

Automatic Intermodal Loading Unit Identification using Computer Vision: A Scoping Review

Emre Gülsoylu^{a,*}, Alhassan Abdelhalim^b, Derya Kara Boztas^c, Ole Grasse^c,
Carlos Jahn^c, Simone Frintrop^a, Janick Edinger^b

^a*University of Hamburg, Department of Informatics, Computer Vision Group, Hamburg, Germany*

^b*University of Hamburg, Department of Informatics, Distributed Operating Systems
Group, Hamburg, Germany*

^c*Hamburg University of Technology, Institute of Maritime Logistics, Hamburg, Germany*

Abstract

The standardisation of Intermodal Loading Units (ILUs), such as containers, semi-trailers and swap bodies, has revolutionised global trade yet their efficient and robust identification remains a critical bottleneck in high-throughput ports and terminals. This paper reviews 63 empirical studies that propose computer vision (CV) based solutions. It covers the last 35 years (1990-2025), tracing the field's evolution from early digital image processing (DIP) and traditional machine learning (ML) to the current dominance of deep learning (DL) techniques. While CV offers cost-effective alternatives for other types of identification techniques, its development is hindered by the lack of publicly available benchmarking datasets. This results in high variance for the reported results such as end-to-end accuracy ranging from 5 % to 96 %. Beyond dataset limitations, this review highlights the emerging challenges especially introduced by the shift from character-based text recognition to scene-text spotting and the integration of mobile cameras (e.g. drones, sensor equipped ground vehicles) for dynamic terminal monitoring. To advance the field, the paper calls for standardised terminology, open-access datasets, shared source code, while outlining future research directions such as contextless text recognition optimised for ISO6346 codes.

Keywords: intermodal loading unit identification, container code recognition, text detection, text recognition, logistics automation

*indicates corresponding author (emre.guelsoylu@uni-hamburg.de).

1. Introduction

The evolution of cargo transportation has been significantly transformed by the concept of containerisation. Before, cargo handling was labour-intensive, neither standardised nor suitable for automatisisation. Handling and storage arrangements were costly and required special attention [1, 2], resulting in terminal congestion. Containerisation has facilitated automation and improved operational efficiency in cargo handling [3], while also significantly contributed to the growth of global trade [4]. The standardisation of freight handling has not only led to the *shipping container* but also led to a variety of standardised types of Intermodal Loading Units (ILUs).

The term ILU has been introduced to describe units designed to be easily transferred between different modes of transport, such as vessels, trains, and trucks, without the need to unload and reload the contents. The loading units, such as containers, interchangeable bodies (swap bodies) and semi-trailers, fall under this term [5]. These units are typically built according to specifications defined in standards such as ISO668 [6]. These specifications ensure compatibility of ILUs with different transport systems [7] for smooth movement of goods along supply chains [8]. A unique identification code, following the ISO6346 [9] standard, is assigned to each ILU for tracking purposes. This code consists of four letters, a six-digit number and a check digit, which together indicate the owner and type of the ILU.

As global trade expands, many leading ports handle tens of millions of containers annually, with some operating at or above their nominal design capacity [10]. In this high throughput, keeping track of ILUs is important to ensure efficient management [11]. The often identically looking ILUs can exclusively be distinguished by their ID codes. Therefore, a reliable system that can rapidly recognise these codes is indispensable. Standardisation and digitalisation enabled the development of various techniques for ILU identification. Among these, Computer Vision (CV)-based identification code recognition stands out as a cost-effective alternative to RFID, which achieves only about 70 % accuracy and entails high installation and maintenance costs [12, 13]. In contrast, CV-based methods can be implemented using a range of camera setups from high-end static Optical Character Recognition (OCR) gates [14], that can identify entering and exiting ILUs [15, 16], to more affordable and flexible options such as using the existing CCTV infrastructure, cameras mounted on aerial or ground vehicles, or even mobile devices, which significantly reduce the need for specialised hardware. Furthermore, CV-based identification plays a crucial role in enhancing security by providing accurate verification for customs clearance, which can support in the correct declaration of goods [17, 18, 19].

Although CV-based methods reduce human-related errors and increase the processing speed [20, 21, 22, 23] there are several challenges: Containers often arrive in

rapid succession, will be stacked for storage, and can be positioned at angles that partially occlude their ID codes. Environmental factors such as rain, fog, poor lighting, or glare from sunlight can degrade image quality, while harsh conditions, such as dirt, rust, or physical damage, further complicate automated ID code recognition [14, 24]. In addition, variations in font style, colour and placement of ID codes across different operators and regions create inconsistencies that reduce the reliability of automatic recognition systems.

At transfer points of the terminal to other modes of transport, each ILU must be identified and cross-checked with its associated documentation before entry or exit. When automated systems fail to achieve reliable recognition, manual verification becomes necessary, further slowing down the process. Even a brief slowdown at a busy gate can cause truck queues stretching hundreds of meters [25], disrupting yard operations and increasing turnaround times for transport vehicles. Extended waiting times contribute to congestion which increases fuel consumption and greenhouse gas emissions from idling trucks, amplifying the environmental footprint of port activities [26]. Therefore, efficient ILU identification is essential for overcoming these operational bottlenecks and sustaining the growth of containerised transport.

Despite its crucial role in improving transportation and trade security, to the best of our knowledge, there are not any literature reviews on ILU identification. To fill this gap, we prepared this literature review by following Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines [27] with the intention of identifying the available publications and proposed concepts.

In summary, there are 4 contributions of this review:

- Catalogue all empirical studies on Intermodal Loading Unit (ILU) identification using Computer Vision (CV) methods between 1990 and 2025.
- Re-evaluate the terminology used in ILU identification field and provide clear definitions.
- Synthesise an overview of Intermodal Loading Unit (ILU) identification methods, image acquisition settings, datasets, and evaluation metrics used in these studies.
- Highlight research gaps and suggest future priorities.

2. Terminology

This review distinguishes the roles of *Digital Image Processing (DIP)*, *Machine Learning (ML)*, and *Deep Learning (DL)* in the context of CV-based ILU identi-

cation to ensure precise use of technical terms and prevent misinterpretation across disciplines. Therefore, we provide the definitions of the general approaches:

Digital Image Processing (DIP): Classical, rule-based image operations (e.g., denoising [28], contrast normalisation [29], edge detection [30, 31], morphologic operations [32], thresholding [33, 34], geometric transforms [35]). These operations can be further extended for extracting hand-crafted features, such as edge detection [36, 30], Scale-Invariant Feature Transform (SIFT) [37] or Histogram of Gradients (HoG) [38]. It requires no training but is sensitive to illumination, motion blur, occlusions, or viewpoint changes [39, 40, 41]. The use cases of DIP methods in ILU identification are image enhancement [42], text localisation and character segmentation [43, 44], without relying on data-driven patterns.

Machine Learning (ML): While DIP methods rely on fixed, hand-crafted rules and deterministic operators (e.g. edge detectors, thresholding, morphologic filtering), ML introduces adaptability by learning from data. ML methods use features engineered via DIP (such as HoG, SIFT) as input to classifiers that can discriminate among classes (e.g. character classes, object vs. non-object, scene categories). Common models include Support Vector Machines (SVMs) [45, 46, 47], Random Forests [48], decision trees [49] and k-Nearest Neighbors [50]. These reduce manual rule-tuning and can generalize better under variations in viewpoint, scale, and lighting, but still depend critically on quality of the features and consistency in preprocessing [51].

Deep Learning (DL): A subfield of ML using deep neural networks, such as Convolutional Neural Networks (CNNs) [52] or transformers [53] that learn features and classifiers jointly and end-to-end. Although DL requires substantial training data and computational resources it automatise the feature engineering step. For similar tasks depending on similar features, deep learning methods can benefit from transfer learning [54] and fine-tuning [55].

Besides the terminology for general approaches, we would like to provide definitions for CV-specific terminology:

Object detection: Localising (drawing bounding boxes around) and classifying (assigning labels to) objects within an image. Unlike image classification (which labels are for the entire image), object detection handles multiple objects per image, providing both **what** and **where** the objects are [56, 57].

Instance segmentation: Localising (pixel-level regions, drawing polygons around) and classifying individual objects within an image. Unlike object detection, which uses bounding boxes, instance segmentation assigns pixel-level class labels and provides precise shapes of the objects [58].

Character segmentation: Localising and segmenting characters in an image, typically as a prior step for recognition [59].

Character recognition: Classifying a single already localised character into its alphanumeric label. This problem is often tackled as an image classification task where the input is a cropped character patch and the output is the label for this single character patch [60].

Character detection: Localising (drawing bounding boxes around) and classifying (assigning labels to) characters within an image. Similar to the object detection, character detection provides information about **where** individual characters are and **which** characters they are [61].

(Scene-)Text detection: Localising regions containing text in an image (e.g., street signs,) without extracting its textual content. Outputs bounding boxes (often quadrilaterals) around words or lines. Unlike character detection, it focuses on full words or lines, which makes it more robust against distractors [62].

