

Advanced Assisted Car Driving in Low-light Scenarios

Francesco Rundo¹^a, Roberto Leotta², Angelo Messina^{3,4}^b and Sebastiano Battiato²^c

¹*STMicroelectronics, ADG Central R&D, Catania, Italy*

²*Department of Mathematics and Computer Science, University of Catania, Catania, Italy*

³*STMicroelectronics, Catania, Italy*

⁴*National Research Council, Institute for Microelectronics and Microsystems (IMM), Catania, Italy*

Keywords: ADAS, Automotive, Deep Learning.

Abstract: The robust identification, tracking and monitoring of driving-scenario moving objects represents an extremely critical task in the safe driving target of the latest generation cars. This accomplishment becomes even more difficult in a poor light driving scenarios such as driving at night or in rough weather conditions. Since the driving detected objects could represent a significant collision risk, the aim of the proposed pipeline is to address the issue of real time low-light driving salient objects detection and tracking. By using a combined time-transient non-linear deep architecture with convolutional network embedding self attention mechanism, the authors will be able to perform a real-time assessment of the low-light driving scenario frames. The downstream deep backbone learns such features from the driving frames thus improved in terms of light exposure in order to identify and segment salient objects. The implemented algorithm is ongoing to be ported over a hybrid architectures consisting of an embedded system with SPC5x Chorus device with an automotive-grade system based on STA1295 MCU core. The collected experimental results confirmed the effectiveness of the proposed approach.

1 INTRODUCTION

Autonomous Driving (AD) or Assisted Driving (ADAS i.e. Advanced Driver Assisting Systems) scenarios are considered very complex environments as often contain multiple inhomogeneous objects that move at different speeds and directions (Heimberger et al., 2017; Horgan et al., 2015). Both for self-driving vehicles as well as for Assisted driving, it is critical to have a well lit sampled driving scene frames as most of the classical computer vision algorithms degrade significantly in the absence of adequate lighting (Pham et al., 2020). More specifically, inadequate or insufficient light exposure in the driving scene does not allow the intelligent semantic segmentation algorithms to identify and track objects or a deep classifier to discriminate one object class rather than another. This poor efficiency of automotive-based computer vision algorithms is mainly due to poor significant pixel-based information i.e. limited discriminating vi-

sual features (Pham et al., 2020) due to video frames containing poorly lit driving scenes. Autonomous driving (AD) and driver assistance (ADAS) systems require high levels of robustness both in performance and fault-tolerance, often requiring high levels of validation and testing before being placed on the market (Rundo et al., 2021a). The authors have already deeply investigated the specifications and issues of ADAS technologies (Rundo et al., 2021a; Rundo et al., 2018b; Trenta et al., 2019; Rundo et al., 2019a; Rundo et al., 2020b; Rundo et al., 2020c; Rundo et al., 2019b; Conoci et al., 2018; Rundo, 2021). Considering what introduced, the authors of this contribution explore a further critical issues in the automotive field (both AD and ADAS) related to driving scenarios in the lack of adequate lighting. This scientific contribution is organized as follow: the next section “Related works” include the analysis of the state-of-the-art in the field of intelligent solutions for low-light driving scenario enhancement while section 3 “The Proposed Pipeline” describes in detail the proposed solution. Finally, the section 4 will report the experimental outcome of the designed pipeline while section 5 “Discussion and Conclusion” includes such descrip-

^a  <https://orcid.org/0000-0003-1766-3065>

^b  <https://orcid.org/0000-0002-2762-6615>

^c  <https://orcid.org/0000-0001-6127-2470>

tion of the main advantages of the proposed pipeline with some ideas for the future development.

2 RELATED WORKS

To improve the light-exposure of sampled car driving video frames, such classic image processing approaches showed limited effective performance. With the advent of the recent techniques of deep learning, several intelligent solutions have been implemented to perform complex and adaptive image processing tasks. With specific focus to the artificial light enhancement approaches applied to video and image processing, an interesting method was proposed by Qu et al in (Qu et al., 2019). They proposed a deep generative architecture i.e. the Cycle Generative Adversarial Networks (CycleGAN) combined with an additional discriminators. The authors tested their solution for addressing the task of the autonomous robot navigation retrieving very promising results. In (Chen et al., 2020) the authors analyzed an interesting framework based on bio-inspired solutions. The authors designed a bio-inspired solution which leverage the working-flow of the human retina. They proposed an event-based neuromorphic vision system suitable to convert asynchronous driving events into synchronous image or grid-like representations for subsequent tasks such as object detection and tracking. An innovative light sensitive cells which contain millions of hardware photo-receptors combined with an intelligent deep algorithm has been proposed in (Chen et al., 2020) to overcome the low-light driving issues. The authors in (Rashed et al., 2019) proposed a deep network named FuseMODNet to cover the issue of low-light driving scenario in ADAS applications. They proposed a robust and real-time Deep Convolutional Backbone for Moving Object Detection (MOD) under low-light conditions by capturing motion information from both camera and LiDAR device. They obtained in testing session a promising 10.1% relative improvement on common Dark-KITTI dataset, and a 4.25% improvement on standard KITTI dataset (Rashed et al., 2019).

