

Article

High-Accuracy Deep Learning-Based Detection and Classification Model in Color-Shift Keying Optical Camera Communication Systems

Francisca V. Vera Vera ¹, Leonardo Muñoz ¹, Francisco Pérez ¹, Lisandra Bravo Alvarez ^{1,*}, Samuel Montejo-Sánchez ^{2,*}, Vicente Matus Icaza ³, Lien Rodríguez-López ⁴ and Gabriel Saavedra ¹

¹ Department of Electrical Engineering, Universidad de Concepción, Edmundo Larenas 219, Concepción 4030000, Chile; franciscavera92@gmail.com (F.V.V.V.); lemunoz2017@udec.cl (L.M.); francisperez@udec.cl (F.P.); gasaavedra@udec.cl (G.S.)

² Instituto Universitario de Investigación y Desarrollo Tecnológico, Universidad Tecnológica Metropolitana, Ignacio Valdivieso 2409, Santiago 8940000, Chile

³ Institute for Technological Development and Innovation in Communications, Universidad de Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain; vicente.matus@ulpgc.es

⁴ Facultad de Ingeniería, Universidad San Sebastián, Lientur 1457, Concepción 4030000, Chile; lien.rodriguez@uss.cl

* Correspondence: lisanbravo@udec.cl (L.B.A.); smontejo@utem.cl (S.M.-S.)

Abstract

The growing number of connected devices has strained traditional radio frequency wireless networks, driving interest in alternative technologies such as optical wireless communications (OWC). Among OWC solutions, optical camera communication (OCC) stands out as a cost-effective option because it leverages existing devices equipped with cameras, such as smartphones and security systems, without requiring specialized hardware. This paper proposes a novel deep learning-based detection and classification model designed to optimize the receiver's performance in an OCC system utilizing color-shift keying (CSK) modulation. The receiver was experimentally validated using an 8×8 LED matrix transmitter and a CMOS camera receiver, achieving reliable communication over distances ranging from 30 cm to 3 m under varying ambient conditions. The system employed CSK modulation to encode data into eight distinct color-based symbols transmitted at fixed frequencies. Captured image sequences of these transmissions were processed through a YOLOv8-based detection and classification framework, which achieved 98.4% accuracy in symbol recognition. This high precision minimizes transmission errors, validating the robustness of the approach in real-world environments. The results highlight OCC's potential for low-cost applications, where high-speed data transfer and long-range are unnecessary, such as Internet of Things connectivity and vehicle-to-vehicle communication. Future work will explore adaptive modulation and coding schemes as well as the integration of more advanced deep learning architectures to improve data rates and system scalability.

Keywords: convolutional neural network (CNN); deep learning; optical camera communication (OCC)



Academic Editors: Borja Genoves Guzman and Maximo Morales-Céspedes

Received: 31 July 2025

Revised: 27 August 2025

Accepted: 29 August 2025

Published: 2 September 2025

Citation: Vera, F.V.V.; Muñoz, L.; Pérez, F.; Alvarez, L.B.; Montejo-Sánchez, S.; Icaza, V.M.; Rodríguez-López, L.; Saavedra, G. High-Accuracy Deep Learning-Based Detection and Classification Model in Color-Shift Keying Optical Camera Communication Systems. *Sensors* **2025**, *25*, 5435. <https://doi.org/10.3390/s25175435>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Every year, the technology and communications landscape undergoes exponential growth, driven by the increasing demand for high-speed Internet services. However, while this surge in demand emphasizes the need to improve network capacity [1], saturation of the

radio frequency (RF) spectrum presents significant challenges [2]. These factors underscore the imperative need to explore new technologies [3] that, although not directly competing with RF in terms of speed, can effectively complement its use in various scenarios [4].

In this context, Optical Wireless Communication (OWC) has emerged as an innovative and viable alternative to the saturated RF spectrum, distinguished by its operation within an unlicensed spectrum and its potential to deliver significantly greater bandwidth than traditional RF systems [5]. Within the realm of OWC, Optical Camera Communication (OCC) technology has experienced substantial technological advancements and renewed research interest. OCC is incorporated in the IEEE 802.15.7 standard [6], which reinforces its feasibility for practical applications and its potential for large-scale adoption in specific environments. This standard introduces Color-Shift Keying (CSK) modulation, employed in this study for its capability to transmit bit streams by varying the colors of a light source.

OCC utilizes hardware from mobile devices to capture video, which acts as a receiver and provides connectivity [7]. Employing light-emitting diodes for data transmission capitalizes on pre-existing infrastructure, thereby significantly reducing deployment costs [8]. Furthermore, OCC derives considerable advantage from continuous advancements in image processing and Deep Learning (DL) technologies to enhance the accuracy and reliability of the receiver [9]. The application of Deep Learning techniques in OCC has been demonstrated to markedly reduce errors in the identification and tracking in receiver devices, even under challenging conditions such as abrupt lighting changes or the presence of obstacles [10].

