https://doi.org/10.2352/EI.2025.37.3.MOBMU-324

Intelligent Autonomous Vehicles (IAV) using Artificial Intelligence focusing on perception

Nithin Jayagovindan, SRH University of Applied Sciences, Ernst-Reuter-Platz 10, 10587 Berlin, Germany, nithinj@live.com

Alexander I. Iliev, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. Georgi Bonchev Str., Block 8, 1113 Sofia, Bulgaria; SRH University of Applied Sciences, Ernst-Reuter-Platz 10, 10587 Berlin, Germany, ailiev@berkeley.edu

Abstract

Artificial Intelligence (AI) contributes significantly to the development of autonomous vehicles in an unmatched way. This paper outlines techniques and algorithms for the implementation of Intelligent Autonomous vehicles (IAV) leveraging AI algorithms for traffic perception, decision-making and control in autonomous vehicles through merging traffic scenario detection, traffic lane detection, semantic segmentation, pedestrian detection, and traffic sign classification and detection. The modern computer vision and deep neural networks-based algorithms enable the real-time analysis of different vehicle data through artificial intelligence. The vehicle dynamics are constituted through AI in vehicle control systems for increased safety and efficiency to ensure that they are optimized with time. In addition, the paper will also discuss challenges and possible future directions, underscore how AI has the potential of driving autonomous vehicles towards safer and more reliable as well as intelligent transportation systems. This is the hope of the future whereby mobility is intelligent, sustainable, and accessible with the combination of AI with autonomous vehicles.

Keywords:

Intelligent Autonomous Vehicles, Efficient Neural Networks, YOLO, CNN, Kalman Filter, AI, Computer Vision

Introduction

Artificial Intelligence represents one of the most significant and transformative technological advancements of our era. Plausible intelligent machines that can think, learn, and converse with humans are no more only in science fiction. New smart self-driving cars can drive across a remote road at night. AI-powered robots can learn new motor skills by experimenting. Moreover, AI powered Intelligent Autonomous vehicles (IAV) can reduce the number of road accidents, increase transportation and taxi productivity as well as new mobility services.

Autonomous vehicles are driverless cars or self-driving cars that have no need for direct human control using human like perception. The vehicles are equipped with state-of-the art technologies and architectures aspects as sensors, cameras, radar, lidar, GPS and artificial intelligence algorithms to detect their

The research and testing of autonomous vehicles are still progressing despite the obstacles. Numerous businesses, such as major digital companies and conventional manufacturers, are actively developing self-driving car technology, and certain regions have already seen a limited rollout of autonomous vehicles for testing. Autonomous vehicles might be a big part of transportation in the future if technology advances and legal obstacles are removed. Numerous benefits that autonomous cars provide could transform the transportation sector and advance society at large.

With an emphasis on the latest architecture and technology, difficulties, and societal effects, this paper seeks to construct intelligent autonomous vehicles and investigate the mutually beneficial interaction between Artificial Intelligence and autonomous vehicles. Through a thorough analysis of the intricate workings of AI integration into autonomous vehicles, we hope to obtain a thorough grasp of the possible advantages and disadvantages of this revolutionary technology.

The integration and inclusion of AI in autonomous vehicles involves sophisticated algorithms, deep learning models, and computer vision systems, all working in tandem to emulate human-like decision-making capabilities and perception. This paper will scrutinize the core AI components, including computer vision, deep learning, and decision-making algorithms, which collectively enable autonomous vehicles to perceive their complex traffic surroundings, make informed decisions, and navigate complex traffic scenarios. Furthermore, the societal implications of self-driving cars or autonomous vehicles go beyond technology and engineering. The AI powered vehicles have the potential to reshape environments, transportation systems and even social norms. To summarize the combination of AI and vehicles represents a milestone, in transportation technology's evolution. As we embark on this transformative journey it is essential to examine how AI and self-driving cars interact. This research aims to contribute to the expanding knowledge, in this field by shedding light on advancements challenges faced and impacts associated with the rise of AI driven autonomous vehicles.

