A CNN-RNN Hybrid Approach for Polish License Plate Recognition: Harnessing Transfer Learning and Real-World Validation

Gergő B. Békési* and Péter Ekler
Department of Automation and Applied Informatics,
Faculty of Electrical Engineering and Informatics,
Budapest University of Technology and Economics, Budapest, Hungary
**Corresponding author: Gergő B. Békési (bekesi.gergo@aut.bme.hu)*

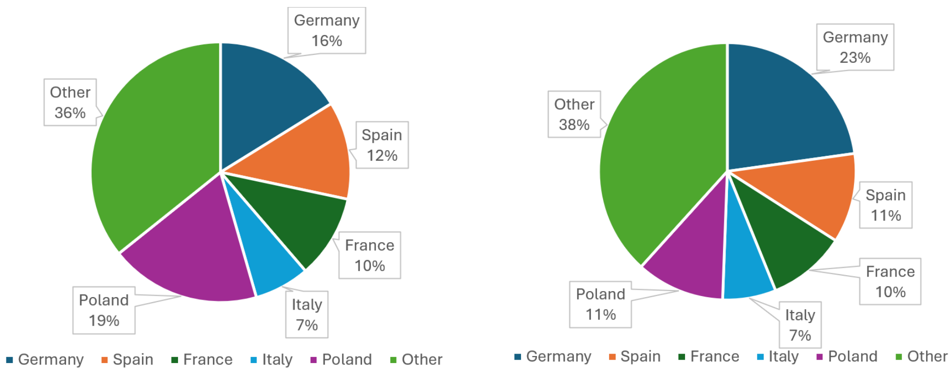
Submitted: Jan 1, 2025 **Accepted:** May 15, 2025 **Published:** Jun 23, 2025

Abstract Automated license plate recognition (LPR) systems have garnered substantial attention within the field of intelligent transportation systems, owing to their pivotal role in facilitating toll collection, enhancing traffic management, and ensuring operational efficiency. Despite recent breakthroughs in convolutional and recurrent neural network architectures, Polish LPR remains underexplored, with most existing approaches relying on conventional optical character recognition. This study proposes a hybrid convolutional neural network – recurrent neural network (CNN-RNN) model equipped with a Thin-Plate Spline (TPS) transformation module, a ResNet-based feature extractor, a bidirectional Long Short-Term Memory (LSTM) sequence model, and an attention-based decoder to address the unique challenges of Polish license plates. The model is trained on a high-difficulty dataset, comprising real-world images without explicit character-level bounding boxes. Empirical evaluations underscore the efficacy of the proposed system, with competitive accuracy and normalized edit distance scores achieved on Polish, Czech, Hungarian, and Slovak datasets. Additionally, transfer learning from closely related Central European plate formats to Polish data demonstrates marked improvements in convergence and overall performance. Further validation on a challenging video-based dataset reveals the robustness of the proposed approach, evidencing its potential applicability in real-world scenarios and highlighting majority voting as an effective strategy to enhance system reliability under variable conditions.

Keywords: Polish license plate recognition, CNN-RNN, transfer learning, majority voting.

1. Introduction

Automated systems significantly enhance operational efficiency, mitigate human error, and facilitate expedited decision-making processes. Within the domain of transportation engineering, active traffic management strategies employ real-time data acquisition and predictive algorithms to optimise roadway performance and alleviate congestion through the dynamic regulation of traffic flow, particularly during peak hours or in response to incidents [30]. Intelligent transportation systems integrate advanced communication and sensor technologies to enhance traffic operations. These systems enable functionalities such as vehicle-to-infrastructure communication and centralized monitoring, thereby reducing delays and improving overall safety [1]. Automated lane control managed by centralized lane coordination mechanisms, facilitate the adaptive allocation of road lanes based on traffic demand. This approach enhances roadway capacity and



a Distribution in million vehicle-kilometres. **b** Distribution in thousand basic transport operations.

Fig. 1. Road freight transport by distance in the EU, Norway, and Switzerland [10].

mitigates bottlenecks during periods of high demand [23]. Furthermore, dynamic traffic flow optimisation utilises sophisticated modelling techniques and real-time data inputs to adjust parameters such as speed limits and lane assignments, thereby minimizing travel times and reducing fuel consumption [50]. Traffic incident management systems leverage automated alert technologies and resource coordination to enable prompt responses to roadway incidents, thus reducing disruptions and improving traffic safety [24]. Adaptive traffic signal control systems, employing machine learning algorithms, dynamically adjust signal timings to accommodate real-time traffic conditions, effectively reducing delays and emissions [44].

The exponential growth of global transportation highlights the pivotal role of automation in state-managed road infrastructure. This extends to freight transportation, where automated toll collection systems, leveraging license plate recognition technology, ensure precise and reliable monitoring [13]. These systems are integral to optimising operational efficiency and sustaining the economic viability of road networks amidst rising transportation demands.

Within the European Union, freight transportation, as measured in vehicle-kilometres, demonstrates significant variability among member states. As depicted in Figure 1, Poland ranks highest in vehicle-kilometres for freight transport and is also dominant in the number of basic transport operations, underscoring its pivotal role within the European logistics network.

Poland's strategic geographic location and well-developed infrastructure position it as a vital transit hub, facilitating connections between Eastern and Western Europe. The dataset also includes Norway and Switzerland, which, while not EU members, maintain close economic ties with the Union. Notably, no data for Liechtenstein was available in the Eurostat database [10]. Therefore, the implementation of efficient LPR systems is particularly critical in Poland, as these systems play a vital role in ensuring the seamless management of transit traffic and optimising the utilisation of road infrastructure within the European logistics network.

The structure of the remainder of this paper is as follows. In Section 2, an overview is provided of state-of-the-art methodologies for LPR, analysing the unique challenges associated with recognising Polish license plates and positioning Polish LPR-focused projects within the broader context of related research. Section 3 introduces an approach to address a gap in the Polish LPR literature by integrating RNNs with CNNs. It also describes the dataset used in this study, characterized by its high level of difficulty, and assesses the efficacy of transfer learning in applying machine learning models trained on license plate datasets from other countries. Section 4 presents the experimental results, while Section 5 concludes the study by summarising the key findings and their implications.

2. LPR and Polish LPR

Polish LPR differs from the LPR systems of other nations not only in terms of license plate format but also in the technologies employed. In this section, the state-of-the-art is introduced, the unique features of the Polish license plate format are highlighted, and the characteristics of Polish LPR projects are described.

2.1. State-of-the-art in LPR

Over the past decade, LPR has evolved from relying on handcrafted features and rule-based heuristics – such as edge detection [35], morphological operations [49], and character segmentation using thresholds [20] – to employing deeply learned representations implemented through various neural network architectures.

While handcrafted methods relied heavily on prior knowledge and manually tuned parameters to detect and recognise license plates under constrained conditions, deep learning-based techniques learn hierarchical features directly from data, enabling improved robustness and generalization across varying environments.

Contemporary LPR solutions can generally be categorized into three main deep learning paradigms: CNN-based pipelines [39], RNN-based frameworks [31], and single-shot detection methods, such as those built on You Only Look Once (YOLO) [2] or Single Shot Detector (SSD) [40] architectures. Each approach entails distinct trade-offs in terms of model complexity, speed, accuracy, and annotation requirements [22].

CNN-based LPR solutions typically follow a pipeline in which character recognition occurs after an initial license plate detection stage. Early methods in this category often employed separate subsystems: one CNN for plate detection, another for character segmentation, and a final CNN for character classification. This step-by-step strategy first localizes the plate in the image, then segments individual characters, and finally classifies each character independently [39].

These CNN architectures leverage learned visual features extracted directly from raw pixel data. They are proficient in accommodating variations in character fonts, plate backgrounds, and moderate viewpoint changes. However, a critical limitation is their reliance on explicit character segmentation. This segmentation step usually requires tightly bounded boxes or heuristics to isolate each character before classification, thereby increasing data annotation complexity and the risk of segmentation errors [33]. Moreover, while CNN-based methods have demonstrated state-of-the-art accuracy under controlled conditions, they can falter when facing substantial geometric distortions or highly non-uniform illumination.

RNN-based LPR architectures are designed to overcome some of the inherent limitations of purely CNN-driven pipelines, particularly regarding character segmentation [7]. Instead of explicitly segmenting characters, RNN-based methods conceptualize the recognition task as a sequence labelling problem. The network ingests either raw plate images or CNN-extracted features and generates a sequence of predicted characters. This approach eliminates the need for character-level bounding boxes since only a correct global sequence label is required [31]. As a result, annotation becomes considerably simpler, as labelling the entire license plate string suffices.

RNN-based models effectively handle variations in character length and spacing while preserving contextual dependencies between characters [16]. This leads to enhanced robustness, particularly for plates with irregular spacing or slight rotations. However, RNN-based solutions may introduce a higher computational overhead, and their performance can depend on the quality of the initial features, often provided by a preceding CNN. Despite these computational costs, these models gracefully circumvent the segmentation challenge, frequently improving robustness and reducing manual annotation efforts [47].