In (Deng et al., 2020) the authors implemented a Retinex decomposition based solution for low-light image enhancement with joint decomposition and denoising. Preliminary the authors analyzed a new joint decomposition and denoising enhancement U-Net network (JDEU). The JDEU network was trained with low-light images only, and high quality normal-light images were used as reference to decompose the desired reflectance component for noise removal. The so computed reflectance and the adjusted illumina-

tion are reconstructed to produce the enhanced image. Experimental results based on LOL dataset confirmed the effectiveness of the pipeline proposed in (Deng et al., 2020). In (Pham et al., 2020) Pham et al investigated the usage of Retinex theory as effective tool for enhancing the illumination and detail of images. They collected a Low-Light Drive (LOL-Drive) dataset and applied a deep retinex neural network, named Drive-Retinex, which was validated using this dataset. The deep Retinex-Net consists of two sub-networks: Decom-Net (decomposes a color image into a reflectance map and an illumination map) and Enhance-Net (enhances the light level in the illumination map). The performed several experimental sessions which confirmed that the proposed method was able to achieve visually appealing low-light enhancement.

Szankin et al analyzed in (Szankin et al., 2018) the application of low-power system for road condition classification and pedestrian detection in challenging environments, including low-light driving. The authors investigated the influence of various factors (lightning conditions, moisture of the road surface and ambient temperature) on the system capability to properly detect the pedestrian and road in the driving scenario. The implemented system was tested on images acquired in different climate zones. The solution they reported in (Szankin et al., 2018) confirmed a precision and recall indexes above 95% in challenging driving scenarios. More details in (Szankin et al., 2018). In (Yang et al., 2021) another interesting approach based on Retinex theory was presented.

The aforementioned scientific approaches have allowed some improvement in driving video frames enhancement associated with a poorly lit driving scenarios. Anyway, the analyzed solutions often fail to find an optimal trade-off between accuracy and time-performance due to the complexity of the designed underlying hardware framework needed to host their proposed algorithms (Rashed et al., 2019; Deng et al., 2020; Pham et al., 2020; Szankin et al., 2018; Yang et al., 2021). The method herein proposed tries to balance the above items, providing robust assessment with acceptable time-performance outcomes. The next section will introduce and detail the proposed pipeline.

3 THE PROPOSED PIPELINE

The target of this scientific contribution is the design of a robust and effective pipeline that allows a robust and intelligent object detection in a poorly lit driving scenarios. The proposed approach is schematized in

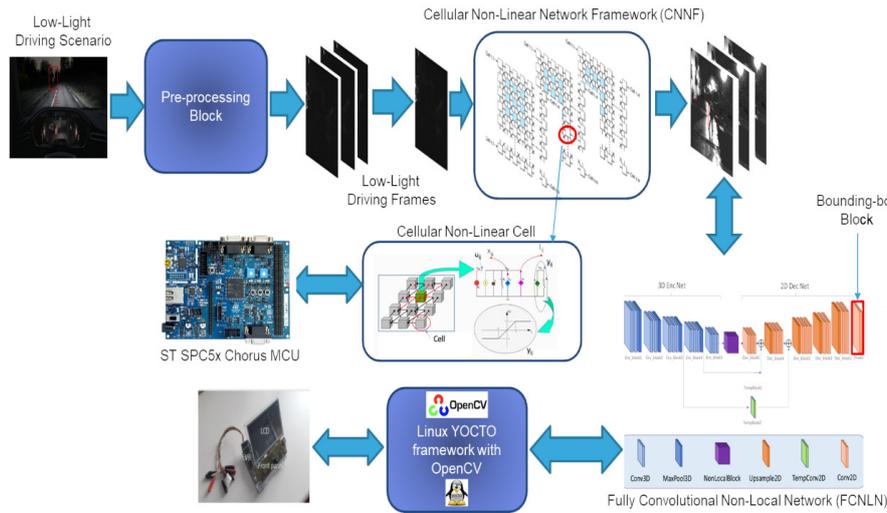


Figure 1: The schematic overview of the proposed pipeline.

Fig. 1. As showed in that figure, the implemented solution is composed by three sub-systems: the first sub-system is the “Pre-processing block”, the second one is the “Deep Cellular Non-Linear Network Framework” and finally there is the “Fully Convolutional Non-Local Network”. More details in the next sections.

