2021 9th European Workshop on Visual Information Processing (EUVIP) | 978-1-6654-3230-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/EUVIP50544.2021.9484049

# REPORT ON THE EUVIP 2021 PANEL SESSION

# VISUAL INFORMATION PROCESSING AND MACHINE LEARNING AT THE CROSSROADS

***Session chairs:*** *M. Deriche[1], M. Mitrea[2]*

***EUVIP2021 general chairs:*** *A. Beghdadi[3], F. Alaya Cheikh[4]*

[1] KFUPM, Dhahran, KSA ; [2] IMT-TSP, France
[3] Université Sorbonne Paris Nord, France; [4] Norwegian University of Science and Technology, Norway

## CONTEXT

Over the last few years, we have witnessed a significant interest in developing powerful tools able to answer the ever-growing demands from traditional consumers to advanced technological companies, spanning applications from multimedia to medicine, and from geosciences to space exploration, to mention a few. In all these applications, the human perception of the environment has been a key factor. Within this panel session hosted at EUVIP 2021, we have discussed the synergies between recent machine learning advances and visual perception required to help the development of powerful visual information processing techniques. This session was designed around some key questions that were addressed by three panelists of different and complementary scientific and technical culture and experience. The main objective was to share the experience of these experts, coming from different backgrounds, with the audience on topical issues that concern the evolution of the themes of the EUVIP workshop directly related to the analysis, processing and coding of visual information and applications and how to associate and integrate new approaches and technologies coming from machine learning. At the end of this session, the aim was also to highlight a number of converging points of view, which would give a more or less realistic idea of the promising avenues for future research.

## PANELISTS

**Fernando Pereira**
IEEE Fellow, EURASIP Fellow, IET Fellow
Professor at Instituto de Telecomunicações, Lisbon, Portugal;
Expertise: Visual Information Coding and Representation

**Louis Chevallier**
Principal Scientist at Interdigital, R&I ISL, Rennes, France
Expertise: Computer Vision and Machine Learning

**Stefan Winkler**
IEEE Fellow
Deputy Director at AI Singapore and Associate Professor at the National University of Singapore;
Expertise: Vision modeling and perceptual approaches for video processing and analysis

## 1. PERCEPTION BASED MACHINE LEARNING

Human vision imposed itself as a very powerful information processing system that makes our interaction with the world around us meaningful. Yet, despite the extensive research efforts in trying to understand how such a system works, the underlying fundamentals and operational principles of visual information processing for humans and their relations to how the human brain operates are still far from being fully explored. On the one hand, we are still unable to precisely locate where and how, along the vision channel between eyes and cortex, a real-life scene is progressively interpreted into distinct objects and regions. On the other hand, human visual perception can still fail to accurately interpret dynamic scenes and images hus leading to a large variety of misinterpretations, like the visual illusions, for instance. In this world, we want to develop human made systems (e.g. machine learning (ML) or artificial intelligence systems) that mimic how the visual information is extracted and interrelated by humans, but at the same time avoid some of the human visual system misinterpretations.

A number of major directions can be followed to understand such a convoluted world and this session is an attempt to do that based on the panelists expertise stemming from different backgrounds (both academia and industrial) and from different image processing research fields.

The exchanges concerned the evolution of the themes of the EUVIP workshop directly related to analysis, processing and coding of visual information and its applications and how to associate and integrate new approaches and technologies coming from machine learning. At the end of this session, the aim was also to highlight a number of converging points of view, which would give a more or less realistic idea of the promising avenues for future research.

During the EUVIP 2021 Discussion Panel Session, three major issues were elaborated, these are presented in the subsequent sub-sections.

### 1.1. Machine Learning for a Better Future

Over the last few years, ML has proven itself to be a powerful methodological framework, able not only to process both data and models, but also to relate them to knowledge. Thus, reliable solutions are devised for major human challenges, like autonomous navigation technology or computer-based diagnosis and intervention in medicine, to mention but two. We should be open to embrace such a revolution that brings together experts from diverse technology areas including signal processing, machine learning, computer vision, etc.

ML based technologies are used to solve difficult problems to ultimately achieve a better life for humans. It is important to understand that ML technology goes beyond improving accuracy of systems. A good example is that of face recognition which traditionally belongs to the classification techniques, but which is also expected to properly meet requirements related to gender equality, privacy issues, or ethics. Such important issues relate to what is now called explainable ML. How much do we know about how ML systems work? Do we trust the training process and the models themselves? Can we explain all of the facets of ML technology?