Several papers in the scientific literature have proposed deep learning-based algorithms to enhance OCC [11–14]. In [11], an OCC technique for smart factory systems is introduced, which employs an LED array as the transmitter and utilizes On-Off keying modulation. Artificial intelligence is incorporated for LED detection, resulting in a significant improvement in performance compared to traditional methods. By optimizing parameters such as shutter speed, camera focal length, and appropriate channel coding, the system enables stable communication links over distances of up to 7 m. Conversely, ref. [12] proposes the design and implementation of a real-time OCC system capable of operating efficiently under high mobility conditions. For this purpose, the YOLOv8 object detection algorithm is employed, which allows for accurate identification of an LED array used as the emission source. The authors of [13] proposed a display-to-camera optical communication system that uses complementary color barcodes in conjunction with deep neural networks to achieve seamless transmission and reliable communication during normal video playback. This system employs the YOLO model to continuously detect the barcode region on electronic displays and utilizes convolutional neural networks to accurately identify pilot symbols and data embedded in the received images. Furthermore, ref. [14] reports a study on an OCC-based vehicle-to-vehicle communication system using LED arrays as transmitters and cameras as receivers. In addition, other works have addressed the design of OCC systems capable of maintaining reliable communication in dynamic conditions, such as vehicular and underwater environments. For instance, ref. [15] applied deep reinforcement learning to achieve ultra-reliable, low-latency vehicular links, ref. [16] developed a channel-adaptive decoding method for underwater OCC, and [17] employed machine learning to meet uRLLC requirements in vehicular networks.

Faced with the unresolved challenge of ensuring stability and accuracy at practical distances and dynamic conditions in OCC systems, this work proposes an approach based on the efficient use of deep learning that goes beyond detecting signal changes or black-and-white symbols by innovatively leveraging the information contained in color. The main contributions of this paper are as follows:

- The implementation and validation of a novel, high-accuracy deep learning-based receiver architecture for an OCC system. Unlike previous approaches that focus on detecting symbol transitions [18], our proposal uses color as the primary information

carrier, focusing on classification of color symbols through the YOLOv8 model, applied for the first time in this context.

- The experimental evaluation over communication distances ranging from 30 cm to 3 m, ensuring the results' applicability to real-world scenarios and confirming system robustness.
- The integration of advanced data augmentation techniques, including noise addition and overlaying real-world environmental images, to improve robustness and generalization.
- A comprehensive hyperparameter study assessing whether YOLOv8's default settings are optimal, further verifying the model's suitability for this specific application.

The remainder of this paper is structured as follows: Section 2 describes the methodology, presenting the proposed Deep Learning-based classifier, detailing the phases of data collection, data preprocessing, model selection and training, as well as validation and adjustment of hyperparameters. Section 3 presents the experiments and discusses the results obtained in the validation of the proposal. Finally, Section 4 summarizes the main conclusions and outlines potential future research directions for further extension.

2. Deep Learning-Based Classifier

In this section, the design of the system architecture is presented. The challenges inherent to symbol classification in this specific context are addressed. This involves considering aspects such as ambient illumination, channel distance, symbol variability, and algorithm robustness against possible distortions or interferences.

2.1. Data Collection

The experimental setup used to generate the dataset is shown in Figure 1. It consists of a transmitter (LED matrix and microcontroller) and a receiver (camera and processing unit), as described below.

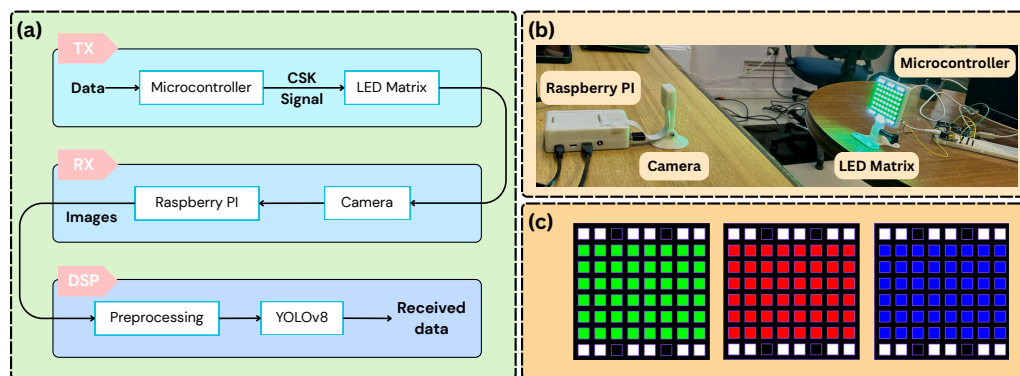


Figure 1. System architecture and experimental setup: (a) Corresponding block diagram illustrating the system components and data flow; (b) experimental setup used for data acquisition; (c) colors on the LED matrix.

An LED matrix controlled using a microcontroller is used as a transmitter. As a modulation format, we used 8-Color Shift Keying (8-CSK) with colors: yellow, blue, white, cyan, magenta, orange, red, and green. This follows the IEEE 802.15.7-2011 standard [19], which sets design rules for the 8-CSK constellation to achieve reliable performance. In 8-CSK, every symbol carries 3 bits of information. All symbols used the same spatial arrangement in the LED matrix. To reduce the blooming effect, common in OCC, each symbol includes white LEDs at the top and bottom rows (see Figure 1c). In addition, the

microcontroller applies gamma correction to adjust for the nonlinear behavior of the LEDs, using the following transformation on each RGB channel:

$$C_{out} = 255 \left(\frac{C_{in}}{255} \right)^{\gamma}, \quad (1)$$

where C_{out} and C_{in} represent the output and input intensity values of the red, green, and blue channels, respectively. These values are constrained to the range $[0, 255]$. In this work, the gamma correction factor is set to $\gamma = 3$. Each transmission sequence consisted of consecutively displaying the eight color symbols on the LED matrix at a frequency of 10 Hz, with each symbol active for a fixed time duration. In our work, eight colors for CSK modulation were chosen to balance data rate and detection accuracy. While it is possible to further expand the color alphabet, this introduces inherent challenges, particularly in distinguishing between visually similar colors, as shown in Table 1.

