Deep Learning Methods Applied in Computer Vision

Ing. Alžbeta Turecková, Ph.D.

Doctoral Thesis Summary

Tomas Bata University in Zlín

Tomas Bata University in Zlín
Faculty of Applied Informatics

Doctoral Thesis Summary

Deep Learning Methods Applied in Computer Vision

Využití metod hlubokého učení v počítačovém vidění

Author:	**Ing. Alžběta Turečková, Ph.D.**
Study Program:	P3902 / Engineering Informatics
Study Field:	3902V023 / Engineering Informatics
Supervisor:	doc. Ing. Zuzana Komínková Oplatková, Ph.D.
Reviewers:	prof. Ing. Jan Platoš, Ph.D.
	doc. Mgr. Petr Hurtik, Ph.D.
	doc. RNDr. Martin Kotyrba, Ph.D.

Zlín, August 2023

Published by **Tomas Bata University in Zlin** in edition **Doctoral Thesis Summary** in 2023.

Klíčová slova: *Hluboké učení, Detekce objektů, Obrazová data s vysokým rozlišením, Dataset Tomato360, Mechanismy pozornosti, Artificial Size Slicing Aided Fine Tuning (ASSAFT), Artificial Size Slicing Aided Hyper Inference (ASSAHI), Odhad výnosů plodin, Tvorba datasetu*

Key words: *Deep Learning, Object Detection, High-resolution Images, Tomato360 Dataset, Attention Mechanisms, Artificial Size Slicing Aided Fine Tuning (ASSAFT), Artificial Size Slicing Aided Hyper Inference (ASSAHI), Crop Yield Estimation, Real-world Dataset Creation*

The full version of dissertation thesis is available in the Library of TBU in Zlín.

ISBN 978-80-7678-191-7

ABSTRACT

This Doctoral Thesis investigates the significant role of data handling in the practical application of deep learning techniques for object detection in high-resolution images. The study examines the impact of attention mechanisms and introduces novel data processing methodologies, namely Artificial Size Slicing Aided Fine Tuning (ASSAFT) and Artificial Size Slicing Aided Hyper Inference (ASSAHI). Despite the potential of attention mechanisms observed in medical imaging, the practical application of similar principles in the custom-made Tomato360 dataset does not prove to be beneficial. On the other hand, a substantial improvement in object detection performance in the Tomato360 dataset was achieved through the newly proposed ASSAFT and ASSAHI techniques. The research underlines the challenges of deploying deep learning techniques in real-world scenarios; concretely, the final proposed solution is utilized and evaluated for estimating crop yields in tomato greenhouses.

ABSTRAKT

Tato disertační práce zkoumá významnou roli zpracování dat při praktickém použití technik hlubokého učení pro detekci objektů v obrazech s vysokým rozlišením. Práce zkoumá dopad mechanismů pozornosti a představuje nové metody zpracování dat, konkrétně Artificial Size Slicing Aided Fine Tuning (ASSAFT) a Artificial Size Slicing Aided Hyper Inference (ASSAHI). Přes úspěšné použití mechanismů pozornosti při zpracování medicínských dat, praktické uplatnění podobných principů v nově vytvořeném datasetu Tomato360 se neukázalo prospěšné. Na druhou stranu, významné zlepšení kvality detekce objektů v datasetu Tomato360 bylo dosaženo prostřednictvím nově navržených technik ASSAFT a ASSAHI. Práce dokumentuje výzvy spojené s nasazením technik hlubokého učení v reálných aplikacích; konkrétně je finální navržené řešení využito pro odhad sklizně rajčat ve skleníku.

TABLE OF CONTENTS

1 INTRODUCTION

This dissertation explores the field of computer vision, focusing on the challenge of small object detection in high-resolution images, specifically in the context of agriculture. Given the vast amount of image/video data generated daily, manual analysis becomes impracticable, reinforcing the need for automatic processing systems. Computer vision, through artificial intelligence algorithms, mimics human visual capabilities to analyze and understand this data.

Despite many advancements achieved in the computer vision field, small object detection remains a significant challenge due to its low contrast, common occlusion, and complex arrangements. This thesis aims to address this issue by developing novel approaches that improve precision and efficiency by including attention mechanisms into deep convolutional neural networks and devising a unique processing pipeline for high-resolution images and small object detection. Aside from that, the innovations were tested on a newly created custom Tomato360 dataset from an agriculture context.

The text of this doctoral thesis summary is divided into key sections, beginning with an exploration of the current "State of the Art" in computer vision and its application to agriculture. It then outlines the specific "Aims of the Dissertation," followed by the "Methodology," detailing innovative techniques developed. The "Results" section presents the findings, including the attention mechanism analysis and estimation of tomato crop yields. The text concludes with discussions on the fulfillment of the research aims, its impact on science and practice, and the overall conclusions drawn from the study.

2 STATE OF THE ART

This section provides an introduction to the computer vision problems faced in this work. Some of the most relevant research papers are cited as an illustration of the state-of-the-art, but for a more detailed description of the current research state, please refer to the Literature Review in the Doctoral Thesis.