(Scene-)Text recognition: Extracting textual information from detected text regions by converting them into machine-readable characters or words [63, 64].

Text spotting: End-to-end method that jointly detects and recognises text in a single step [65, 61].

2.1. Terminology Beyond OCR Gates

Initially, CV methods were applied at terminal gates, where the task is to automatically identify loading units as they entered or exited the facility. These gates are among the most vulnerable and critical points of a terminal, making reliable identification essential. The task performed there was to capture an image of the loading unit and read its identification code through DIP [44]. In the literature, this task is often referred to as Automatic Container Code Recognition (ACCR) [66, 67, 68, 69, 70, 71, 72] or transportation unit (TU) identification [73]. The focus of ACCR is on using CV to read the ISO6346 identification codes from containers at the point of entry. These entrances remained as the critical bottlenecks in terms of operational efficiency and congestion impact, so their optimisation becomes one of the key challenges in terminal management [74].

However, the term ACCR is not precise. First, the word container is too narrow, since not only containers but also semi-trailers and swap bodies carry ISO6346-compliant codes (as defined in EN 13044 [75]). Second, the word *recognition* does not fully capture the actual tasks involved. In computer vision, recognition usually means extracting text from a clean image, especially for the mobile cameras. The term *recognition* refers to extraction of text from an image in the CV field. However, this presupposes that the visual input image consists solely of the target text. This assumption is rarely valid in the real-world scenarios involving surveillance cameras or mobile cameras, where text is typically embedded within complex and cluttered

scenes. In these cases, an additional stage, *text detection* is required to localise the text regions before recognition can be performed [76]. Over time, these methods were extended beyond the gates. With advancements in computer vision and robotics, identification is now performed at quay cranes [77], yard blocks [78], and movable objects such as reach stackers [79]. These new use cases introduce more variation and complexity than the original gate-based setting.

For these reasons, we propose the term Intermodal Loading Unit (ILU) identification as an inclusive designation for the task of determining which specific ILUs with ISO 6346-compliant ID codes are visible in an image. This term includes various loading units and also for the CV terminology such as ILU detection/segmentation stage introduced in recent papers [73, 80], scene-text detection and recognition. A term that covers all these stages aligned with CV terminology is important, as it would address shared challenges across the CV applications. For example, using a common terminology can link the relevant studies to the broader problems such as small object segmentation [81] or small text detection [82] which are actively studied in general purpose CV methods. Moreover, a common terminology can help in finding the distinct aspects of ILU identification. Unlike in most common text recognition needs, where a language model based encoder (e.g. TrOCR [83]) is useful due to contextual cues present in the natural language, the contextless nature of the target text in ILU identification makes this approach ineffective. Such models tend to overfit to frequently occurring text instances in training data [84], limiting their applicability in this context. Adopting the term ILU identification can provide conceptual clarity and establish a coherent basis for researchers and practitioners.

3. Methodology

We finalised our search for articles in August 2025, using Google Scholar and dblp-Computer Science Bibliography to find scientific articles on ILU identification using CV or DIP techniques. Our search terms included *container code recognition*, *trailer code recognition*, and *identification of containers and trailers*. As discussed in Section 2, the terminology used in this field is not standardised and it directly affects our search methodology. We checked the cited papers and 'related articles' in Google Scholar and Research Rabbit to expand our pool of eligible articles.

As the first review on ILU identification, we applied inclusive eligibility criteria to capture the full scope of available evidence. These criteria were specified *a priori* before screening began. They follow the Population-Concept-Context (PCC) logic and the methodological guidance in [27]. Our review includes articles written in English, published from 1990 to 2025, that follow a scientific structure. Both

peer-reviewed articles and preprints are included as long as they report quantitative results, allowing us to catalogue the latest research findings. The final selection of articles and data charting were determined through consensus among the authors, following a review of titles, abstracts, and full texts.

From the 63 eligible articles the following data items were extracted into a spreadsheet¹: journal or conference details, contributions, methods used, method class (DIP, ML, or DL), dataset specifics (availability, number of images, diversity in lighting and weather), image collection region and camera position such as fixed (CCTV) or moving (aerial or ground vehicle), evaluation metrics, results, deployment status, suggested future directions and funding. The resulting table was then analysed and discussed by the authors.

4. Results and Discussion

Our search yielded 91 articles published between 1990 and 2025. During the screening process, articles were excluded due to being duplicates ($n = 2$), not reporting quantitative results ($n = 6$), written in a non-English language ($n = 5$), focusing solely on container detection or integrity assessment rather than identification via ISO6346 compliant codes ($n = 6$), solutions for other industrial text recognition applications such as wagon code recognition ($n = 5$), unavailability of their full text ($n = 2$), or being patents ($n = 2$).

Table 1 presents the characteristics of the eligible articles. The data indicate a consistent increase in publications over time, especially from 2011 onwards. This trend reflects growing interest in automatic ILU identification systems, likely driven by advancements in CV as well as greater computational power affordable to major terminals and ports.

Asia contributes 79.71 % of the publications. This region handles high volumes of maritime traffic requiring efficient logistics and ILU identification systems. Countries such as China, Taiwan, South Korea, Vietnam, and Japan play significant roles in global trade, making effective logistics and supply chain management essential. Substantial investment in research and development by Asian governments and companies further supports this trend [85, 86].

Public funding is reported as the main source, supporting about two out of five of the publications, indicating strong governmental backing. This support is often motivated by the potential for technological advancements to benefit national infrastructure and economic growth [87, 88, 89]. Public funding is particularly important

¹<https://cloud.uni-hamburg.de/s/FHqaGLR4Ccy8ceG>

Table 1: Eligible article characteristics ($n = 63$). Five articles involve international collaborations, which results in the count of 69 rather than the 63, in geographic regions section.

Category	Characteristic	Count	Percentage
Year of publication	1990–1995	2	3.17 %
	1996–2000	1	1.59 %
	2001–2005	5	7.94 %
	2006–2010	3	4.76 %
	2011–2015	10	15.87 %
	2016–2020	18	28.57 %
	2021–2025	24	38.10 %
Geographic region	Asia	55	79.71 %
	Europe	8	11.59 %
	North America	3	4.35 %
	Africa	2	2.90 %
	Oceania	1	1.45 %
Funding source	Not reported	32	50.79 %
	Public	24	38.10 %
	Public-Industry Collab.	4	6.35 %
	Industry	3	4.76 %
Rank	Unranked	32	50.79 %
	Q1 - Scimago	8	12.70 %
	A - CORE	5	7.94 %
	Q2 - Scimago	5	7.94 %
	B - CORE	4	6.35 %
	Q3 - Scimago	4	6.35 %
	Q4 - Scimago	4	6.35 %
	C - CORE	1	1.59 %

for the logistics sector that operates with low profit margins [90, 91]. Industry funding and public-industry collaborations constitute one ninth of the publications, suggesting that while there is some industry involvement, companies largely rely on public sector investments to drive technological development. However, the presence of pure industry funding indicates a genuine market demand for companies to invest in research to remain competitive [92, 11].

To analyse the rankings of journals and conferences we followed the rankings of Scimago² for journals and CORE³ for conferences. Table 1 shows that more than half of the articles on automatic ILU identification task are published in unranked outlets, which is not uncommon for applied or niche technical topics [93].

Among the ranked publications, Q1-Scimago and A-CORE have the highest numbers, suggesting that the topic is relevant for high-quality research venues. However, the majority of publications are spread across lower-tier or unranked venues. This distribution implies that the field could benefit from a greater emphasis on more novel contributions, increased comparability, and reproducibility of the results. This could be achieved through publicly available benchmark datasets and repositories that showcase the source code and the model weights. By adopting such research standards, ILU identification tasks with CV-based methods could be more relevant for high-ranked publication outlets.

Overall, the characteristics of the reviewed articles highlight a growing research field characterised by substantial activity in Asia and a predominance of public funding. Expanding funding sources and promoting industry collaboration may further advance automatic ILU identification through CV technologies.