Table 1. RGB values of the different classes.

Class	R	G	B
Yellow	255	255	0
Blue	0	0	255
White	255	255	255
Cyan	0	255	255
Magenta	255	0	255
Orange	255	165	0
Red	255	0	0
Green	0	255	0

The receiver was a V2 camera connected to a Raspberry Pi 4. This camera integrates a Sony IMX219 CMOS image sensor [20]. The camera was configured to record video at 30 frames per second (FPS), resulting in an oversampling factor of 1.5, considering a transmitter symbol rate of 10 Hz. The captured videos contain frames where each symbol is visible under different conditions.

Five-second videos of the symbol sequences were recorded under varied experimental conditions to introduce diversity in the dataset. The parameters adjusted in each experiment were:

- Channel distance: 50 cm, 100 cm, 150 cm, 300 cm.
- Camera exposure time: 500 μ s, 1000 μ s, 4000 μ s, 6000 μ s.
- Angle between TX and RX: aligned and unaligned matrix.
- Controlled light environment: dark room and illuminated room.

Each video was processed frame by frame. Frames affected by symbol transitions or motion artifacts (such as frame 2 in Figure 2) were manually discarded, retaining only clean frames that represent each class [21]. Due to oversampling, an average of three usable images per class was extracted from each sequence. In total, 3,247 images were collected for training and validation. A separate test set was later acquired under the same parameter ranges, but at different random combinations of distance, angle, and exposure time within the laboratory. This ensured the test set included unseen samples that still followed the same distribution as the training data.



Figure 2. Frames captured from the cyan and magenta classes when oversampling.

To provide a comprehensive overview of the experimental conditions, the key system parameters used for data collection are summarized in Table 2.

Table 2. Key System Parameters for Experimental Setup.

Parameter Category	Specifics
Transmitter (TX)	
LED Matrix	8 × 8 LED array
Modulation Scheme	Color-Shift Keying (CSK) with 8 levels
Transmission Frequency	10 Hz
Receiver (RX)	
Camera Type	CMOS V2 with Sony IMX219 sensor
Frame Rate	30 fps
Channel & Environment	
Communication Distance	30 cm to 3 m
Lighting Conditions	Dark room, illuminated room
TX-RX Alignment	Aligned and unaligned

2.2. Data Preprocessing

Data pre-processing consists of four stages: data augmentation, dataset splitting, image resizing, and pixel value normalization. The data augmentation technique was applied directly to the raw data, i.e., to the data extracted from the captured videos, without having received any additional operation beyond the removal of non-useful frames. For this, the Albumentations library was used to first define the transformations and then apply them to all the collected images. In addition to these transformations, two additional components were incorporated: the addition of white Gaussian noise with different standard deviations (σ), and the superimposition of an office image on the images of the training, validation, and test sets.

Figure 3 illustrates the two transformations performed, which were applied randomly to the images. Thus, the training and validation samples were duplicated and, subsequently, one of three operations was applied: Albumentations transformations, noise addition (with mean 0 and a standard deviation randomly selected from the values (0.7, 1, 1.5, 2, 5), or office image overlay. All these operations have the same probability of being applied to images duplicated from the original data sets. The use of the overlay image was intended to increase the variability in the dataset, thereby enabling the model to learn to recognize the LED matrix in a wider variety of scenarios. This additional variability was necessary due to the exposure times used.

The samples intended for model evaluation were also duplicated; however, only the noise addition and image overlay operations were applied to them, excluding the Albumentations transformations, since these were used exclusively to improve model learning during training. The data splitting consisted of allocating 80% of the data for training and the remaining 20% for validation, from the 6494 samples obtained after data augmentation.

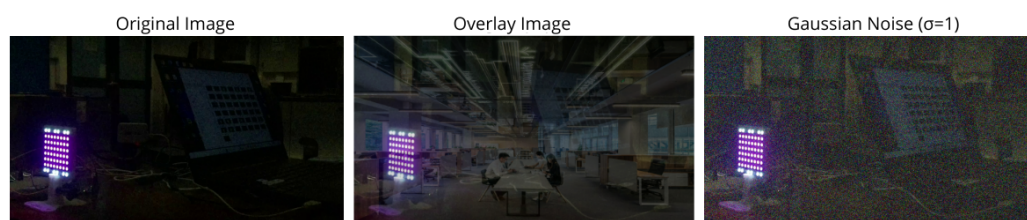


Figure 3. Transformations performed on the images in the test set.

In the data resizing stage, a hyperparameter, known as *imgsz*, is set with the help of the YOLOv8 network, which is responsible for resizing the images to a predefined size. During the training of the deep learning model, YOLOv8 allows flexibility in the input sizes by automatically performing the resizing [22]. Normalization is an essential preprocessing technique that adjusts the pixel values of the images to a standard range, thus facilitating faster convergence during training and improving model performance. Normalization is automatically and seamlessly integrated as part of the preprocessing stage in YOLOv8 during model training. This automated preprocessing ensures that the input images are prepared consistently and properly before being processed by the CNN. To evaluate the balance of the generated and split dataset, the number of samples per class in each set is obtained, as shown in Table 3.

Table 3. Number of samples per class in the datasets.