Time traffic detection for perception (IAV)

The application of YOLOv8 algorithm for real-time traffic detection and perception demonstrated impressive results. The algorithm consistently identified and localized various traffic objects, including vehicles, pedestrians, and cyclists, in each frame of the input dataset or video stream. Latest YOLO model exhibited high accuracy and precision in traffic objects detection. The bounding boxes generated around the objects closely matched their actual positions, showcasing the algorithm's capability to provide precise spatial localization.

First step is to divide the input image into a G×G grid. For each grid cell, run a CNN that predicts y of the following form:

$$y = [p_c, b_x, b_y, b_h, b_w, c1, c2..., c_p...]^T \in R^{G\times G\times k\times(5+p)} \quad (1)$$

where: Pc is the probability of detecting an object bx, by, bh, bw are the properties of the detected bouding box, c1, ..., cp is a one-hot representation of which of the p classes were detected, and k is the number of anchor boxes. Run the non-max suppression algorithm to remove any potential duplicate overlapping bounding box.Unlike traditional CNN, Yolo detects an object in a picture and also predicts probability of object and where it is located.

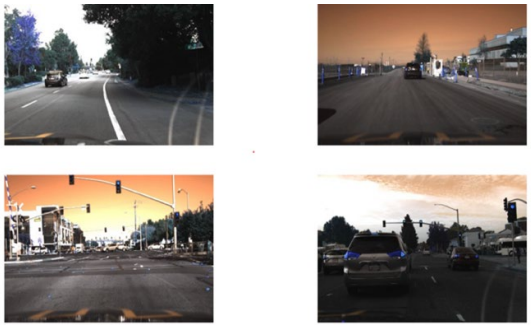


Figure 1. *Traffic detection random inputs.*

Figure 2. *Traffic detection using YOLO v8.*

As real-time processing is critical for traffic detection and perception, YOLOv8 might be a more suitable choice due to its single-pass architecture compared to other models like faster RCNN.

Semantic segmentation for perception (IAV)

In terms of real time processing capability, ENet's efficiency in terms of model size and architecture design facilitated real-time processing of high-resolution images. The algorithm maintained a high frames-per-second (FPS) rate, making it well-suited for applications requiring timely and responsive traffic perception. Also, in this research ENet demonstrated high accuracy in semantic segmentation, effectively distinguishing between various classes such as road, vehicles, pedestrians, and background. The segmentation maps generated by ENet reflected precise understanding of the scene, contributing to a detailed perception of the traffic environment. Below are the outputs of the semantic segmentation using CITYSCAPES open dataset.

Table 1. Architecture of Enet (deep neural network)

Name	Type	Output size
Initial		16 x 256 x 256
bottleneck 1.0	downsampling	64 x 128 x 128
4 x bottleneck1.x		64 x 128 x 128
bottleneck2.0	downsampling	128 x 64 x 64
bottleneck2.1		128 x 64 x 64
bottleneck2.2	dilated 2	128 x 64 x 64
bottleneck2.3	asymmetric 5	128 x 64 x 64
bottleneck2.4	dilated 4	128 x 64 x 64
bottleneck2.5		128 x 64 x 64
bottleneck2.6	dilated 8	128 x 64 x 64
bottleneck2.7	asymmetric 5	128 x 64 x 64
bottleneck2.8	dilated 16	128 x 64 x 64
Repeat section 2, without bottleneck2.0		
bottleneck4.0	upsampling	64 x 128 x 128
bottleneck4.1		64 x 128 x 128
bottleneck4.2		64 x 128 x 128
bottleneck50	upsampling	16 x 256 x 256
bottleneck5.1		16 x 256 x 256
fullconv		C x 512 x 512

Table 1 displays the architecture of our network. The initial digit following each block name and the horizontal lines in the table indicate that it is broken into multiple stages. For a sample input image resolution of 512 by 512, the output sizes are shown. We approach ResNets from the perspective that they have a convolutional filter extension that divides from the single main branch, and then, as seen in Figure 3b, combine back using an element wise addition. Three convolutional layers make up each block: a core convolutional layer (conv in Figure 3b), a 1 × 1 expansion, and a 1 × 1 projection that lowers the dimensionality. Between each convolution, we sandwich PReLU and Batch Normalization. A max pooling layer is added to the main branch if downsampling is the reason for the bottleneck.