2.1 Basic Computer Vision Tasks

Computer vision aims to mimic human image interpretation through increasingly complex tasks: classification, object detection, semantic segmentation, instance segmentation, and panoptic segmentation, as illustrated in Fig. 2.1 with an image of dogs in a park.

Classification is the simplest task. An image of dogs in a park from our example would be labeled entirely as 'dogs' or 'park'. In 2012, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) served as a catalyst for progress, and an introduction of deep convolutional neural networks (CNN) [11] has contributed to significant progress in the computer vision area. Current image classification models achieve even higher consistency than various human operators [1].

In object detection, objects within an image are identified, located, and labeled with bounding boxes. In the dogs in a park image, each dog would be located and encased in a rectangle. This task saw a revolution in accuracy with the introduction of the R-CNN framework [9], combining region proposal algorithms with CNNs.

Semantic segmentation utilizes a different approach for understanding the image by classifying each pixel. In our example, each pixel could be classified as 'dogs', 'grass', or 'tree'. This approach was revolutionized by the introduction of fully convolutional networks (FCNs) [15].

Instance segmentation merges object detection and semantic segmentation, distinguishing different instances of the same class. In our example, it could differentiate and segment each pixel of each individual 'dog'.

Lastly, panoptic segmentation [13] combines semantic and instance segmentation, providing a holistic understanding of the image information. All pixels are labeled according to their semantic class, and individual instances are identified, resulting in a detailed image representation.

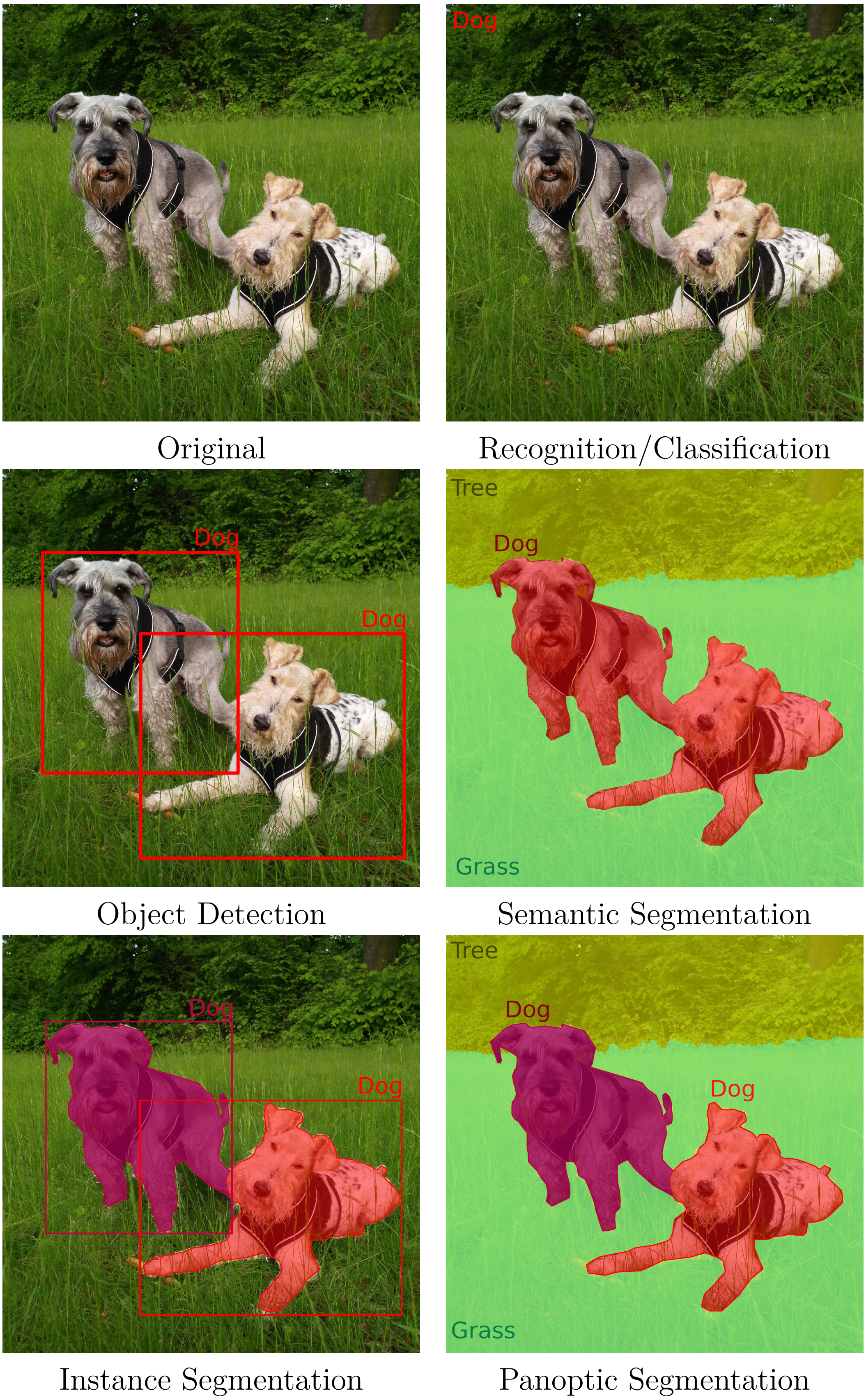

Fig. 2.1 Illustration of the fundamental computer vision tasks from a computer science point of view.