Class	Training	Validation	Test
0: Yellow	646	162	288
1: Blue	630	158	294
2: White	670	168	290
3: Cyan	657	165	290
4: Magenta	646	162	292
5: Orange	664	166	290
6: Red	649	163	292
7: Green	630	158	296
Total	5192	1302	2332

2.3. Model Selection and Training

To effectively address the challenge of classifying CSK symbols in real-world Optical Camera Communication (OCC) environments, this work employs a robust deep learning approach based on the YOLOv8 family of convolutional neural networks (CNNs), renowned for its exceptional balance between speed and accuracy. A central component of our methodology is the use of transfer learning: instead of training a model from scratch, which would require a vast amount of labeled data, we leverage a YOLOv8 model pre-trained on the large-scale ImageNet dataset. This enables the model to retain its powerful, generalized feature extraction capabilities while being fine-tuned for the specific task of CSK symbol classification. For our task, we adapted its main backbone, based on CSP-Darknet, as a feature extractor and complemented it with a classification head capable of recognizing the eight color symbol classes in our dataset. This backbone integrates optimized convolutional modules, Cross Stage Partial layers for improved efficiency, and Darknet Bottleneck residual connections. Finally, the Spatial Pyramid Pooling Fast layer was replaced with a dedicated classification layer. This final layer transforms the extracted features into output predictions. Although originally designed to support up to 1000 output classes, the model automatically adjusts its final linear layer to 8 during training, matching the number of CSK symbol classes present in the dataset.

2.4. Validation and Adjustment of Hyperparameters

A model was initially trained using the default hyperparameters of YOLOv8, with validation enabled. Subsequently, a manual tuning of the key hyperparameters such as the initial learning rate (lr0), optimizer choice (AdamW, SGD), and dropout rate was performed (see Table 4), seeking a balance between accuracy, stability, and generalizability. Two values per parameter were defined, and an exhaustive search was applied to train nine different combinations, including the model with default values. Comparing the results, it was observed that the model with adjusted hyperparameters does not represent an improvement compared with the default hyperparameters. Finally, the best model was retrained for 100 epochs to evaluate its performance robustly on the test set. Figure 4 shows the loss and the validation results during training of the models used to adjust the hyperparameters.

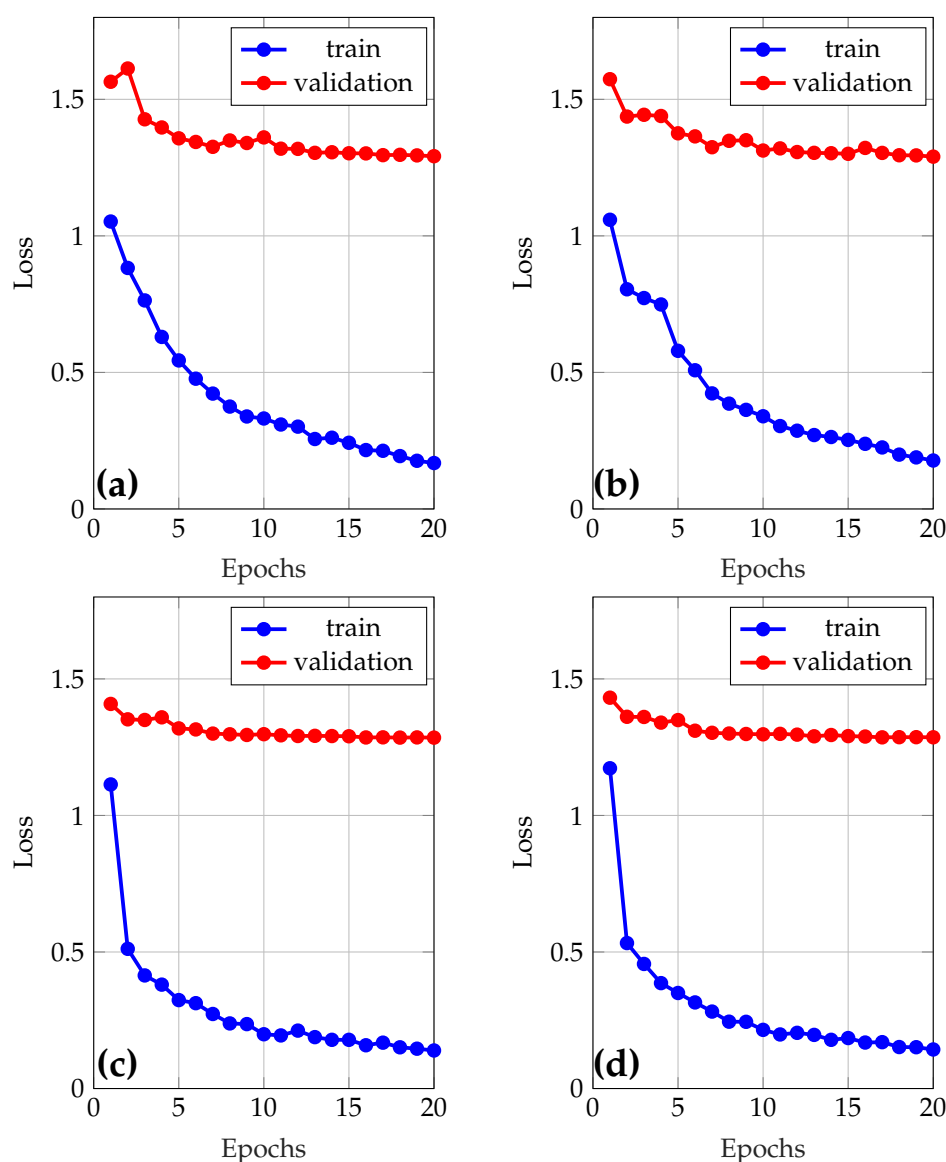


Figure 4. Learning and loss curves during model training. (a) Optimizer = AdamW, lr0 = 0.01, dropout = 0.0; (b) Optimizer = AdamW, lr0 = 0.01, dropout = 0.2; (c) Optimizer = AdamW, lr0 = 0.001, dropout = 0.0; (d) Optimizer = AdamW, lr0 = 0.001, dropout = 0.2.

