Damian Campo, David Martin, and Lucio Marcenaro

Unsupervised Anomaly Detection Using Intelligent and Heterogeneous Autonomous Systems

Highlights from the 2020 IEEE Signal Processing Cup student competition

The IEEE Signal Processing Cup (SP Cup) is a challenge that is organized yearly by the IEEE Signal Processing Society (SPS) Student Services Committee at ICASSP. Similar to the Video and Image Processing Cup (organized at ICIP), the researchers in the signal processing field propose different topics for the challenge. The best one is selected based on votes from the chairs of the technical committees of the IEEE SPS. The criteria used for the selection are the overall proposal quality, the topic's originality, the feasibility of the challenge, and the appropriateness for ICASSP. Additionally, the experience of the proposing team and the interest of the different technical committees are essential factors in the selection.

The topic for the challenge is typically selected at least six months before the actual conference. In this time frame, data sets for the challenge are released by the organizing team, and the real challenge begins with different selection phases that allows only the three best teams to participate in the conference's final challenge.

The challenge is organized with the undergraduates in mind. The primary objective of the SP Cup is to introduce bachelor's or master's students to the signal processing field of research, involving them in the design and implementation of real working solutions for major open problems in the community. Each team can be composed of a minimum of three to a maximum of 10 students. They can be supported by more senior researchers (one faculty member per team is mandatory, and a graduate student is permitted) to guide them in implementing the solution for the challenge.

Every year, several hundred undergraduate students participate in the first phase of the challenge and submit their results to access the final event at the conference. Evaluating all the received submissions and supporting all the teams during the different phases of the challenge are complex and time-consuming activities carried out by the organizing team. However, thanks to continuous efforts, it is possible to introduce new students to the research field. The SP Cup shows students how real-world problems can be tackled by applying computational algorithms and techniques that are often studied theoretically (not practically) in their academic courses. The idea of "learning by doing" is an essential concept behind the SP Cup initiative: the theoretical study that has practical feedback allows us to understand better and memorize faster. Moreover, through the SP Cup, undergraduates can practically exploit the benefits of teamwork and face the difficulties and constraints of interacting with other students to solve a real-world signal processing task.

The teams are awarded prizes of US$5,000; US$2,500; and US$1,000 for first, second, and third place, respectively. All of the SP Cup prizes are funded by MathWorks, Inc., which also offers complimentary MathWorks Student Competitions Software for use in the competition. The three selected teams are also supported by the SPS with travel grants to attend the final competition at ICASSP or ICIP. Because all ICASSP events were organized virtually due to travel restrictions caused by the COVID-19 pandemic, this version of the challenge (SP Cup 2020) was carried out via a video meeting.

The topic selected for the seventh edition of the SP Cup at ICASSP 2020 was "Unsupervised Abnormality Detection by Using Intelligent and Heterogeneous Autonomous Systems." The organizing team was composed of researchers from the Carlos III University of Madrid, Spain, and the University of Genoa, Italy. The topic of this challenge is an active research field that involves both the SPS and Intelligent Transportation Systems (ITS) Societies and aims at exploring the importance of systems that recognize anomalies incrementally in a safe, robust, and unsupervised manner. The main idea consists of contributing to advances toward unresolved ground-breaking problems. The following example can explain the foundations of this idea: if a child observes a bike only once, then the child can recognize all the bikes in the world.

Digital Object Identifier 10.1109/MSP.2020.3002482
Date of current version: 2 September 2020

 1053-5888/20©2020IEEE

New perspectives have been explored by adopting advanced frameworks that embed intelligent systems into ground or aerial vehicles. For example, in [1], the authors use reinforcement learning to adapt a quadrotor to the task at hand, improving its performance while requiring little domain knowledge to be explicitly encoded there. In recent studies [2], it has been demonstrated that intelligent systems must learn to solve different tasks in diverse environments, focusing attention on the most relevant task and environmental features for solving a problem. An intelligent system should retain the knowledge learned in previous tasks and use it when required. The nature of knowledge should be explainable and transferable to other agents. In [3], the authors point out that there are still many open problems and assumptions to address, where a transfer learning framework would serve as an intelligent base platform for monitoring, problem solving, and general decision support in heterogeneous dynamic environments, leading to a new generation of human-interactive robots.