2.2 Applying Computer Vision to Agriculture

Precision agriculture optimizes crop management by accurately monitoring plant health to guide fertilization and pest control strategies. Essential to this is gathering detailed information about plants or field sections regularly [8]. Despite the wealth of data gathered through various sensors monitoring the plants, augmenting this information with new sources can always improve the accuracy of decision-making processes.

This thesis introduces a novel deep learning-based methodology for detecting and counting tomato fruits using computer vision. This approach aids in precise harvest prediction. Accurate harvest prediction is critical for tomato cultivation as deviations between predicted and actual harvest can lead to significant commercial losses and logistical problems. The goal of this work is to manage large-scale fruit counting in commercial tomato greenhouses.

Fruit detection has seen extensive research, with significant advancements highlighted by [24, 10]. While traditional image processing techniques have limitations, the rise of AI has allowed machine learning to excel in agricultural computer vision tasks. Various methods have been explored for fruit detection. YOLOv3-tiny architecture [23] has been adapted for real-time tomato detection with notable results, achieving 91.92% F1-score. The YOLO-tomato model [14], which applies circular bounding boxes for detection, improves the F1-score to an impressive 93.91%.

However, most research focuses on direct tomato detection in isolated images of tomato trusses and fails to account for the challenges of capturing the whole tomato plant. The authors of [17] work with images of the whole tomato plant from a greenhouse, similar to this work. Although the authors provide the yield mapping of the whole image row by stitching the images together, the detection results were evaluated only on single images achieving an F1-score of 83.67% but failing to provide the final number of fruits in the tomato row.

Large-scale fruit counting has also been attempted in [19], but the uniform background and non-clustered nature of pears simplify the process compared to the tomato greenhouse environment. A more comparable study used deep CNN networks for detection, counting, and maturity assessment of cherry tomatoes in multi-spectral images [4]. Similar to the paper [19], authors of [4] apply the DeepSORT algorithm to track the tomatoes in a video. The value F1-score is not stated; the IDF1-score achieved by the best solution is 51.4%. Proving that the process of tomato counting in a tomato greenhouse environment is very challenging.

In response, the author of this thesis simplifies the process by using high-resolution images of entire tomato rows, which are then sliced and analyzed by an object detector. This approach was presented first in [A1] and effectively avoids object tracking difficulties and thus improves final counting accuracy.

The state-of-the-art section about Applications of Deep Learning in Agriculture in the dissertation includes a thorough review of the latest applications of deep learning in agriculture, providing the necessary context for understanding the innovations introduced.

2.3 Small Object Detection in High-Resolution Images

Small object detection in high-resolution images poses a significant challenge in computer vision. This issue is relevant in satellite imagery analysis and surveillance systems but also in medical imaging or agriculture. Fig. 2.2 illustrates examples of small object detection problems, including a greenhouse tomato row from a dataset created in this thesis.

The scale difference between the entire image and target objects presents difficulties for conventional CNNs due to factors like reduced feature representation of small objects, overwhelming background features, and increased computational demands.

To mitigate these issues, one solution could be integrating attention mechanisms into model architecture [18, 26]. Another typical approach, common especially in satellite data processing, utilizes the cropping of high-resolution images into sequential subregions or chips for detection [2]. This work aims

