AIMH Lab for the Industry

Fabio Carrara, Luca Ciampi, Marco Di Benedetto, Fabrizio Falchi, Claudio Gennaro, Fabio Valerio Massoli, Giuseppe Amato

Artificial Intelligence for Media and Humanities laboratory
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", CNR
<name.surname>@isti.cnr.it

Abstract

In this short paper, we report the activities of the Artificial Intelligence for Media and Humanities (AIMH) laboratory of the ISTI-CNR related to Industry. The massive digitalization affecting all the stages of product design, production, and control calls for data-driven algorithms helping in the coordination of humans, machines, and digital resources in Industry 4.0. In this context, we developed AI-based Computer-Vision technologies of general interest in the emergent digital paradigm of the fourth industrial revolution, focusing on anomaly detection and object counting for computer-assisted testing and quality control. Moreover, in the automotive sector, we explore the use of virtual worlds to develop AI systems in otherwise practically unfeasible scenarios, showing an application for accident avoidance in self-driving car AI agents.

1 Introduction

In the emergent Industry 4.0 paradigm, Artificial Intelligence (AI) is destined to play an essential role in several aspects of new highly-digital industrial strategies, such as production processes, quality control, logistics, and sustainability. The massive digitalization affecting all the stages of product design, production, and control calls for data-driven algorithms helping in the coordination of humans, machines, and digital resources in the new industry.

In this context, the AIMH laboratory employs its vision-based data-driven AI systems expertise to research and develop general-interest technologies for industrial applications. Our research activities focused on improving computer-assisted quality control methods from visual data (images and videos), proposing novel automatic object counting and anomaly detection methodologies. Moreover, in the context of the automotive industry, our activity focused on the adoption of virtual worlds to accelerate the development of next-generation self-driving cars; the use of synthetic scenarios permits us to train and evaluate novel data-driven algorithms for accident detection and avoidance that would be practically unfeasible to develop in real-world scenarios. The following section describes in detail the activities conducted by our laboratory in these contexts.

2 Research and Applications

2.1 Anomaly Detection

Anomalies are ubiquitous in industrial manufacturing and can express an unexpected event due to incomplete knowledge about the data distribution or an unknown process that suddenly comes into play and distorts the observations. Usually, due to such events' rarity, to train deep learning (DL) models on the anomaly detection (AD) task, practitioners only rely on "normal" data, i.e., non-anomalous samples, and let learning models infer the distribution beneath the input data.

In [Carrara *et al.*, 2021], we propose CBiGAN — a novel reconstruction-based method for anomaly detection in images. In reconstruction-based methods, the distribution of non-anomalous images is learned using a generative model. Anomalous samples are detected if they do not fit the trained model, i.e., the model cannot correctly reconstruct the anomalous part of the image. Our proposal introduces a consistency constraint as a regularization term in both the encoder and decoder of a Bidirectional GAN (BiGAN) generative model. The proposed model exhibits fairly good modeling power and reconstruction consistency capability (see Figure 1). We evaluate the proposed method on MVTec AD — a real-world benchmark for unsupervised anomaly detection on high-resolution images — and compare it against standard baselines and state-of-the-art approaches. Experiments show that the proposed method improves the performance of BiGAN formulations by a large margin and performs comparably to expensive state-of-the-art iterative methods while reducing the computational cost.

In [Massoli *et al.*, 2021], we instead propose a novel framework, named multilayer one-class classification (MOCCA), to train and test DL models on the AD task. Specifically, we applied our approach to autoencoders. A key novelty in our work stems from the explicit optimization of the intermediate representations for the task at hand. Indeed, differently from commonly used approaches that consider a neural network as a single computational block, i.e., using the output of the last layer only, MOCCA explicitly leverages the multilayer structure of deep architectures. Each layer's feature space is optimized for AD during training, while in the