ITS are able to react, make decisions, and change their behavior in response to dynamic environment perception during operation [4]–[7]. Accordingly, autonomous systems with self-awareness capabilities identify, differentiate, and classify states and progressively determine actions. The work presented in [8] discusses an algorithm that allows an autonomous system to self-learn, meaning that a robot can determine the optimal sequences of actions needed to avoid obstacles while wandering in a dynamically changing environment without prior knowledge about the anticipated obstacles or states. A recent self-learning model for robots is also proposed in [9], where the learning mechanism adjusts the sensorimotor mapping at every learning step, helping the robot to choose motions. Such a robot is endowed with self-learning and self-organizing abilities that allow it to learn different skills (such as keeping itself balanced) in an unsupervised way by interacting with the environment.

SP Cup 2020 focuses on the automatic detection of anomalies in an unsupervised way based on multisensory data coming from an agent observing its surroundings. In this edition, the teams have the task of designing algorithms that must be capable of determining the status behavior of an autonomous ground or aerial vehicle that observes and interacts with the environment.

The challenge has motivated all the participants to create innovative contributions to the field of autonomous systems. Their proposed algorithms use normal, known data to infer anomalies on unlabeled new data automatically. One initial step toward decision making in autonomous systems is the understanding of the data in terms of normal or abnormal information in a time series of multisensory data. The detection of anomalies is a topic that comprises several different fields, such as signal processing, intelligent systems, machine learning, and data fusion from smart sensors, and it can be applied to diverse platforms. Particularly, SP Cup 2020 considered autonomous ground and aerial vehicles as application cases. The sensory data provided in this challenge consist of the video and inertial measurement unit (IMU) data from a drone that performs a series of tasks.

The members of the organizing committee, who belong to the IEEE SPS, the IEEE ITS Society (ITSS), and the IEEE Autonomous System Initiative (ASI), have organized this seventh edition of the SP Cup: an exciting unsupervised abnormality detection challenge, where ground or aerial autonomous systems interplay with the environment to discover abnormalities automatically.

Tasks, resources, and evaluation criteria

At SP Cup 2020, experiments with a drone that observes its surroundings were provided to the different teams. This aerial vehicle uses a PixHawk 2 Cube flight controller. It is a powerful controller with IMU sensors and also has internal software that uses the MAVLink communications protocol, which is compatible with the Robot Operating System (ROS) framework. The aerial platform has a digital camera from the manufacturer Basler, model number acA2040-35gc. The color camera uses a Sony sensor IMX265 model based on CMOS technology. The advantage of this camera is that it can be powered by power over Ethernet, i.e., the camera uses the same port for communication and power (Figure 1).

This competition's goal has been to detect anomalies in the aerial system's behavior based on embedded sensory data in real time. The competition has consisted of the following stages: phase one, an open competition in which all the teams can participate, and phase two (the final competition), where the three teams with the highest performance in the open challenge were selected as finalists and were invited to participate in the final event.

Phase one of the open competition was designed to give teams the data sets needed to familiarize themselves with the proposed challenge. Accordingly, the provided data sets were divided into two groups: experiments with only normal data and experiments with mixed normal-abnormal data. The data sets were ROS based, with IMU and video camera-synchronized data.

The students' main task consisted of taking the data (the IMU and video camera information) from the experiments containing only normal data and create/train models to differentiate between normal and abnormal data in the experiments that presented mixed information. The proposed challenge falls in the category of semisupervised learning, in which training data

FIGURE 1. The captive drone used at SP Cup 2020 to capture all of the data sets, which were composed of the video and IMU data.

contains only normal instances without any anomalies, and the testing data have mixed information. Figure 2 shows both the video camera data and IMU information coming from the drone while it experiences normal and mixed normal/abnormal situations.

To solve the challenge, the students took the provided data sets and proposed methods that facilitate the real-time detection of anomalies based on multisensory data. Participants could use different methods and algorithms from the state of the art to reach the goal of the challenge.