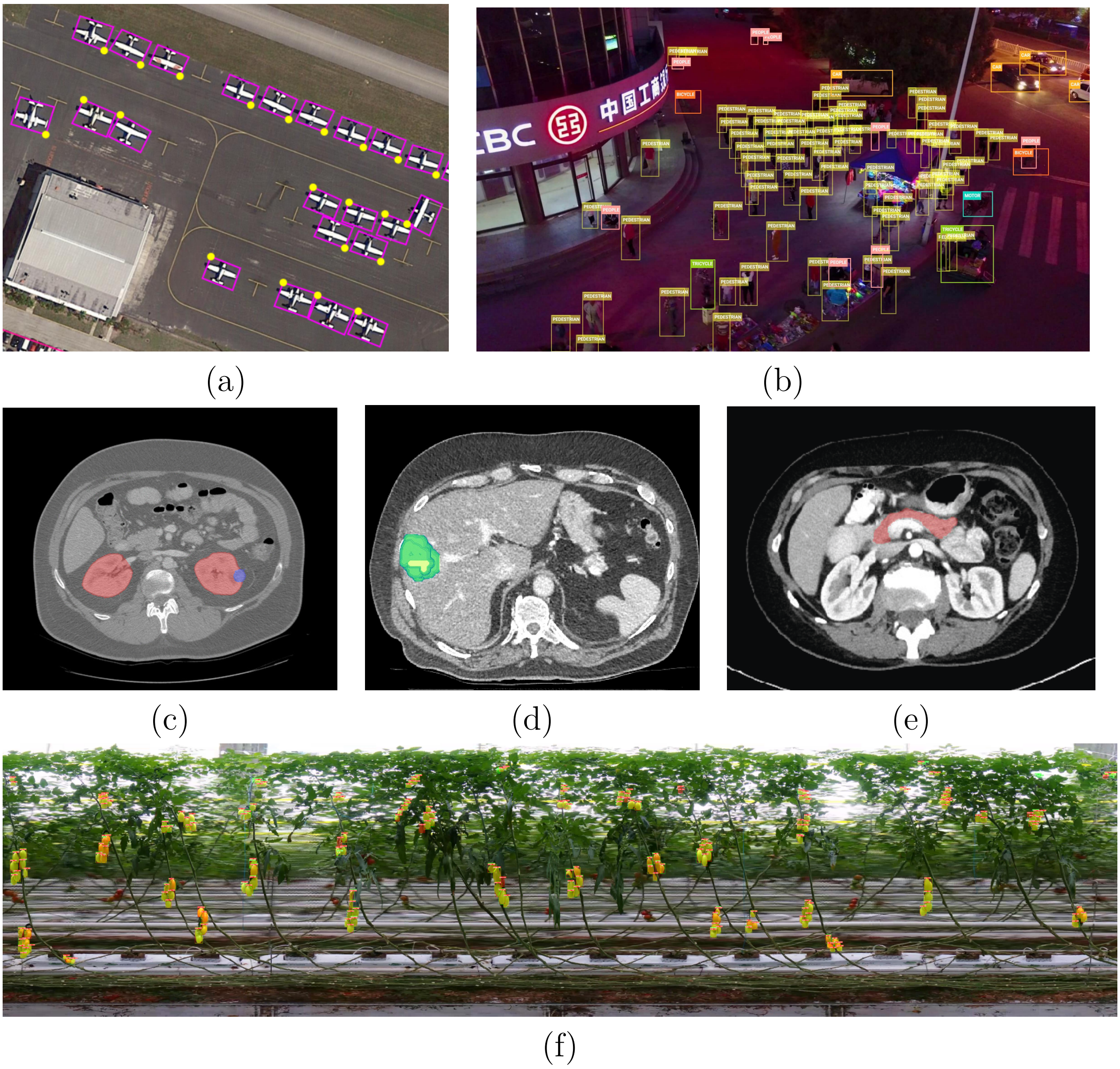

Fig. 2.2 Example images representing the challenging small object detection, (a) satellite image: DOTA dataset [6], (b) surveillance: Visdrone dataset [25], (c-e) - medical images: KiTS dataset [16], LiTS dataset [3], pancreas segmentation [22], (f) Tomato row in a greenhouse: dataset created and analyzed in this thesis.

to improve the accuracy and efficiency of small object detection within high-resolution images by exploring both those methods. Relevant research is detailed in the sections "Attention Mechanism in CNN" and "Processing High-Resolution Images" of the dissertation thesis.

3 AIMS OF DOCTORAL THESIS

The following items are proposed as aims of the dissertation:

1. **Appraise the current state of the research area:** Specifically, deep learning methods applied in computer vision with a particular focus on small object detection and segmentation in high-resolution images.

2. **Develop and curate a custom dataset in a tomato greenhouse:** The creation of a custom, real-world dataset aims to demonstrate the transfer of AI technologies from theory to practical implementation. This involves acquiring, collecting, and labeling high-resolution images that capture the challenges specific to this domain.

3. **Investigate and compare the effectiveness of attention mechanisms:** Explore possibilities of incorporating attention mechanisms into different convolutional neural network (CNN) architectures. Compare their performance in terms of accuracy and computational efficiency.

4. **Develop an enhanced deep learning pipeline:** Design and develop a novel processing pipeline tailored to handle the challenging task of small object detection in high-resolution images.

5. **Evaluate the proposed pipeline on the custom dataset:** Apply and test the developed processing pipeline on the custom dataset from the tomato greenhouse. Measure its performance against existing standard techniques used for small object detection. Assess and compare the proposed pipeline's accuracy, robustness, and efficiency.

6. **Analyze the impact and practicality of the proposed methods:** Conduct a comprehensive analysis to understand the impact of incorporating attention mechanisms and the newly developed processing pipeline on small object detection in high-resolution images. Evaluate their practicality in real-world scenarios, considering factors such as computational requirements, scalability, and generalizability.

4 METHODOLOGY

In order to develop a successful deep convolutional neural network model, an extensive and complex workflow is necessary. The quality of the established pipeline frequently has a significant impact on the final model results [12]. The Methodology part of the dissertation thesis, therefore, summarizes the essential building blocks of a robust workflow, discussing the actual trends in the application area of computer vision. On top of that, a novel method for small object detection in high-resolution images is proposed. The proposed methodology combines two novel techniques: Artificial Size Slicing Aided Fine Tuning (ASSAFT) and Artificial Size Slicing Aided Hyper Inference (ASSAHI). Those innovative techniques are shortly described in the following subsections.