4.1. Methods

The development of methods for ILU identification has undergone a significant transformation over three decades, as illustrated in Figure 1. For two decades between 1990 and 2010, the field relied predominantly on DIP techniques combined with early ML methods and to a lesser degree, DIP methods alone. This mirrors broader trends in the CV literature [94]. In this period, researchers mainly followed a three-stage approach: 1) Removal of the background using thresholding [95, 43], edge detection [43, 96] and morphological operations [97, 96]. 2) After segmenting the characters, each was classified by methods like template matching [98, 99, 100, 96], or traditional ML methods like Multi-Layer Perceptron (MLP) [101, 97, 44] or fuzzy logic [102]. 3) Finally, the predictions on the individual characters were postprocessed

²<https://www.scimagojr.com/>

³<https://www.core.edu.au/icore-portal>

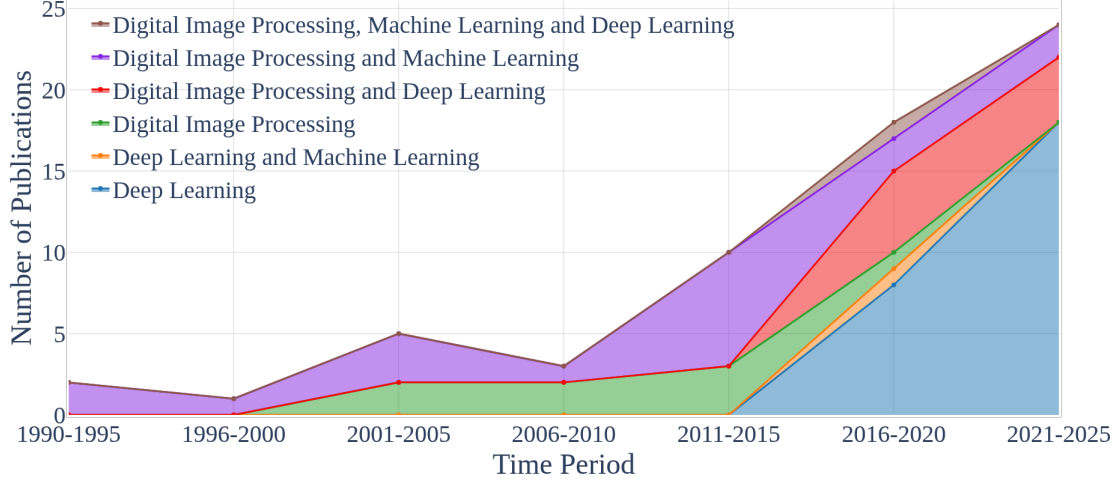


Figure 1: Approach categories used for ILU identification task over the years.

to output a full ID code. These methods performed adequately in controlled image acquisition systems like OCR-gates [103] but struggled with challenging conditions, such as variability of brightness [68] and damages [104, 105] on the ILU images.

In the subsequent five years, a modest increase in the use of ML can be observed, particularly in hybrid DIP and ML approaches [106, 107, 108, 109, 110, 12]. This reflects the growing recognition of ML’s potential for enhancing robustness in character and symbol recognition tasks [111] in CV. However, the field lagged several years behind the broad adoption of DL for other applications, despite its transformative influence on CV [112, 113] from 2012 onwards (e.g., the emergence of AlexNet [114]). This delay may be attributed to the scarcity of labelled ILU identification datasets, limited computational infrastructure, and the conservative approach for adopting new technologies in the logistics industry.

The five-year interval from 2016 to 2020 marks a turning point, with a notable adaptation of DL-based methods. Publications using only DL-based approaches rise sharply, as well as the hybrid strategies combining DIP, ML, and DL. This phase was driven by advances in GPU hardware [115, 116], and the availability of open-source frameworks such as TensorFlow [117] and PyTorch [118]. Deep architectures, particularly CNNs [119], rapidly proved their superiority for conducting ILU identification under challenging conditions [72, 120]. As a consequence, purely DIP-based approaches declined significantly, often limited to pre- or post-processing roles within hybrid pipelines [68, 80, 20, 121, 70].

ILU identification with fixed cameras often resembles OCR from scanned docu-

ments, as the background is mostly uniform. Segmenting and recognising characters is a sensible approach in such cases. Until 2016, all methods developed for ILU identification focused on character-level localisation and recognition, and this approach is still in use with DL methods [122, 123, 124]. However, with the increasing use of mobile cameras (e.g. vehicle-mounted cameras), the task has evolved to resemble scene-text detection and recognition. As a result of this, the trend is shifting towards using comprehensive scene-text detection and recognition methods rather than relying on character detection and recognition.

By 2021–2025, DL-based approaches had established themselves as the predominant paradigm for ILU identification. Even hybrid DIP-based methods virtually disappeared, indicating that the DL models now offer both the required accuracy and operational robustness [66, 125, 126]. The relative infrequency of hybrid approaches suggests that most effective solutions now rarely require DIP or traditional ML methods, except in specialised scenarios such as real-time constraints [104, 16, 18, 127, 124, 15].

4.2. Main Approaches for Deep Learning

Observing the growing interest of DL methods, we now focus on the CV tasks relevant for extracting textual information. OCR techniques are designed to identify and localize individual characters in an image. They are used predominantly for document text recognition and perform well under standard conditions, such as uniform backgrounds and consistent lighting (e.g., scanned documents). However, these techniques may fall short when dealing with *text-in-wild* scenarios [128], where the background and lighting are highly diverse and the text orientation is arbitrary.

For tasks involving the diverse image properties of natural scenes, such as identifying signs in urban areas [129], two primary subtasks are distinguished. Scene-text detection focuses on identifying and localizing text regions within an image without interpreting their content, while scene-text recognition converts these localized regions into machine-readable characters [130, 131, 132]. The combination of these tasks is known as scene-text spotting, which handles text detection and recognition in an end-to-end manner. These methods are generally faster and more efficient due to their unified feature extraction module [133, 134].

Focusing solely on character detection and recognition can be insufficient when image conditions are highly diverse, such as varying lighting and backgrounds. Relying on character detection in diverse scenes can increase false positives, as many patterns may look similar to letters. Additionally, DL methods require vast amounts of data, and character-level annotation is more expensive than word-level or line-level annotation due to its granularity.

Table 2: Main approaches to ILU identification. The table reports how often each combination appears in the surveyed literature. ● = detection, ● = recognition.

Object	Character	Text	Count	Percentage
○	○	●●	20	54.1%
○	○	●	4	10.8%
○	○	●●	3	8.1%
○	●	○	2	5.4%
○	●●	○	2	5.4%
○	●●	●	2	5.4%
●	○	●●	1	2.7%
○	●	●	1	2.7%
○	○	●	1	2.7%
○	●	○	1	2.7%

This shift from character detection and recognition to text detection and recognition is shown in Table 2, where the majority of articles employ DL methods for both scene-text detection and recognition tasks. The share of hybrid approaches using non-DL methods for one of these steps are minimal. Examples include using traditional techniques for text detection such as cropping a predefined area for searching ID codes [127] or for character recognition, template matching after a DL model detects text areas [135]. Instead, researchers largely favour more comprehensive pipelines that rely on DL in each step. This shift towards unified DL-based approaches for both scene-text detection and recognition highlights their potential. However, to further enhance efficiency, researchers should focus more on text spotting [136, 134].

In ILU identification, text recognition remains the greatest challenge. Environmental conditions and technical issues such as unreliable hardware or connectivity problems between devices can cause delays at the gate. These delays increase waiting times for stakeholders, lead to congestion, add to environmental pollution, and create further setbacks for subsequent loading units. When recognition errors occur, the risk of incorrect inventory lists rises, which not only requires time-consuming manual corrections in the handling process but can also raise security concerns [14]. To address these challenges more effectively, end-to-end approaches have been proposed. By combining detection and recognition into a single forward pass with a unified feature extraction module, they streamline the process and improve overall efficiency.

Few studies introduce object detection as a prior step to scene-text detection and



Figure 2: Sample images for fixed perspective (left) [72], ground vehicle perspective (middle) [73], aerial perspective (right) [84].

recognition [73, 80]. Detecting the ILU first reduces the search area for the text detection step and enables association of detected and recognised IDs with other objects, thereby improving scene understanding [73].

4.3. Datasets

Datasets form the backbone of reliable ILU detection. The quality, diversity, and annotation of the underlying data strongly influence how well models perform once deployed in real-world port environments. Carefully curated datasets with consistent labelling not only enable robust training and evaluation, but also determine whether methods can generalise beyond laboratory settings.

Understanding the properties of datasets is therefore essential for evaluating the performance and applicability of CV-based ILU identification systems. Below, we summarise the main characteristics and settings of the datasets used in the eligible studies, highlighting camera positioning, diversity, and availability.

4.3.1. Image Acquisition Setups

We categorised cameras as fixed or mobile cameras based on their mounting and usage. Fixed cameras include OCR gates, CCTV cameras, and crane-mounted cameras. These are installed at permanent locations such as terminal gates, yards, or crane structures, providing stable, controlled views for capturing ILU codes. Their main advantage is delivering consistent, high-quality images ideal for computationally efficient methods like classical DIP and traditional ML, as long as ILUs are present at predefined areas. Although the installation and operation of these systems involve substantial infrastructure investment, fixed cameras often struggle to capture sufficient feature information in large and complex environments, resulting in reduced accuracy and reliability of detection and recognition based on visual neural networks [137].

Table 3: Summary of dataset characteristics: camera position and dataset availability.

	Category	Count	Percentage
Camera position	Only Fixed	47	74.60 %
	Fixed and Ground	10	15.87 %
	Fixed, Aerial and Ground	3	4.76 %
	Only Ground	1	1.58 %
	Only Aerial	1	1.58 %
	Not reported	1	1.58 %
Availability	Not public	54	85.71 %
	Fully available	3	4.76 %
	Partially available	3	4.76 %
	Upon request	3	4.76 %
	Behind paywall	1	1.58 %

In contrast, mobile cameras are mounted on unmanned aerial vehicles (UAVs), reach stackers, terminal trucks, or other industrial vehicles. For ground-level perspectives, we also considered handheld images of ILUs, as they closely resemble those captured by vehicle-mounted systems. Unlike fixed setups, mobile cameras provide flexibility by moving across the terminal and capturing a wider range of perspectives. This mobility enables real-time port monitoring and localisation of each ILU in real-world coordinates, as well as monitoring their positions by handling vehicles such as reach stackers [138, 139]. While they cover wider spatial areas and provide richer scene diversity, mobile cameras introduce more complex image variability, requiring advanced processing methods and diverse publicly available datasets.