The evaluation criteria of phase one consists of a series of points that measure the generation of innovative and functional solutions for identifying normal and abnormal behaviors from multisensory information. Accordingly, the following parameters were taken into consideration by a research judging panel for evaluating the proposals from the participants (maximum score: 10 points):

- the quality of their project and the effort given during the open competition to reach advances in this research field (0–2 points)
- the novelty and the research advances of the created algorithms and models for unsupervised abnormality detection in the provided data sets. A detailed report must be provided (0–2 points).
- the results of the performance using ROS-based data sets with IMU and camera-synchronized data. An executable task with a specific user interface or a MATLAB implementation must be provided to demonstrate the provided results (0–2 points).
- The authors must provide a full conference paper using the IEEE format (six pages). 4.a: clarity of the presentation (0–1 point); 4.b: the theoretical contribution (0–1 point); 4.c: the technological contribution (0–1 point); 4.d: importance of the results (0–1 point).

The research judging panel for evaluating phase one included two reviewers who assessed each team's performance based on the aforementioned point scale. From such an evaluation, the three teams with the highest scores were invited to participate in the final competition. Figure 3 depicts the results of the 10 best teams during phase one. As shown, the maximum possible score is 20 points (10 points per reviewer).

Technical highlights

In this section, we provide an overview of the most innovative ideas presented by the different teams throughout the competition. Because the challenge included a drone's multisensory information,