4.1 Slicing Aided Fine-tuning and Hyper Inference

In various fields, including satellite imagery analysis, surveillance systems, medical imaging, and agriculture, there can be situations where processed images may be too large for a CNN model to process in one go. This problem, relating to small object detection in high-resolution images, is detailed in section 2.3. A common approach to address this issue is to either downsample the images or process the image in patches, which are sub-images extracted from the larger image. These sub-images can overlap, and their size can vary depending on the model and its application.

While downsampling or using larger patches allows the model to capture more contextual information, this comes at the cost of reduced detail. On the other hand, processing smaller patches provides high-resolution details but at the expense of losing broader contextual information. This necessitates finding a balance between the two extremes. Patch cropping, for instance, is commonly employed in medical image segmentation, which usually deals with extensive and often multimodal data. Models in this area are typically trained on specific patch sizes that are tailored to the data and the application at hand.

The term 'Slicing Aided Fine-tuning' was introduced in paper [2] to describe a data processing pipeline that augments training images with overlapping cropped image patches. This process enriches the training data to include both original resolution images and image patches or crops in preparation for 'Slicing Aided Hyper Inference', where the input image is predicted whole, and also the image is cut into several image patches, which are predicted too. The final prediction then gathers all the predictions together. The problem of multiple prediction merging is described in the following section 4.1.3.

The best prediction performance might be achieved by applying different cropping sizes, though this method significantly increases computational requirements for both training and inference, as it enlarges the training dataset size and requires the model prediction multiple times: once for the original image and then for each image patch.

4.1.1 Artificial Size Slicing Aided Fine-tuning

Certain situations necessitate using image patches for both training and prediction as the input image data is too large for one-time model processing, and simple downsampling can degrade the image to the extent that object detection becomes impossible. This is the case with the custom-made Tomato360 dataset addressed in this work. In a recent study [A1], the author of this thesis explores the impact of varying image patch sizes on prediction accuracy.

To generalize the patch-cropping process, a proposal is made to crop artificial-sized image patches centered around object groups, effectively utilizing the fact that tomatoes are growing in trusses. This strategy of augmenting the training data with these patches enhances the model's ability to discern and localize overlapping tomatoes.

During the training phase, generating these artificial-sized image patches is relatively straightforward. Fig. 4.1 provides a schematic of a slightly more complicated Artificial Size Slicing Aided Hyper Inference process discussed in the following section 4.1.2, but visualize the idea of mask creation and patches cropping utilized in ASSAFT, too.

In the case of training data, where instance segmentation masks are available (or the area of boxes in scenarios lacking instance segmentation annotation), a foreground segmentation map was created from all object's instance masks. A binary dilation operation was employed to cluster objects into larger groups, and image patches were cropped in accordance with the position of each connected group. This method not only provides the model with detailed information about small objects in the image but also effectively avoids cutting objects during the patch cropping process.

However, implementing this solution poses a couple of challenges. Firstly, the dilation operation can be computationally demanding in scenarios involving numerous small objects in high-resolution images, especially with larger dilation operator sizes. To improve efficiency, the input image is downscaled first, the dilation operation is applied, and the output is subsequently upscaled back to the original resolution. This streamlined the process without causing significant damage to the final output. Secondly, the risk of cropping too small image patches around isolated objects in the image is mitigated by imposing a minimum crop size, ensuring such small objects were cropped with a broader context around them.

4.1.2 Artificial Size Slicing Aided Hyper Inference

In accordance with ASSAFT proposed in section 4.1.1, a similar extension utilizing artificial size image patches for Slicing Aided Hyper Inference (SAHI) is provided and named Artificial Size Slicing Aided Hyper Inference (ASSAHI).

To specify the placement of these artificially-sized patches, a mask that identifies object group positions within the original image is required. For this purpose, a semantic segmentation deep convolutional neural network might be trained. This network doesn't require ultra-precise segmentation annotations but should reliably detect clusters of smaller objects utilizing a greater context of the image. Hence, it should be trained on complete input images or relatively large image patches if slicing is necessary.

Once the input image mask is generated, slicing follows the same rules established in the ASSAFT. The mask undergoes dilation, and image patches are situated around each detected object/group within the masked image. Those patches are predicted by the model and combined in the final prediction. The entire ASSAHI procedure is graphically represented in Fig. 4.1.

ASSAHI presents two key benefits. Firstly, it feeds the model with detailed information about smaller objects within the image, enabling more precise detection, which is especially beneficial in scenarios where objects lay close to each other or even overlap. Concurrently, it effectively prevents objects from being cropped/segmented into parts during patch slicing. Furthermore, in a dataset where object groups sparsely populate the input data, ASSAHI can save a substantial amount of computation resources, which would otherwise be spent by processing many (empty) small image patches sliced by the standard SAHI method, all while maintaining necessary detail for precise small object detection.