Table 4. Hyperparameters to be adjusted in YOLOv8.

Hyperparameter	Default	Values	Definition
lr0	0.01	0.01, 0.001	Initial learning rate. Adjusting this value is crucial to the optimization process, influencing how quickly the model weights are updated.
optimizer	auto	AdamW, SGD	Choice of optimizer for training. Options include SGD, AdamW, etc. Affects convergence speed and stability.
dropout	0.0	0.0, 0.2	Dropout rate for regularization in classification tasks, preventing overfitting by randomly omitting units during training.

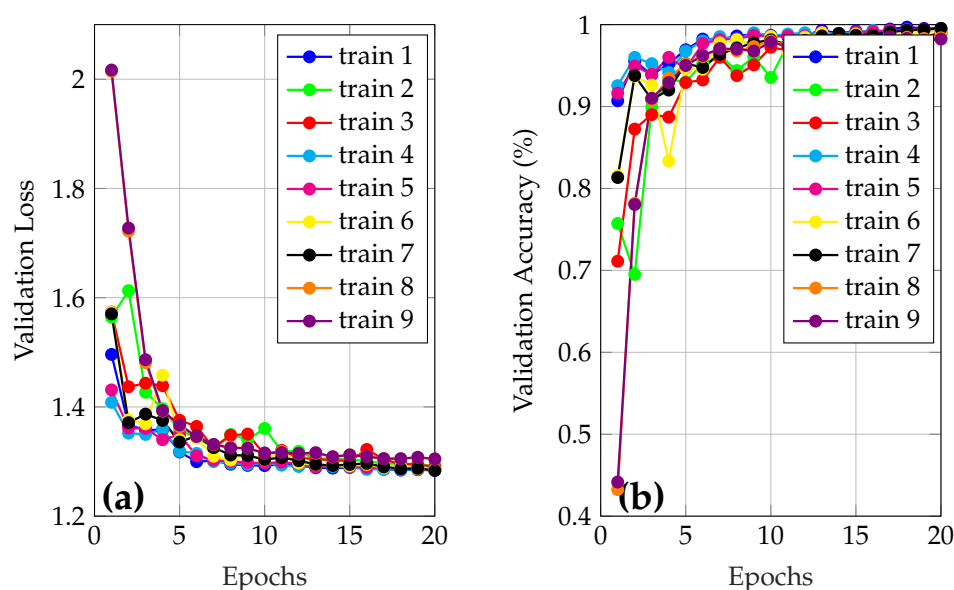
3. Results

This section details the performance of our YOLOv8-based classifier in the OCC system, beginning with a summary of the key parameters used to ensure reproducibility.

Table 5 presents the loss and accuracy results in the validation set for the various training runs performed during this process. This table facilitates the identification of the best hyperparameter settings for the classification task. In this table, the lowest loss value and the highest level of accuracy achieved during the epochs of the validation stage are presented. The validation curves (accuracy and loss vs epochs) for the reported data are shown in Figure 5.

Table 5. Validation loss and accuracy results when training with different hyperparameter configurations.

Train	Optimizer	lr0	Dropout	Loss	Accuracy (%)
1	AdamW	0.000714	0.0	1.2837	99.69
2	AdamW	0.01	0.0	1.2918	99.69
3	AdamW	0.01	0.2	1.2902	99.00
4	AdamW	0.001	0.0	1.2850	99.39
5	AdamW	0.001	0.2	1.2862	99.46
6	SGD	0.01	0.0	1.2852	99.23
7	SGD	0.01	0.2	1.2838	99.54
8	SGD	0.001	0.0	1.3046	98.46
9	SGD	0.001	0.2	1.3049	98.46

**Figure 5.** (a) Loss and (b) accuracy validation curves for trained models.

In Figure 5a, the blue curve (train 1), corresponds to the best model and presents a low loss in most epochs, starting at 1.4961 and reaching a minimum of 1.2837. The curve shows a decline until the sixth epoch, where the loss stabilizes around 1.3. On the other hand, in Figure 5b, it is observed that the accuracy starts 90.71%, and achieves 99.69% after 20 epochs. This indicates, at first glance, that the model has outstanding performance even with a reduced number of iterations.

The best model from Figure 5 was selected (train 1), and was re-trained using the best hyperparameter configuration. This process generated the training and validation loss graphs, as well as the validation accuracy, which are shown in Figures 6a and 6b, respectively. Training of the final model, minimum loss of 1.27 and maximum accuracy of 99.85% were achieved. This accuracy value will be used by YOLO as the criterion to select the best model, which will be used in the test set.