Single-shot detectors, including YOLO and SSD, represent another prominent approach to LPR. Unlike multi-stage pipelines, single-shot methods attempt detection and, in some cases, recognition within a single feed-forward pass [5]. YOLO-based architectures divide the image into a grid and directly predict bounding boxes and class probabilities [2], thereby achieving remarkable inference speed. SSD similarly applies a single forward pass to detect objects at multiple scales, which can be advantageous for handling plates of varying sizes and positions [40].

A key advantage of single-shot methods is their computational efficiency, enabling near real-time recognition on standard GPU hardware. Their speed is appealing for real-world applications requiring high throughput, such as toll booths, traffic monitoring, or large-scale surveillance systems. However, these models can struggle with very small plates, especially when the field of view is extensive, or when plate distortion is significant. Although single-shot methods can be adapted to handle character recognition, this often involves treating each character as a separate object [27]. Without careful calibration or sophisticated refinement steps, their accuracy may not match that of specialised CNN- or RNN-based solutions, particularly under challenging conditions.

CNN-based and RNN-based approaches can produce highly accurate results but may be slower due to multi-stage processing or the overhead associated with recurrent computations. In contrast, single-shot models deliver improved speed, which is invaluable for real-time applications, albeit potentially at the expense of slightly lower accuracy in complex scenarios. Furthermore, CNN-driven and single-shot pipelines typically require bounding boxes for both plates and individual characters, whereas RNN-based methods greatly simplify annotation by relying solely on the license plate's textual content. All three approaches address variations in plate appearance, lighting, and orientation. CNN-based methods depend on pre-processing and carefully engineered pipelines, RNN-based methods accommodate irregular spacing and character sequences without explicit segmentation, and single-shot detectors excel at rapid processing but may require meticulous tuning to handle small plates and severe distortions effectively.

The choice among CNN-based, RNN-based, or single-shot solutions ultimately depends on the priorities of the application. CNN architectures remain strong candidates when high accuracy is paramount and per-character bounding box annotations are readily available. RNN-based approaches simplify labelling and segmentation tasks, rendering them highly attractive in diverse and distorted plate conditions. YOLO and SSD models stand out in scenarios where speed is critical, even if doing so involves a trade-off against absolute accuracy.

2.2. Polish LPR

Polish license plates adhere to the standardized European format while incorporating unique features defined by national regulations. This subsection provides an overview of the Polish license plate standard and analyses how Polish LPR systems diverge from the global state-of-the-art approaches.

2.2.1. Polish License Plate Format

Polish vehicle registration plates are standardized under national regulations, ensuring uniformity and comprehensive traceability throughout the country. Contemporary registration plates comply with the European Union's standardized aesthetic guidelines, featuring black alphanumeric characters embossed on a white, reflective background, along with a blue European stripe displaying the "PL" country code and the EU stars. This harmonized design has been adopted to enhance visibility, interoperability, and consistency within the broader European context.

A hierarchical geographic encoding scheme underpins the Polish plate system. Each plate's prefix encodes its origin at both the voivodeship (province) and county levels.

Fig. 2. A Polish license plate from the dataset.

Specifically, the initial letter denotes the voivodeship, while subsequent letters specify the county. Additional letters may follow to represent larger urban centres or historically significant administrative divisions. These area codes are succeeded by a series of digits, and in some instances, additional letters, forming a unique alphanumeric identifier for each vehicle. To address the diminishing pool of available combinations in densely populated regions, supplementary leading characters have been introduced, thus expanding the combinatory space without necessitating a comprehensive overhaul of the national registration format. A Polish license plate, illustrating this hierarchical geographic encoding scheme, is presented in Figure 2.

In addition to standard passenger vehicle plates, a range of specialised categories is employed to accommodate varying vehicle types and operational contexts. Different configurations and dimensions are used for motorcycles, mopeds, and agricultural vehicles, often employing smaller, two-line layouts. Reduced-size plates exist for vehicles imported from the United States and Japan to ensure proper fit. Moreover, vehicles utilising electric or hydrogen propulsion are issued plates with a green background, a feature that aids in regulating privileges such as bus lane access. Other specialised variants—such as those for historical cars, temporary registrations, vehicle testing, professional dealerships, diplomatic missions, military units, the police, and other government services—adhere to specific colour schemes, character arrangements, and symbolic elements. In doing so, the system accommodates diverse transportation needs while maintaining a coherent and functionally robust identification framework [34].

2.2.2. Polish LPR Projects

Despite the widespread adoption of CNNs, single-shot methods, and CNN-RNN hybrid approaches in contemporary LPR systems, the domain of Polish LPR exhibits a distinct reliance on conventional OCR-based methodologies. While CNNs are occasionally integrated within OCR pipelines, primarily for character recognition tasks, and single-shot detection frameworks are sporadically employed, the utilisation of RNNs remains notably absent across the reviewed Polish LPR projects.

Polish researchers have demonstrated notable successes in various computer vision applications—ranging from object detection [45] and face recognition [32] to medical image analysis [26] and aerial surveillance [12]. Nevertheless, these achievements have not been systematically extended to the LPR domain, perhaps due to the long-standing availability of commercial and open-source OCR-based solutions. Consequently, large-scale, specialized approaches for license plate recognition have thus far received comparatively less attention. This subsection reviews the existing works, illustrating how classical segmentation and recognition pipelines dominate the Polish domain and highlighting the conspicuous absence of RNN-based methods.

Kluwak et al. [25] employ multi-frame recognition and object tracking to enhance single-frame license plate recognition. Although the authors introduce a novel mechanism for combining results from consecutive video frames, known as In Track Clustering Correction (ITCC), their approach relies on established computer vision techniques through the default settings of OpenALPR [38], which is a computer vision based software for LPR tasks written in C++. While the authors do not provide a detailed description of preprocessing, one may infer that it involves binarization and filtering to address noise under adverse weather conditions, followed by morphological operations such as dilation and erosion and subsequent histogram transformations. These steps likely facilitate character segmentation through the detection of connected components, often referred to as tracks of recognised characters, and are then followed by CNN-based character recognition. Subsequently, the outputs of conventional optical character recognition are refined through a clustering-based voting process across frames. Although alternative strategies could be employed for single-frame LPR in the presented pipeline, the authors continue to use these standard methods in their implementation.

Janowski et al. [18] investigate the effectiveness of LPR in video streamed under constrained networking conditions, focusing on both human and automatic recognition systems. The study examines two custom-built LPR algorithms, the Labelling and Artificial Neural Networks (LANN) and the Periodic Walsh Piecewise-Linear Descriptors (PWPLD). The LANN algorithm follows a traditional pipeline, involving preprocessing with Otsu binarization and morphological operations, license plate localization through connected component analysis, segmentation for character extraction, and recognition via a neural network trained on scaled character images. The PWPLD algorithm incorporates preprocessing with contour-based features, license plate detection using brightness regularities, and recognition through decision-tree rules generated by See5 [41] software. Both methods were evaluated against human recognition, demonstrating that while human accuracy exceeded that of the algorithms, the LANN method performed comparably under certain conditions aided by adaptive feedback mechanisms and parameter tuning. The authors measured performance under both compressed conditions (CC) and no compression (NC), using accuracy and the $(1 + \mathrm{LED})^{-1}$ as evaluation measures, where LED is the Levenshtein distance [29] sum to character number sum ratio.

Kekez [21] proposes a multi-stage approach for real-time LPR that utilises a cascade of classifiers to achieve high accuracy while minimizing training data requirements. The process begins with preprocessing the license plate image, which includes greyscale conversion, binarization, negation, and morphological operations to enhance the plate's features. Character segmentation is performed using connected component analysis, with optional dilation or erosion to address challenges such as poor lighting or occlusions. The character recognition stage employs a cascade of classifiers, starting with a neural network for initial classification, followed by parallel Random Forest, decision tree, or rule-based classifiers to refine results. A final stage applies contextual rules to improve accuracy further, such as disambiguating similar characters based on their position within the plate. The proposed system achieves over 98% recognition accuracy by leveraging small, country-specific training datasets derived from legal specifications rather than real license plate images, offering a computationally efficient alternative to deep learning methods like YOLO, which the author also highlights as an important architecture in the field.

Leszczuk et al. [28] propose an objective evaluation methodology for assessing video quality tailored to automatic LPR under various challenging conditions. The approach integrates a dataset of video frames annotated with ground truth coordinates, simulating real-world scenarios such as occlusions, motion blur, and lighting variations. License plate recognition is performed using the aforementioned OpenALPR library, which combines traditional image processing techniques with OCR to detect and classify license plates. The recognition pipeline utilises confidence scores and country-specific settings to improve accuracy. The study highlights the influence of video distortions on LPR performance, such as Gaussian noise, defocus, and JPEG compression, and demonstrates the limitations of standard video quality metrics in recognition tasks. The methodology achieves an F-measure of 0.777, underscoring its potential as a robust framework for evaluating and enhancing video-based LPR systems in diverse operational conditions.