On the other hand, ASSAHI may encounter difficulties in datasets where objects don't group or, conversely, form excessively large groups. In the former situation, a singular small object could potentially be overlooked by the primary segmentation model, thus being excluded from subsequent higher-resolution processing and missed entirely in the final prediction. In the latter case of extensive object groups, such a large group may be accommodated into a single patch whose resolution might exceed the model's handling capabilities. Consequently, the patch would be downscaled during the data loading process, risking the loss of details necessary for distinguishing smaller objects.

A balanced combination of ASSAHI and SAHI techniques appears to be the best solution. Then ASSAHI assists with detailed small object detection while SAHI ensures comprehensive coverage of the entire input image with relatively large patches. Crop sizes in SAHI and ASSAHI parameter setups must always be adjusted to suit the specific dataset at hand. The exploration of these techniques applied to the Tomato360 dataset is presented in the results part of this work, under section 5.3.

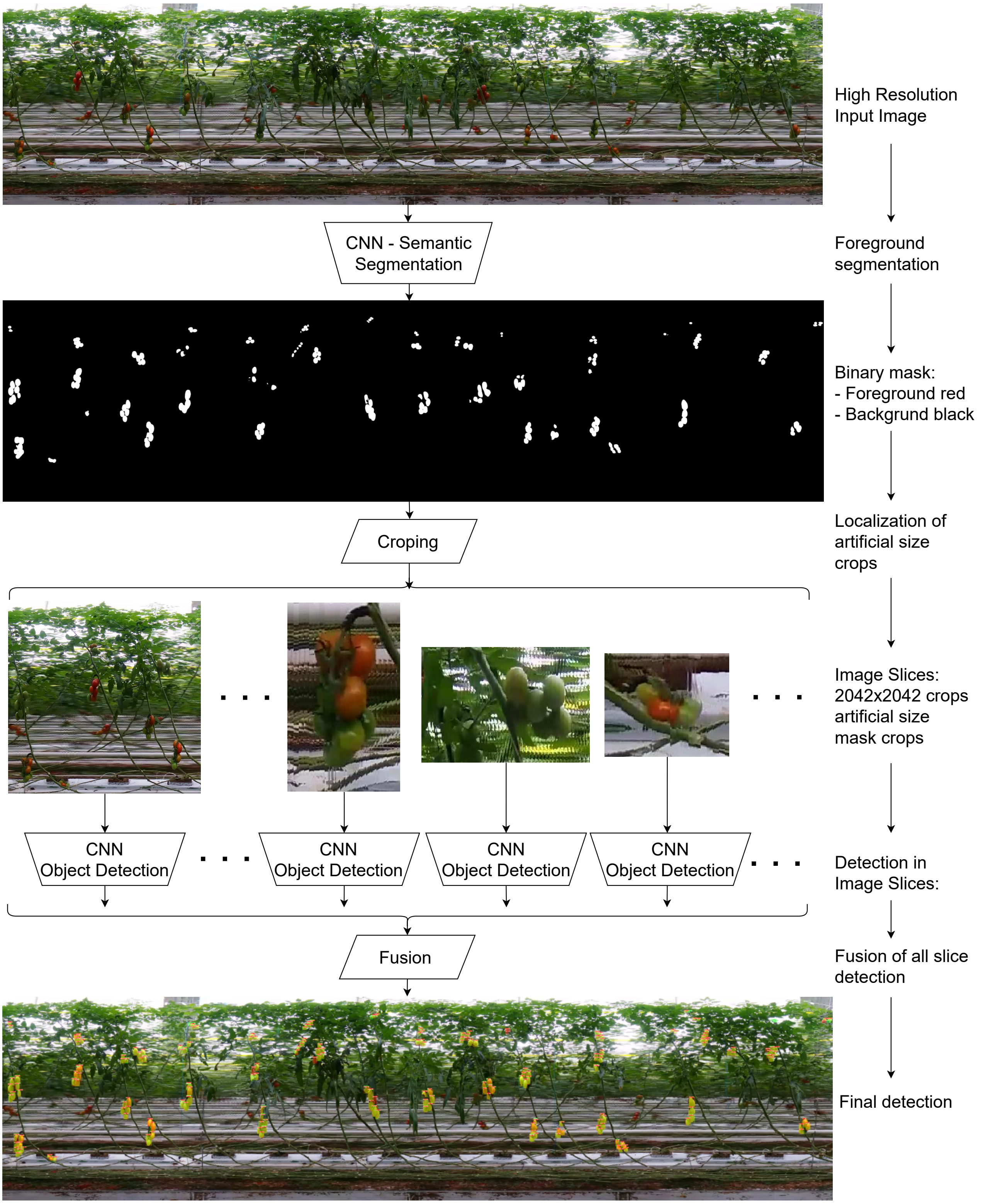

Fig. 4.1 Visualization of ASSAHI - Artificial Size Slicing Aided Hyper Inference procedure proposed in this thesis.

4.1.3 Image stitching

While merging patches' prediction is a relatively straightforward process in semantic segmentation, it is quite complex for object detection tasks. Objects on the border may be detected multiple times or even fragmented into two or more parts if they are positioned on a crop border. There are several potential strategies for merging or suppressing repeated predictions in object detection. Generally speaking, the process has two main parameters: the match metrics and the post-processing algorithm.

