networks. He is a co-director of the Dalle Molle Institute for Artificial Intelligence Research in Manno, in the district of Lugano, in Ticino in southern Switzerland.

René Vidal: Mathematics of Deep learning

The past few years have seen a dramatic increase in the performance of recognition systems thanks to the introduction of deep networks for representation learning. However, the mathematical reasons for this success remain elusive. For example, a key issue is that the neural network training problem is non- convex, hence optimization algorithms may not return a global minima. In addition, the regularization properties of algorithms such as dropout remain poorly understood. The first part of this tutorial will overview recent work on the theory of deep learning that aims to understand how to design the network architecture, how to regularize the network weights, and how to guarantee global optimality. The second part of this tutorial will present an analysis of the optimization and regularization properties of dropout for matrix factorization. Examples from neuroscience and computer vision will also be presented.

Prof. René Vidal, Professor at JHU, USA, and Chief scientist at NORCE. He is the Herschel Seder Professor of Biomedical Engineering and the Inaugural Director of the Mathematical Institute for Data Science at The Johns Hopkins University. He has secondary appointments in Computer Science, Electrical and Computer Engineering, and Mechanical Engineering.

MACHINE LEARNING AND VISION FOR INDUSTRIAL APPLICATIONS

MALVIC

Thomas Bäck: Industrial Optimization and the Search for New Algorithms

Direct global optimization algorithms based on some instance of evolutionary computation have shown big successes in a wide range of application domains, for example engineering design optimization. When problem dimensionality is small (n < 20), so-called efficient global optimization (EGO) is also a very suitable class of algorithms, and I will introduce a generalization of the concept of an acquisition function in EGO that automatically handles the exploration – exploitation tradeoff.

In automatic machine learning, the optimization of hyperparameters (also called the algorithm configuration problem) is currently of considerable interest. I will briefly explain this problem and then provide some examples illustrating that this task can be handled by direct global optimization algorithms as well. While it is commonly applied to machine learning algorithms, algorithm configuration for evolution strategies is a new application domain. I will give a simple example how a combinatorial design space of 4.608 configuration variants of evolution strategies can be created and investigated using data mining. This kind of "combinatorial algorithmics" provides an opportunity for discovering the unexplored areas of the optimization algorithm design space. Finally, I provide a quick idea of an extension of EGO for the combined algorithm selection and hyperparameter optimization (CASH) task in machine learning.

To conclude, I return to engineering design optimization tasks, one in wing design and one in ship design. Both are multi-objective, both use a variant of efficient global optimization, and the first focuses on modeling user preferences in objective space while the second learns internal models of the constraints using radial basis functions. Both aim at illustrating today's requirements in engineering design applications.Thomas Bäck, Professor at Leiden University (The Netherlands) and Chief Scientist at NORCE. He **is** head of the Natural Computing Research Group and Director of Education at the Leiden Institute of Advanced Computer Science (LIACS). He received his PhD in Computer Science from Dortmund University, Germany, in 1994. He has been Associate Professor of Computer Science at Leiden University since 1996 and full Professor for Natural Computing since 2002.

Daniel Cremers: Deep Visual SLAM

Visual Simultaneous Localization and Mapping (SLAM) is of utmost importance to autonomous systems and augmented reality. I will discuss direct methods for visual SLAM (LSD SLAM and DSO) that recover camera motion and 3D structure directly from brightness consistency thereby providing better performance in terms of precision and robustness compared to classical keypoint-based techniques.

Moreover, I will demonstrate how we can leverage the predictive power of self-supervised deep learning in order to significantly boost the performance of direct SLAM methods. The resulting methods D3VO allow us to track a single camera with a precision that is on par with state-of-the-art stereo-inertial odometry methods.

Lastly, I will introduce MonoRec - a deep network that can generate faithful dense reconstruction of the observed world from a single moving camera.

Prof. Daniel Cremers is Professor of Informatics and Mathematics at TU Munich and Germany. He is one of the leading experts in computer vision, machine learning & deep networks with focus on mathematical image analysis (segmentation, motion estimation, multiview reconstruction, visual SLAM). In December 2010 he was listed among "Germany's top 40 researchers below 40" (Capital). On March 1st 2016, Prof. Cremers received the Gottfried Wilhelm Leibniz Award, the biggest award in German academia.

MACHINE LEARNING AND VISION FOR INDUSTRIAL APPLICATIONS

MALVIC

Horst Bischof: Understanding Activities in an Industrial Context

This talk will highlight some recent work in the area of understanding actions and human activities. Special emphasis will be devoted to sequence segmentation and recognition of complex (long-term) Activities and domain adaptation. Examples from real world applications will illustrate the presented methods.

Prof. Horst Bischof, Professor, and Vice-Rector TU Graz, Austria, is vice rector for research and Professor at the Institute for Computer Graphics and Vision at the Graz University of Technology, Austria. He has more than 750 publications with notable works on object recognition, visual learning, on-line and life-long learning, motion and tracking, visual surveillance and biometrics and medical computer vision.