The initial level, depicted in Figure 3a, consists of a single block. Except for Stage 3's omission of the 0th bottleneck and lack of initial downsampling of the input, Stages two and three are structurally identical to Stage one. There are five bottleneck blocks in Stage 1. These initial three stages make up the encoder. Stages 4 and 5 belong to the decoder.

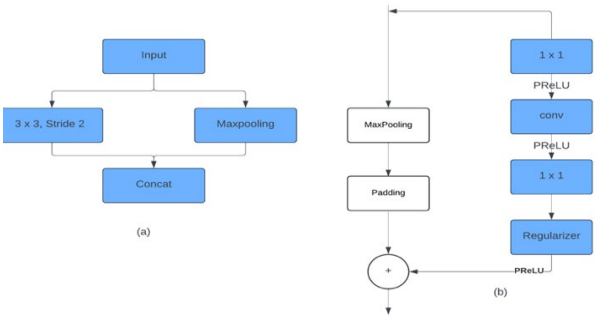


***Figure* 3.** (a) ENet initial block and (b) ENet bottleneck module [5].

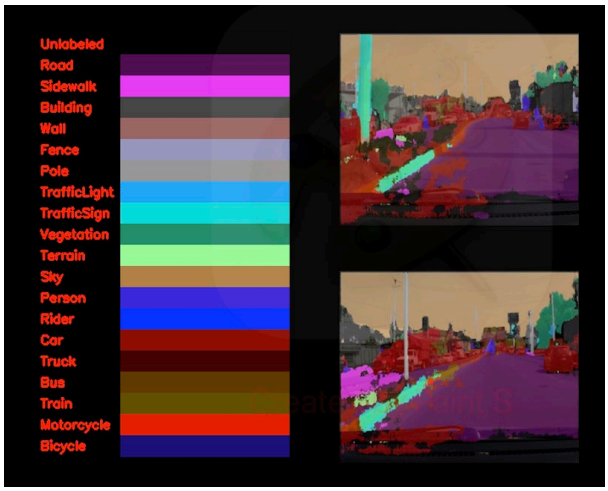


Figure 4. *Pixel classification and semantic segmentation results.*

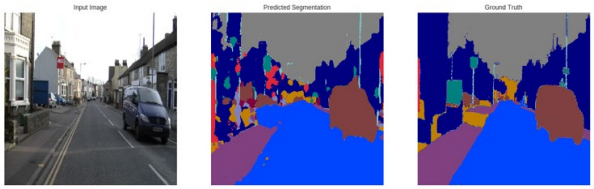


Figure 5. *Predicted segmentation and ground truth.*

In the above figure, predicted segmentation is the result of the semantic segmentation model's inference on an input image. It entails giving each pixel in the image a class name, so segmenting the image into areas that correspond to several object classes. Ground truth (above figure) in semantic segmentation refers to the manually annotated segmentation of the same image, provided by human annotators. These annotations serve as the "truth" against which the model's predictions are compared.

1 Pedestrian detection for perception (IAV)

The proposed algorithms in this research exhibited high accuracy in detecting pedestrians within the traffic scene. The algorithm successfully identified and localized pedestrians, contributing to a reliable perception of the surrounding environment. The system demonstrated real-time processing capabilities, maintaining high frames-per-second (FPS) rates during video analysis. This is crucial for applications requiring timely detection of pedestrians in dynamic traffic scenarios. The MeanShift algorithm conducts tracking repeatedly by computing a centroid based on probability values in the current tracking area. The Kalman filter in computer vision can improve the accuracy of a monitored object's position estimate.

Prediction step: The current state Xt+1 is estimated based on the previous state Xt using motion model F (transition state matrix) and process noise v:

$$xt+1=Fxt + v \quad (2)$$

The above equation for the new state contains information about one's position Pt and velocity Vt and can be written as below:

$$pt+1=pt + vt * \Delta t \quad (3)$$

$$vt+1= vt \quad (4)$$

The measure of the estimated state accuracy is given by

$$Pt+1 = F\ pt\ Ft + Q \quad (5)$$

To obtain the new measurements z of the bound boxes from the pedestrian detection module and calculating measurement update y:

$$Y = Z – Hxt+1 \quad (6)$$

The Kalman filter predicts the position of an object based on historical measurements. The prediction can then be modified using actual data. However, it is limited to using this for a single item. We therefore need a single Kalman filter for each object that will be under observation. The green box with the thin border in the result below represents the detected contour. The cyan box surrounded by a thick border represents the Kalman-corrected MeanShift tracking localization. Additionally, the center point shown by the blue dot is predicted by the Kalman filter. Below images are generated as output using the proposed algorithm.