Match metrics identify potential detections for merging or suppression. The threshold value of match metrics is crucial; a higher value allows only highly similar predictions to be merged. The post-processing algorithm, on the other hand, dictates the sequence in which potential detections are processed and how the final detection instance is formulated.

There are two common match metrics. Intersection over union (IOU) is a term used to describe the extent of overlap of two bounding boxes. The greater the region of overlap, the greater the IOU. The metric is defined as:

$$IOU = \frac{Area\ of\ intersection}{Area\ of\ union} \tag{4.1}$$

Intersection over a smaller area (IOS) is very similar to the IOU metric; only the area of the intersection is divided by the area of the smaller of the two boxes.

The thesis also details three variants of the post-processing algorithms that were utilized in the experiments: Greedy non-maximum suppression (NMS), Non-maximum merging (NMM), and Greedy non-maximum merging (GREEDYNMM). In a recent study [A1], the author of this thesis explores the influence of the parameters mentioned above on the final prediction precision in Tomato360 dataset.

5 RESULTS

Mainly the results relevant to the created Tomato360 dataset as a main linking component of the thesis are presented in the following text. For the full overview of all realized case studies, please refer to the full text of the dissertation.

5.1 Dataset Tomato360

The creation of the Tomato360 dataset represents a fundamental contribution of this dissertation. Developed as part of a collaborative project supported by the Technology Agency of the Czech Republic, the dataset provides high-resolution images of tomato rows in a greenhouse. It facilitates precise counting of tomato fruits for accurate yield predictions, essential for supply chain logistics and delivery contracts in commercial agriculture.

The dataset serves two purposes. First, it showcases the translation of AI technologies from theoretical constructs to practice, shedding light on challenges and solutions encountered. Second, it offers a benchmark for evaluating the performance of deep learning models, particularly outside the domain of standard public datasets, thereby examining their robustness and generalizability.

5.1.1 Data Collection and Annotation

The source videos for the Tomato360 dataset were collected using a 360-degree camera (Ricoh Theta Z1) in a hydroponic greenhouse Bezdinek over 2020-2022. The dataset's name stems from this camera choice, as it provided the wide field of view necessary to capture the vertical range of tomato plants (up to 5 meters in height, but with narrow alleys), which traditional techniques struggled with. Importantly, this approach separates data acquisition from processing, allowing easier automation of data collection.