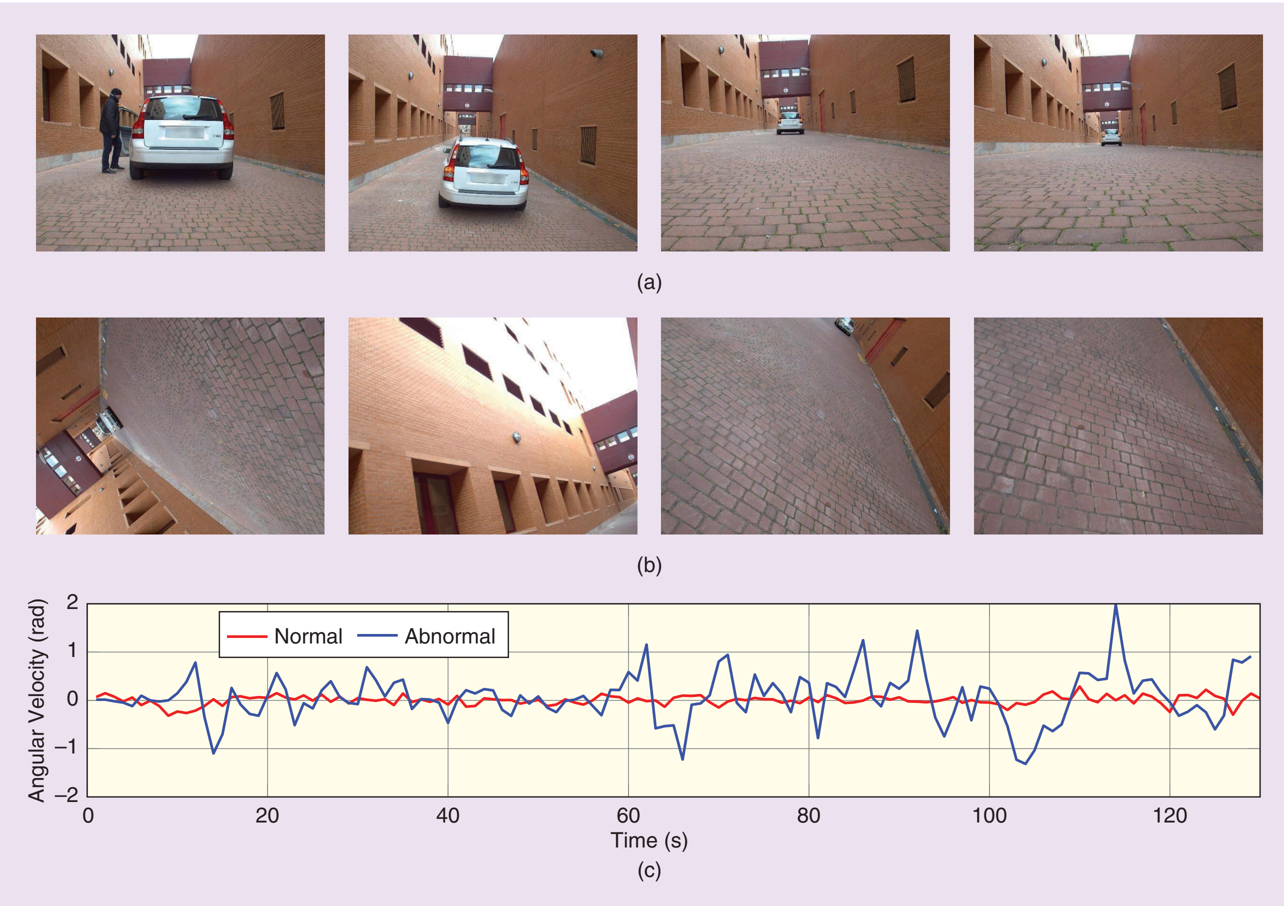

FIGURE 2. The IMU signals and images recorded by the drone while observing a vehicle under normal and abnormal conditions. (a) The images captured during a normal drone's flight, (b) the images captured during an abnormal drone's flight, and (c) the angular velocity signals (based on the IMU information) for normal (red) and abnormal (blue) drone flights.

the teams needed to deal with two different sample rates coming from the IMU and the video data. For solving this problem, some of the teams decided to split the anomaly detection into two parts, one dedicated to the IMU information and the other to the images. Other groups employed some preprocessing steps for combining the effects of multisensory data in different ways, such as interpolating feature spaces [10] or associating multiple samples into a single observation.

For dealing with the images, some groups first considered preprocessing techniques such as demosaicing [11] to improve the quality of images and the calculation of optical flows [12] to enrich their predictive models. Deep learning algorithms were employed to handle the video data in both representation (feature extraction) and forecasting (prediction of next frames) tasks. Accordingly, the use of convolutional neural networks [13] and recurrent neural networks such as long short-term memory (LSTM) were popular techniques used in the competition. In particular, works based on existing network architectures such as ResNet [14] and SegNet [15] have offered distinguishable performances. Additionally, the proposed methods based on state-of-the-art strategies, such as those in [16]–[18], have demonstrated usefulness for detecting anomalies.

Different approaches were taken for measuring anomalies in an online fashion, i.e., as the observations from the IMU and the camera are acquired. The reconstruction error (loss) from the LSTM autoencoders was a popular choice for measuring anomalies. Other works have used Euclidean and probabilistic distances as metrics to identify abnormal data. It is worth mentioning that an innovative approach that defined anomalies at different levels of abstraction (as well as the winner's method), which employed a kernel function for embedding the data in a reduced dimension, generated a set of kernel matrices [19] living in a non-Euclidean space. Then, they used a Riemannian distance between the transformed data for detecting anomalies.

Final competition

The three highest-scoring teams from the open competition stage (phase one) were selected as finalists. They were invited to compete in the final stage of the challenge (phase two) at 2020 ICASSP held 4–8 May (virtually). In phase two, the finalists submitted a 10-min video in which they described the technical aspects and the advantages/limitations of their approaches. Their videos were played live at the event and were followed by a round of questions from an evaluating committee that included the technical organizers of the challenge together with Prof. Raed Shubair (Massachusetts Institute of Technology, Cambridge) and Prof. Martin Haardt (Ilmenau University of Technology, Germany).

The virtual modality of SP Cup 2020 provided the opportunity for a broad audience to attend and participate in the event by formulating questions for the finalists. Undoubtedly, the fact that the session at the conference was fully virtual also introduced several additional technical issues that had to be addressed in advance: the video contributions from organizers and team members were prerecorded and uploaded in advance. Nonetheless, there were some compatibility problems with the platform selected for content broadcasting. In addition, the question-and-answer sessions were not as interactive as they