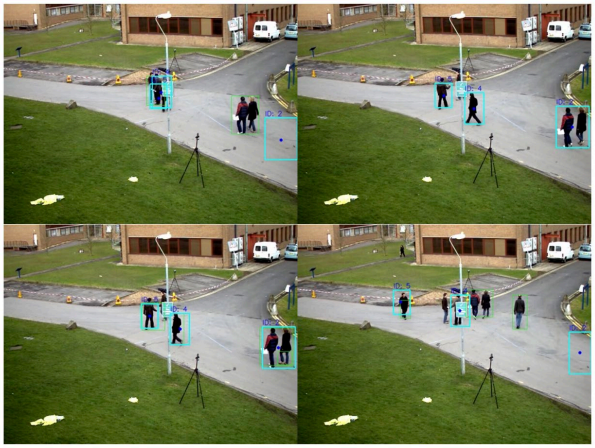


Figure 6. *Pedestrian detection using Artificial Intelligence.*

Traffic sign classification for perception (IAV)

The deep learning model achieved high accuracy in classifying various types of traffic signs, including regulatory, warning, and informational signs. The classification and detection results shown below demonstrated the model's proficiency in recognizing and assigning the correct labels to different signs.

Convolutional Neural Network (CNN) for traffic sign detection in this paper, we would typically represent the CNN layers and the operations mathematically. Below is a general formula representation:

Given an input image I with dimensions H×W×D (Height, Width, Depth), and a convolutional kernel K of size kH×kW×D, the output feature map F for the i-th layer can be expressed as:

$$Fi=\sigma(Ii-1*Ki+bi) \quad (7)$$

Where, Ii−1 is the input to the i-th layer, ∗ denotes the convolution operation, Ki is the convolutional kernel for the i-th layer, bi is the bias term for the i-th layer, σ is the activation function, typically ReLU (Rectified Linear Unit).

If we apply a pooling operation, typically max pooling, to reduce the spatial dimensions:

$$Pi=\text{maxpool}(Fi, pH, pW) \quad (8)$$

where: Pi is the output of the pooling layer, pH, pH and pW are the height and width of the pooling window. After flattening the output of the final convolutional or pooling layer, it is fed into a fully connected (dense) layer:

$$Oi=\sigma(Wi\cdot Pi-1+bi) \quad (9)$$

where: Wi is the weight matrix of the fully connected layer, Pi−1 is the flattened input vector from the previous layer, bi is the bias term, Oi is the output of the fully connected layer. For classification, the final output layer is typically a softmax layer that gives the probabilities for each traffic sign class:

$$y\hat{}=\text{softmax}(OL) \quad (10)$$

where: OL is the output of the last fully connected layer, y^ is the predicted probability distribution over the classes.