3.1 The Pre-processing Block

As reported in Fig. 1, the first sub-system of the proposed approach is the Pre-processing block. The aim of this block is the normalization of the sampled driving video. Specifically, the sampled low-light driving frames (captured with a low frame-rate video-device having frame-per-second (fps) in the range 40 – 60) will be resized (with a classic bi-cubic algorithm) to a size of 256×256 . For this purpose, a simple gray-level automotive-grade video-camera would be enough even though a similar colour camera, could be used with a downstream classic YCbCr conversion from which the gray-level luminance “Y” will be extracted. The single sampled driving frame $I_k(x, y)$ thus pre-processed will be buffered in the memory area of the micro ST SPC5x MCU and gradually processed by the next Deep Cellular Non-Linear Network sub-system.

3.2 The Deep Cellular Non-linear Network Framework

As showed in Fig. 1, the pre-processed low-light frames $I_k(x, y)$ will be further processed by the the Deep Cellular Non-Linear Network Framework. The goal of this block is the artificial increase of the light-

exposure of the sampled video frames representative of the driving scenario in order to get a robust object detection or segmentation. A brief introduction about the Time-transient Deep Non-linear Network as knows as Cellular Neural (or Non-Linear) Network (TCNN). The first architecture of the Cellular Neural (or Nonlinear) Network (CNN) was firstly proposed by L.O Chua and L. Yang (Chua and Yang, 1988a; Chua and Yang, 1988b). The CNNs architecture could be defined as high speed local interconnected computing array of analog processors known as “cells” (Chua and Yang, 1988b).

Many applications and extensions have been proposed in scientific literature (Conoci et al., 2017; Mizutani, 1994). The core of the CNN backbone is the cell. The CNNs processing is configured through the instructions provided by the so-called cloning templates (Chua and Yang, 1988a; Chua and Yang, 1988b). Each cell of the CNN array may be considered as dynamical system which is arranged into a topological 2D or 3D structure. The CNN cells interact each other within its neighborhood defined by heuristically ad-hoc defined radius. Each CNN cell has an input, a state and an output as Piece-Wise-Linear(PWL) state re-mapping. The CNNs can be implemented with analog discrete components or by means of U-VLSI technology performing near real-time analogic processing task. Some stability results and consideration about the dynamics of the CNNs can be found in (Cardarilli et al., 1993).

In the CNN paradigm different heuristic relationship cell-models between state, input and neighborhood can be defined. Consequently, the CNN architecture can be considered as a system of cells (or dynamical neurons) mapped over a normed space S_N

(cell grid), which is a discrete subset of \mathbb{R}^n (generally $n \leq 3$) with distance function $d: S_N \rightarrow N$. The CNNs cells are identified by indices defined in a space-set P_i . Neighborhood function $N_r(\cdot)$ can be defined with the following Eq. 1 and Eq. 2:

$$N_r: P_i \rightarrow P_i^\theta \quad (1)$$

$$N_r(k) = \{P|d(i, j) \leq r_c\} \quad (2)$$

where θ depends on r_c (neighborhood radius) and on spatial-geometry representation of the grid.

Cells are multiple input – single output nonlinear processors. The CNNs can be implemented as single layer or multi-layers so that the cell grid can be e.g. a planar array (with rectangular, square, octagonal geometry) or a k-dimensional array (usually $k \geq 3$), generally considered and realized as a stack of k-dimensional arrays (layers). For the pipeline herein described, a further extension of the Chua's CNN is proposed i.e. the time Transiert CNN (TCNN). Specifically, a new cloning template matrix $A_4(i, j; k, l)$ as detailed in Eq. 3 is added in the classical CNN dynamical paradigm.

$$\begin{aligned} C \frac{dx_{ij}(t, t_k)}{dt} &= \frac{-1}{R_x} x_{ij} \\ &+ \sum_{C(k, l) \in N_r(i, j)} A_1(i, j; k, l) y_{kl}(t, t_k) \\ &+ \sum_{C(k, l) \in N_r(i, j)} A_2(i, j; k, l) u_{kl}(t, t_k) \\ &+ \sum_{C(k, l) \in N_r(i, j)} A_3(i, j; k, l) x_{kl}(t, t_k) \\ &+ \sum_{C(k, l) \in N_r(i, j)} A_4(i, j; k, l) (y_{ij}(t), y_{kl}(t), t_k) \\ &+ I \end{aligned} \quad (3)$$

$(1 \leq i \leq M, 1 \leq j \leq N)$

where,

$$\begin{aligned} y_{ij}(t) &= \frac{1}{2} (|x_{ij}(t, t_k) + 1| - |x_{ij}(t, t_k) - 1|) \\ N_r(i, j) &= \{C_r(k, l); (\max(|k - i|, |l - j|) \leq r_c, \\ &\quad 1 \leq k \leq M, \\ &\quad 1 \leq l \leq N)\} \end{aligned}$$

In Eq. 3 the $N_r(i, j)$ represents the neighborhood of each cell $C(i, j)$ with radius r_c . The terms x_{ij} , y_{ij} , u_{ij} and I are respectively: the state, the output and the input of the cell $C(i, j)$ while $A_1(i, j; k, l)$, $A_2(i, j; k, l)$, $A_3(i, j; k, l)$ and $A_4(i, j; k, l)$ are the cloning templates suitable to define the TCNNs processing task. As described by the researchers (Chua and Yang, 1988b;

Conoci et al., 2017; Mizutani, 1994; Cardarilli et al., 1993; Roska and Chua, 1992; Arena et al., 1996) through the numerical configuration of the cloning templates as well as the bias I , it is possible to configure the type of processing provided by the CNNs.