Takeo Kanade:

Prof. Takeo Kanade, Professor, Carnegie Mellon University, USA, is a Japanese computer scientist and one of the world's foremorst scientists in computer vision. He has more than 300 publications and 20 patents and with notable works including Lucas-Kanade method, face detector, Tomasi-Kanade factorization method…etc.

Marius Leordeanu: Mining for Meaning. From Vision to Language

Prof. Marius Leordeanu, Professor, Politehnica University of Bucharest, is also a Senior Scientist of the Romanian Academy (IMAR). He holds a PhD from the Robotics Institute of CMU and Bachelor's in Computer Science and Mathematics in 2003, from Hunter College of the City University of New York.

Stefano Soatto: Learning Representations

Representations are functions of the data that are "useful" for a task. Of all functions, one wishes to design or learn those that contain all the "information" in the data, and none of the variability that is irrelevant to the task. Depending on how one defines and measures "useful" and "information", different notions of representations can be instantiated. What are the relationships among those? Are there common principles behind the different tools and models? Is there a common notion of "optimality" that emerges from all formalisms? If so, are such optimal representations computable? If not, can they be approximated? If such representations are learned using "past data" (training set), can we predict how well they will perform on "future data" (test set)?

These questions have nothing to do with Deep Learning, but understanding them sets the stage for the second part of the lecture. In Deep Learning, we are given a training set, and we minimize a loss function that, at least at face value, knows nothing about "future data". Just like the activations of a network in response to a test datum can be understood as a representation of future data, the parameters (weights) of a network can be understood as a representation of the past training set. What properties should the weights exhibit that can be optimized during training, which ensure that desirable properties of the activations emerge? Is there something special in deep neural networks that addresses this issue of generalization? Do these properties translate to a variational principle? Does this principle have anything to do with optimality of representations? Can they be imbued into the optimization we use to train deep networks?

In this lecture we will derive a theory of representation that is the first to address these questions for deep learning. The question the theory answers is: "What are the functions of given (past) data one can compute, so that the resulting representation of future data is best for the task at hand?" What it does not address is what happens when the task is not completely specified beforehand. Furthermore, we will dive deep into how such representations can be computed in practice, and what to do when the task is not specified at the outset.

Prof. Stefano Soatto**,** Professor, UCLA, Director of Applied Science, Amazon AI, is a professor of computer science with notable works in Computer Vision and Nonlinear Estimation and Control Theory. vision, sound, touch) to interact with humans and the environment.

MACHINE LEARNING AND VISION FOR INDUSTRIAL APPLICATIONS

Xin Yao: Ensemble Approaches to Class Imbalance Learning

Many real world classification problems have highly imbalanced and skew data distributions. In fault diagnosis and condition monitoring for example, there are ample data for the normal class, yet data for faults are always very limited and costly to obtain. It is often a challenge to increase the performance of a classifier on minority classes without sacrificing the performance on majority classes. This talk discusses some of the techniques and algorithms that have been developed for class imbalance learning, especially through ensemble learning.

First, the motivations behind ensemble learning are introduced and the importance of diversity highlighted.

Second, some of the challenges of multi-class imbalance learning and potential solutions are presented. What might have worked well for the binary case do not work for multiple classes anymore, especially when the number of classes increases.

Third, online class imbalance learning will be discussed, which can be seen as a combination of online learning and class imbalance learning. Online class imbalance learning poses new research challenges that still have not been well understood., let alone solved, especially for imbalanced data streams with concept drift.

Fourth, the natural fit of multi-objective learning to class imbalance learning is pointed out. The relationship between multi-objective learning and ensemble learning will be discussed. Finally, future research directions will be given.

Xin Yao**,** Chair Professor, Southern University of Science and Technology, Shenzhen, China. He is also a part-time Professor of Computer Science at the University of Birmingham, UK. His major research interests include evolutionary computation, ensemble learning and search-based software engineering. His work won the 2001 IEEE Donald G. Fink Prize Paper Award; 2010, 2015 and 2017 IEEE Transactions on Evolutionary Computation Outstanding Paper Awards; 2010 BT Gordon Radley Award for Best Author of Innovation (Finalist); 2011 IEEE Transactions on Neural Networks Outstanding Paper Award; and many other best paper awards. He received a prestigious Royal Society Wolfson Research Merit Award in 2012 and the IEEE CIS Evolutionary Computation Pioneer Award in 2013. He was recently selected to receive the 2020 IEEE Frank Rosenblatt Award.

