An Analysis of Data-Centric Artificial Intelligence in Computer Vision Applications

Daniel Canedo
danielduartecanedo@ua.pt

Petia Georgieva
petia@ua.pt

António Neves
an@ua.pt

IEETA/DETI
Universidade de Aveiro
Aveiro, 3810-193, Portugal

Abstract

Deep learning is witnessing rapid advancements, majorly impacting the computer vision field. However, as the complexity of these algorithms increases, their demand for data grows exponentially. As a result, there is an increasing emphasis on data-centric artificial intelligence in deep learning. In computer vision, data is primarily comprised of images and videos, forming datasets that are crucial inputs for deep learning algorithms during the learning process. However, these datasets can often be limited in size, biased, inadequate, and lacking proper labeling, particularly in the domain of computer vision where data collection, storage, labeling, and processing require substantial infrastructure and human resources. Consequently, researching how to tackle data collection, data quality, data generation, and data processing is of utmost importance in this field. This work explores an application of data-centric artificial intelligence in three distinct domains within computer vision: facial expression recognition, dirt detection in the context of intelligent robotics, and archaeological site detection. Given the distinct nature of the data involved with these applications, the objective of this work is to provide an analysis on how to conduct data management depending on the computer vision application.

1 Introduction

With the fast advancement of computational specifications, there has been a corresponding development of larger and more efficient deep learning algorithms. Nevertheless, it is important to note that as these algorithms increase in size, their demand for data also escalates. For instance, in the study conducted by Alwosheel et al. (2018) [1], the authors recommend a ratio of fifty training samples for each adjustable parameter in the network, and nowadays artificial intelligence (AI) networks can encompass millions of adjustable parameters. Consequently, deep learning is shifting towards data-centric AI approaches recently. This type of approach focuses its time and effort on meticulously preparing the most optimal data to be fed by deep learning algorithms. This shift is aligned with the ongoing trend of streamlining the interaction with AI [2].

Currently, a significant portion of researchers engage with deep learning at a higher level, as the requirement for writing extensive and complex code to train and deploy an AI model continues to diminish. As the barrier of entry continues to decrease, even individuals with limited experience can readily engage with this technology. The primary challenges that they may encounter lie in data preparation and hyperparameter fine-tuning. This work addresses those challenges for three distinct computer vision applications: facial expression recognition (FER), dirt detection, and archaeological site detection.

2 Facial Expression Recognition

For this computer vision application, the impact of data pre-processing algorithms on deep learning for FER was studied [3, 4]. A simple Convolutional Neural Network (CNN) was proposed to conduct this study. Figure 1 illustrates this CNN.

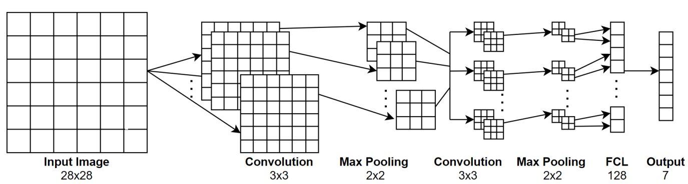

Figure 1: Proposed CNN architecture for FER.

Several pre-processing algorithms were applied in an incremental order to the Cohn-Kanade Dataset (CK+) [5], starting from face rotation correction, followed by face cropping, intensity normalization, histogram equalization, and ending with image smoothing. Each pre-processing step created a dataset, which was then used to train the CNN to observe the impact each pre-processing step has on the learning process. In the end, the attention maps from the last layer were obtained for the testing set using the CNN trained with raw images and the CNN trained with fully pre-processed images. The latter learned facial features better, as can be seen in Figure 2, which was reflected in a better prediction accuracy: 93.90% versus 71.22%.

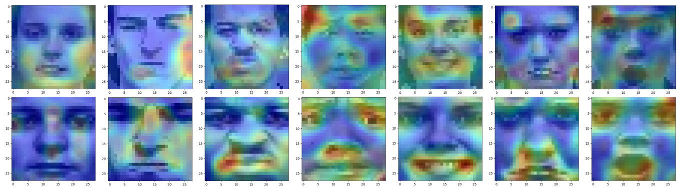

Figure 2: Average attention maps. The first row represents the CNN trained with raw images and the second row represents the CNN trained with pre-processed images. From left to right: neutral, anger, disgust, fear, happiness, sadness, and surprise.