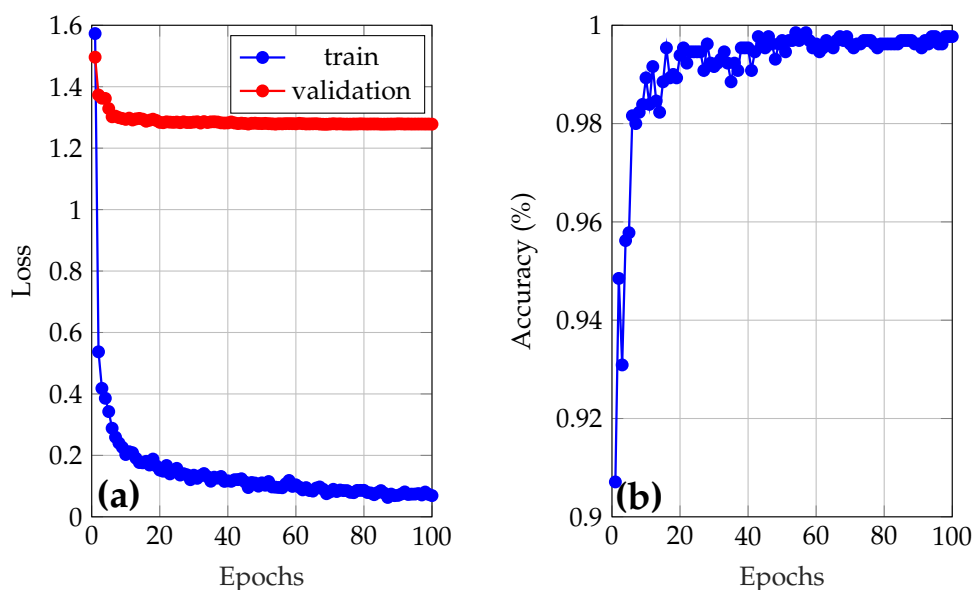


Figure 6. (a) Loss and (b) accuracy of the final model.

When analyzing the learning curves of the best model, trained for 100 epochs, in Figure 6a, a marked discrepancy between training loss and validation loss is observed. The training loss decreases rapidly, reaching very low values, while the validation loss remains significantly higher in comparison. However, it is seen that both losses could continue to decrease by further training the model using more epochs, suggesting that the model is not overfitting. Figure 6b shows a high validation accuracy throughout the epochs. The accuracy increases rapidly during the first epochs and stabilizes at values close to 98–99%. This high validation accuracy, combined with a high validation loss, could preliminarily indicate that the model makes correct predictions in most cases, but with low confidence in its classifications. That is, the model assigns lower probabilities to the correct classes, which is penalized by the loss function without affecting the overall accuracy. However, when analyzing the probabilities associated with each sample during validation, it was confirmed that this is not the case, as all samples were classified with reliabilities higher than 99%.

To evaluate the performance of the final model, a separate test set was used, different from the training and validation data. This set includes subsequently collected images covering various experimental conditions such as channel distances of 60 cm, 70 cm, 110 cm, and 250 cm, as well as different combinations of exposure time, environment variations, illumination levels, and transmitter positions. Additionally, these images were

also processed to duplicate their quantity and. Noise and a different overlay image were added to the duplicated set.

Predictions were made using the final model, with the results summarized in the confusion matrix shown in Figure 7. From the confusion matrix, it is observed that most predictions are concentrated on the main diagonal, indicating correct classification in the vast majority of cases. Classes such as green, blue, and cyan show near-perfect accuracy, with no errors or only one incorrect prediction. However, some specific confusions are observed, especially in the white class, which was misclassified as yellow in eight cases and as cyan in ten, suggesting some difficulty for the model in distinguishing between light spectrum colors or colors possibly influenced by similar lighting conditions. Minor errors are also recorded in other classes, such as yellow and magenta, although without significantly affecting the overall performance of the system.

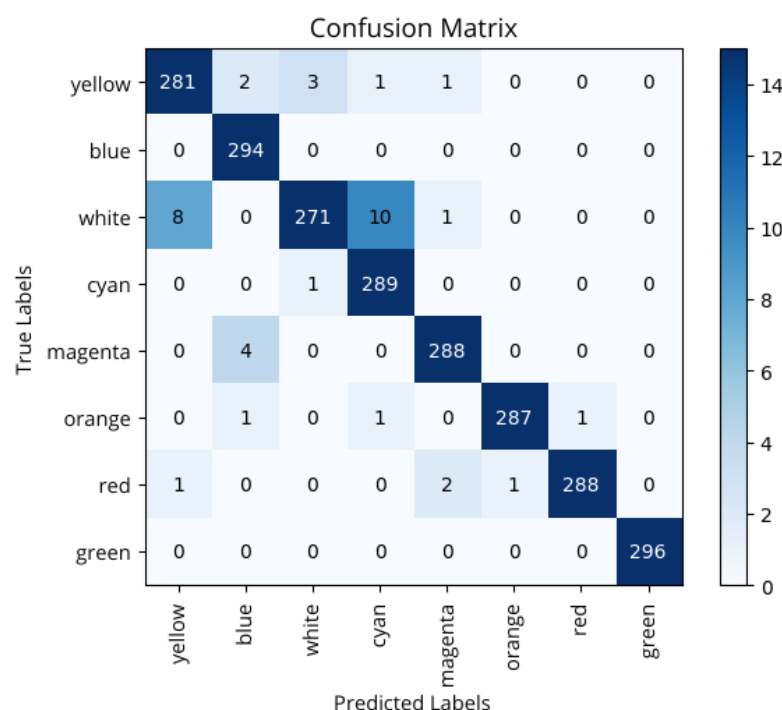


Figure 7. Confounding matrix before adding variability to the test set.

It can be concluded that very good results are obtained, in general; however, some exceptions exist, especially among classes with visually similar samples, such as white and yellow, cyan and white, or blue and magenta. This is primarily due to the phenomenon of channel crosstalk, which causes interference between color channels as a result of spectral overlap in the LED array and the camera's limitations to distinguish close tones. Additionally, factors like ambient lighting and noise in the capture system contribute to the camera registering color mixtures rather than pure tones, leading to confusion between these classes when interpreted by the YOLOv8 model.