In this contribution, the authors propose a Time-transient Deep CNN (TCNN) i.e. a non-linear network which dynamically evolves in a short time range, i.e. during the transient $[t, t_k]$. Normally, CNN evolves up to a defined steady-state (Chua and Yang, 1988b; Conoci et al., 2017; Mizutani, 1994; Cardarilli et al., 1993). The dynamic mathematical model of the TCNN is reported in 3. Specifically, the input visual 256×256 low-light driving frame $I_k(x, y)$, will be fed as state x_{ij} and input u_{ij} of the TCNN $D_k(x, y)$. Each setup of the cloning templates and bias $A_1(i, j; k, l)$, $A_2(i, j; k, l)$, $A_3(i, j; k, l)$, $A_4(i, j; k, l)$, I will allow to retrieve a specific augmentation-enhancement of the input frame in order to provide an artificial enhancement of the low-light frame. As reported in (Chua and Yang, 1988a; Chua and Yang, 1988b; Conoci et al., 2017; Mizutani, 1994; Cardarilli et al., 1993; Roska and Chua, 1992; Arena et al., 1996), defined a specific CNNs target to be reached, there is no analytic-deterministic algorithm to retrieve the coefficients of the correlated cloning templates. There are several database (Mizutani, 1994; Cardarilli et al., 1993; Roska and Chua, 1992; Arena et al., 1996) but each processing task need ad-hoc cloning templates numerical configuration. Therefore, by means of heuristically driven optimization tests, we tried to setup a cloning templates configurations that were able to artificially improve the lighting conditions of the input sampled driving frames. The following cloning templates configuration were used as final setup of the proposed TCNN: $A_1 = [0.03, 0.03, 0.03; 0.02, 0.02, 0.025; 0.25, 0.25; 0.25]$; $A_2 = [0.01, 0.01, 0.01; 0.01, 0.01, 0.01; 0.04, 0.04; 0.04]$; $A_3 = 0$; $A_4 = 0$; $I = -0.55$. The output of the previous TCNN will be further processed by another TCNN configured as follow: $A_1 = [0.75, 0, 0; 0.75, 2.5, 0.75; 0.75, 0, -0.75]$; $A_2 = [0.01, 0.01, 0.01; 0.01, 0.01, 0.01; 0.04, 0.04; 0.04]$; $A_3 = 0$; $A_4 = 0$; $I = -0.75$ and $A_1 = [0.75, 0.75, 0.75; 0.75, 2.5, 0.75; 0.75, 0 - 0.75]$; $A_2 = [0.01, 0.01, 0.01; 0.01, 0.01, 0.01; 0.04, 0.04, 0.04]$; $A_3 = [0.75, 0.75, 0.75; 0.75, 2.5, 0.75; -0.55, 0, -0.55]$; $A_4 = 0$; $I = -0.75$. In the Fig. 2 we report such instances of the TCNN light enhancement, with a detail of each improvement occurred at each TCNN processing as per aforementioned cloning setup.

As shown in Fig. 2, the TCNN-based framework is able to significantly improve the light exposure of the source low-light driving video frames.

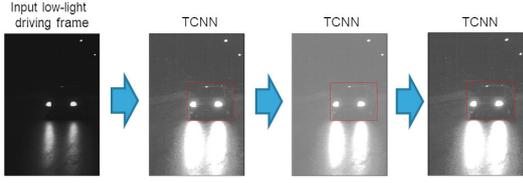


Figure 2: An instance of the low-light image enhancement performed by the proposed TCNN.

3.3 The Fully Convolutional Non-local Network

The target of this block is the salient detection and segmentation (bounding box) of the light-enhanced input driving video frames. As highlighted in Fig. 1, the output of the previous TCNN block will be fed as input to this Fully Convolutional Non-Local Network (FCNLN) sub-system.