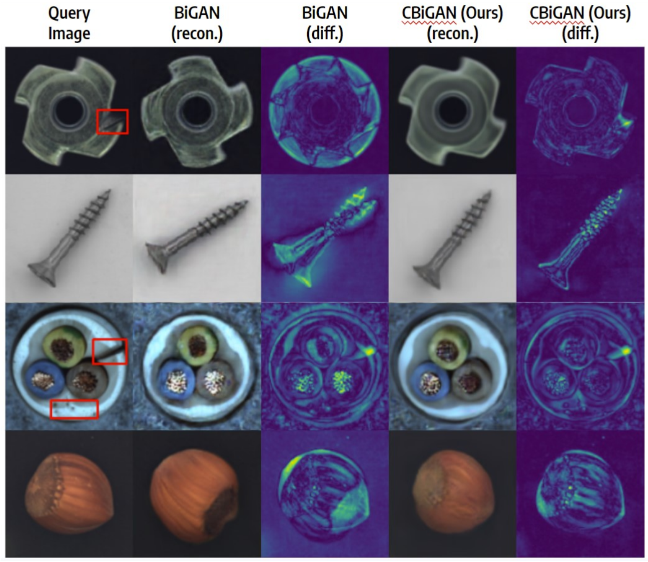

Figure 1: **Visual anomaly detection for quality-control in industrial manufacturing.** We show our reconstruction-based anomaly detection model (CBiGAN) vs previous GAN-based state of the art BiGAN. Columns report input images (1st col., red boxes indicate anomalies), reconstructions (2nd & 4th col.), and per-pixel absolute differences (3rd & 5th col.). Our improved reconstruction ability achieves a lower false positive detection rate and a more precise anomaly localization. Image Courtesy of [Carrara *et al.*, 2021].

test phase, the deep representations extracted from the trained layers are combined to detect anomalies. With MOCCA, we split the training process into two steps. First, the autoencoder is trained on the reconstruction task only. Then, we only retain the encoder tasked with minimizing the L2 distance between the output representation and a reference point, the anomaly-free training data centroid, at each considered layer. Subsequently, we combine the deep features extracted at the various trained layers of the encoder model to detect anomalies at inference time. To assess the performance of the models trained with MOCCA, we conduct extensive experiments on publicly available datasets, namely CIFAR10, MVTec AD, and ShanghaiTech. We show that our proposed

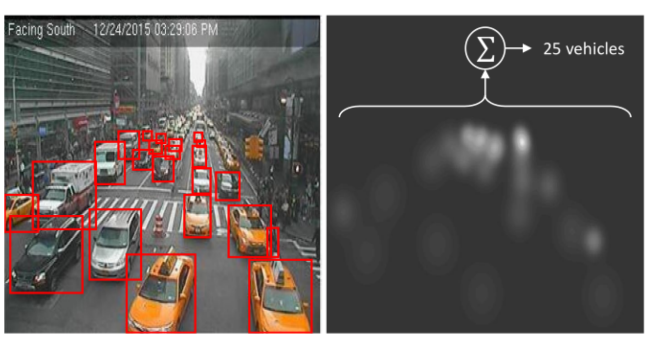

Figure 2: **Visual Counting in Dense Scenarios.** An example of counting by density estimation. The predicted density map (left) is integrated to obtain an estimation of the number of objects present in the scene. Image Courtesy of [Ciampi *et al.*, 2021b].