Wroblewski [46] introduces a real-time algorithm for recognising Polish car license plates, achieving a recognition rate of 60-70%. The method follows a three-stage process: plate localization, symbol segmentation, and symbol recognition. Plate localization involves detecting and isolating the license plate area from the input image using a modified edge detection algorithm that emphasizes vertical edges, as these are prominent in license plate regions. Symbol segmentation extracts individual characters from the plate image, while recognition is performed using a neural network-based approach. The algorithm is designed for cost efficiency, requiring only the computational power of a personal computer.

Łubkowski and Laskowski [52] investigate the quality and reliability of automatic LPR systems in the context of diverse environmental and technical conditions. Their analysis focuses on critical factors such as light intensity, image resolution, and camera angles, and how these influence the performance of license plate detection and recognition. The automatic LPR workflow encompasses image acquisition, localization of the license plate, segmentation of individual characters, and their subsequent recognition, facilitated by optical character recognition and syntactic validation algorithms. The researchers employed the Simple-LPR application in a controlled experimental setup, testing its functionality under varying scenarios. The results reveal that the system achieves reliable recognition under favourable lighting conditions, specifically light intensities ranging from 5000 to 7000 lux, and at standard image resolutions. However, when using low-resolution images performance declines significantly.

Jureczko and Uherek [19] propose a video-based automatic LPR system utilising CNNs for accurate and efficient character recognition. The system employs Tiny YOLO for vehicle detection and localization within video frames, isolating regions of interest (ROIs) containing license plates. These ROIs undergo image processing steps, including Gaussian blur, thresholding, and contour detection using OpenCV [4], to extract license plates. Character recognition is performed with the Tesseract OCR [42] library, enhanced by multiple thresholding iterations and a voting mechanism to ensure consistent results. To optimise processing, a dictionary-based structure tracks previously recognised plates, avoiding redundant computations. Tests demonstrated the system's ability to accurately recognise license plates in controlled lighting conditions, processing frames at 0.62 seconds per frame. While effective in stable environments, the system's performance is hindered by adverse weather conditions and inconsistent lighting, highlighting the need for further improvements for robust, real-world applications.

The reviewed projects highlight a predominant reliance on OCR-based solutions and a notable absence of RNN-based approaches in Polish LPR systems. This study aims to introduce a CNN-RNN-based solution while acknowledging that current Polish LPR solutions primarily focus on OCR. The proposed model is designed to be potentially integrable into widely popular frameworks such as the EasyOCR pipeline [17]. Nonetheless, the central focus of this research is the training of the CNN-RNN model. It aims to contribute an original Polish LPR solution anchored in a CNN-RNN approach, thereby addressing the notable gap in current research and paving the way for further advancements in accurate, efficient, and versatile license plate recognition for Poland. Additionally, this study investigates how models trained on non-Polish license plates can be effectively fine-tuned to adapt to Polish license plates, emphasizing the importance of adaptation and transfer learning in this domain.

3. Methods and data

In the subsequent subsections, we present the overall approach for the proposed Polish LPR system, beginning with a detailed explanation of the hybrid CNN-RNN method. This covers the role of a Thin-Plate Spline transformation, a ResNet-based feature extractor, and an attention-based decoder within the recognition pipeline. We then introduce the data used to train and validate our model, emphasizing the challenges inherent in real-world license plate images and the reasons behind choosing the data resource.

3.1. Methods

A typical CNN-RNN approach for text recognition, including license plate recognition, typically involves three primary stages: feature extraction, sequence modelling, and prediction [51]. In certain scenarios, an additional preliminary transformation stage may be incorporated before feature extraction to address geometric distortions. When significant geometric variations such as curved text or perspective distortions are anticipated, researchers often include this optional transformation stage at the beginning of the pipeline. Methods such as Spatial Transformer Networks are frequently employed to rectify input images, thereby simplifying subsequent processing and enhancing overall recognition performance [37].

In the standard architecture, regardless of whether the transformation stage is included, a CNN is initially used to extract high-level features that minimize irrelevant elements such as background clutter or font variations. These features are then processed by an RNN, typically a Gated Recurrent Unit (GRU) [9], a LSTM network [43], or a bidirectional LSTM [15], to capture contextual relationships between characters within the textual sequence, such as the structural patterns of a license plate. Finally, the text is recognised using a decoder, which may utilise either Connectionist Temporal Classification (CTC) [8] to align input frames with output labels through a collapsing strategy or attention-based mechanisms that dynamically focus on specific segments of the feature sequence.

In the present work, we employ a variant of the CNN-RNN-based pipeline [3] that leverages a TPS transformation module, a ResNet-based [14] feature extractor, a bidirectional Long Short-Term Memory (BiLSTM) [11] sequence model, and an attention-based decoder to recognise license plates under various geometric and visual distortions. Mathematically, let $\mathbf{x} \in \mathbb{R}^{H \times W}$ denote the input license plate image, represented as a single-channel grayscale image of height H and width W. Our approach first applies a TPS Spatial Transformer Network (TPS-STN) that learns a parametric transformation to rectify distorted inputs. We define a set of F fiducial control points $(C_i^x, C_i^y)_{i=1}^F$ in normalized coordinates on the input image. The TPS transformation warps $\mathbf{x}$ to a canonical form $\mathbf{x}' \in \mathbb{R}^{H' \times W'}$ via the mapping:

$$T_{\text{TPS}} : (u, v) \mapsto (u', v') = \phi\left((u, v) \mid \theta\right), \tag{1}$$

where ϕ is obtained by solving a system of radial basis functions centered at the fiducial points, and θ encapsulates the learnable parameters of the localization sub-network. By aligning the text region into a more "straightened" coordinate frame, the TPS module simplifies subsequent recognition.

Once the image is rectified, it is passed to the ResNet-based feature extractor. Denoting the rectified image as $\mathbf{x}'$, a series of convolutional and downsampling layers $f_{\text{ResNet}}(\cdot)$

yield a deep feature map:

$$\mathbf{F} = f_{\text{ResNet}}(\mathbf{x}') \in \mathbb{R}^{H_f \times W_f \times C_f}, \tag{2}$$

where H_f, W_f, and C_f respectively denote the height, width, and number of channels in the final feature representation. The ResNet backbone is composed of stacked residual blocks, which improve gradient flow and mitigate vanishing gradients, facilitating feature learning even in complex visual environments such as license plates with varied fonts or cluttered backgrounds.

Next, we reshape the spatial dimensions $H_f \times W_f$ of $\mathbf{F}$ into a sequence of feature vectors along the width dimension:

$$\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_T = \text{reshape}(\mathbf{F}), \tag{3}$$

where $T = H_f \times W_f$. These vectors are then fed into a bidirectional LSTM:

$$\mathbf{h}_t = \text{BiLSTM}(\mathbf{f}_t, \mathbf{h}_{t-1}), \tag{4}$$

where $\mathbf{h}_t \in \mathbb{R}^d$ is the hidden state at step t, combining both forward and backward passes. The LSTM gating mechanisms selectively retain or discard information, allowing the model to capture dependencies among character regions. The bidirectionality is particularly useful for recognizing alphanumeric patterns, as it captures context from both directions.

Finally, an attention-based decoder predicts the output character sequence. Denote the full set of BiLSTM hidden states as $\mathbf{H} = \mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_T$. At decoding step k, the attention mechanism computes a context vector $\mathbf{c}_k$ as:

$$\alpha_{k,t} = \text{softmax}\left(\mathbf{w}^\top \tanh\left(\mathbf{W}_h \mathbf{h}_t + \mathbf{W}_s \mathbf{s}_{k-1}\right)\right), \tag{5}$$

$$\mathbf{c}_k = \sum \alpha_{k,t} \mathbf{h}_t, \tag{6}$$

where $\alpha_{k,t} \in [0, 1]$ are the attention weights, $\mathbf{s}_{k-1}$ is the previous decoder state, and $\mathbf{w}$, $\mathbf{W}_h$, and $\mathbf{W}_s$ are learnable parameters. The decoder state then updates via:

$$\mathbf{s}_k = \text{LSTMCell}(\mathbf{c}_k, \mathbf{s}_{k-1}). \tag{7}$$

producing the distribution over possible next characters. By focusing on the most relevant portions of the feature sequence at each decoding step, attention avoids the assumption of rigid alignment, as in standard CTC, and is more robust to variable character spacing, occlusions, or partial distortions frequently found in license plate scenarios. The block diagram of the proposed architecture is illustrated in Figure 3.