The sampled driving scene video frames will be processed by ad-hoc designed 3D to 2D Semantic Segmentation Fully Convolutional Non-Local Network as reported in Fig. 1. Through a semantic segmentation of the driving context, thanks to the encoding / decoding architecture of designed deep backbone, the saliency map of the driving scene will be reconstructed. This saliency map will be used by the bounding-box block which reconstruct the segmentation Region of Interest (ROI). The proposed FCNLN architecture is structured as follows. The encoder block (3D Enc Net) processes the space-time features of the captured driving scene frames and it is made up of 5 blocks. The first two blocks includes (for each block) two separable convolution layers with $3 \times 3 \times 2$ kernel filter followed by a batch normalization, ReLU layer and a downstream $1 \times 2 \times 2$ max-pooling layer. The remaining three blocks (for each block) includes two separable convolution layers with $3 \times 3 \times 3$ kernel filter followed by a batch normalization, another convolutional layer with $3 \times 3 \times 3$ kernel, batch normalization and ReLU with a downstream $1 \times 2 \times 2$ max-pooling layer.

Before to feed the so processed features to the Decoder side, a Non-Local self attention block is embedded in the proposed backbone. Non-local blocks have been recently introduced (Wang et al., 2018), as very promising approach for capturing space-time long-range dependencies and correlation on feature maps, resulting in a sort of “self-attention” mechanism (Rundo et al., 2020a). Non-local blocks take inspiration from the non-local means method, extensively applied in computer vision (Wang et al., 2018; Rundo et al., 2020a). Self-attention through non-local blocks aims to enforce the model to extract correlation among feature maps by weighting the averaged sum

of the features at all spatial positions in the processed feature maps (Wang et al., 2018).

In our pipeline, non-local blocks operate between the 3D encoder and 2D decoder side respectively. The mathematical formulation of non-local operation is reported. Given a generic deep network as well as a general input data x , the employed non-local operation computes the corresponding response y_i (of the given Deep architecture) at a i location in the input data as a weighted sum of the input data at all positions $j \neq i$:

$$y_i = \frac{1}{\Psi(x)} \sum_{\forall j} \xi(x_i, x_j) \beta(x_j) \quad (4)$$

With $\xi(\cdot)$ being a pairwise potential describing the affinity or relationship between data positions at index i and j respectively. The function $\beta(\cdot)$ is, instead, a unary potential modulating ξ according to input data. The sum is then normalized by a factor $\Psi(x)$. The parameters of ξ , β and Ψ potentials are learned during model’s training and defined as in the following Eq. 5:

$$\xi(x_i, x_j) = e^{\Gamma(x_i)^T \Phi(x_j)} \quad (5)$$

Where Γ and Φ are two linear transformations of the input data x with learnable weights W_Γ and W_Φ :

$$\begin{aligned} \Gamma(x_i) &= W_\Gamma x_i \\ \Phi(x_j) &= W_\Phi x_j \\ \beta(x_j) &= W_\beta x_j \end{aligned} \quad (6)$$

For the $\beta(\cdot)$ function, a common linear embedding (classical $1 \times 1 \times 1$ convolution) with learnable weights W_β is employed. The normalization function $\Psi(x)$ is detailed in the following Eq. 7.

$$\Psi(x) = \sum_{\forall j} \xi(x_i, x_j) \quad (7)$$

The above mathematical formulation of Non-Local features processing is named “Embedded Gaussian” (Conoci et al., 2017). The output of Non-Local processing of the encoded features will be fed in the decoder side of the pipeline. The Decoder backbone (2D Dec Net) is composed according to the encoder backbone for up-sampling and decoding the visual Non-Local features of the encoder. The output of the so designed FCNLN is the feature saliency map of the acquired scene frame i.e. the segmented area of the most salient object. The Fig. 3 shows some instance of the processing output of the proposed FCNLN.

The Bounding-box block will define the bounding area around the saliency map by means of an enhanced minimum rectangular box criteria (increased by 20% along each dimension) which is able to enclose the salience area. The Fig. 4 shows an example of an automatically generated bounding-box (red rectangle).

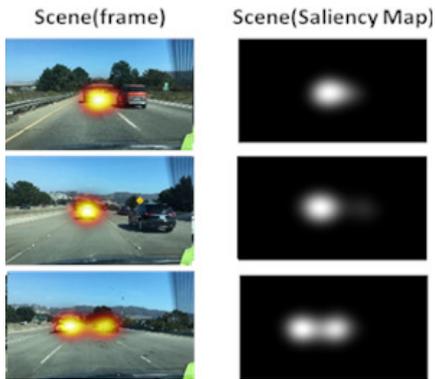


Figure 3: Saliency analysis of the video representing the driving scene.

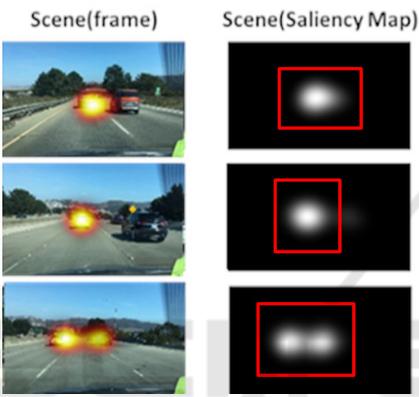


Figure 4: Bounding-box segmentation (red rectangular box) of the saliency map.