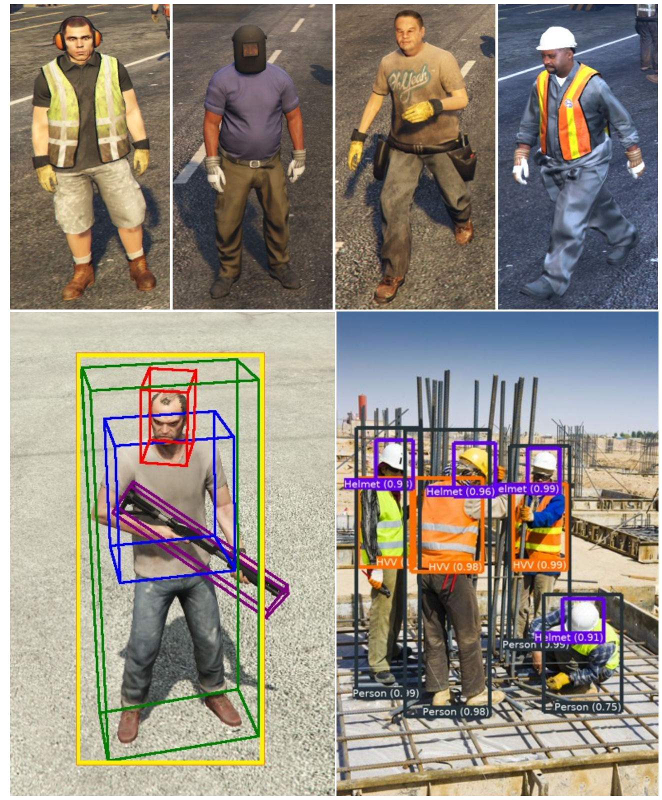

Figure 3: **Learning PPE detection in virtual worlds for ensuring compliance with safety and security rules**: *Top:* examples of safety equipment objects depicted by the virtual engine and detected by our system; virtual data comprise most of the training set of the detection neural network. *Bottom Left:* the 3D and 2D bounding boxes provided by the virtual engine used to automatically obtain annotated data. *Bottom Right:* examples of object detected by our system on a real-world image. Image Courtesy of [Di Benedetto *et al.*, 2021].

method reaches comparable or superior performance to state-of-the-art approaches available in the literature.

2.2 Visual Counting

The counting task estimates the number of objects instances, like objects, people, or vehicles, in still images or video frames. Its interdisciplinary and widespread applicability also drew the attention of industrial manufacturing processes (specifically quality-control and testing stages) that can benefit from this technology. Current solutions are formulated as supervised deep learning-based problems belonging to one of two main categories: counting by *detection* and counting by *regression*. Detection-based approaches require prior detection of single instances of objects. On the other hand, regression-based techniques establish a direct mapping between image features and the number of objects in the scene, either directly or via the estimation of a density map that is then integrated to obtain the number of the objects. We provide an example of counting by density estimation in Fig. 2. Regression techniques show superior performance in crowded and highly-occluded scenarios but often lose the