This architecture was trained on Central European license plates, specifically from Poland, as well as from other Visegrád Group countries (Czech Republic, Hungary,

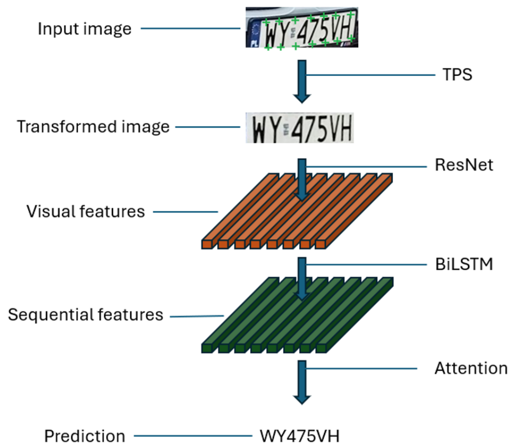


Fig. 3. Block diagram of the proposed architecture.

and Slovakia). The aim was to investigate whether transfer learning, applied to models initially trained on license plates from some geographically proximate countries, enhances performance. The characteristics of the dataset are detailed in the subsequent subsection.

The initial training phase consisted of 50 epochs using the Adadelta [48] optimisation algorithm. Adadelta was selected for this text recognition task due to its adaptive learning rate properties, which are particularly effective in scenarios where feature scaling and regularization are critical. A starting learning rate of 1.0 was employed, with a decay rate of 0.95 to ensure gradual adjustment of the learning rate during training. The epsilon parameter of Adadelta, set to 10^{-8}, played a key role in maintaining numerical stability, especially when updating the learning rate for sparse gradient scenarios. Gradient clipping was applied with a threshold of 5, which helps prevent exploding gradients, thereby stabilizing the training process.

Following the base training, a transfer learning phase, referred to here as retraining, was conducted for 30 epochs using the same hyperparameters. This phase started with the models trained on the specific datasets from the Czech Republic, Hungary, and Slovakia, and continued by retraining them on the Polish dataset. This approach aimed to refine the model's performance by leveraging domain-specific features from the respective datasets.

To ensure practical applicability in real-world scenarios, the developed models are designed to operate sequentially over temporally ordered image frames, such as those extracted from video streams. This temporal deployment capability allows the recognition pipeline to process each frame individually, yielding a predicted text output per frame. In the present study, this approach was employed during the testing phase:

after the transfer learning stage, the best-performing model – selected based on its validation accuracy – was evaluated on video sequences from the aforementioned Parking Database [28]. For each frame in a sequence, the model predicted a license plate text; the final recognised plate was then determined by selecting the most frequently predicted string across all frames (a strategy commonly referred to as majority voting, where the mode of individual predictions is taken as the final decision). This strategy enhances robustness against frame-level noise, motion blur, and partial occlusions, which are common in surveillance footage. The effectiveness of this method is illustrated later in Subsection 4.4.

The CNN-RNN architecture offers strong capabilities for sequential text recognition; its novelty in this work lies in its application to the recognition of Polish license plates – a task that has received limited attention in prior research. By embedding this architecture within a majority voting framework for video-based inference, the method demonstrates practical robustness in dynamic, real-world scenarios.

3.2. Data

The dataset utilised in this study originates from a platform dedicated to vehicle license plates and related topics, including their history, production technologies, usage, and statistical distribution. It serves as an extensive online repository where users can upload and share images of vehicles and license plates. The platform fosters an engaged community of enthusiasts who discuss vehicle characteristics, evaluate uploaded images, and exchange observations. Although license plates can be searched based on their alphanumeric identifiers, the website adheres to strict privacy policies and does not associate license plates with personal information about vehicle owners.

The dataset employed in this research specifically includes license plate images from Poland and the countries examined for transfer learning and was curated from publicly accessible resources. The images feature a wide range of real-world scenarios, including variations in lighting conditions, viewing angles, and plate appearances. Notably, this dataset does not include character-level bounding boxes or individual character annotations. Instead, each license plate is annotated with a single text label representing the complete license plate string. This characteristic aligns well with the requirements of RNN-based architectures, which excel in sequence labelling tasks by processing the entire plate string as a single sequence rather than relying on explicit character segmentation.

The absence of character-level bounding boxes, while simplifying annotation efforts, presents unique challenges for automated LPR systems. These challenges are particularly pronounced in datasets with diverse plate designs, irregular spacing, and distortions. Consequently, this dataset is considered more demanding than those providing detailed character-level annotations. However, its characteristics render it highly suitable for RNN-based methodologies, which inherently bypass the segmentation requirement by focusing on sequence-level recognition.

Fig. 4. Examples of the images from the dataset.

The utilisation of this dataset supports the development of Polish CNN-RNN hybrid models, emphasizing sequence labelling as a robust approach to LPR. Furthermore, leveraging this dataset underscores the importance of adapting algorithms to real-world, imperfect conditions, enhancing the applicability and generalizability of the resulting models to practical scenarios. For each of the four countries under study—Poland, the Czech Republic, Hungary, and Slovakia—exactly 1000 images were available, which were divided into a training-validation split of 800 and 200 images, respectively. Although the data source hosts a substantially larger number of license plate images overall, it only provides up to 1,000 images per country in publicly accessible form. Some examples from the dataset, illustrating its diversity and relevance to the proposed approach, are presented in Figure 4.

In designing the training set, special consideration was given to the distinct format and characteristics of Polish license plates, which differ from those of the other countries. For instance, Polish plates often feature regional codes encoded in their prefixes, and the overall alphanumeric structure may include unique character arrangements not found elsewhere. To capture these idiosyncrasies, the Polish subset was curated to include diverse plate designs (e.g., specialized plates for electric vehicles, two-line motorcycle plates, and regional variations). These differences also guided architectural decisions—most notably, the use of a sequence-based approach that can handle variable text lengths and character sets. By explicitly accounting for these features, the model is better equipped to recognise Polish plates accurately while still leveraging cross-country similarities for transfer learning and overall robustness.

While Polish license plates share certain similarities with other European formats – such as Czech plates, which also encode geographic information in the initial characters – no large-scale quantitative analysis has yet been devoted specifically to Polish LPR, unlike, for instance, the Czech case [36]. Consequently, this study addresses an important gap by investigating how domain knowledge from related Central European plate formats can be leveraged via transfer learning. Moreover, the CNN-RNN architecture is particularly well-suited for capturing positional dependencies, including region codes often found at the beginning of Polish plates. By examining how effectively these sequence-based features can enhance Polish LPR, the study underscores both the unique

Fig. 5. Realistic challenges of the dataset.

aspects of Polish license plates and the broader applicability of such a model for other European systems with similar encoding schemes.

In addition to training and evaluating Polish CNN-RNN hybrid models, this dataset was also used to test the best-performing architecture through majority voting on video sequences from the Parking Database, consisting of 22 videos recorded by parking lot cameras. The Parking Database videos, freely available to the research community, provide challenging real-world conditions for evaluating license plate recognition models. Each video includes corresponding ground truth annotations and vehicle registration plates as reference data. This video dataset was accessed directly without applying any preprocessing steps or distortion models as outlined in Leszczuk et al. [28].

Moreover, it must be noted that there are realistic images in the training and evaluation set, not just in the test set. Figure 5 showcases some of the more challenging license plate images within our dataset, captured under suboptimal lighting, precipitation, or partial occlusion. These conditions reflect real-world limitations that can hinder accurate recognition and are therefore crucial to include for robust model training and evaluation. By incorporating examples of low-quality plates, our dataset ensures that the proposed approach is tested against a broad spectrum of difficulties. Furthermore, using majority voting over consecutive frames helps mitigate transient noise from these adverse conditions, improving overall recognition stability.

4. Results

This section presents the evaluation of the proposed CNN-RNN-based LPR architecture. First, the performance is assessed for models trained on Polish, Czech, Hungarian, and Slovak datasets over 50 epochs in Subsection 4.1. Two measures are used: accuracy – the percentage of license plates fully correctly predicted – and normalized edit distance (normED) – which measures similarity between the predicted and ground-truth plate strings. Next, the effects of retraining are investigated by adapting models initially trained on Czech, Hungarian, and Slovak license plates to Polish data in Subsection 4.2.

Following that the robustness to weather conditions is examined in Subsection 4.3. Finally, Subsection 4.4 reports the results of testing on real-world video sequences, where majority voting is applied to enhance recognition stability.

4.1. Baseline training on individual datasets

Each of the four baseline models was trained for 50 epochs on its corresponding national dataset (Polish, Czech, Hungarian, and Slovak). Let N denote the total number of license plates in the validation set. The accuracy measure is defined as:

$$\text{Acc} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(\hat{y}_i = y_i), \tag{8}$$

where $\hat{y}_i$ represents the predicted license plate for the i-th sample, y_i is the corresponding ground-truth label, and $\mathbb{I}(\cdot)$ is the indicator function that returns 1 if the predicted label matches the ground truth exactly, and 0 otherwise.