As shown in Table 3 the most common collection setting is the fixed CCTV cameras at OCR-gates [140, 123, 141] or cranes (74.60 %) [80, 108] and a mixture of these images with mobile phone snapshots (15.87 %) [67]. However, vehicle-mounted cameras are increasingly used alongside fixed cameras to capture wider areas and provide information across the terminal [142, 84, 122, 73].

4.3.2. Diversity

Most datasets consist of images gathered from the real-world settings without controlled environments such as the World’s busiest ports like Shanghai [16], Singapore [44], Hong Kong [143], Kaohsiung [13], and Hamburg [73], as well as smaller inland ports like Dresden and Riesa [84]. Besides the terminal-specific data collection, web scraping were also used [125, 72]. In addition, controlled environments such

as the Shanghai Maritime University test facility [42] were used for data collection.

Other datasets were created with images taken in Malaysia [66], Sweden [123], Vietnam [18, 124, 127, 68], Thailand [20], South Korea [142, 71, 144, 108, 102], Japan [99], and Italy [145]. However, datasets consisting of images from multiple locations are rare. For DL methods, capturing the nuances of ILUs across different countries and their respective ports can be important.

The number of images used in the experiments of the eligible articles ranges from a minimum of 30 images to a maximum of 34000, with a median of 1050 images. Since the DIP-based methods does not require a huge amount of data, researchers who tackled this problem before DL, experimented with a lower number of images compared to recent studies. However, with the emergence of ML and DL, the need for data has increased significantly. Most of the studies have addressed this need by pre-training their models on general-purpose scene-text detection and recognition datasets such as ICDAR2015 [146], CTW1500 [147], and SynthText [148], then fine-tuning on smaller ILU datasets.

Early datasets already included efforts to represent diverse conditions in order to evaluate generalisation. Lighting variability is the most discussed factor, with datasets often including images taken in bright, low-light, and uneven lighting scenarios [100, 43]. Considering how the earlier, thresholding-based methods may get affected by small changes in light intensity, diversifying the lighting conditions in the experiments was a major focus [145, 97, 143, 96]. However, currently lighting conditions are classified with more general classes: Day and night settings are frequently included to test the adaptability of the systems [104, 122, 73, 16].

Weather conditions such as rain, snow, and fog are also represented, as they can significantly affect image quality and ILU identification performance [73, 67, 126, 143]. Additionally, some data sets feature images with motion blur, camera noise, dirty, rusty, damaged, or partially occluded ID codes to further challenge algorithms with the exceptional conditions that exist in real-world settings [104, 73, 105, 126, 72]. Rarely were these conditions simulated to increase the data sets [125].

4.3.3. Availability

Table 3 shows the availability of the datasets used by the identified articles. The significant majority of the datasets are not publicly available, which creates a substantial barrier for research and development in this area. The lack of availability also prevents the reproducibility of results and fair comparison of the proposed methods. Further details for the publicly available datasets are shown in Table 4.

Among these, the *TRansportation Unit Detection and Identification (TRUDI)* dataset [73] stands out for its annotation density. It offers 35,034 instances across 733

Table 4: Publicly available ILU identification datasets.

Dataset	Images	Instances	Conditions	Location
Shipping container code [23]	3600	4000	Day/night, Rainy, snowy Dirty, damaged ID	Not reported
Container Number-OCR [105]	2851	2851	Low/high light Foggy Dirty, damaged ID	Tianjin Port
TRUDI [73]	733	35034	Day/night Rainy, snowy, foggy Motion blur Dirty, damaged ID	Multiple countries, Multiple terminals

images, covering five classes: `Container`, `trailer`, `tank container`, `ID code`, and `logo`. The *Shipping Container Code* dataset [23] leads in the number of images, and the *Container Number-OCR* dataset [105], though smaller (2,851 images from Tianjin Port), offers similar challenging conditions as others: variable weather conditions and degraded ID codes (dirt, rust, damage). Uniquely, *TRUDI* introduces motion blur as an additional challenging condition, as most of its images were captured via UAVs, reach stackers, and terminal trucks.

4.4. Evaluation

To evaluate the proposed ILU identification methods, several evaluation metrics are commonly used. As there are multiple stages of ILU identification using CV methods, such as text detection and text recognition, each stage can be evaluated individually. However, end-to-end accuracy or success rate, is the most relevant metric due to its practical importance. It is the most strict metric as one misrecognised character is enough for failed identification.

Since the ILU identification problem is mostly solved with multiple stages such as text detection and text recognition, some articles only report the results of individual stages and do not include end-to-end results. However, the results for individual stages offer limited insight for the applicability of the method for the ILU identification task.

Precision and recall are the common metrics for these individual stages. These metrics show the balance between false positives and false negatives, with precision

focusing on the accuracy of positive predictions and recall on the model’s ability to identify all relevant instances. The F1-score, or H-mean, harmoniously combines precision and recall into a single metric, offering a balanced view of the model’s performance.

For object detection or text detection tasks, mean average precision (mAP) and average precision (AP) are commonly used, measuring precision across various intersection over union (IoU) levels. The IoU metric is important for assessing the overlap between predicted and actual bounding boxes. As for the recognition, besides precision, recall and F1-score, other metrics such as the character error rate (CER) is used. CER measures the proportion of characters that were incorrectly predicted compared to the total number of characters in the reference text, using the Levenshtein-distance [149].

Speed and efficiency are evaluated using metrics such as frames per second (FPS), and inference time, which assess the model’s operational speed. Model complexity is assessed through the number of parameters, giga-floating point operations per seconds (GFLOPs), and model size, indicating the computational demands and storage requirements.

Although the recent studies use the same, established metrics adapted from general purpose text detection and recognition task, putting the reported results for ILU identification into a frame is not possible with the limited amount of publicly available datasets. Due to different datasets used in the experiments, the reported end-to-end accuracy of the proposed methods ranges from 5 % [142] to 96,17 % [144]. This level of variance can be observed in every metric, including the processing speed.

4.5. Future Research Directions

In the literature, future directions for the ILU identification task mainly focus on enhancing the datasets by including more images taken from diverse environments such as various distances, angles, and under different weather and lighting conditions. This expansion aims to cover multiple ports and scenarios to improve model generalisation. However, increasing the diversity of proprietary datasets would not solve the fundamental problem with the datasets for ILU identification. Recent papers acknowledge the lack of benchmark datasets and its negative effects on the field [122, 73]. Given that creating a new dataset for downstream CV tasks are tedious, time consuming, and costly [150, 151], alternatives such as synthetic data generation can be considered [152, 148], e.g., by using a game engine [153].

Handling challenging conditions, such as low-light scenarios, occlusions, and damaged ID codes, is another direction pointed by the researchers, with techniques proposed to manage partial occlusions and degraded text [105]. Current solutions are

relying on "off-the-shelf" super-resolution methods such as *Super-Resolution-CNN* [154]. Literature on scene-text super-resolution has focused primarily on natural language, leaving an opportunity for research specifically on contextless text such as ILU codes. Alternatively, multi-view recognition is also identified as a future direction due to its potential against degraded ID codes [23, 12] as multiple images of a single ILU are captured from different angles in this approach.

Improving image quality through preprocessing is another focus area. Suggested methods include image enhancement techniques such as noise reduction [17], or the use of Generative Adversarial Networks (GANs) [155, 156] to address issues such as shadows, rust, and dirt [23, 105]. Postprocessing steps are also recommended to verify recognition results by using the check digit and the owner code. [15, 157, 69, 72].

For DL-based methods, refining network architectures to increase the receptive field so that the model can *see* more of the scene in a single pass. It is important for detecting containers appears to be small due to perspective [122]. Another future work suggestion is incorporating attention mechanisms to boost recognition accuracy [158]. The development of unified, end-to-end models is suggested to streamline processing pipelines [73, 140, 105, 157].

The contextless nature of ISO6346 codes makes scene-text spotting task harder as there are no natural language cues [159]. Therefore, directly applying current scene-text recognition models that rely on semantic context, is not the most suitable approach, as they usually use a language model or at least a decoder module pretrained on natural language to correct the initial recognition result [160, 83]. However, incorporating complex language models increases model complexity, while not yielding any benefits for contextless recognition tasks [159]. Researching on the contextless scene-text recognition models is an opportunity which may result in novel and more efficient methodologies. These new methods can also be useful for similar applications including International Union of Railways (UIC)-wagon codes, or registration plates.

Improvements for real-time processing are also highlighted, with a focus on optimising algorithms for mobile devices and enhancing systems to work in real-time, potentially through video stream analysis [123, 42].