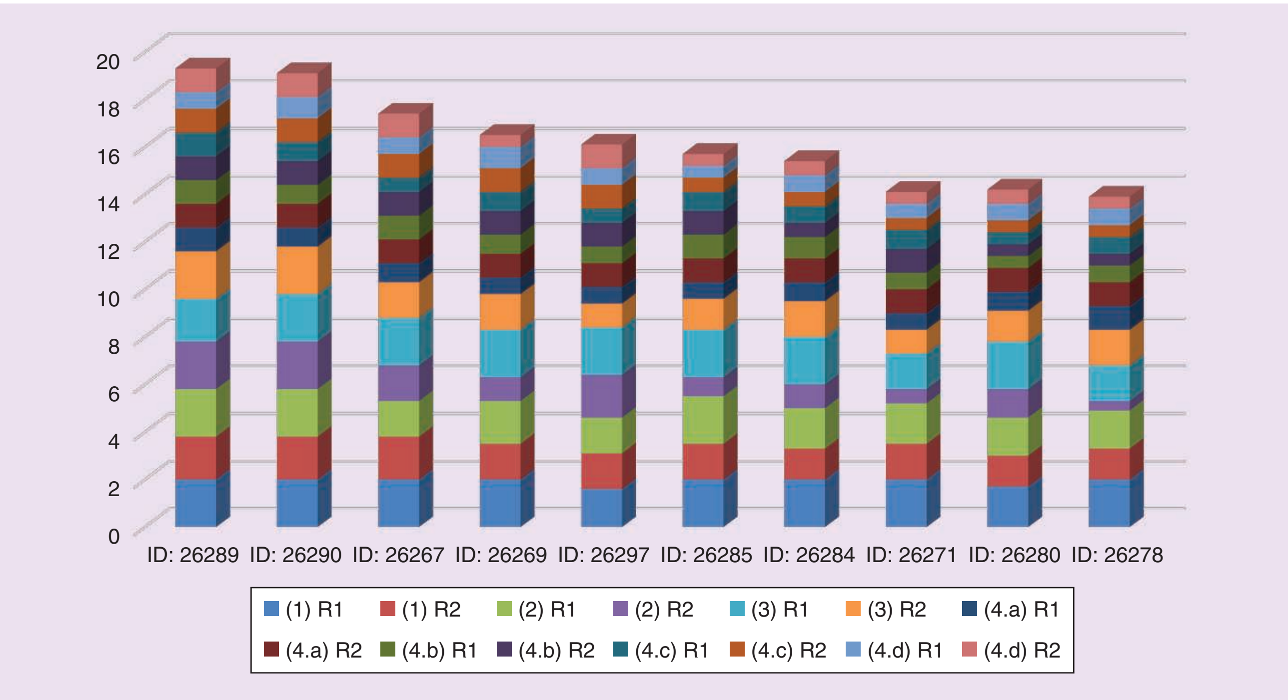

FIGURE 3. The results of the 10 best teams during phase one at SP Cup 2020.

would have been in person. In phase two, the three final teams were able to show the capabilities of the developed algorithms using their models to confront the proposed real problem in autonomous systems. The final evaluation criteria (10 points maximum per final committee member) were as follows:

- how the team has faced the problem of unsupervised abnormality detection (0–2 points)
- the clarity and quality of the information provided in the video (0–2 points)
- the originality of the method and the importance of the obtained results from a technological and scientific viewpoint (taking into consideration the provided article and the live video) (0–2 points)
- the capability of understanding the limitations and assumptions of the proposed approach (from the questions to and answers from the judges and the audience) (0–2 points)
- understanding of possible future directions to improve their own methods and the identification of potential applications (from the questions to and answers from the judges and the audience) (0–2 points).

The finalists

The results of the three groups' virtual presentation led to the following final scores.

SIPL, Technion (Grand-prize winner)

- *Affiliation*: Israel Institute of Technology
- *Supervisor*: Yair Moshe
- *Tutor*: Pavel Lifshits
- *Students*: David Ben-Said, Samuel Sendrowicz, and Theo Adrai.

Icarus Inhibition (First runner-up)

- *Affiliation*: University of Moratuwa, Sri Lanka
- *Supervisor*: Chamira Edussooriya
- *Tutor*: Dumindu Tissera
- *Students*: Suman Navaratnarajah, Yasintha Supun Madhushanike Wackwella Gamage, Thiru Thillai Nadarasar Bahavan, Vinu Vihan Maddumage, Gershom Seneviratne, Damitha Senevirathne, Isuru Wijesiri, Sulhi Cader, and Suchitha Dehigaspitiya.