3 Dirt Detection

For this computer vision application, a vision system for a floor-cleaning robot was developed [6, 7], enabling it to optimize their navigation and analyzing the surrounding floor, leading to a reduction on power, water and chemical products' consumption. The main contributions of this work are in the data collection, generation, and annotation. A self-calibration algorithm was implemented to stabilize image intensity acquired by the digital cameras and improve the robustness of the vision system. A dataset was artificially generated, and publicly available datasets were annotated [8] with three classes: solid dirt, liquid dirt, and scuff marks. Data augmentation was performed with two goals in mind: increase the data and balance the classes. Finally, everything was combined into a dataset that was used to train a YOLOv5 model, which is an object detection algorithm, for dirt detection. Figure 3 illustrates this process.

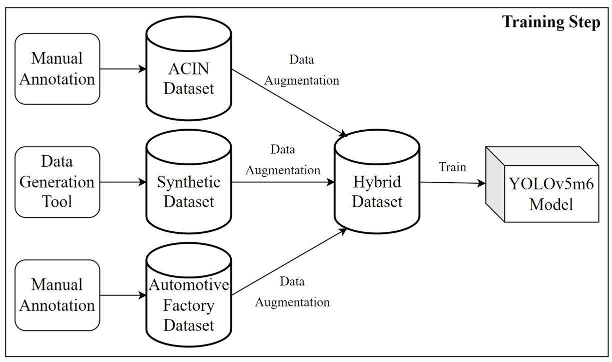

Figure 3: Vision system pipeline of the floor-cleaning robot.

When operating, the robot parameters, such as water, detergent, and speed, are adjusted based on the dirty area outputted by the vision system. Figure 4 illustrates this process.

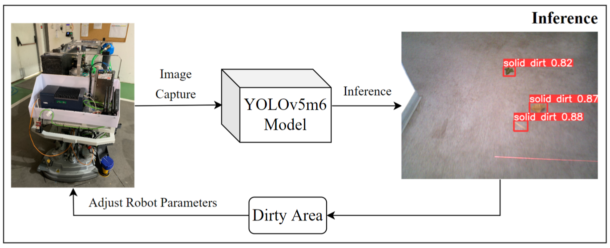

Figure 4: Vision system pipeline of the floor-cleaning robot.

Combining different types of data, artificially increasing floor variety, and understanding that scuff marks are a relevant type of dirt despite being overlooked in the literature, were crucial data-centric practices to achieve better performance on dirt detection.

4 Archaeological Site Detection

In this computer vision application, the goal was to uncover unknown burial mounds in Alto-Minho [9] in the context of the Odyssey project. The Comunidade Intermunicipal do Alto Minho (CIM Alto Minho) provided us with the airborne Light Detection and Ranging (LiDAR) data from 2018 (2 points per m²) covering this region (2220 km²). A visualization technique was applied to the 1-meter LiDAR-derived Digital Terrain Models (DTMs), namely the Local Relief Model (LRM). Software capable of dealing with such high-resolution images was developed, automatically dividing them into smaller images around annotated burial mounds, creating a conventional object detection dataset. Figure 5 illustrates this process.

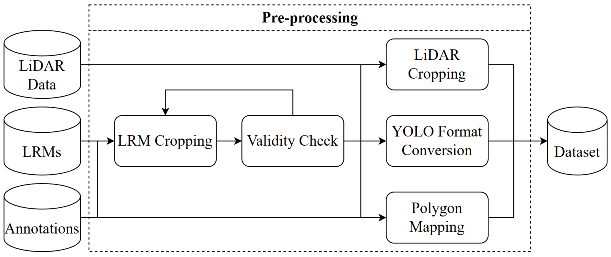

Figure 5: Data pre-processing pipeline.

The dataset was augmented with a copy-paste object embedding technique, relying on the Land-Use and Occupation Charter (LBR) of Portugal, 2018, to paste burial mounds onto probable regions. Figure 6 illustrates this process.

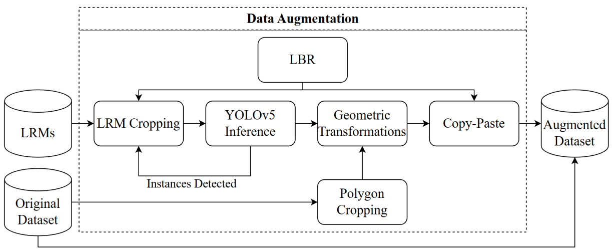

Figure 6: Data augmentation pipeline.