The proposed FCNLN architecture has been validated and tested on the DHF1K dataset (Min and Corso, 2019) retrieving the following performance: Area Under the Curve: 0.899; Similarity: 0.455; Correlation Coefficient: 0.491; Normalized Scanpath Saliency: 2.772. Unfortunately, more efficient deep architectures have the disadvantage of being complex and difficult to be hosted into automotive-grade embedded systems (STMicroelectronics, 2018; Rundo et al., 2021b; Rundo et al., 2020b).

4 EXPERIMENTAL RESULTS

To test the proposed pipeline, we arranged separate validation for each of the implemented subsystem. Specifically, for the visual saliency assessment, the author has extracted images from the following dataset: Oxford Robot Car dataset (Maddern et al., 2017) and the Exclusively Dark (ExDark) Image Dataset (Loh and Chan, 2019). This so composed dataset contains more than 20 million images having an average resolution greater than 640×480 . We se-

lect 15000 driving frames of that dataset to compose the training set. Moreover, a further testing and validation sessions have been made over such custom dataset. The authors have splitted the dataset as follow: 70% for training as well as 30% for testing and validation of the proposed approach.

The FCNLN has been trained with a mini-batch gradient descent with Adam optimizer and initial learning rate of 0.01. The deep model is implemented using Pytorch framework. Experiments were carried out on a server with Intel Xeon CPUs equipped with a Nvidia GTX 2080 GPU with 16 Gbyte ad memory video. The collected experimental results have been reported in the following Table 1 and Table 2. In particular, Table 1 shows the results obtained without low-light enhancement, Table 2 shows the results obtained with the low-light enhancement.

Table 1: Benchmark comparison with similar pipeline without low-light enhancement.

| Pipeline | Number of Detected Objects | Average Degree of Confidence | Accuracy |
|-----------------------------------|----------------------------|------------------------------|----------|
| Ground Truth | 12115 | 1.0 | 100% |
| Proposed | 9995 | 0.9001 | 82.501% |
| Yolo V3 | 7991 | 0.8851 | 65.959% |
| FCN with DenseNet-201 as backbone | 8337 | 0.8912 | 68.815% |
| Faster-R-CNN (ResNet as backbone) | 8009 | 0.8543 | 66.108% |
| Mask-R-CNN (ResNet as backbone) | 9765 | 0.8991 | 80.602% |

Table 2: Benchmark comparison with similar pipeline with low-light enhancement.

| Pipeline | Number of Detected Objects | Average Degree of Confidence | Accuracy |
|-----------------------------------|----------------------------|------------------------------|----------|
| Ground Truth | 12115 | 1.0 | 100% |
| Proposed | 10222 | 0.9055 | 84.374% |
| Proposed without Non-Local Block | 10002 | 0.9541 | 82.558% |
| Yolo V3 | 9991 | 0.8991 | 82.468% |
| FCN with DenseNet-201 as backbone | 9123 | 0.9001 | 75.303% |
| Faster-R-CNN (ResNet as backbone) | 9229 | 0.8998 | 76.178% |
| Mask-R-CNN (ResNet as backbone) | 10125 | 0.9112 | 83.574% |

As showed in Table 1, the proposed pipeline outperforms the compared deep architecture in terms of object detection accuracy in low-light driving scenarios. The enhanced performance made by the contribution of the designed TCNN processing is reported in Table 2.

From Table 2 it is highlighted a significant increase in the performance of the TCNN enhanced pipeline. The proposed whole pipeline was able to perform better than the others also in terms of classification and this would seem to be related to the action of Non-Local self-attention blocks as the network without these blocks degrades in performance (see ablation reported in Table 2). The Fig. 5 reports some instances of the enhanced driving video frames

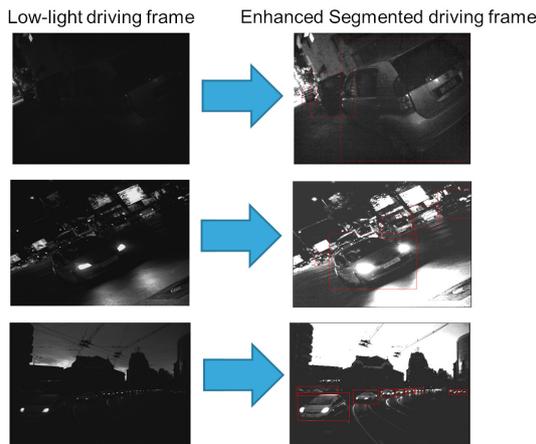


Figure 5: Some instances of the low-light driving frames with corresponding enhanced and segmented frames.

with embedded bounding box of the light-enhanced detected salient objects.