### 1.2. Deployment of Machine Learning in Industry

It is important to note that while we can find countless ML based algorithms developed for any typical applications, many of such algorithms require heavy computations. The acceptance of ML technology in the real world depends upon making such algorithms explainable and practically implementable over a large variety of platforms, including the resource restricted ones, e.g. mobile platforms or edge. As an example, three applications that attracted a lot of attention among scientists in the multimedia industry are: high resolution media, face tracking in 3D, and indexing. While upscaling is a traditional image and video processing application, ML and in particular Deep Learning (DL) based approaches have been shown to be able to combine the traditional pixel-based scaling with semantics. The fusion of content semantics with model-based algorithms can deliver an enhanced quality of upscaled media. Such enhanced quality is directly related to perceived quality as appreciated by humans. Tracking of faces over 3D videos has also been an excellent application in which DL and transfer learning can be used to develop powerful self-supervised ML-based techniques that can accurately track faces across video data. Finally, a third application in industry that has attracted a lot of attention is that of multimodal indexing. An example of such an application can be that of searching for images using text and image description over a mobile phone. Among the major challenges that need to be considered is the complexity of algorithms so that battery power is not depleted quickly. In summary, it is important to trust that ML and DL did come to our rescue in addressing many practical problems in the multimedia industry, but the computational complexity is still a major challenge.

### 1.3. Visual Information Processing: Humans *vs.* Machines

Visual information processing offers an ideal environment to bring together the perception of the real world by humans and machines. There are some similarities between these two perceptions including the concept of feature extraction, the concept of sparsity, the concept of neural processing, but there are also a number of differences, related to hardware performance and complexity, the concept of generalization, adversarial examples, etc.

Another important issue that needs more attention relates to the potential synergies between mathematical models and physical representation of the real world in order to guide the processes of training, learning and testing. In this respect, several concepts did not reach their maturity among the ML community, as for example: symbolic representation, logic, semantics, reasoning, etc. A final issue is the concept of excellence and/or acceptability. Is the concept of good quality or good performance and good model the same across humans and machines? Do the metrics used for measuring performance in machines truly reflect what we as humans perceive as quality? Many of these issues have been discussed extensively in the literature but we are still a long way from reaching that level of maturity in ML systems.

## 2. HUMAN PERCEPTION VS. MACHINE PERCEPTION

Developed around five main questions, the exchanges among the panelists and audience are integrated, summarized and organized as follows.

*Q1. Are there any common fundamentals between how humans understand and analyze visual information and how machines process visual scenes?*

ML models mimic the human vision mechanisms and, according to each type of solution, there is an explicit and/or an implicit trend for including in the ML models layers corresponding to the HVS. Yet, there is a fundamental difference: human brain learns in a different way; based on a series of predictions each “error” becomes a new prediction, letting us assess just the difference between our prediction and the particular case we encountered. A second difference is that machines still cannot learn from what other machines learned. This does not relate to federative/transfer learning but rather to the lack of explanation and of semantic generalization. From the understanding point of view, ML still misses fundamental cultural aspects like ethics, for instance.

*Q2. How can we preserve deductive and deep reasoning in the field of visual information processing and analysis?*

In this respect, the main challenge would be to pass from data-based learning to model-assisted learning. While this would bring machine and human learning closer to each other, it would also have a large variety of practical advantages: explainability, lower computing/storage resources, and better performance.

*Q3 What can we hope to preserve and evolve from research based on physical observations and mathematical modeling in the field of visual information processing?*

Not only that ML will be challenged to evolve from data towards models, but these models are expected to be multimodal: physical reasoning should be crossed with psychological, causal, spatio-temporal, etc.

*Q4. How can the field of visual information processing benefit from the advances made in computing power and in machine learning?*

At least at a first glance, ML and computing power go together hand in hand: intuitively, the larger the computing power and the database are, the more accurate the results will be. Yet, this phenomenon should be carefully controlled from the ethical point of view: otherwise, there is a huge risk related to the aggregation of the ML-based decisions on the largest industrial players.

*Q5. What can we advise young researchers who are starting their doctoral training or early career in this world invaded by the machine learning wave?*

Current day DL frameworks allow a new PhD student to set-up a preliminary testbed virtually instantaneous (in a 1 to 3 days lap). Yet, this does not solve in any way the R&D work he/she is expected to do. The mathematical models beneath still require understanding, analysis and evolution, the experimental framework still requires proper design, the experimental data still need to sanitized and to be checked for their relevance, the experiments still need to be properly conducted and the conclusions to be drawn.

## 3. CONCLUSION

In summary, as expected the discussion confirmed that ML is really a powerful technology, but many theoretical and practical aspects are still not fully understood in order to make ML based technology more widely spread and accepted among the scientific community and related industry. As many of the previous scientific innovation waves that our community has witnessed, e.g., wavelet, fractals, compressive sensing and neural networks, during the last four decades, DL is a disruptive approach but still considered as another wave that visual information processing experts should master and learn to adapt and use effectively in their research field. To conclude, this relatively short debate has the merit of having shown the need to further develop the points raised in a more in-depth reflection targeted at some of the panelists' fields of expertise.