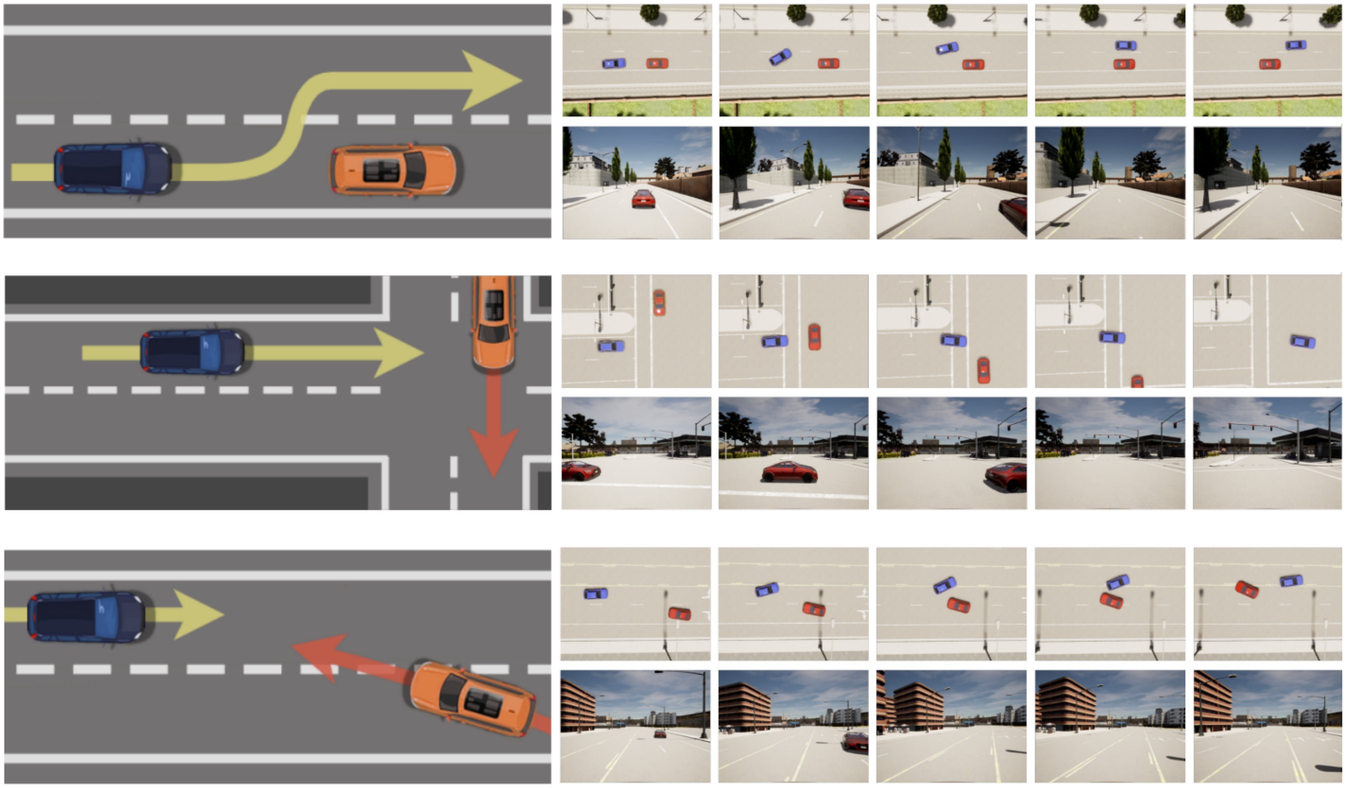

Figure 4: **Accident avoidance agent trained in a virtual world**: *Left:* Imminent accident scenarios used to evaluate the system, the agent is represented by the blue car. In the top scenario, the agent must avoid the preceding vehicle that suddenly brakes. The middle scenario involves the prevention of a crash at an intersection. In the bottom scenario, a vehicle loses control and invades the agent's lane, requiring a quick avoidance maneuver. *Right:* Frame sequences illustrating the agent reaction. We show the bird's eye view (above) and agent's camera (below), with time increasing left to right for each sequence. Image Courtesy of [Merola *et al.*, 2022].

ability to locate objects precisely.

The AIMH laboratory proposes AI-based visual counting solutions belonging to both categories. In [Amato *et al.*, 2018] and [Ciampi *et al.*, 2021a], we introduce a system able to detect and count the vehicles present in parking lots. The AI algorithm analyzes images gathered from smart cameras, directly running on them. In another work [Amato *et al.*, 2019a], we instead propose an AI-based application responsible for localizing and counting cars in images taken from drones. We also present a detect-and-count AI algorithm able to localize and count pedestrians present in images coming from video surveillance systems [Ciampi *et al.*, 2020a] [Amato *et al.*, 2019b]. The main peculiarity of this approach is that it exploits synthetic data coming from a video-game engine for the training phase. On the other hand, concerning the counting by regression techniques, we propose an AI-based solution able to estimate traffic density and counting vehicles present in urban scenarios [Ciampi *et al.*, 2020b; Ciampi *et al.*, 2021b]. This approach exploits an *Unsupervised Domain Adaptation* strategy in charge of adapting the algorithm to new scenarios never seen during the training phase.

2.3 PPE detection

Deep learning has achieved impressive results in many machine learning tasks such as image recognition and computer vision. However, its applicability to supervised problems is constrained by the availability of high-quality training data consisting of large numbers of humans annotated examples (e.g., millions). To overcome this problem, the AI world is currently increasingly exploiting artificially generated images or video sequences using realistic photo rendering engines such as those used in entertainment applications. In this way, we can easily create large sets of training images to train deep learning algorithms.

In [Di Benedetto *et al.*, 2021; Di Benedetto *et al.*, 2019], we generated photo-realistic synthetic image sets to train deep learning models to recognize the correct use of personal protection equipment (PPE, e.g., worker safety helmets, high visibility vests, ear protection devices) during at-risk work activities (see Figure 3). Then, we performed the adaptation of the domain to real-world images using a minimal set of real-world images. We demonstrated that training with generated synthetic sets and applying domain adaptation is an effective solution for applications with no available training set.

2.4 Self-Driving Cars

Self-driving systems have recently received massive attention in both academic and industrial contexts, leading to major improvements in standard navigation scenarios typically identified as well-maintained urban routes.

Critical events like road accidents or unexpected obstacles, however, require the execution of specific emergency actions that deviate from the ordinary driving behavior and are therefore harder to incorporate in the system. In [Merola *et al.*, 2022], we propose a system that is specifically built to take control of the vehicle and perform an emergency maneuver in case of a dangerous scenario. The presented architecture is based on a deep reinforcement learning algorithm, trained in a simulated environment and using raw sensory data as input. We evaluate the system's performance on several typical pre-accident scenario (see Figure 4) and show promising results, with the vehicle being able to consistently perform an avoidance maneuver to nullify or minimize the incoming damage.