The samples in each class were correctly classified with high accuracy and, overall, the model demonstrated a high generalization ability with an accuracy of 98.4% in the test set. The confusion matrix, therefore, not only confirms the strong overall performance of the model but also provides a nuanced understanding of its limitations. The strong concentration of correct predictions along the main diagonal, notably for classes like blue and green (with 294 and 296 correct predictions, respectively), demonstrates effective generalization under real testing conditions. However, the confusion between white, cyan, and yellow highlights the inherent physical constraints of the optical camera communication

system, such as channel crosstalk and sensor saturation, when processing colors requiring high intensity across multiple RGB channels. Minor misclassifications, like magenta being confused with blue, further reinforce these.

The strong overall accuracy and robust performance of our model against varying distances and lighting conditions hold significant implications for its practical deployment. A symbol recognition accuracy of approximately 98.4% translates directly into an extremely low data transmission error rate, which is a critical requirement for safety-sensitive applications where reliability is paramount. In vehicle-to-vehicle (V2V) communication, for example, our system could be used to reliably classify emergency signals, such as the flashing lights of an ambulance or police car, or the red light of a traffic signal, providing a robust communication channel in complex and dynamic environments. Similarly, in industrial IoT scenarios, this high reliability is fundamental for the correct execution of machine-to-machine commands, preventing costly operational errors. Therefore, our results not only confirm the model's technical efficacy but also validate its potential as a low-cost, high-reliability solution for real-world applications where low-data-rate, high-precision signaling is required.

In addition, system robustness in real-world applications could be further improved through specific design improvements. For instance, the use of light diffusers in vehicular and other outdoor scenarios would mitigate channel crosstalk, allowing color separation under conditions such as direct sunlight or car headlight interference. Similarly, adaptive gain control mechanisms in the camera receiver would allow automatic adjustment to varying illumination, thereby ensuring stable detection in dynamic outdoor environments. At the communication layer, advanced error correction coding could compensate for residual symbol misclassifications, while temporal and spatial filtering would stabilize detection in settings with strong motion or background variability. These enhancements, although not implemented in this work, are challenges observed in V2V and industrial IoT scenarios.

4. Conclusions

This paper evaluated the performance of the YOLOv8 model in symbol classification within optical camera communication environments. The results consistently demonstrated strong performance in all scenarios analyzed, achieving an accuracy of 98.4% on the test set, indicating the effective generalizability of the model. Incorporating noise transformations and varied data augmentation techniques applied to training and validation images was essential for developing a robust and resilient model. These strategies substantially improved generalizability, allowing the model to reach values above 93% across all evaluation metrics for each symbol class, even when the test set included noise and previously unseen samples.

From the outset, the model was designed and trained to handle diverse and complex conditions by simulating realistic environments during the training phase, which is essential for its successful deployment and practical application. Misclassified samples were primarily caused by channel crosstalk; specifically, the RGB color information captured by the image sensor is affected by overlapping signals from the red, green, and blue channels. This interference causes certain colors in the LED matrix, which should ideally appear as distinct hues (e.g., white), to be perceived as different colors, such as yellow or cyan. Such distortions highlight the inherent challenges of color-based modulation schemes in OCC systems.

For future research, two main directions are recommended. First, explore the implementation of the YOLOv8 model for real-time detection and sensing in OCC applications, which could offer valuable tools for automated monitoring and the management of low-latency optical communication systems. Second, developing a more diverse and comprehensive dataset is essential. This dataset should encompass variations such as

different background textures, a wider range of channel distances, diverse ambient lighting conditions, and various indoor/outdoor scenarios, which will allow the model to generalize across a broader spectrum of real-world environments.

Author Contributions: This manuscript was prepared by several authors, with each author's specific contributions outlined below. Conceptualization, F.V.V.V., L.B.A. and G.S.; methodology, F.V.V.V., L.M., F.P. and V.M.I.; software, F.V.V.V., L.M., F.P. and L.R.-L.; validation, L.B.A., S.M.-S., V.M.I. and G.S.; formal analysis, L.B.A., F.V.V.V., and V.M.I.; investigation, F.V.V.V., L.M., F.P., L.B.A. and L.R.-L.; resources, S.M.-S., L.R.-L. and G.S.; data curation, F.V.V.V., L.M. and F.P.; writing—original draft preparation, L.B.A.; writing—review and editing, F.V.V.V., S.M.-S., L.R.-L. and G.S.; supervision, L.B.A., S.M.-S., V.M.I. and G.S.; project administration, S.M.-S., L.R.-L. and G.S.; funding acquisition, S.M.-S., L.R.-L. and G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by ANID FONDECYT Regular 1231826 and ANID FONDECYT Regular 1241977.