5. Conclusion

This review provides the first synthesis of CV-based ILU identification research by analysing 63 empirical studies published over three decades. There is a clear upward trend for the number of publications with the rise of DL methods and Asia's dominance as a global trade hub, with 79.71 % of studies originating from the region.

Public funding supports over a third of research highlighting a reliance on public investment for transportation related infrastructure. Policymakers could foster progress by funding collaborative initiatives and public-private partnerships, which currently account for just 6.35 % of studies.

Although there are articles published in top journals and conferences, most studies were published in unranked outlets. The main reason for this is the direct application of general purpose CV methods to the ILU identification task. This lack of methodological contribution can be increased by adapting existing methods with modifications that addresses challenges, such as contextless text spotting and real-time processing constraints.

Methodologically, the field has shifted from classical image processing to DL, currently used in more than 90 % of recent studies, thanks to GPU advancements and frameworks like PyTorch. This shift comes with the need of huge datasets to train DL models. Data limitations are the biggest obstacle with 85.71 % of datasets being private and most studies relying on controlled image acquisition setups. The median dataset size is too small for robust DL training, forcing researchers to adapt general text detection and recognition models. Another critical point for improving the quality and credibility of publications is using or developing benchmark datasets. The lack of benchmark dataset makes performance comparisons difficult, leaving gaps in understanding which methods work best in which setting.

The research on ILU identification mainly focuses on fixed camera settings such as OCR-gates. However, recent studies show a shift from fixed to mobile cameras, such as drones or vehicle mounted cameras. This shift can spark more innovation due to new challenges related to dynamic scenes from the technical perspective [84] as well as social perspective such as privacy issues [161, 162, 163]. Camera mounted vehicles offer new subtasks for the ILU identification task such as pose estimation of the ILUs by predicting the real-world coordinates (position) and heading (orientation). This way, a precise monitoring of the terminal is possible.

To conclude, we suggest emphasising standardisation for evaluation with publicly available benchmark datasets, contextless text spotting architectures, real-time processing on edge devices and more focus on vehicle mounted camera setups for a more comprehensive ILU identification and terminal monitoring.

Acknowledgements

The project is supported by the German Federal Ministry for Digital and Transport (BMDV) in the funding program Innovative Hafentechnologien II (IHATEC II) with the funding number 19H23002C.

Disclosure statement

The authors report there are no competing interests to declare.

References

- [1] B. Behdani, B. Wiegman, V. Roso, H. Haralambides, Port-hinterland transport and logistics: emerging trends and frontier research, *Maritime Economics & Logistics* 22 (2020) 1–25.
- [2] P. M. Panayides, Economic organization of intermodal transport, *Transport Reviews* 22 (4) (2002) 401–414.
- [3] United Nations, Review OF Maritime Transport, https://unctad.org/system/files/official-document/rmt2024_en.pdf, accessed: 2025-09-19 (2024).
- [4] M. Levinson, Introduction to The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger, Princeton University Press, 2008.
- [5] L. Filina-Dawidowicz, M. Kostrzewski, [The Complexity of Logistics Services at Transshipment Terminals](#), in: *Energies*, Vol. 15, MDPI, 2022, p. 1435. doi: 10.3390/en15041435.
URL <https://www.mdpi.com/1996-1073/15/4/1435>
- [6] [Series 1 Freight Containers – Classification, Dimensions and Ratings](#) (2020).
URL <https://www.iso.org/standard/79529.html>
- [7] EURLEX, Intermodal transport: intermodal loading units (ILUs), <https://eur-lex.europa.eu/EN/legal-content/summary/intermodal-transport-intermodal-loading-units-ilus.html>, accessed: 2025-07-13 (2007).
- [8] H. A. Van Klink, G. C. van Den Berg, Gateways and intermodalism, *Journal of transport geography* 6 (1) (1998) 1–9.
- [9] I. O. for Standardization, Freight containers — Coding, identification and marking, Standard, International Organization for Standardization, Geneva, CH (2022).

- [10] A.-K. Lange, A. Schwientek, C. Jahn, Reducing truck congestion at ports—classification and trends, in: Proceedings of the hamburg international conference of logistics (HICL), epubli, 2017, pp. 37–58.
- [11] F. Parola, M. Risitano, M. Ferretti, E. Panetti, The drivers of port competitiveness: a critical review, *Transport Reviews* 37 (1) (2017) 116–138.
- [12] M. Chen, W. Wu, X. Yang, X. He, Hidden-Markov-model-based segmentation confidence applied to container code character extraction, *IEEE Transactions on intelligent transportation systems* 12 (4) (2011) 1147–1156.
- [13] S.-C. Tai, T.-Y. Chang, S.-Y. Wang, A method of container codes detection and segmentation by using improved projection algorithm, in: 2011 9th World Congress on Intelligent Control and Automation,, IEEE, 2011, pp. 587–592.
- [14] H. Garmouch, O. Abdoun, Enhancing crane and gate OCR efficiency at container terminal using a hybrid genetic algorithm and neural network model: case study of tangier med port, *Discover Applied Sciences* 7 (7) (2025) 714.
- [15] Y. Li, H. Li, G. Gao, Towards end-to-end container code recognition, *Multi-media Tools and Applications* 81 (11) (2022) 15901–15918.
- [16] J. Zhao, N. Jia, X. Liu, G. Wang, W. Zhao, A Practical Unified Network for Localization and Recognition of Arbitrary-oriented Container code and type, *IEEE Transactions on Instrumentation and Measurement* (2024).
- [17] S. Hlabisa, R. L. Khuboni, J.-R. Tapamo, Automated shipping container code localization and recognition using deep learning, in: 2024 Conference on Information Communications Technology and Society (ICTAS), IEEE, 2024, pp. 198–203.
- [18] D.-D. Ngo, H.-T. Pham, T.-T. Be, V.-B. Nguyen, M.-H. Le, et al., A Vision-based Container-Code Checking System: Case Study at International Terminal, in: 2023 International Workshop on Intelligent Systems (IWIS), IEEE, 2023, pp. 1–6.
- [19] G.-R. Huang, S.-F. Su, Q. Chen, Y.-S. Lin, Study on content-independent feature matching systems for container images, in: 2013 International Conference on System Science and Engineering (ICSSE), IEEE, 2013, pp. 31–36.

- [20] C. Roeksukrungrueang, T. Kusonthammrat, N. Kunapronsujarit, T. N. Aruwong, S. Chivapreecha, An implementation of automatic container number recognition system, in: 2018 international workshop on advanced image technology (IWAIT), IEEE, 2018, pp. 1–4.
- [21] W. Zhang, L. Zhu, L. Xu, J. Zhou, H. Sun, X. Liu, Deep learning based container text recognition, in: 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, 2019, pp. 69–74.
- [22] C.-C. Hsu, Y.-Z. Yang, A. Chang, S. S. Morsalin, G.-T. Shen, L.-S. Shiu, Automatic Recognition of Container Serial Code, in: 2023 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan), IEEE, 2023, pp. 257–258.
- [23] R. Khuboni, S. Hlabisa, J. R. Tapamo, Confidence-Guided Shipping Container Code Recognition Using Deep Learning, Available at SSRN 5021624 (2025).
- [24] K. Iwata, E. Ishidera, T. Yamaai, Y. Satoh, H. Tanaka, K. Takahashi, A. Furuhata, Y. Tanabe, H. Matsumura, Guidelines for External Disturbance Factors in the Use of OCR in Real-World Environments, arXiv preprint arXiv:2504.14913 (2025).
- [25] L. N. Spasovic, Y. He, D. Besenski, et al., Quantifying Impact of Port Truck Traffic on Highway Operations Using GPS-Based Speed Data, Tech. rep., Rutgers University. Center for Advanced Infrastructure and Transportation (2015).
- [26] G. Giuliano, T. O’Brien, Reducing port-related truck emissions: The terminal gate appointment system at the Ports of Los Angeles and Long Beach, Transportation Research Part D: Transport and Environment 12 (7) (2007) 460–473.
- [27] A. C. Tricco, E. Lillie, W. Zarin, K. K. O’Brien, H. Colquhoun, D. Levac, D. Moher, M. D. Peters, T. Horsley, L. Weeks, et al., PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation, Annals of internal medicine 169 (7) (2018) 467–473.
- [28] A. Buades, B. Coll, J.-M. Morel, A review of image denoising algorithms, with a new one, Multiscale modeling & simulation 4 (2) (2005) 490–530.
- [29] R. C. Gonzalez, Digital image processing, Pearson education india, 2009.