MACHINE LEARNING AND VISION FOR INDUSTRIAL APPLICATIONS

MALVIC

PRACTICAL TALKS

Mathias Grundmann: Live Perception for Mobile & Web

In this talk, I will present Machine Learning (ML) solutions for Live Perception developed by Google Research for mobile and web. Live Perception or Viewfinder ML comes with major technical challenges to enable ML on-device, in real-time and with low-latency. Once solved it enables applications like virtual beauty try-on in YouTube, AR Effects in Duo, gesture controls of devices and view-finder tracking for Google Lens and Translate. In this talk, we will cover the core-recipes behind Google's Live Perception solutions, from model design to enabling ML infrastructure like MediaPipe and TFLite GPU acceleration. In particular we will be covering Face Meshes and iris tracking, Hand tracking and gesture control, body tracking for fitness applications and 3D object detection. The covered solutions are also available to the research and developer community via MediaPipe, —an open source, cross platform framework for building perception pipelines for mobile, web, desktop and python.

Matthias Grundmann is a Research Director in Google Research working in the area of Computer Vision, Machine Learning and Computational Video. He is leading a vertical team of ~40 Research and Software Engineers with focus on Machine Learning solutions for Live Perception (low-latency, on-device and real-time). His team develops high-quality, cross-platform ML solutions (MediaPipe) driven by GPU/CPU accelerated ML inference (TFLite GPU and XNNPack) for mobile and web. Among the rich portfolio of technologies his team develops are solutions for hand and body tracking, high-fidelity facial geometry and iris estimation, video segmentation, 2D object and calibration-free 6 DOF camera tracking, 3D object detection, Motion Photos and Live Photo stabilization.

Matthias received his Ph.D. from the Georgia Institute of Technology in 2013 for his work on Computational Video with focus on Video Stabilization and Rolling Shutter removal for YouTube. His work on Rolling Shutter removal won the best paper award at ICCP, 2012. He was recipient of the 2011 Ph.D. Google Fellowship in Computer Vision.

Stefano Soatto: Learning Representations

Representations are functions of the data that are "useful" for a task. Of all functions, one wishes to design or learn those that contain all the "information" in the data, and none of the variability that is irrelevant to the task. Depending on how one defines and measures "useful" and "information", different notions of representations can be instantiated. What are the relationships among those? Are there common principles behind the different tools and models? Is there a common notion of "optimality" that emerges from all formalisms? If so, are such optimal representations computable? If not, can they be approximated? If such representations are learned using "past data" (training set), can we predict how well they will perform on "future data" (test set)?

These questions have nothing to do with Deep Learning, but understanding them sets the stage for the second part of the lecture. In Deep Learning, we are given a training set, and we minimize a loss function that, at least at face value, knows nothing about "future data". Just like the activations of a network in response to a test datum can be understood as a representation of future data, the parameters (weights) of a network can be understood as a representation of the past training set. What properties should the weights exhibit that can be optimized during training, which ensure that desirable properties of the activations emerge? Is there something special in deep neural networks that addresses this issue of generalization? Do these properties translate to a variational principle? Does this principle have anything to do with optimality of representations? Can they be imbued into the optimization we use to train deep networks?

In this lecture we will derive a theory of representation that is the first to address these questions for deep learning. The question the theory answers is: "What are the functions of given (past) data one can compute, so that the resulting representation of future data is best for the task at hand?" What it does not address is what happens when the task is not completely specified beforehand. Furthermore, we will dive deep into how such representations can be computed in practice, and what to do when the task is not specified at the outset.

MACHINE LEARNING AND VISION FOR INDUSTRIAL APPLICATIONS

Prof. Stefano Soatto, Professor, UCLA, Director of Applied Science, Amazon AI. He is a professor of computer science with notable works in Computer Vision and Nonlinear Estimation and Control Theory (vision, sound, touch) to interact with humans and the environment.

Fridtjof Stein: Looking Far Ahead … Perception Challenges in the Field of Autonomous Trucking

Trucks are special in several aspects. Therefore it is only partially possible to transfer an existing sensor set of a robo taxi to an autonomous truck. In this talk I will focus on the specific sensor challenges in the context of the different modalities. I will elaborate on hardware- and software topics.

Dr. Fridtjof Stein is a senior scientist at Daimler truck within the field of perception. He works for about three decades at Daimler in the field of autonomous driving in public traffic including real-time vision especially in the fields of stereo vision, optical flow, object detection, and ground modeling in the automotive domain.

Gal Chechik: Practical talk

Gal Chechik, Nvidia.

MACHINE LEARNING AND VISION FOR INDUSTRIAL APPLICATIONS

BUSINESS TALKS

Success Stories – How did AI Shape Your Busines

Introduction by chair Anne Grete Ellingsen, project manager national European Digital innovation Hub candidate. Short presentation of the EUs program on investment in digital infrastructure and the benefit for SMEs and start ups

Crayon presented by Geir Gulliksen

Intelecy presented by Espen Davidsen

Aquabyte presented by Trude Jansen Hagland

Rocket farm presented by Dan Peter Rye Moen

Idean presented by Lars Petter Aase