5 DISCUSSION AND CONCLUSION

The ability of our proposed pipeline to assess the overall low-light driving has been confirmed by the performance results described in the previous section. Compared with other methods in the literature, the implemented system shows competitive advantages mainly related to the underlying hosting hardware. Our approach overcomes the main drawbacks of the similar solutions as it uses only the TCNN block for performing ad-hoc pre-processing of the source low-light driving frames. This effective pipeline is currently being ported to an embedded system based on the STA1295 Accordo5 SoC platform produced by STMicroelectronics with a software framework embedding a distribution of YOCTO Linux O.S (STMicroelectronics, 2018) and OpenCV stack. We are working to enhance the pipeline by means of adaptive domain adaptation approaches in order to improve the overall robustness of the proposed intelligent pipeline (Rundo et al., 2019c; Rundo et al., 2019d; Rundo et al., 2018a; Banna et al., 2018; Banna et al., 2019).

ACKNOWLEDGEMENTS

This research was funded by the National Funded Program 2014-2020 under grant agreement n. 1733, (ADAS + Project). The reported information is covered by the following registered patents: IT Patent Nr. 102017000120714, 24 October 2017. IT Patent

Nr. 102019000005868, 16 April 2018; IT Patent Nr. 102019000000133, 07 January 2019.

REFERENCES

- Arena, P., Baglio, S., Fortuna, L., and Manganaro, G. (1996). Dynamics of state controlled cnns. In *1996 IEEE International Symposium on Circuits and Systems. Circuits and Systems Connecting the World. IS-CAS 96*, volume 3, pages 56–59. IEEE.
- Banna, G. L., Camerini, A., Bronte, G., Anile, G., Addeo, A., Rundo, F., Zanghi, G., Lal, R., and Libra, M. (2018). Oral metronomic vinorelbine in advanced non-small cell lung cancer patients unfit for chemotherapy. *Anticancer research*, 38(6):3689–3697.
- Banna, G. L., Olivier, T., Rundo, F., Malapelle, U., Fraggetta, F., Libra, M., and Addeo, A. (2019). The promise of digital biopsy for the prediction of tumor molecular features and clinical outcomes associated with immunotherapy. *Frontiers in medicine*, 6:172.
- Cardarilli, G., Lojacono, R., Salerno, M., and Sargeni, F. (1993). Vlsi implementation of a cellular neural network with programmable control operator. In *Proceedings of 36th Midwest Symposium on Circuits and Systems*, pages 1089–1092. IEEE.
- Chen, G., Cao, H., Conradt, J., Tang, H., Rohrbein, F., and Knoll, A. (2020). Event-based neuromorphic vision for autonomous driving: a paradigm shift for bio-inspired visual sensing and perception. *IEEE Signal Processing Magazine*, 37(4):34–49.
- Chua, L. O. and Yang, L. (1988a). Cellular neural networks: Applications. *IEEE Transactions on circuits and systems*, 35(10):1273–1290.
- Chua, L. O. and Yang, L. (1988b). Cellular neural networks: Theory. *IEEE Transactions on circuits and systems*, 35(10):1257–1272.
- Conoci, S., Rundo, F., Fallica, G., Lena, D., Buraoli, I., and Demarchi, D. (2018). Live demonstration of portable systems based on silicon sensors for the monitoring of physiological parameters of driver drowsiness and pulse wave velocity. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–3. IEEE.
- Conoci, S., Rundo, F., Petralta, S., and Battiato, S. (2017). Advanced skin lesion discrimination pipeline for early melanoma cancer diagnosis towards poc devices. In *2017 European Conference on Circuit Theory and Design (ECCTD)*, pages 1–4. IEEE.
- Deng, J., Pang, G., Wan, L., and Yu, Z. (2020). Low-light image enhancement based on joint decomposition and denoising u-net network. In *2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*, pages 883–888. IEEE.
- Heimberger, M., Horgan, J., Hughes, C., McDonald, J., and Yogamani, S. (2017). Computer vision in automated