- [30] J. Canny, A computational approach to edge detection, *IEEE Transactions on pattern analysis and machine intelligence* (6) (2009) 679–698.
- [31] B. Jähne, H. Haußecker, P. Geißler, *HANDBOOK OF COMPUTER VISION AND APPLICATIONS*, Academic Press, 1999.
- [32] J. Serra, L. Vincent, An overview of morphological filtering, *Circuits, Systems and Signal Processing* 11 (1) (1992) 47–108.
- [33] N. Otsu, et al., A threshold selection method from gray-level histograms, *Automatica* 11 (285-296) (1975) 23–27.
- [34] M. Sezgin, B. I. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, *Journal of Electronic imaging* 13 (1) (2004) 146–168.
- [35] R. Szeliski, *Computer vision: algorithms and applications*, Springer Nature, 2022.
- [36] H. Spontón, J. Cardelino, A review of classic edge detectors, *Image Processing On Line* 5 (2015) 90–123.
- [37] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International journal of computer vision* 60 (2) (2004) 91–110.
- [38] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, Vol. 1, Ieee, 2005, pp. 886–893.
- [39] E. Kurimo, L. Lepistö, J. Nikkanen, J. Grén, I. Kunttu, J. Laaksonen, The effect of motion blur and signal noise on image quality in low light imaging, in: *Scandinavian Conference on Image Analysis*, Springer, 2009, pp. 81–90.
- [40] F. Dufaux, *Grand challenges in image processing* (2021).
- [41] F. Li, J. Li, Edge detection of motion-blurred images based on gans., in: *British Machine Vision Conference (BMVC)*, 2022, p. 266.
- [42] Y. Zhu, Q. Zhang, W. Zeng, A New Model of Adaptive Container Image Pre-processing, in: *2015 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering*, Atlantis Press, 2015.

- [43] H. Zhiwei, L. Jilin, M. Hongqing, L. Peihong, A new localization method for container auto-recognition system, in: International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003, Vol. 2, IEEE, 2003, pp. 1170–1172.
- [44] H. C. Lui, C. M. Lee, F. Gao, Neural network application to container number recognition, in: Proceedings., Fourteenth Annual International Computer Software and Applications Conference, IEEE, 1990, pp. 190–195.
- [45] A. E. Ghahnavieh, A. Amirkhani-Shahraki, A. A. Raie, Enhancing the license plates character recognition methods by means of svm, in: 2014 22nd Iranian Conference on Electrical Engineering (ICEE), IEEE, 2014, pp. 220–225.
- [46] N. Sharma, M. A. Haq, P. K. Dahiya, B. Marwah, R. Lalit, N. Mittal, I. Keshta, Deep learning and svm-based approach for indian licence plate character recognition., Computers, Materials & Continua 74 (1) (2023).
- [47] W.-C. Liu, C.-H. Lin, A hierarchical license plate recognition system using supervised k-means and support vector machine, in: 2017 International Conference on Applied System Innovation (ICASI), IEEE, 2017, pp. 1622–1625.
- [48] Z. Akhtar, R. Ali, Automatic number plate recognition using random forest classifier, SN Computer Science 1 (3) (2020) 120.
- [49] C.-M. Tsai, F. Y. Shih, An efficient detection and recognition system for multiple motorcycle license plates based on decision tree, International Journal of Pattern Recognition and Artificial Intelligence 36 (05) (2022) 2250022.
- [50] A. K. Sahoo, Automatic recognition of indian vehicles license plates using machine learning approaches, Materials Today: Proceedings 49 (2022) 2982–2988.
- [51] S. Du, M. Ibrahim, M. Shehata, W. Badawy, Automatic license plate recognition (alpr): A state-of-the-art review, IEEE Transactions on circuits and systems for video technology 23 (2) (2012) 311–325.
- [52] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (2002) 2278–2324.
- [53] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

- [54] L. Y. Pratt, Discriminability-based transfer between neural networks, *Advances in neural information processing systems* 5 (1992).
- [55] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *nature* 521 (7553) (2015) 436–444.
- [56] R. Girshick, J. Donahue, T. Darrell, J. Malik, Region-based convolutional networks for accurate object detection and segmentation, *IEEE transactions on pattern analysis and machine intelligence* 38 (1) (2015) 142–158.
- [57] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [58] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [59] R. G. Casey, E. Lecolinet, A survey of methods and strategies in character segmentation, *IEEE transactions on pattern analysis and machine intelligence* 18 (7) (2002) 690–706.
- [60] J. Denker, W. Gardner, H. Graf, D. Henderson, R. Howard, W. Hubbard, L. D. Jackel, H. Baird, I. Guyon, Neural network recognizer for hand-written zip code digits, *Advances in neural information processing systems* 1 (1988).
- [61] K. Wang, B. Babenko, S. Belongie, End-to-end scene text recognition, in: *2011 International conference on computer vision*, IEEE, 2011, pp. 1457–1464.
- [62] B. Epshtein, E. Ofek, Y. Wexler, Detecting text in natural scenes with stroke width transform, in: *2010 IEEE computer society conference on computer vision and pattern recognition*, IEEE, 2010, pp. 2963–2970.
- [63] Q. Ye, D. Doermann, Text detection and recognition in imagery: A survey, *IEEE transactions on pattern analysis and machine intelligence* 37 (7) (2014) 1480–1500.
- [64] B. Shi, X. Bai, C. Yao, An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (11) (2017) 2298–2304. [doi:10.1109/TPAMI.2016.2646371](https://doi.org/10.1109/TPAMI.2016.2646371).

- [65] P. Lyu, M. Liao, C. Yao, W. Wu, X. Bai, Mask textspotter: An end-to-end trainable neural network for spotting text with arbitrary shapes, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 67–83.
- [66] L. J. Lau, L. T. Lim, Y. Tew, Modelling Studies of Automatic Container Code Recognition System for Real Time Implementation, in: 2024 IEEE Symposium on Industrial Electronics & Applications (ISIEA), IEEE, 2024, pp. 1–6.
- [67] M. Yu, S. Zhu, B. Lu, Q. Chen, T. Wang, A Two-Stage Automatic Container Code Recognition Method Considering Environmental Interference, *Applied Sciences* 14 (11) (2024) 4779.
- [68] D. Nguyen, D. Nguyen, T. Nguyen, K. Ngo, H. Cao, T. Vuong, T. Quan, Automatic container code recognition using MultiDeep pipeline, in: International Conference on Computational Collective Intelligence, Springer, 2020, pp. 139–153.
- [69] W. Zhiming, W. Wuxi, X. Yuxiang, Automatic container code recognition via Faster-RCNN, in: 2019 5th International Conference on Control, Automation and Robotics (ICCAR), IEEE, 2019, pp. 870–874.
- [70] L. Cao, Z. Gai, E. Liu, H. Gao, H. Li, L. Yang, H. Li, Automatic container code recognition system based on geometrical clustering and spatial structure template matching, in: International Conference in Communications, Signal Processing, and Systems, Springer, 2017, pp. 2198–2204.
- [71] Y. Yoon, K.-D. Ban, H. Yoon, J. Kim, Automatic container code recognition from multiple views, *Etri Journal* 38 (4) (2016) 767–775.
- [72] A. Verma, M. Sharma, R. Hebbalaguppe, E. Hassan, L. Vig, Automatic container code recognition via spatial transformer networks and connected component region proposals, in: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2016, pp. 728–733.
- [73] E. Gülsoylu, A. Kelm, L. Bengtson, M. Hirsch, C. Wilms, T. Rolff, J. Edinger, S. Frintrop, TRUDI and TITUS: A Multi-Perspective Dataset and A Three-Stage Recognition System for Transportation Unit Identification, *arXiv preprint arXiv:2508.02372* (2025).
- [74] M. Balliauw, H. Meersman, E. Van de Voorde, T. Vanelslander, Towards improved port capacity investment decisions under uncertainty: a real options approach, *Transport Reviews* 39 (4) (2019) 531–552.

- [75] International Union of Combined Road-Rail Transport Companies (UIRR), [New Markings of Intermodal Loading Units in Europe](#), in: UIRR Brochure: ILU Code (EN 13044 Marking of Intermodal Loading Units), Brussels, Belgium, 2011, published May 2011.
URL https://www.kombiverkehr.de/_Resources/Persistent/1/d/9/c/1d9cd757961710526f0e89aa8d90cef42c013200/UIRR%20Brochure%20ILU%20Code.pdf
- [76] S. Liu, Z. Guo, Y. Li, C. Wang, L. Chen, Z. Sun, Y. Zhou, B. Xu, [A Survey of Text Detection and Recognition Algorithms Based on Deep Learning Technology](#), in: Neurocomputing, Elsevier, 2023. doi:10.1016/j.neucom.2023.126113.
URL <https://www.sciencedirect.com/science/article/pii/S0925231223008251>
- [77] J. Wang, M. He, Y. Zhang, Z. Zhang, O. Postolache, C. Mi, High-precision pose measurement of containers on the transfer platform of the dual-trolley quayside container crane based on machine vision, *Sensors* 25 (9) (2025) 2760.
- [78] T. N. Cuong, S.-S. You, G.-S. Cho, B. Choi, H.-S. Kim, N. Q. Vinh, J.-H. Yeon, Safe operations of a reach stacker by computer vision in an automated container terminal, *Alexandria Engineering Journal* 109 (2024) 285–298.
- [79] F. Hakimi, T. Khaled, M. Al-Kharaz, A. C. F. Gouabou, K. Amzil, Towards a digital twin modeling method for container terminal port, *Procedia Computer Science* 246 (2024) 3113–3121.
- [80] J. Wu, J. Guo, X. Li, Key methods of recognizing container number automatically using video stream in intelligent tally, in: *Chinese Intelligent Systems Conference*, Springer, 2019, pp. 10–17.
- [81] C. Mun, S. Lee, Y. Uh, J. Choe, H. Byun, Small objects matters in weakly-supervised semantic segmentation, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 414–423.
- [82] Y. Zheng, Y. Xie, Y. Qu, X. Yang, C. Li, Y. Zhang, Scale robust deep oriented-text detection network, *Pattern Recognition* 102 (2020) 107180.
- [83] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, F. Wei, Trocr: Transformer-based optical character recognition with pre-trained models, in: *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37, 2023, pp. 13094–13102.