To capture partial string mismatches, the normalized edit distance is employed:

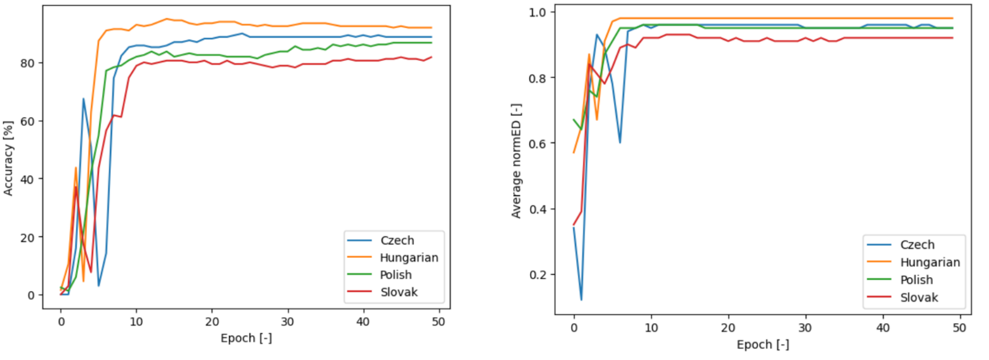
$$\text{normED} = 1 - \frac{d_{\text{lev}}(\hat{y}_i, y_i)}{\max(|\hat{y}_i|, |y_i|)}, \tag{9}$$

where $d_{\text{lev}}(\hat{y}_i, y_i)$ is the Levenshtein distance between the predicted and ground-truth strings, and $|\cdot|$ denotes string length. Higher normED values indicate greater similarity between $\hat{y}_i$ and y_i.

These two evaluation metrics were selected to jointly reflect the dual nature of the task, which lies at the intersection of pattern recognition and text recognition. Accuracy, a standard metric in the pattern recognition domain, quantifies the proportion of completely correct predictions and is widely used in the literature, facilitating comparison not only with prior work in this domain but also across different application areas. Its inclusion enables us to position the difficulty of the task within a broader machine learning context.

However, in the specific case of license plate recognition, where predictions are structured character sequences, exact-match accuracy alone may not sufficiently capture the quality of near-correct outputs. Therefore, normED, a well-established metric in the text recognition domain, is also adopted. Unlike simple character-level accuracy, normED penalizes both over- and under-predictions, and—by normalizing the Levenshtein distance—enables fair evaluation across license plates of varying lengths. This makes it particularly suitable for sequence-based recognition tasks, providing a more nuanced and stringent assessment of prediction errors.

It is also worth noting that, in the context of Polish license plate recognition, earlier studies have employed alternative text-level metrics such as the inverse of $(1 + \text{LED})$



a The accuracy curves. **b** The normED curves.

Fig. 6. Training curves for baseline training.

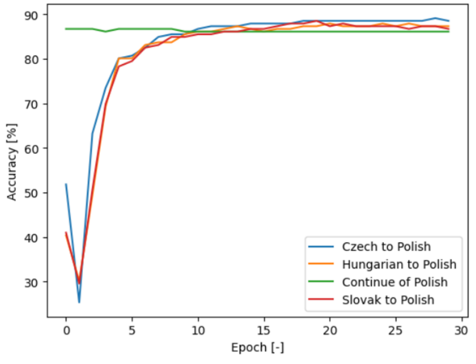
and character-level accuracy, as discussed in Subsection 2.2.2. To ensure comparability with these prior works, we computed these metrics for our models as well, and the corresponding results are presented in Subsection 4.2.

Figures 6a and 6b show the accuracy and average normED curves, respectively, over 50 epochs for the four baseline models. From Figure 6a it results that the Hungarian dataset achieves the highest final accuracy (95.0%), followed by the Czech (89.9%), Polish (87.3%), and Slovak (84.1%) datasets. The Hungarian model also displays relatively steady convergence. In contrast, the Czech accuracy curve undergoes more pronounced fluctuations in the early epochs before stabilizing. The curves for the Polish and Slovak sets are not as smooth as the one for Hungarian data but exhibit fewer fluctuations than the Czech curve overall.

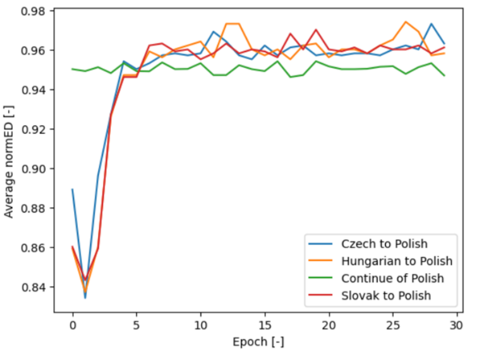
As shown in Figure 6b, the normED measure follows a similar pattern. The Hungarian model reaches approximately 98.3%, the Czech model 96.0%, the Polish model 95.8%, and the Slovak model 92.3%. Notably, the differences in normED across the four countries are smaller than those observed for accuracy, indicating that even when the entire license plate is not predicted perfectly, the predicted string remains relatively close to the ground truth. Overall, the Polish model consistently ranks near the middle of the performance range and demonstrates stable, incremental improvements throughout training.

4.2. Transfer learning

A transfer learning phase was conducted to examine whether knowledge learned from one Central European dataset could be transferred to Polish license plate recognition. Specifically, the final weights of three baseline models (Czech, Hungarian, and Slovak)



a The accuracy curves.



b The normED curves.

Fig. 7. The curves of retraining.

were used as initialization for Polish LPR training. Each retraining session ran for an additional 30 epochs under the same hyperparameter settings.

Figures 7a and 7b depict the accuracy and normED curves, respectively, for these retrained models. There is also a Continue of Polish curve included to illustrate the effect of extending training solely on the Polish dataset without external initialization. This provides a direct comparison between simple prolonged training and transfer learning, highlighting the relative benefits of incorporating prior knowledge from related domains.

All three retraining curves in Figure 7a exhibit smoother convergence compared to their respective baselines. The continuous Polish training does not show a significant change in terms of accuracy. The final accuracy values after 30 epochs are:

- Czech-based retraining: 89.2%,
- Hungarian-based retraining: 88.0%,
- Continued training on Polish: 87.2%,
- Slovak-based retraining: 88.6%.

A similar trend is observed in Figure 7b, although with a slight decrease in the Polish-only case for the normED measure. The final normED values are:

- Czech-based retraining: 97.3%,
- Hungarian-based retraining: 97.4%,
- Continued training on Polish: 94.7%,
- Slovak-based retraining: 97.0%.

So, these are the last values of the series of epochs. Importantly, continuing training on the Polish dataset for an additional 30 epochs without introducing new knowledge led to a slight decline in both validation accuracy and normED (from 87.3% to 87.2% and from 95.8% to 94.7%, respectively). This suggests that further training on the

Tab. 1. Performance before and after transfer learning on Polish license plate data.

Data	**Accuracy**	**Accuracy increment**	**normED**	**normED increment**
Continued training on Polish	87.2%	−0.1%	94.7%	−1.1%
Czech to Polish	89.2%	+1.9%	97.3%	+1.5%
Hungarian to Polish	88.0%	+0.7%	97.4%	+1.6%
Slovak to Polish	88.6%	+1.3%	97.0%	+1.2%

same dataset may induce overfitting, whereby the model becomes too specialized to the training data and fails to generalize better to the validation set.

To demonstrate the impact of transfer learning, Table 1 presents a direct comparison between the performance of models trained from scratch on Polish data and those retrained using transfer learning from related Central European license plate formats (Czech, Hungarian, and Slovak). The results show consistent improvements across all evaluation measures after applying transfer learning, confirming its effectiveness in enhancing recognition accuracy and robustness.

It is worth noting that although the differences between the various transfer learning approaches are small, they result in a significant reduction in normed error, defined as 1-normED. For example, the original error for Polish license plates was 4.2%, which could be reduced by 28-38%.

These results suggest that transferring knowledge from closely related Central European plate formats (Czech, Hungarian, and Slovak) to Polish data can be beneficial, as all three retrained models converge to similar performance levels. Although the differences among them are relatively small, the Czech-based retraining yields marginally higher accuracy on the Polish validation set, while the Hungarian-based retraining achieves a slightly higher normED.

Table 2 provides a comparison between the presented results and those reported in the literature, using shared evaluation measures where possible.

As shown in Table 2, the proposed approach consistently outperforms most of the results reported in the literature for Polish license plate recognition. While several earlier studies focus primarily on image quality evaluation or system architecture without reporting concrete recognition measures – such as the works by Leszczuk et al. [28], Łubkowski et al. [52], and Jureczko et al. [19] – those that do provide quantitative performance data are included in this comparison.

Most of the studies utilise accuracy as the primary evaluation measure, typically under varying environmental conditions. This makes it possible to conduct a meaningful comparison with the results of the presented method. In this regard, the proposed approach outperforms the results of Kluwak et al. [25], Janowski et al. [18], and Wróblewski [46] in general scenarios. The only exception is Kluwak et al. [25], who report

Tab. 2. Overview of the proposed and literature results in Polish LPR.