- parking systems: Design, implementation and challenges. *Image and Vision Computing*, 68:88–101.
- Horgan, J., Hughes, C., McDonald, J., and Yogamani, S. (2015). Vision-based driver assistance systems: Survey, taxonomy and advances. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pages 2032–2039. IEEE.
- Loh, Y. P. and Chan, C. S. (2019). Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42.
- Maddern, W., Pascoe, G., Linegar, C., and Newman, P. (2017). 1 year, 1000 km: The oxford robotcar dataset. *The International Journal of Robotics Research*, 36(1):3–15.
- Min, K. and Corso, J. J. (2019). Tased-net: Temporally-aggregating spatial encoder-decoder network for video saliency detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2394–2403.
- Mizutani, H. (1994). A new learning method for multi-layered cellular neural networks. In *Proceedings of the Third IEEE International Workshop on Cellular Neural Networks and their Applications (CNNA-94)*, pages 195–200. IEEE.
- Pham, L. H., Tran, D. N.-N., and Jeon, J. W. (2020). Low-light image enhancement for autonomous driving systems using driveretinet-net. In *2020 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, pages 1–5. IEEE.
- Qu, Y., Ou, Y., and Xiong, R. (2019). Low illumination enhancement for object detection in self-driving. In *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1738–1743. IEEE.
- Rashed, H., Ramzy, M., Vaquero, V., El Sallab, A., Sistu, G., and Yogamani, S. (2019). Fusemodnet: Real-time camera and lidar based moving object detection for robust low-light autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0.
- Roska, T. and Chua, L. O. (1992). Cellular neural networks with non-linear and delay-type template elements and non-uniform grids. *International Journal of Circuit Theory and Applications*, 20(5):469–481.
- Rundo, F. (2021). Intelligent real-time deep system for robust objects tracking in low-light driving scenario. *Computation*, 9(11):117.
- Rundo, F., Banna, G. L., Prezzavento, L., Trenta, F., Conoci, S., and Battiato, S. (2020a). 3d non-local neural network: A non-invasive biomarker for immunotherapy treatment outcome prediction. case-study: Metastatic urothelial carcinoma. *Journal of Imaging*, 6(12):133.
- Rundo, F., Conoci, S., Banna, G. L., Ortis, A., Stanco, F., and Battiato, S. (2018a). Evaluation of levenberg-marquardt neural networks and stacked autoencoders clustering for skin lesion analysis, screening and follow-up. *IET Computer Vision*, 12(7):957–962.
- Rundo, F., Conoci, S., Battiato, S., Trenta, F., and Spampinato, C. (2020b). Innovative saliency based deep driving scene understanding system for automatic safety assessment in next-generation cars. In *2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, pages 1–6. IEEE.
- Rundo, F., Conoci, S., Spampinato, C., Leotta, R., Trenta, F., and Battiato, S. (2021a). Deep neuro-vision embedded architecture for safety assessment in perceptive advanced driver assistance systems: the pedestrian tracking system use-case. *Frontiers in neuroinformatics*, 15.
- Rundo, F., Leotta, R., and Battiato, S. (2021b). Real-time deep neuro-vision embedded processing system for saliency-based car driving safety monitoring. In *2021 4th International Conference on Circuits, Systems and Simulation (ICCSS)*, pages 218–224. IEEE.
- Rundo, F., Petralia, S., Fallica, G., and Conoci, S. (2018b). A nonlinear pattern recognition pipeline for ppg/ecg medical assessments. In *Convegno Nazionale Sensori*, pages 473–480. Springer.
- Rundo, F., Rinella, S., Massimino, S., Coco, M., Fallica, G., Parenti, R., Conoci, S., and Perciavalle, V. (2019a). An innovative deep learning algorithm for drowsiness detection from eeg signal. *Computation*, 7(1):13.
- Rundo, F., Spampinato, C., Battiato, S., Trenta, F., and Conoci, S. (2020c). Advanced 1d temporal deep dilated convolutional embedded perceptual system for fast car-driver drowsiness monitoring. In *2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, pages 1–6. IEEE.
- Rundo, F., Spampinato, C., and Conoci, S. (2019b). Ad-hoc shallow neural network to learn hyper filtered photoplethysmographic (ppg) signal for efficient car-driver drowsiness monitoring. *Electronics*, 8(8):890.
- Rundo, F., Trenta, F., Di Stallo, A. L., and Battiato, S. (2019c). Advanced markov-based machine learning framework for making adaptive trading system. *Computation*, 7(1):4.
- Rundo, F., Trenta, F., di Stallo, A. L., and Battiato, S. (2019d). Grid trading system robot (gtsbot): A novel mathematical algorithm for trading fx market. *Applied Sciences*, 9(9):1796.
- STMicroelectronics (2018). STMicroelectronics ACCORDO 5 Automotive Microcontroller. <https://www.st.com/en/automotive-infotainment-and-telematics/sta1295.html>. (accessed on 01 June 2021).
- Szankin, M., Kwaśniewska, A., Ruminski, J., and Nicolas, R. (2018). Road condition evaluation using fusion of multiple deep models on always-on vision processor. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pages 3273–3279. IEEE.
- Trenta, F., Conoci, S., Rundo, F., and Battiato, S. (2019). Advanced motion-tracking system with multi-layers deep learning framework for innovative car-driver drowsiness monitoring. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–5. IEEE.
- Wang, X., Girshick, R., Gupta, A., and He, K. (2018). Non-local neural networks. In *Proceedings of the IEEE*

conference on computer vision and pattern recognition, pages 7794–7803.

Yang, W., Wang, W., Huang, H., Wang, S., and Liu, J. (2021). Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086.