- [84] J. Teegen, A. Kelm, O. Grasse, M. Hillemann, E. Gülsoylu, S. Frintrop, Drone-Based Identification of Containers and Semi-Trailers in Inland Ports, in: 12th International Conference on Logistics and Maritime Systems, 2024.
- [85] J. S. L. Lam, T. Notteboom, The greening of ports: a comparison of port management tools used by leading ports in asia and europe, *Transport Reviews* 34 (2) (2014) 169–189.
- [86] P. T.-W. Lee, K.-C. Hu, T. Chen, External costs of domestic container transportation: Short-sea shipping versus trucking in taiwan, *Transport Reviews* 30 (3) (2010) 315–335.
- [87] B. Slack, A. Frémont, Transformation of port terminal operations: from the local to the global, *Transport Reviews* 25 (1) (2005) 117–130.
- [88] Y. Xie, C. Zhang, Q. Lai, China’s rise as a major contributor to science and technology, *Proceedings of the National Academy of Sciences* 111 (26) (2014) 9437–9442.
- [89] B. Owens, China’s research clout leads to growth in homegrown science publishing, *Nature* 630 (8015) (2024) S2–S4.
- [90] G. Zhang, J. Xu, Y. Wang, A study on the promotion effect of government guidance on the construction of a national unified market logistics channel, *Plos one* 18 (11) (2023) e0293969.
- [91] L. Hausmann, A. Mittal, G. Rome, [Lessons from growth outperformers in logistics](https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/lessons-from-growth-outperformers-in-logistics), accessed: 2025-08-06 (January 2021).
URL <https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/lessons-from-growth-outperformers-in-logistics>
- [92] K. Kwon, S. Jun, Y.-J. Lee, S. Choi, C. Lee, Logistics technology forecasting framework using patent analysis for technology roadmap, *Sustainability* 14 (9) (2022) 5430.
- [93] D. Chavarro, P. Tang, I. Ràfols, Why researchers publish in non-mainstream journals: Training, knowledge bridging, and gap filling, *Research policy* 46 (9) (2017) 1666–1680.

- [94] W. Hsieh, Z. Bi, J. Liu, B. Peng, S. Zhang, X. Pan, J. Xu, J. Wang, K. Chen, C. H. Yin, et al., Deep Learning, Machine Learning–Digital Signal and Image Processing: From Theory to Application, arXiv preprint arXiv:2410.20304 (2024).
- [95] S. Gu, X. Luo, An efficient algorithm of container code characters location, in: 2010 2nd International Conference on Computer Engineering and Technology, Vol. 6, IEEE, 2010, pp. V6–180.
- [96] J.-H. Jang, J.-H. Ku, K.-S. Hong, J.-H. Kim, Two-stage recognition of freight train ID number under outdoor environment, in: Proceedings of 3rd International Conference on Document Analysis and Recognition, Vol. 2, IEEE, 1995, pp. 986–990.
- [97] I. S. Igual, G. A. Garcia, A. P. Jimenez, Preprocessing and recognition of characters in container codes, in: 2002 International Conference on Pattern Recognition, Vol. 3, IEEE, 2002, pp. 143–146.
- [98] B. Shi, N. Yu, J. Xu, Q. Zhao, Extraction and recognition alphabetic and digital characters on industrial containers, in: 2009 International conference on computational intelligence and security, Vol. 1, IEEE, 2009, pp. 340–343.
- [99] S. Kumano, K. Miyamoto, M. Tamagawa, H. Ikeda, K. Kan, Development of a container identification mark recognition system, Electronics and Communications in Japan (Part II: Electronics) 87 (12) (2004) 38–50.
- [100] Z. He, J. Liu, H. Ma, P. Li, A new automatic extraction method of container identity codes, IEEE Transactions on intelligent transportation systems 6 (1) (2005) 72–78.
- [101] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors, nature 323 (6088) (1986) 533–536.
- [102] K.-B. Kim, Y. W. Woo, H.-K. Yang, An intelligent system for container image recognition using ART2-based self-organizing supervised learning algorithm, in: Asia-Pacific Conference on Simulated Evolution and Learning, Springer, 2006, pp. 897–904.
- [103] K. Moszyk, M. Deja, M. Dobrzynski, [Automation of the Road Gate Operations Process at the Container Terminal—A Case Study of DCT Gdańsk SA](#), in: Proceedings of Sustainability, MDPI, 2021, p. 6291. doi:10.3390/su13116291. URL <https://www.mdpi.com/2071-1050/13/11/6291>

- [104] L. Yao, C. Tang, Y. Wan, Advanced Text Detection of Container Numbers via Dual-Branch Adaptive Multi-Scale Network, *Applied Sciences* 15 (3) (2025) 1492.
- [105] Z. Lin, C. Dong, Y. Wan, Research on Intelligent Recognition Algorithm of Container Numbers in Ports Based on Deep Learning, in: *International Conference on Intelligent Computing*, Springer, 2024, pp. 184–196.
- [106] C.-S. Fahn, B.-Y. Su, M.-L. Wu, Vision-based identification extraction for freight containers, in: *2015 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, IEEE, 2015, pp. 51–57.
- [107] L.-d. Chen, W.-m. Zeng, N.-z. Wang, Container-code pattern recognition based on attribute grid computing, in: *2013 IEEE International Conference on Granular Computing (GrC)*, IEEE, 2013, pp. 33–37.
- [108] K.-m. Koo, E.-y. Cha, A novel container ISO-code recognition method using texture clustering with a spatial structure window, *International Journal of Advanced Science and Technology* 41 (4) (2012) 83–92.
- [109] W. Wu, Z. Liu, M. Chen, X. Yang, X. He, An automated vision system for container-code recognition, *Expert Systems with Applications* 39 (3) (2012) 2842–2855.
- [110] H.-C. Chen, C.-K. Chen, F.-Y. Hsu, Y.-S. Lin, Y.-T. Wu, Y.-N. Sun, A computer vision system for automated container code recognition, in: *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, Vol. 1, 2011.
- [111] V. K. Chauhan, S. Singh, A. Sharma, HCR-Net: A deep learning based script independent handwritten character recognition network, *Multimedia Tools and Applications* 83 (32) (2024) 78433–78467.
- [112] C. B. Murthy, M. F. Hashmi, N. D. Bokde, Z. W. Geem, Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—A comprehensive review, *Applied sciences* 10 (9) (2020) 3280.
- [113] C. Angelica, H. Purnama, F. Purnomo, et al., Impact of computer vision with deep learning approach in medical imaging diagnosis, in: *2021 1st international conference on computer science and artificial intelligence (ICCSAI)*, Vol. 1, IEEE, 2021, pp. 37–41.

- [114] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems* 25 (2012).
- [115] W. Jeon, G. Ko, J. Lee, H. Lee, D. Ha, W. W. Ro, Deep learning with GPUs, in: *Advances in computers*, Vol. 122, Elsevier, 2021, pp. 167–215.
- [116] M. Pandey, M. Fernandez, F. Gentile, O. Isayev, A. Tropsha, A. C. Stern, A. Cherkasov, The transformational role of GPU computing and deep learning in drug discovery, *Nature Machine Intelligence* 4 (3) (2022) 211–221.
- [117] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, software available from tensorflow.org (2015).
URL <https://www.tensorflow.org/>
- [118] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library, *arXiv preprint arXiv:1912.01703* (2019).
- [119] Y. LeCun, Y. Bengio, et al., Convolutional networks for images, speech, and time series, *The handbook of brain theory and neural networks* 3361 (10) (1995) 1995.
- [120] W. Al-Khawand, S. Kadry, R. Bozzo, K. Smaili, 8-neighborhood variant for a better container code extraction and recognition, *International Journal of Computer Science and Information Security* 14 (4) (2016) 182.
- [121] Y. Liu, T. Li, L. Jiang, X. Liang, Container-code recognition system based on computer vision and deep neural networks, in: *AIP conference proceedings*, Vol. 1955, AIP Publishing LLC, 2018, p. 040118.
- [122] T. Liu, X. Wu, F. Li, Lightweight container number recognition based on deep learning, *International Journal of System Assurance Engineering and Management* 16 (3) (2025) 1058–1071.