Acknowledgments: Gabriel Saavedra thanks Instituto Milenio de Investigación en óptica ICN17-012 and ANID FONDECYT Regular 1231826, Samuel Montejó-Sánchez thanks ANID FONDECYT Regular 1241977.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Haas, H.; Yin, L.; Wang, Y.; Chen, C. What is LiFi? *J. Light. Technol.* **2015**, *34*, 1533–1544. [\[CrossRef\]](#)
2. Ahmed, T.; Liu, H.; Gayah, V.V. OCC-MP: A Max-Pressure framework to prioritize transit and high occupancy vehicles. *Transp. Res. Part C Emerg. Technol.* **2024**, *166*, 104795. [\[CrossRef\]](#)
3. Mohsan, S.A.H.; Amjad, H. A comprehensive survey on hybrid wireless networks: Practical considerations, challenges, applications and research directions. *Opt. Quantum Electron.* **2021**, *53*, 523. [\[CrossRef\]](#)
4. Bravo Alvarez, L.; Montejó-Sánchez, S.; Rodríguez-López, L.; Azurdia-Meza, C.; Saavedra, G. A Review of Hybrid VLC/RF Networks: Features, Applications, and Future Directions. *Sensors* **2023**, *23*, 7545. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Lisandra, B.A.; Samuel, M.S.; Lien, R.L.; José, N.K.; David, R.G.; Gabriel, S. Enhanced Network Selection Algorithms for IoT-Home Environments With Hybrid VLC/RF Systems. *IEEE Access* **2024**, *12*, 108942–108952. [\[CrossRef\]](#)
6. *IEEE Std 802.16-2004*; IEEE Standard for Local and Metropolitan Area Networks-Part 16: Air Interface for Fixed Broad-Band Wireless Access Systems. IEEE: Piscataway, NJ, USA, 2004.
7. Pan, J.; Wang, Z.; Wang, L. Co-Occ: Coupling Explicit Feature Fusion with Volume Rendering Regularization for Multi-Modal 3D Semantic Occupancy Prediction. *IEEE Robot. Autom. Lett.* **2024**, *9*, 5687–5694. [\[CrossRef\]](#)
8. Dao, N.N.; Do, T.H.; Cho, S.; Dustdar, S. Information Revealed by Vision: A Review on the Next-Generation OCC Standard for AIoV. *IT Prof.* **2022**, *24*, 58–65. [\[CrossRef\]](#)
9. Nguyen, H.; Utama, I.B.K.Y.; Jang, Y.M. Enabling Technologies and New Challenges in IEEE 802.15.7 Optical Camera Communications Standard. *IEEE Commun. Mag.* **2023**, *62*, 90–95. [\[CrossRef\]](#)
10. Lyu, X.; Dai, P.; Li, Z.; Yan, D.; Lin, Y.; Peng, Y.; Qi, X. Learning a Room with the Occ-SDF Hybrid: Signed Distance Function Mingled with Occupancy Aids Scene Representation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023; pp. 8940–8950.
11. Nguyen, H.; Jang, Y.M. An Experimental Demonstration of 2D-MIMO Based Deep Learning for OCC System. In Proceedings of the 2024 Fifteenth International Conference on Ubiquitous and Future Networks (ICUFN), Budapest, Hungary, 2–5 July 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 165–168.
12. Sitanggang, O.S.; Nguyen, V.L.; Nguyen, H.; Pamungkas, R.F.; Faridh, M.M.; Jang, Y.M. Design and implementation of a 2D MIMO OCC system based on deep learning. *Sensors* **2023**, *23*, 7637. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Kim, M.T.; Kim, B.W. DeepCCB-OCC: Deep Learning-Driven Complementary Color Barcode-Based Optical Camera Communications. *Appl. Sci.* **2022**, *12*, 11239. [\[CrossRef\]](#)
14. Sun, X.; Shi, W.; Cheng, Q.; Liu, W.; Wang, Z.; Zhang, J. An LED Detection and Recognition Method Based on Deep Learning in Vehicle Optical Camera Communication. *IEEE Access* **2021**, *9*, 80897–80905. [\[CrossRef\]](#)
15. Islam, A.; Thomos, N.; Musavian, L. Achieving uRLLC with Machine Learning Based Vehicular OCC. In Proceedings of the GLOBECOM 2022—2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 4–8 December 2022.

16. Aberathna, T.; Kumarage, M.; Atthanayake, N. Channel Adaptive Decoding in Underwater Optical Camera Communication Systems. In Proceedings of the 2025 5th International Conference on Advanced Research in Computing (ICARC), Belihuloya, Sri Lanka, 19–20 February 2025.
17. Islam, A.; Thomos, N.; Musavian, L. Deep Reinforcement Learning-Based Ultra Reliable and Low Latency Vehicular OCC. *IEEE Trans. Commun.* **2025**, *73*, 3254–3267. [[CrossRef](#)]
18. Fernandes, D.; Matus, V.; Figueiredo, M.; Alves, L.N. Asynchronous encoding scheme for optical camera communication system using two-dimensional transmitter. In Proceedings of the 2023 South American Conference On Visible Light Communications (SACVLC), Santiago, Chile, 8–10 November 2023.
19. *IEEE Std 802.15.7-2011*; IEEE Standard for Local and Metropolitan Area Networks—Part 15.7: Short-Range Wireless Optical Communication Using Visible Light. IEEE: Piscataway, NJ, USA, 2011.
20. Matus, V.; Teli, S.R.; Guerra, V.; Jurado-Verdu, C.; Zvanovec, S.; Perez-Jimenez, R. Evaluation of Fog Effects on Optical Camera Communications Link. In Proceedings of the 2020 3rd West Asian Symposium on Optical and Millimeter-wave Wireless Communication (WASOWC), Tehran, Iran, 24–25 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–5.
21. Onodera, Y.; Hisano, D.; Maruta, K.; Nakayama, Y. First Demonstration of 512-Color Shift Keying Signal Demodulation Using Neural Equalization for Optical Camera Communication. In Proceedings of the Optical Fiber Communication Conference, San Diego, CA, USA, 5–9 March 2023; Optica Publishing Group: Washington, DC, USA, 2023; p. Th3H-7.
22. Terven, J.; Córdova-Esparza, D.M.; Romero-González, J.A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.