A YOLOv5 model was trained using the final dataset, and subsequently, used to infer new burial mounds in Alto-Minho. A post-processing validation step was applied to the inferences, using the LBR to discard inferences in less probable regions, and using the raw LiDAR data to train a Local Outlier Factor (LOF) model which objective is to discard inferences that have a different 3D morphology than the known burial mounds. Figure 7 illustrates the inference and post-processing validation process. The results underwent digital validation conducted by four archaeologists, revealing a precision of 72.53%. Figure 8 visually depicts several newly discovered burial mounds in the Alto-Minho region.

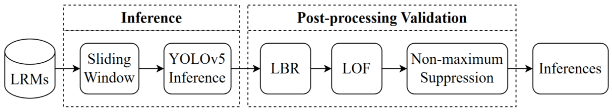

Figure 7: Inference and post-processing validation pipeline.

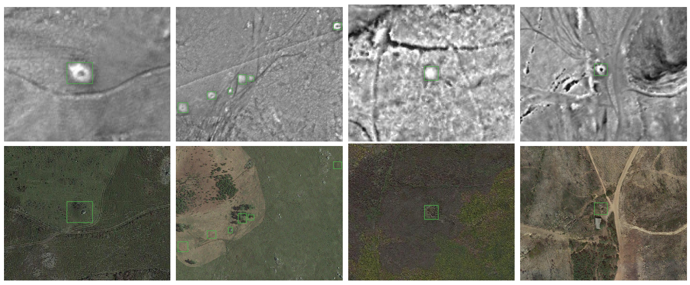

Figure 8: Validated inferences. On the first row, the LRM images. On the second row, the corresponding Google Satellite images.

5 Conclusion

With the recent increase in accessibility and usability of AI, the task of managing data to train and deploy AI models becomes one of the focal points to improve model efficiency. In this work, three distinct computer vision applications are addressed with data-centric AI. Each computer vision has its own challenges regarding data quality and availability. Diagnosing and tackling those challenges to reach high quality datasets has a substantial impact on AI performance, as illustrated in this work.

References

[1] Ahmad Alwosheel, Sander van Cranenburgh, and Caspar G Chorus. Is your dataset big enough? sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of choice modelling*, 28:167–182, 2018.

[2] Ekaba Bisong and Ekaba Bisong. Google colaboratory. *Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners*, pages 59–64, 2019.

[3] Daniel Canedo and António JR Neves. The impact of pre-processing algorithms in facial expression recognition. In *Thirteenth International Conference on Machine Vision*, volume 11605, pages 495–502. SPIE, 2021.

[4] Daniel Canedo and António JR Neves. Facial expression recognition using computer vision: A systematic review. *Applied Sciences*, 9(21):4678, 2019.

[5] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 ieee computer society conference on computer vision and pattern recognition-workshops*, pages 94–101. IEEE, 2010.

[6] Daniel Canedo, Pedro Fonseca, Petia Georgieva, and António JR Neves. A deep learning-based dirt detection computer vision system for floor-cleaning robots with improved data collection. *Technologies*, 9(4):94, 2021.

[7] Daniel Canedo, Pedro Fonseca, Petia Georgieva, and António J Neves. An innovative vision system for floor-cleaning robots based on yolov5. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 378–389. Springer, 2022.

[8] Andreas Grünauer, Georg Halmetschlager-Funek, Johann Prankl, and Markus Vincze. The power of gmms: Unsupervised dirt spot detection for industrial floor cleaning robots. In *Towards Autonomous Robotic Systems: 18th Annual Conference, TAROS 2017, Guildford, UK, July 19–21, 2017, Proceedings 18*, pages 436–449. Springer, 2017.

[9] Daniel Canedo, João Fonte, Luis Gonçalves Seco, Marta Vázquez, Rita Dias, Tiago Do Pereiro, João Hipólito, Fernando Menéndez-Marsh, Petia Georgieva, and António JR Neves. Uncovering archaeological sites in airborne lidar data with data-centric artificial intelligence. *IEEE Access*, 2023.