- [123] A. Diaz-Diaz, F. Parrilla, R. De La Iglesia, R. Barea, L. M. Bergasa, Wagon and Container Codes Detection and Recognition based on YOLOv8, in: 2024 7th Iberian Robotics Conference (ROBOT), IEEE, 2024, pp. 1–6.
- [124] H.-S. Nguyen, C.-D. Huynh, N.-Q. Bui, Digital transformation for shipping container terminals using automated container code recognition, TELKOMNIKA (Telecommunication Computing Electronics and Control) 21 (3) (2023) 535–544.
- [125] D. Yang, G. Wang, M. Liu, S. Yue, H. Zhang, X. Chen, M. Zhang, Lightweight container code recognition based on multi-reuse feature fusion and multi-branch structure merger, Journal of Real-Time Image Processing 20 (6) (2023) 108.
- [126] R. Zhang, Z. Bahrami, Z. Liu, A vertical text spotting model for trailer and container codes, IEEE Transactions on Instrumentation and Measurement 70 (2021) 1–13.
- [127] N. T. Hoa, N. Van Tao, Build an Effective System for Container Code Recognition, in: P. T. Nghia, V. D. Thai, N. T. Thuy, L. H. Son, V.-N. Huynh (Eds.), Proceedings of ICTA 2023: Advances in Information and Communication Technology, Vol. 847 of Lecture Notes in Networks and Systems, Springer, Cham, 2023. doi:10.1007/978-3-031-49529-8_4.
- [128] X. Chen, L. Jin, Y. Zhu, C. Luo, T. Wang, Text recognition in the wild: A survey, ACM Computing Surveys (CSUR) 54 (2) (2021) 1–35.
- [129] C. Zhang, W. Ding, G. Peng, F. Fu, W. Wang, Street view text recognition with deep learning for urban scene understanding in intelligent transportation systems, IEEE Transactions on Intelligent Transportation Systems 22 (7) (2020) 4727–4743.
- [130] Y. Zhu, C. Yao, X. Bai, Scene text detection and recognition: Recent advances and future trends, Frontiers of Computer Science 10 (1) (2016) 19–36.
- [131] S. Long, X. He, C. Yao, Scene text detection and recognition: The deep learning era, International Journal of Computer Vision 129 (1) (2021) 161–184.
- [132] F. Naiemi, V. Ghods, H. Khalesi, Scene text detection and recognition: a survey, Multimedia Tools and Applications 81 (14) (2022) 20255–20290.
- [133] M. Jaderberg, A. Vedaldi, A. Zisserman, Deep features for text spotting, in: European conference on computer vision, Springer, 2014, pp. 512–528.

- [134] X. Zhang, Y. Su, S. Tripathi, Z. Tu, Text spotting transformers, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 9519–9528.
- [135] L. Mei, J. Guo, Q. Liu, P. Lu, A novel framework for container code-character recognition based on deep learning and template matching, in: 2016 International conference on industrial informatics-computing technology, intelligent technology, industrial information integration (ICIICII), IEEE, 2016, pp. 78–82.
- [136] S. Long, S. Qin, D. Panteleev, A. Bissacco, Y. Fujii, M. Raptis, Towards end-to-end unified scene text detection and layout analysis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 1049–1059.
- [137] Z. Shen, J. Wang, Y. Zhang, L. Zheng, C. Mi, Y. Shen, [Identification and Positioning Method of Bulk Cargo Terminal Unloading Hopper Based on Monocular Vision Three-Dimensional Measurement](#), Journal of Marine Science and Engineering 12 (8) (2024) 1282. doi:10.3390/jmse12081282. URL <https://doi.org/10.3390/jmse12081282>
- [138] G. Fancello, P. Serra, A. Schintu, A. Zoratti, et al., Performance evaluation of a tracking system for intermodal traffic: an experimentation in the Tyrrhenian area, EUROPEAN TRANSPORT/TRASPORTI EUROPEI (76-Special Issue 1 “TIS Roma 2019”) (2020) 1–12.
- [139] J. Muñuzuri, L. Onieva, P. Cortés, J. Guadix, Using IoT data and applications to improve port-based intermodal supply chains, Computers & Industrial Engineering 139 (2020) 105668.
- [140] Y. Xu, Z. Liang, Y. Liang, X. Li, W. Pan, J. You, Z. Long, Y. Zhai, A. Genovese, V. Piuri, et al., Data-Driven Container Marking Detection and Recognition System With an Open Large-Scale Scene Text Dataset, IEEE Transactions on Emerging Topics in Computational Intelligence (2024).
- [141] C. Mi, L. Cao, Z. Zhang, Y. Feng, L. Yao, Y. Wu, A port container code recognition algorithm under natural conditions, Journal of Coastal Research 103 (SI) (2020) 822–829.
- [142] H. Choi, J. Jeong, C. Lee, S. Yun, K. Bang, J. Byun, Design and Implementation for BIC Code Recognition System of Containers using OCR and CRAFT

in Smart Logistics, WSEAS Transactions on Computer Research 11 (2023) 62–72.

- [143] J. C. Lee, Automatic character recognition for moving and stationary vehicles and containers in real-life images, in: IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339), Vol. 4, IEEE, 1999, pp. 2824–2828.
- [144] R. Sharma, S. R. Lee, A novel container iso code localization using an object clustering method with opencv and visual studio application, International Journal of Image, Graphics and Signal Processing 5 (7) (2013) 54.
- [145] M. Goccia, M. Bruzzo, C. Scagliola, S. G. Dellepiane, Recognition of Container Code Characters through Gray-Level Feature Extraction and Gradient-Based Classifier Optimization, in: International Conference on Document Analysis and Recognition (ICDAR), 2003, p. 973.
- [146] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou, S. Ghosh, A. Bagdanov, M. Iwamura, J. Matas, L. Neumann, V. R. Chandrasekhar, S. Lu, et al., ICDAR 2015 competition on robust reading, in: 2015 13th international conference on document analysis and recognition (ICDAR), IEEE, 2015, pp. 1156–1160.
- [147] L. Yuliang, J. Lianwen, Z. Shuaitao, Z. Sheng, Detecting curve text in the wild: New dataset and new solution, arXiv preprint arXiv:1712.02170 (2017).
- [148] A. Gupta, A. Vedaldi, A. Zisserman, Synthetic data for text localisation in natural images, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2315–2324.
- [149] V. Lcvenshtcin, Binary coors capable or ‘correcting deletions, insertions, and reversals, in: Soviet physics-doklady, Vol. 10, 1966.
- [150] W. Wu, Y. Zhao, M. Z. Shou, H. Zhou, C. Shen, Diffumask: Synthesizing images with pixel-level annotations for semantic segmentation using diffusion models, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 1206–1217.
- [151] J. G. Tejero, M. S. Zinkernagel, S. Wolf, R. Sznitman, P. Márquez-Neila, Full or Weak annotations? An adaptive strategy for budget-constrained annotation campaigns, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 11381–11391.

- [152] T. Guan, W. Shen, X. Yang, X. Wang, X. Yang, Bridging synthetic and real worlds for pre-training scene text detectors, in: European Conference on Computer Vision, Springer, 2024, pp. 428–446.
- [153] S. Long, C. Yao, Unrealtext: Synthesizing realistic scene text images from the unreal world, arXiv preprint arXiv:2003.10608 (2020).
- [154] C. Dong, C. C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE transactions on pattern analysis and machine intelligence 38 (2) (2015) 295–307.
- [155] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Advances in neural information processing systems 27 (2014).
- [156] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434 (2015).
- [157] C. Li, S. Liu, Q. Xia, H. Wang, H. Chen, Automatic container code localization and recognition via an efficient code detector and sequence recognition, in: 2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), IEEE, 2019, pp. 532–537.
- [158] X. Feng, Z. Wang, T. Liu, Port container number recognition system based on improved YOLO and CRNN algorithm, in: 2020 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), IEEE, 2020, pp. 72–77.
- [159] S. Fang, H. Xie, Y. Wang, Z. Mao, Y. Zhang, Read like humans: Autonomous, bidirectional and iterative language modeling for scene text recognition, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 7098–7107.
- [160] Y. Zhang, C. Liu, J. Wei, X. Yang, Y. Zhou, C. Ma, X. Ji, Linguistics-aware masked image modeling for self-supervised scene text recognition, in: Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 9318–9328.
- [161] T. Gruchmann, A. Jazairy, Big brother is watching you: Examining truck drivers’ acceptance of road-facing dashcams, Transportation Research Part F: Traffic Psychology and Behaviour 111 (2025) 316–330.

- [162] Y. Mekdad, A. Aris, L. Babun, A. El Fergougui, M. Conti, R. Lazzeretti, A. S. Uluagac, A survey on security and privacy issues of uavs, *Computer networks* 224 (2023) 109626.
- [163] V. Chang, P. Chundury, M. Chetty, Spiders in the sky: User perceptions of drones, privacy, and security, in: *Proceedings of the 2017 CHI conference on human factors in computing systems*, 2017, pp. 6765–6776.