Publication	**Measure**	**Result**	**Note**
Kluwak et al. [25]	Accuracy ITCC Accuracy	53.12%-84.10% 62.83%-90.96%	The accuracy ranges reflect variation across datasets under different environmental conditions. ITCC aggregates and corrects predictions across multiple frames, resulting in better performance.
Janowski et al. [18]	LANN CC accuracy PWPLD CC accuracy Human CC accuracy LANN CC $(1+\mathrm{LED})^{-1}$ PWPLD CC $(1+\mathrm{LED})^{-1}$ LANN NC $(1+\mathrm{LED})^{-1}$ PWPLD NC $(1+\mathrm{LED})^{-1}$	4.89% 1.67% 70.33% 63% 57% 93% 81%	Out of 900 sequences, the LANN and PWPLD methods correctly recognised only 44 and 15 cases, respectively, while human subjects achieved 633 correct responses. Due to the low accuracy of automatic methods, alternative measures were introduced to enable more nuanced performance comparison.
Kekez [21]	Character accuracy Character accuracy with correction	88.7% 99%	Based on number plate and character format correction rules, post-processing was implemented for the output of the recognition algorithm.
Wróblewski [46]	Accuracy	60%-70%	The accuracy range reflects variation in image quality.
Czech to Polish	Accuracy normED $(1+\mathrm{LED})^{-1}$ Character accuracy	89.2% 97.3% 96.8% 96.6%	Pretrained on Czech license plates and retrained on Polish ones.
Hungarian to Polish	Accuracy normED $(1+\mathrm{LED})^{-1}$ Character accuracy	88.0% 97.4% 97.0% 96.9%	Pretrained on Hungarian license plates and retrained on Polish ones.
Slovak to Polish	Accuracy normED $(1+\mathrm{LED})^{-1}$ Character accuracy	88.6% 97.0% 96.6% 96.5%	Pretrained on Slovak license plates and retrained on Polish ones.

higher accuracy for high-quality images under ideal lighting conditions by aggregating predictions across multiple frames using their ITCC technique. However, this specific scenario is addressed separately in the next subsection using a dedicated technique developed within this study. Overall, the presented solution demonstrates superior general performance.

In addition to accuracy, several publications also consider character-level measures. For instance, Janowski et al. [18] employ the $(1+\mathrm{LED})^{-1}$ measure, which is clearly outperformed by the proposed models. Kekez [21] reports a character accuracy of 88.7% under standard conditions, which is also exceeded by the presented approach. Although his post-processed version achieves up to 99% by applying character format correction rules – such as measuring distances between specific points of characters – this relies in part on synthetic data, which reduces the relevance and generalizability of the result, especially in real-world scenarios.

Therefore, in more realistic application scenarios-where image quality, environmental conditions, and plate variability cannot be controlled-the presented approach emerges as the more robust and practically preferable solution.

4.3. Robustness to weather conditions

An additional experiment was conducted to evaluate the robustness of the proposed models in adverse weather and lighting conditions by synthetically modifying the Polish validation set using the Albumentations [6] library. Five image subsets were generated to simulate fog, rain, shadow, snow, and sunflare. Four models were tested on these subsets: one trained exclusively on Polish plates, and three produced by transferring knowledge from other Central European datasets – Czech, Hungarian, and Slovak – and subsequently retraining on Polish plates. Table 3 summarizes the accuracy obtained on these five subsets, while Table 4 reports the corresponding normED.

Overall, fog and rain tend to cause only moderate performance degradation compared to the original test conditions. For instance, when evaluating rain images, the transfer-based Czech to Polish and Hungarian to Polish models each achieve 76.51% accuracy, outperforming the purely Polish-trained model's 73.49%. The normED under rain likewise increases from 92.43% for the Polish to Polish baseline to around 93.3% for

Tab. 3. Accuracy under Adverse Conditions.

Subset	**Continue of Polish (%)**	**Czech to Polish (%)**	**Hungarian to Polish (%)**	**Slovak to Polish (%)**
Fog	73.49	76.51	72.29	71.69
Rain	73.49	76.51	76.51	76.51
Shadow	67.47	71.08	72.29	71.69
Snow	44.58	45.18	46.39	46.39
Sunflare	45.78	48.19	44.58	46.39

Tab. 4. NormED under Adverse Conditions.

Subset	**Continue of Polish (%)**	**Czech to Polish (%)**	**Hungarian to Polish (%)**	**Slovak to Polish (%)**
Fog	92.39	93.45	92.16	92.23
Rain	92.43	93.33	93.36	93.01
Shadow	91.57	92.89	92.92	93.01
Snow	73.67	75.24	73.91	73.74
Sunflare	70.74	72.01	70.92	71.25

Fig. 8. License Plates under Adverse Conditions.

both Czech to Polish and Hungarian to Polish, indicating more precise character-level predictions. Shadow similarly shows notable gains from transfer: its best-performing models attain 72.29% and 71.69% accuracy, respectively, compared to 67.47% for the Polish to Polish baseline.

A different pattern emerges in snow and sunflare conditions, where plate details can be heavily obscured. While accuracy remains relatively low overall, transfer still provides marginal improvements. Under snow, for example, the best-performing transfer-based model reaches 46.39% accuracy, whereas the baseline stays at 44.58%. In sunflare images, Czech to Polish yields 48.19%, exceeding the baseline of 45.78%. Nonetheless, it should be noted that not all transfer-trained models universally surpass the purely Polish model under every challenging scenario: for instance, while Czech to Polish outperforms Polish to Polish under fog, Hungarian to Polish falls slightly behind. Such variation underscores the complexity of weather-induced artifacts and suggests that future efforts could refine domain adaptation strategies or explore specialized attention mechanisms to further strengthen resilience when license plates are partially occluded, blurred, or distorted by inclement weather and harsh lighting. Figure 8 presents examples of each distortion category, illustrating how the applied simulations affect the visibility of license plates.

4.4. Testing on video sequences

To assess real-world applicability, the model that performed best on the Polish validation set-originally trained on Czech license plates and then retrained on Polish-was evaluated on 22 raw video sequences from the publicly available Parking Database [28]. Because these videos were recorded under natural conditions without preprocessing or image enhancements, they provide a more demanding test environment than the still-image validation dataset. As described earlier in Subsection 3.1, the final license plate prediction for each video was determined by applying a majority voting strategy across all frame-level predictions. This means that each frame was independently processed to

Tab. 5. Majority Votes.

Source (src)	**Majority Vote**	**Frequency (%)**	**Second Vote**	**Frequency (%)**	**Third Vote**	**Frequency (%)**
src1	kmi77ev	99.61	kmi77ee	0.39	–	–
src4	kr154gk	81.78	kr154ck	14.10	krise1	0.87
src5	kr439ha	75.72	kr439hal	19.93	kr439ma	4.35
src6	kr537cw	95.37	kr537ch	1.54	kr577ch	0.77
src7	kr539jk	64.77	kr5399k	20.35	kr5393k	2.84
src8	kr642lu	84.43	ar5411	0.94	vr642lu	0.71
src10	kr701kn	96.45	kr701kni	2.37	ka701kn	0.47
src11	kr731gw	93.79	vk731gw	0.59	pr771gm	0.30
src12	kr960mf	46.08	kry60mf	39.43	kr960mp	1.66
src13	kr992es	100.00	–	–	–	–
src15	kr2492k	100.00	–	–	–	–
src18	kr3527l	84.58	kr352yl	1.10	kavvl	0.66
src20	kr6986n	97.69	krr1114	0.29	kr65841	0.29
src21	kr8146m	92.88	kk8146m	2.67	wk8146m	1.78
src22	kr9195n	99.80	xr803lg	0.20	–	–
src23	kr9253g	98.15	kr9257g	1.23	wr9257g	0.21
src24	kr9611u	100.00	–	–	–	–
src25	kr9764s	50.51	kr97645	47.45	kr97649	0.41
src26	kra3hp2	81.13	krajhp2	13.73	krashp2	1.47
src27	kli05659	100.00	–	–	–	–
src28	kwia593	100.00	–	–	–	–
src30	w67045w	58.85	wig7045w	12.17	wug7045w	9.51
Average		86.40		8.10		1.20

yield a character string, and the most frequently occurring string in the entire sequence was selected as the final recognised plate.

A majority voting mechanism was used for each video source ID to stabilize predictions across frames, thereby reducing frame-by-frame recognition errors. In this context, the "Majority Vote" refers to the most frequently predicted license plate string across all frames of a given video, while the "Second Vote" and "Third Vote" denote the next most frequent predictions. Their corresponding frequencies indicate how dominant each prediction was relative to the total number of frames. Table 5 presents the top three majority votes for each source video along with their relative frequencies; the bottom row shows the overall accuracy of the majority-voting approach on all videos. So 22 videos were used in the testing phase and the identifiers of Table 5 follow the original coding scheme of the Parking Database [28], in which video sequences are labelled with non-sequential IDs such as src1, src4, src5, etc. This facilitates traceability and allows readers to map the presented results directly to the corresponding videos in the database.