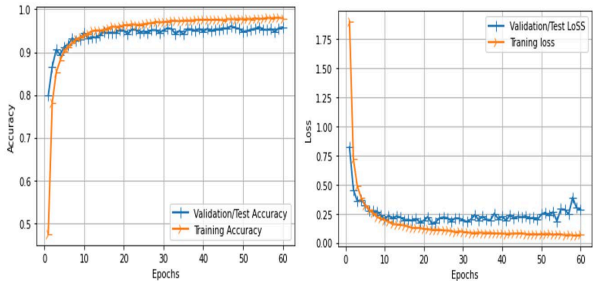


Figure 7. *Test accuracy: 0.9429 and a training loss versus test loss graph.*

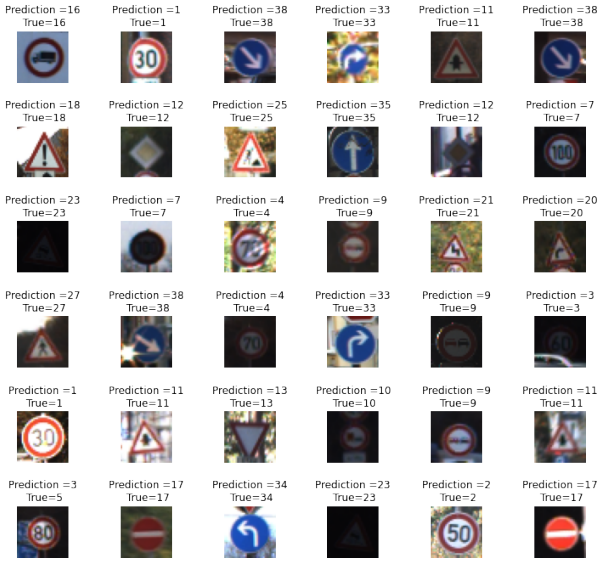


Figure 8. *Traffic sign prediction with respective class Id's.*

By observing above two graphs, the study finds that the proposed model for building IAV is performing very well. The traffic sign detector classifier with 94% accuracy can enhance the traffic perception and can help to build intelligent autonomous vehicles.

Conclusion

Self-driving cars or autonomous vehicles are a cutting-edge invention that will impact our lives in many ways, including how we perceive their usefulness, the ability to reconcile daily life,

advantages, and, of course, drawbacks, such as the lack of completely safe devices and software, legal requirements, data security and protection, network problems, and ethical consideration.

With the creation of cutting-edge deep learning and computer vision frameworks, as well as sophisticated AI models, signs of progress have been largely made in the performance of traffic perception and sensing in different weather conditions. This paper research also shows how to build IAV (Intelligent Autonomous vehicles) with advanced environment perception by implanting real time traffic detection, traffic lane detection, semantic segmentation for traffic scene understanding and pedestrian and traffic sign classification.

The advancements presented in this paper highlight the transformative potential of integrating state of the art AI technologies into Intelligent Autonomous Vehicles (IAVs) for enhanced perception. By leveraging algorithms like YOLOv8 for real-time traffic detection and ENet for semantic segmentation, we have demonstrated significant improvements in accurately detecting and understanding complex traffic environments. These technologies enable autonomous vehicles to perform real-time, precise perception tasks that are crucial for safe and efficient navigation in dynamic traffic scenarios.

Moreover, the comprehensive AI-driven framework proposed in this research, encompassing traffic object detection, pedestrian recognition, and traffic sign classification, contributes to developing safer, more intelligent, and reliable autonomous transportation systems. This work marks a step forward in creating more intelligent mobility solutions, paving the way for the future of autonomous driving. By advancing traffic awareness capabilities, we move closer to a world where autonomous vehicles become integral to sustainable and efficient transportation.

References

[1] Mrinal R. Bachute, Javed M. Subhedar - Autonomous Driving Architectures: Insights of Machine Learning and Deep Learning Algorithms https://doi.org/10.1016/j.mlwa.2021.100164.

[2] Youcef Djenouri AsmaBelhadi Gautam Srivastava DjamelDjenouri Jerry Chun-Wei Lin - Vehicle detection using improved region convolution neural network for accident prevention in smart roads - https://doi.org/10.1016/j.patrec.2022.04.012.

[3] Autonomous Intelligent Vehicles (AIV): Research statements, open issues, challenges and road for future https://doi.org/10.1016/j.ijin.2021.07.002

[4] Explainable Artificial Intelligence for Autonomous Driving: A Comprehensive Overview and Field Guide for Future Research Directions Shahin Atakishiyev, Mohammad Salameh, Hengshuai Yao, Randy Goebel https://doi.org/10.48550/arXiv.2112.11561

[5] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. arXiv preprint arXiv:1606.02147, 2016.

[6] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," arXiv preprint arXiv:1511.00561, 2015.