BUET Andromeda (Second runner-up)

- *Affiliation*: Bangladesh University of Engineering and Technology
- *Supervisor*: Mohammad Ariful Haque
- *Students*: Himaddri Roy, Shafin Bin Hamid, Munshi Sanowar Raihan, Prasun Datta, Ashiqur Rasul, Md. Mushfiqur Rahman, and K.M. Naimul Hassan.

Strengths and weaknesses of the finalists

The three finalists offered interesting high-level solutions to the anomaly-detection problem based on multisensory information. This section discusses the main strengths and weaknesses of each finalist team.

BUET Andromeda (Second runner up)

This team proposed two novel methods for the detection of anomalies, one that dealt with the IMU information and the other with the video data. Their approach was based on an LTSM autoencoder for the sensor data and a convolutional autoencoder based on optical flow information for analyzing the video data. They proposed a parametric anomaly score that took values from 0 (normal data) to 1 (abnormal data). They performed an exciting analysis of their results in a tangible/explicative way, where the obtained anomalies were directly associated with actual drones' behavioral events. One weakness of their approach consisted of the use of an optical flow algorithm that slowed down their method.

Icarus Inhibition (First runner up)

This team proposed a method that analyzed the data from the IMU and the images independently. Their method was based on an LSTM algorithm that used previous information to understand the normal distribution of data. Then, when trying to reconstruct future information, if such data falls inside the 95% range of the confidence of learned distributions, it is classified as normal. Their approach was clearly explained and entirely implemented on the ROS, which was a great step toward a technological/practical on-device implementation. One flaw of their method was the lack of a global continuous-anomaly metric that could have taken into consideration the IMU and video data at the same time.

SIPL (Grand-prize winner)

This team proposed a method that featured strong mathematical foundations where the features coming from the IMU and video data are fused by using a kernel function that transforms them into a Riemannian manifold. The anomalies were then calculated based on the Riemannian distance between the projected points onto the manifold. This group proved the robustness of their solution by satisfactorily testing their algorithm on another data set [20] that considered a multimodal data set (including the measurements from a Kinect camera and a wearable inertial sensor) composed of different human activities. One weakness of their proposed approach was the imbalance of features that came from different data (the video data tend to be more important due to their large number of features).

Acknowledgments

We would like to express our gratitude to 2020 ICASSP for its virtual hosting of the 2020 SP Cup Competition. We also thank the IEEE SPS, the IEEE ITSS, and the IEEE ASI for their endorsements as well as the Intelligent Systems Lab of the University Carlos III of Madrid for providing the data sets.

Authors

Damian Campo (damian.campo@edu.unige.it) received his Ph.D. degree in cognitive environments from the University of Genoa, Italy, in 2018, where he has been a postdoctoral researcher in the Department of Electrical, Electronics, and Telecommunication Engineering and Naval Architecture since that time. He is a member of the IEEE Signal Processing Society.

David Martin (dmgomez@ing.uc3m.es) is a professor at the University Carlos III of Madrid, where he is a member of the Intelligent Systems Lab. He is a member of the IEEE Intelligent Transportation Systems Society (ITSS), the IEEE Signal Processing Society (SPS), and the IEEE SPS-Autonomous Systems Initiative. In 2015, the IEEE Society awarded him as the best reviewer of the 18th IEEE International Conference on ITS.

Lucio Marcenaro (lucio.marcenaro@unige.it) is an assistant professor in the Department of Electrical, Electronics, and Telecommunication Engineering and Naval Architecture, the University of Genoa, Italy. He chairs the Student Services Committee of the IEEE Signal Processing Society, supporting the Video and Image Processing Cup and the IEEE Signal Processing Cup. He is a Member of the IEEE.

References

[1] F. M. R. Silva and M. Veloso, "Towards table tennis with a quadrotor autonomous learning robot and onboard vision," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2015, pp. 649–655. doi: 10.1109/IROS.2015.7353441.

[2] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre et al., "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, 2018. doi: 10.1126/science.aar6404.

[3] T. T. Nguyen, T. Silander, Z. Li, and T.-Y. Leong, "Scalable transfer learning in heterogeneous, dynamic environments," *Artif. Intell.*, vol. 247, pp. 70–94, June 2017. doi: 10.1016/j.artint.2015.09.013.