Overall, the system achieves 86.4% accuracy on these raw test videos, which is about

1.8% lower than its performance on the Polish still-image validation set. This decline is attributed to real-world challenges such as variable lighting, motion blur, and occlusions that are not fully represented in the still-image dataset. Nonetheless, the majority voting method exhibits considerable robustness, demonstrating its ability to accurately identify plates under stable or moderately changing conditions. Moreover, for every tested sequence, the majority vote aligns with the ground truth plate, and thus – when only the ensemble output is considered – the overall accuracy reaches 100%. This means that, despite individual frame-level misclassifications—which are reflected in the presence of second and third votes – the most frequent prediction in each sequence always matches the ground truth.

While overall accuracy provides a broad measure of performance, it is critical to assess the system's behaviour in terms of specific error types, especially in real-world deployment scenarios where the consequences of false positives and false negatives differ substantially. In the context of automated access control, a false negative occurs when an authorised vehicle is incorrectly denied entry, whereas a false positive arises when an unauthorised vehicle is mistakenly admitted. Although both are technically misclassifications, their implications diverge – false negatives cause inconvenience, while false positives may result in security breaches.

To quantify these risks, we consider a typical operational setting in which a vehicle remains within the camera's field of view for approximately 2 seconds. This duration corresponds to $T = 50$ consecutive video frames at 25 frames per second. Each frame is processed independently, and the final decision is determined by a majority vote over all frame-level predictions. Given the observed average per-frame accuracy of $p = 0.864$, the number of correct predictions over the 50 frames can be modelled as a binomial random variable $X \sim \text{Binomial}(n = 50, p = 0.864)$. A false negative corresponds to the event in which fewer than 26 correct predictions are made, i.e., $\mathbb{P}(X < 26)$.

This demonstrates that the likelihood of rejecting an authorised vehicle, when aggregating predictions across 50 frames, is less than $8.4 \cdot 10^{-10}$, which is effectively zero in practice.

To estimate the false positive probability, we adopt a worst-case analysis. We assume that the system consistently misclassifies the observed unauthorised plate in every frame, and furthermore, that all misclassifications result in the same incorrect prediction—one that coincidentally belongs to the authorised set. This corresponds to a per-frame probability of consistent misclassification equal to $p = 1 - 0.864 = 0.136$. The number of times this incorrect (but authorised) plate appears as a prediction over 50 frames is modelled as $Y \sim \text{Binomial}(n = 50, p = 0.136)$, and a false positive occurs if at least 26 such misclassifications arise, i.e., $\mathbb{P}(Y \geq 26)$.

This demonstrates that the likelihood of accepting an unauthorised vehicle, when aggregating predictions across 50 frames, is less than $1.3 \cdot 10^{-10}$, which is effectively zero in practice as well.

These results indicate that, even under highly pessimistic assumptions, the probability of both rejecting an authorised plate and accepting an unauthorised one remains vanishingly small. The use of majority voting across a short temporal window provides not only improved robustness but also strong statistical guarantees against both types of critical classification errors in security-sensitive applications.

5. Conclusion and future work

This study presents a novel CNN-RNN-based framework tailored to license plate recognition in Poland, addressing the absence of recurrent neural network approaches in contemporary Polish LPR literature. By combining a Thin-Plate Spline module, a ResNet feature extractor, a BiLSTM sequence model, and an attention-based decoder, the proposed method capitalizes on sequence labelling to mitigate the segmentation bottlenecks common to traditional optical character recognition workflows. Comparative evaluations across Czech, Hungarian, Polish, and Slovak datasets underscore the model's robust performance, with transfer learning revealing notable improvements when retraining non-Polish models on Polish data. Moreover, testing on real-world parking-lot video sequences validates the adaptability of the proposed solution to dynamic conditions, reinforcing the benefit of majority voting to enhance recognition stability. Ultimately, the findings illustrate that the synergy of CNN-RNN architectures and transfer learning serves as an effective pathway for advancing the performance and resilience of Polish LPR systems.

Future work may also examine a broader spectrum of real-world applications to underscore the significance of rapid and precise license plate recognition. As an increasing number of cities implement environmental zones to mitigate air pollution and traffic congestion, accurate LPR systems can play a critical role in regulating and enforcing vehicle entry based on emissions standards. This functionality will likely become more pertinent as such zoning regulations proliferate in additional regions, including Poland. Beyond environmental monitoring, further applications include integration into advanced parking systems, where the technology could facilitate real-time detection of available spaces, guide drivers accordingly, and streamline the administration of car sharing or rental fleets. In the logistics and supply chain domains, improved plate recognition can enhance fleet tracking and management, while in residential and commercial contexts, the technology may be incorporated into automated gates or smart-home platforms to enable secure, seamless access for approved vehicles. Law enforcement, public safety, and public transportation authorities may likewise benefit from improved accuracy and robustness across large-scale, real-time deployments under diverse environmental conditions. By pursuing domain-specific data augmentation strategies, examining more sophisticated attention mechanisms, and conducting expanded empirical investigations, future research can aim to address the persistent challenges of plate variability and distortion – ultimately fostering faster and more reliable LPR solutions for an increasingly diverse range of use cases.

Acknowledgement

This research was supported by the Wacław Felczak Foundation through the Jagelló grant program, designed to assist doctoral students conducting research related to Poland at Hungarian higher education institutions.

WACŁAW FELCZAK FOUNDATION

Project no. KDP-IKT-2023-900-I1-00000957/0000003 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the Co-operative Doctoral Program [C2269690] funding scheme.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE MINISTRY OF CULTURE AND INNOVATION

References

[1] K. Ahmad, H. Khujamatov, A. Lazarev, N. Usmanova, M. Alduailij, et al. Internet of Things-Aided Intelligent Transport Systems in Smart Cities: Challenges, Opportunities, and Future. *Wireless Communications and Mobile Computing* 2023(1):7989079, 2023. doi:10.1155/2023/7989079.

[2] M. Arsenović, S. Sladojević, A. Anderla, and D. Stefanović. Deep learning driven plates recognition system. In: *17th International Scientific Conference on Industrial Systems*, pp. 1–4, 2017. `https://www.iim.ftn.uns.ac.rs/is17/papers/29.pdf`.

[3] J. Baek, G. Kim, J. Lee, S. Park, D. Han, et al. What is wrong with scene text recognition model comparisons? dataset and model analysis. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4714–4722, 2019. doi:10.1109/ICCV.2019.00481.

[4] G. Bradski. The openCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer* 25(11):120–123, 2000.

[5] A. C. Bukola, P. A. Owolawi, C. Du, and E. Van Wyk. A systematic review and comparative analysis approach to boom gate access using plate number recognition. *Computers* 13(11):286, 2024. doi:10.3390/computers13110286.

[6] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, et al. Albumentations: fast and flexible image augmentations. *Information* 11(2):125, 2020. doi:10.3390/info11020125.

[7] N. Cahyadi, S. Salsabila, S. Raniprima, and V. Monita. Review of license plate recognition techniques with deep learning. *Jurnal Teknologika* 14(2):458–469, 2024. `https://jurnal.wastukancana.ac.id/index.php/teknologika/article/view/415`.

[8] L. T. A. Dang, V. D. Ngoc, P. C. L. Thien Vu, N. N. Truong, and T. D. Trinh. Vietnam vehicle number recognition based on an improved CRNN with attention mechanism. *International Journal of Intelligent Transportation Systems Research* pp. 1–16, 2024. doi:10.1007/s13177-024-00402-7.

[9] P. Das, S. Mitra, S. Chakraborty, M. H. K. Mehedi, M. Y. M. Adib, et al. CNN-GRU based fusion architecture for bengali license plate recognition with explainable AI. In: *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–6, 2023. doi:10.1109/ICCCNT56998.2023.10413582.

[10] European Commission. Road freight transport by type of goods – annual data. Eurostat, 2024. `https://ec.europa.eu/eurostat/databrowser/view/road_go_ta_dc/`. [Accessed: Jan 1, 2025].

[11] A. Graves and J. Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks* 18(5-6):602–610, 2005. doi:10.1016/j.neunet.2005.06.042.

[12] J. Gąsienica-Józkowy, M. Knapik, and B. Cyganek. An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance. *Integrated Computer-Aided Engineering* 28(3):221–235, 2021. doi:10.3233/ICA-210649.

[13] S. Hadavi, H. B. Rai, S. Verlinde, H. Huang, C. Macharis, et al. Analyzing passenger and freight vehicle movements from automatic-number plate recognition camera data. *European Transport Research Review* 12:1–17, 2020. doi:10.1186/s12544-020-00405-x.

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi:10.1109/CVPR.2016.90.

[15] K. Hefnawy, A. Lila, E. Hemayed, and M. Elshenawy. A robust license plate detection and recognition framework for arabic plates with severe tilt angles. *International Journal of Advanced Computer Science & Applications* 15(2), 2024. doi:10.14569/IJACSA.2024.0150287.

[16] C. Y. Huang. A review of deep learning for character recognition area. In: *The 26th International Scientific and Practical Conference "World Problems and Ways of Solving Modern Problems"*, p. 245, 2024. doi:10.46299/ISG.2024.1.26.

[17] Jaided AI. Easyocr (version 1.6.2) [computer software], 2020. `https://github.com/JaidedAI/EasyOCR`. [Accessed: May 27, 2025].