3 Projects

AI4Media

A Centre of Excellence delivering next generation AI Research and Training at the service of Media, Society and Democracy.

NAUSICAA

The NAUtical Safety by means of Integrated Computer-Assisted Appliances 4.0 (NAUSICAA) project aims at creating a system for medium and large boats in which conventional control, propulsion, and thrust systems are integrated with a series of latest generation sensors, such as lidar systems, cameras, radar, marine drones, and aircraft, in order to provide complementary assistance during navigation and mooring. The idea is to develop a solution aimed at increasing safety in the nautical sector through the development, integration, and validation of innovative technologies for the reconstruction and visualization of the surrounding environment.

4 Challenges

Current challenges in the adoption of vision-based AI systems by industries highly overlap with current limitations of deep data-driven models, e.g., the need of non-trivial amounts of labeled data for training and evaluation. Whenever possible, we deem harnessing synthetic data from virtual worlds can push the boundaries of current applications, specially in practically unfeasible scenarios like safety-critical ones, e.g. detection of situation of damage to people or machinery.

References

[Amato *et al.*, 2018] G. Amato, P. Bolettieri, D. Moroni, F. Carrara, L. Ciampi, G. Pieri, C. Gennaro, G. R. Leone, e C. Vairo. A wireless smart camera network for parking monitoring. In *2018 IEEE Globecom Workshops (GC Wkshps)*. IEEE, dec 2018.

[Amato *et al.*, 2019a] G. Amato, L. Ciampi, F. Falchi, e C. Gennaro. Counting vehicles with deep learning in onboard UAV imagery. In *2019 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, jun 2019.

[Amato *et al.*, 2019b] G. Amato, L. Ciampi, F. Falchi, C. Gennaro, e N. Messina. Learning pedestrian detection from virtual worlds. In *International Conference on Image Analysis and Processing (ICIAP)*, pages 302–312. Springer, 2019.

[Carrara *et al.*, 2021] F. Carrara, G. Amato, L. Brombin, F. Falchi, e C. Gennaro. Combining GANs and autoencoders for efficient anomaly detection. In *25th International Conference on Pattern Recognition (ICPR)*, pages 3939–3946. IEEE, 2021.

[Ciampi *et al.*, 2020a] L. Ciampi, N. Messina, F. Falchi, C. Gennaro, e G. Amato. Virtual to real adaptation of pedestrian detectors. *Sensors*, 20(18):5250, sep 2020.

[Ciampi *et al.*, 2020b] L. Ciampi, C. Santiago, J. P. Costeira, C. Gennaro, e G. Amato. Unsupervised vehicle counting via multiple camera domain adaptation. In *Proceedings of the First International Workshop on New Foundations for Human-Centered AI (NeHuAI), Santiago de Compostella, Spain, September 4, 2020*, volume 2659 of *CEUR Workshop Proceedings*, pages 82–85. CEUR-WS.org, 2020.

[Ciampi *et al.*, 2021a] L. Ciampi, C. Gennaro, F. Carrara, F. Falchi, C. Vairo, e G. Amato. Multi-camera vehicle counting using edge-ai. *arXiv:2106.02842*, 2021.

[Ciampi *et al.*, 2021b] L. Ciampi, C. Santiago, J. Costeira, C. Gennaro, e G. Amato. Domain adaptation for traffic density estimation. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications, 2021.

[Di Benedetto *et al.*, 2019] M. Di Benedetto, E. Meloni, G. Amato, F. Falchi, e C. Gennaro. Learning safety equipment detection using virtual worlds. In *2019 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–6. IEEE, 2019.

[Di Benedetto *et al.*, 2021] M. Di Benedetto, F. Carrara, E. Meloni, G. Amato, F. Falchi, e C. Gennaro. Learning accurate personal protective equipment detection from virtual worlds. *Multimedia Tools and Applications*, 80(15):23241–23253, 2021.

[Massoli *et al.*, 2021] F. V. Massoli, F. Falchi, A. Kantarci, Ş. Akti, H. K. Ekenel, e G. Amato. Mocca: Multilayer one-class classification for anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[Merola *et al.*, 2022] F. Merola, F. Falchi, C. Gennaro, e M. Di Benedetto. Reinforced damage minimization in critical events for self-driving vehicless. In *17th Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, 2022.