[4] M. Buehler, K. Iagnemma, and S. Singh, *The DARPA Urban Challenge. Autonomous Vehicles in City Traffic*. Berlin: Springer-Verlag, 2009.

[5] T. Nothdurft, P. Hecker, S. Ohl, F. Saust, M. Maurer, A. Reschka, and J. R. Böhmer, "Stadtpilot: First fully autonomous test drives in urban traffic," in *Proc. 14th Int. IEEE Conf. Intelligent Transportation Systems (ITSC)*, 2011, pp. 919–924. doi: 10.1109/ITSC.2011.6082883.

[6] A. Broggi, P. Cerri, S. Debattisti, M. C. Laghi, P. Medici, M. Panciroli, and A. Prioletti, "PROUD-Public road urban driverless test: Architecture and results," in *Proc. IEEE Intelligent Vehicles Symp.*, 2014, pp. 648–654. doi: 10.1109/IVS.2014.6856478.

[7] M. M. T. S. Sivaraman, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1773–1795, 2013. doi: 10.1109/TITS.2013.2266661.

[8] Y. Yusof, H. A. H. Mansor, and A. Ahmad, "Formulation of a lightweight hybrid ai algorithm towards self-learning autonomous systems," in *Proc. IEEE Conf. Systems, Process and Control (ICSPC)*, 2016, pp. 142–147. doi: 10.1109/SPC.2016.7920719.

[9] X. Zhang, X. Ruan, Y. Xiao, and J. Huang, "Sensorimotor self-learning model based on operant conditioning for two-wheeled robot," *J. Shanghai Jiaotong University (Sci.)*, vol. 22, no. 2, pp. 148–155, 2017. doi: 10.1007/s12204-017-1814-8.

[10] R. Bonetto, M. Soldan, A. Lanaro, S. Milani, and M. Rossi, Seq2Seq RNN based gait anomaly detection from smartphone acquired multimodal motion data. 2019. [Online]. Available: arXiv:1911.08608

[11] H. S. Malvar, L.-w. He, and R. Cutler, "High-quality linear interpolation for demosaicing of Bayer-patterned color images," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, pp. iii–485, May 2004. doi: 10.1109/ICASSP.2004.1326587.

[12] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L 1 optical flow," in *Joint Pattern Recognition Symp.*, 2007, pp. 214–223. doi: 10.1007/978-3-540-74936-3_22.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Advances Neural Information Processing Systems*, 2012, pp. 1097–1105.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[15] V. Badrinarayanan, A. Handa, and R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. 2015. [Online]. Available: arXiv:1505.07293

[16] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," *Neurocomputing*, vol. 262, pp. 134–147, Nov. 2017. doi: 10.1016/j.neucom.2017.04.070.

[17] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed, "DeepAnT: A deep learning approach for unsupervised anomaly detection in time series," *IEEE Access*, vol. 7, pp. 1991–2005, Dec. 2018. doi: 10.1109/ACCESS.2018.2886457.

[18] G. Pang, C. Shen, and A. van den Hengel, "Deep anomaly detection with deviation networks," in *Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining*, 2019, pp. 353–362. doi: 10.1145/3292500.3330871.

[19] R. Bhatia, *Positive Definite Matrices*, vol. 24. Princeton, NJ: Princeton Univ. Press, 2009.

[20] C. Chen, R. Jafari, and N. Kehtarnavaz, "UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2015, pp. 168–172. doi: 10.1109/ICIP.2015.7350781.

Wei Liu, Mohammad Reza Anbiyaei, Xue Jiang,
Lei Zhang, and Lucio Marcenaro

Let There Be a Beam

Highlights from the 2020 IEEE 5-Minute Video Clip Contest

The annual IEEE 5-Minute Video Clip Contest (5-MICC) was launched by the IEEE Signal Processing Society (SPS), and beamforming, which has a wide range of applications in radar, sonar, microphone arrays, radio astronomy, seismology, medical diagnosis and treatment, and wireless communications, was chosen as this year's topic. After two stages of fierce competition, three finalist videos were selected by the organizing committee and placed online for public voting. The first one is about fast beam alignment in millimeter-wave (mm-wave) radios, the second

Digital Object Identifier 10.1109/MSP.2020.3002485
Date of current version: 2 September 2020