[18] L. Janowski, P. Kozłowski, R. Baran, P. Romaniak, A. Glowacz, et al. Quality assessment for a visual and automatic license plate recognition. *Multimedia Tools and Applications* 68:23–40, 2014. doi:10.1007/s11042-012-1199-5.

[19] M. Jureczko and F. Uherek. The automatic license plate recognition system based on video sequences using artificial neural networks. In: *Modelling in Engineering 2020: Applied Mechanics*, pp. 95–106, 2021. doi:10.1007/978-3-030-68455-6_9.

[20] F. Kahraman, B. Kurt, and M. Gökmen. License plate character segmentation based on the Gabor transform and vector quantization. In: *International Symposium on Computer and Information Sciences*, pp. 381–388, 2003. doi:10.1007/978-3-540-39737-3_48.

[21] M. Kekez. An approach to license plate recognition in real time using multi-stage computational intelligence classifier. *International Journal of Electronics and Telecommunications* pp. 275–280, 2023. doi:10.24425/ijet.2023.144361.

[22] M. M. Khan, M. U. Ilyas, I. R. Khan, S. M. Alshomrani, and S. Rahardja. License plate recognition methods employing neural networks. *IEEE Access* 11:73613–73646, 2023. doi:10.1109/ACCESS.2023.3254365.

[23] Z. H. Khattak, B. L. Smith, M. D. Fontaine, J. Ma, and A. J. Khattak. Active lane management and control using connected and automated vehicles in a mixed traffic environment. *Transportation Research Part C: Emerging Technologies* 139:103648, 2022. doi:10.1016/j.trc.2022.103648.

[24] W. Kim and G. L. Chang. Development of a hybrid prediction model for freeway incident duration: A case study in maryland. *International Journal of Intelligent Transportation Systems Research* 10:22–33, 2012. doi:10.1007/s13177-011-0039-8.

[25] K. Kluwak, J. Segen, M. Kulbacki, A. Drabik, and K. Wojciechowski. ALPR-extension to traditional plate recognition methods. In: *Intelligent Information and Database Systems: 8th Asian Conference (ACIIDS 2016)*, pp. 755–764, 2016. doi:10.1007/978-3-662-49390-8_73.

[26] E. Kot, Z. Krawczyk, K. Siwek, L. Królicki, and P. Czwarnowski. Deep learning-based framework for tumour detection and semantic segmentation. *Bulletin of the Polish Academy of Sciences Technical Sciences* p. e136750, 2021. doi:10.24425/bpasts.2021.136750.

[27] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Gonçalves, et al. A robust real-time automatic license plate recognition based on the YOLO detector. In: *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–10, 2018. doi:10.1109/IJCNN.2018.8489629.

[28] M. Leszczuk, L. Janowski, J. Nawała, J. Zhu, Y. Wang, et al. Objective video quality assessment and ground truth coordinates for automatic license plate recognition. *Electronics* 12(23):4721, 2023. doi:10.3390/electronics12234721.

[29] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk* 163(4):845–848, 1965. `http://mi.mathnet.ru/dan31411`. [Accessed: Apr 12, 2025].

[30] D. Li and J. Lasenby. Mitigating urban motorway congestion and emissions via active traffic management. *Research in Transportation Business & Management* 48:100789, 2023. doi:10.1016/j.rtbm.2022.100789.

[31] H. Li, P. Wang, and C. Shen. Toward end-to-end car license plate detection and recognition with deep neural networks. *IEEE Transactions on Intelligent Transportation Systems* 20(3):1126–1136, 2018. doi:10.1109/TITS.2018.2847291.

[32] T. Lindner, D. Wyrwał, M. Białek, and P. Nowak. Face recognition system based on a single-board computer. In: *2020 International Conference Mechatronic Systems and Materials (MSM)*, pp. 1–6, 2020. doi:10.1109/MSM49833.2020.9201668.

[33] S. Z. Masood, G. Shu, A. Dehghan, and E. G. Ortiz. License plate detection and recognition using deeply learned convolutional neural networks. arXiv, arXiv:1703.07330, 2017. doi:10.48550/arXiv.1703.07330.

[34] Minister Infrastruktury. Rozporządzenie ministra infrastruktury z dnia 8 listopada 2024 r. w sprawie rejestracji i oznaczania pojazdów, wymagań dla tablic rejestracyjnych oraz wzorów innych dokumentów związanych z rejestracją pojazdów. Dz.U. 2024 poz. 1709, 2024. `https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20240001709`. (Minster of Infrastructure. Regulation of the Minister of Infrastructure of 8 November 2024 on the registration and marking of vehicles, requirements for license plates and the use of other documents related to vehicle registration).

[35] A. Mousa. Canny edge-detection based vehicle plate recognition. *International Journal of Signal Processing, Image Processing and Pattern Recognition* 5(3):1–8, 2012. `https://staff.najah.edu/media/sites/default/files/Canny_Edge-Detection_Based_Vehicle_Plate_Recognition.pdf`.

[36] D. Moussa, A. Maier, F. Schirrmacher, and C. Riess. Sequence-based recognition of license plates with severe out-of-distribution degradations. In: *International Conference on Computer Analysis of Images and Patterns*, pp. 175–185, 2021. doi:10.1007/978-3-030-89131-2_16.

[37] V. Naosekpam, A. S. Shishir, and N. Sahu. Scene text recognition with orientation rectification via IC-STN. In: *TENCON 2021 – IEEE Region 10 Conference (TENCON)*, pp. 664–669, 2021. doi:10.1109/TENCON54134.2021.9707183.

[38] OpenALPR. Automatic license plate recognition. Rekor, 2015. http://www.openalpr.com/. [Accessed: Apr 12, 2025].

[39] V. H. Pham, P. Dinh, and V. H. Nguyen. CNN-based character recognition for license plate recognition system. In: *Asian Conference on Intelligent Information and Database Systems*, 2018. doi:10.1007/978-3-319-75420-8_56.

[40] P. S. Puri and P. P. Bartakke. SSD based vehicle number plate detection and recognition. In: *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, pp. 1–6, 2023. doi:10.1109/NMITCON58196.2023.10276329.

[41] See5. Data mining tool. Rulequest Research, 2012. http://www.rulequest.com/products.html. [Accessed: Apr 12, 2025].

[42] R. Smith. An overview of the Tesseract OCR engine. In: *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, pp. 629–633, 2007. doi:10.1109/ICDAR.2007.4376991.

[43] S. Srinivasan, D. Prabha, N. M. Raffic, K. G. Babu, S. Thirumurugaveerakumar, et al. Automated vehicle number plate recognition system using convolution long short-term memory technique. In: *Object Detection with Deep Learning Models*, pp. 101–115. Chapman and Hall/CRC, 2022. doi:10.1201/9781003206736-6.

[44] Y. Wang, X. Yang, H. Liang, and Y. Liu. A review of the self-adaptive traffic signal control system based on future traffic environment. *Journal of Advanced Transportation* 2018(1):1096123, 2018. doi:10.1155/2018/1096123.

[45] A. Wilkowski, M. Stefańczyk, and W. Kasprzak. Training data extraction and object detection in surveillance scenario. *Sensors* 20(9):2689, 2020. doi:10.3390/s20092689.

[46] P. Wróblewski. Recognition of polish car license plates. In: *Computer Recognition Systems: Proceedings of the 4th International Conference on Computer Recognition Systems (CORES'05)*, pp. 893–900, 2005. doi:10.1007/3-540-32390-2_105.

[47] F. You, Y. Zhao, and X. Wang. Combination of CNN with GRU for plate recognition. *Journal of Physics: Conference Series* 1187(3):032008, 2019. doi:10.1088/1742-6596/1187/3/032008.

[48] M. D. Zeiler. ADADELTA: An adaptive learning rate method. arXiv, arXiv:1212.5701, 2012. doi:10.48550/arXiv.1212.5701.

[49] H. Zhao, C. Song, H. Zhao, and S. Zhang. License plate recognition system based on morphology and LS-SVM. In: *2008 IEEE International Conference on Granular Computing*, pp. 826–829, 2008. doi:10.1109/GRC.2008.4664738.

[50] J. Zhao, W. Li, J. Wang, and X. Ban. Dynamic traffic signal timing optimization strategy incorporating various vehicle fuel consumption characteristics. *IEEE Transactions on Vehicular Technology* 65(6):3874–3887, 2015. doi:10.1109/TVT.2015.2506629.

[51] L. Zou, Z. He, K. Wang, Z. Wu, Y. Wang, et al. Text recognition model based on multi-scale fusion CRNN. *Sensors* 23(16):7034, 2023. doi:10.3390/s23167034.

[52] P. Łubkowski and D. Laskowski. Assessment of quality of identification of data in systems of automatic licence plate recognition. In: *Smart Solutions in Today's Transport: 17th International Conference on Transport Systems Telematics (TST 2017)*, pp. 482–493, 2017. doi:10.1007/978-3-319-66251-0_39.