[7] Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues https://doi.org/10.1016/j.array.2021.100057

[8] You Only Look Once: Unified, Real-Time Object Detection https://doi.org/10.48550/arXiv.1506.02640

[9] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks https://doi.org/10.48550/arXiv.1506.01497

[10] End to End Learning for Self-Driving Cars https://doi.org/10.48550/arXiv.1604.07316

[11] Traffic Sign Classification Using Deep and Quantum Neural Networks https://doi.org/10.48550/arXiv.2209.15251

[12] Real-time semantic segmentation on FPGAs for autonomous vehicles with hls4ml https://doi.org/10.48550/arXiv.2205.07690

[13] Source: http://pjreddie.com/yolo/

[14] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. arXiv preprint arXiv:1310.1531, 2013.

[15] A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS arXiv:2304.00501v6.

[16] https://docs.ultralytics.com

[17] A. Broggi, P. Cerri, S. Debattisti, M.C. Laghi, P. Medici, D. Molinari, M. Panciroli, A. Prioletti, Proud Public Road urban driverless-car test IEEE Trans. Intell. Transport. Syst., 16 (Dec 2015), pp. 3508-3519.

[18] Autonomous Intelligent Vehicles (AIV): Research statements, open issues, challenges and road for future Amit Kumar Tyagi, S U Aswathy.

[19] Perception and sensing for autonomous vehicles under adverse weather conditions: A survey https://doi.org/10.1016/j.isprsjprs.2022.12.021

[20] An Introduction to Convolutional Neural Networks Keiron O'Shea, Ryan Nash arXiv:1511.08458v2.

[21] M. B. Blaschko and C. H. Lampert. Learning to localize objects with structured output regression. In Computer Vision–ECCV 2008, pages 2–15. Springer, 2008.

[22] L. Bourdev and J. Malik. Poselets: Body part detectors trained using 3d human pose annotations. In International Conference on Computer Vision (ICCV), 2009.

[23] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.
[24] Liu, C.; Li, S.; Chang, F.; Wang, Y. Machine Vision Based Traffic Sign Detection Methods: Review, Analyses and Perspectives. IEEE Access 2019, 7, 86578–86596

[25] Autonomous Intelligent Vehicles (AIV): Research statements, open issues, challenges and road for future Amit Kumar Tyagi, S U Aswathy

[26] A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS arXiv:2304.00501v6

[27] N. V. Sai Prakash Nagulapati, S. R. Venati, V. Chandran and S. R, "Pedestrian Detection and Tracking Through Kalman Filtering," 2022 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2022, pp. 1-6, doi: 10.1109/ESCI53509.2022.9758215.

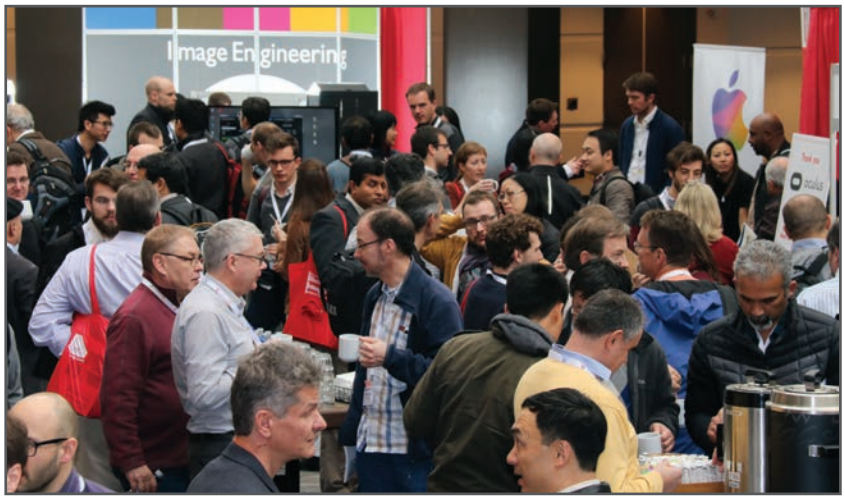
Author Biography

Alexander Iliev is currently a Professor at SRH Berlin University, he is also an Associate Professor at the Bulgarian Academy of Sciences in Sofia, Bulgaria, and a Lead Lecturer at UC Berkeley. He has many peer-revied publications as well as international conferences, patents, and books on topics in areas like Digital Signal Processing and Artificial Intelligence. Nithin Jayagovindan, was an master’s degree student of Big Data and AI at SRH University of Applied Sciences.

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 75 1947 IS&T 2022 imaging.org