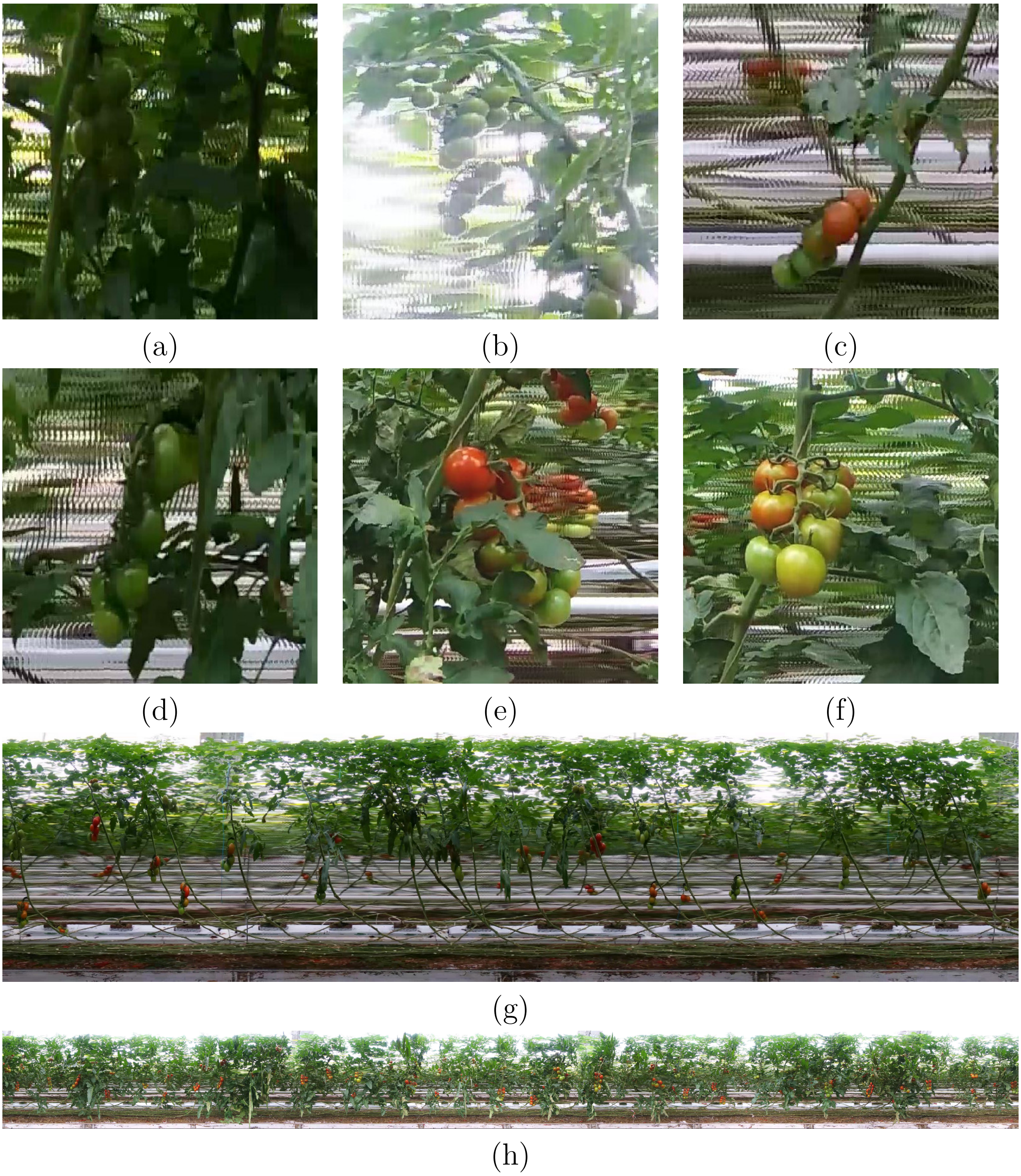

Fig. 5.1 Example images documenting difficulties of Tomato360 dataset: (a) dark, (b) bright, (c) back row tomatoes, (d) fruit overlapping each other, (e) leaf occluding tomatoes, (f) indeterminacy of ripeness, (g-f) overview of whole wide images.

The preprocessing stage involved the conversion of the dual-view video stream into a unified panorama using equirectangular projection. After decomposing the video into frames, a vertical section from each frame was extracted, stitched together to produce an image of the entire row, and adjusted to correct any distortion from the 180-degree view. This method created a comprehensive, high-resolution image where each fruit appeared once, enabling accurate counting.

The annotation tool Labelme was chosen for data annotation. Upon completion, the annotations were exported to COCO format. The final dataset comprised 58 images, each containing tomatoes annotated according to their ripeness stage. In total, 13815 tomatoes were annotated.

5.1.2 Dataset challenges

The Tomato360 dataset introduces several unique challenges. The first is the high-resolution images, reaching up to 20,000 pixels in width but comprising small objects to be detected. Moreover, the images contain low-frequency noise due to frame sampling from the video source. Greenhouse-specific environmental factors also add complexity, such as variable brightness levels caused by the high vertical range in the images. This variation can influence the accuracy of the detection model.

The nature of tomato growth further complicates object detection. Tomatoes grow in trusses, often overlapping each other. It can be frequently obscured by leaves, leading to further detection difficulties. Unripe green tomatoes can blend with leaves, especially in dark or bright image parts. Additionally, tomatoes from the back row sometimes appear in images but should not be included in the count. The mistaken inclusion of back-row tomatoes caused frequent discrepancies among human annotators, too. Lastly, subjective judgments about tomato ripeness add another layer of complexity, as perceptions about a tomato's ripeness can vary significantly among individuals and with the context. Figure 5.1 showcases the difficulties present in the Tomato360 dataset.

5.2 Attention Mechanism Analysis

This section of the dissertation studies the role of attention mechanisms incorporated into deep convolutional neural networks (CNN), particularly focusing on small object detection. The use of attention mechanisms is showcased in two different application segments, but in both cases, the problem includes small, sparsely distributed targets embedded within larger contextual information.

The first part explores the domain of abdominal organs and tumor segmentation, where the tumors represent small and sparsely scattered objects within a larger 3D context. Here, the conventional U-net architectures are bolstered with attention gates to discern the minuscule yet significant tumors situated in specific organs. This enhancement successfully enriches the precision of segmentation, utilizing the surrounding organ context as vital information. For further information about his research, refer to the original paper [A2] published by the author of this thesis or to the corresponding results section in the dissertation.

The following results section in the dissertation investigates the inclusion of spatial attention mechanisms in object detection applied in the context of the Tomato360 dataset. In this case, the small objects of interest are tomatoes within a high-resolution image, which, similar to tumors, are relatively small and variably distributed. Also here, the spatial context, including the plant structure and neighboring tomatoes, plays a crucial role in the detection process, so the attention mechanism presumably might help.

An ablation study has been executed to evaluate the effects of different spatial attention mechanisms [26] on different models' performance. Surprisingly, these architectural choices did not significantly improve the final object detection. Therefore, taking into account the added computational complexities brought by attention module addition, the usage of the attention modules was evaluated as unprofitable in this use-case. For more detail, please refer to the adequate section in the results of the full thesis text.

5.3 Artificial Size Slicing Aided Fine-tuning and Hyper Inference

This section evaluates the newly proposed methodologies - Artificial Size Slicing Aided Fine-tuning (ASSAFT) and Artificial Size Slicing Hyper Inference (ASSAHI) - on the Tomato360 dataset. These methods extend on previous work, Slicing Aided Fine Tuning (SAFT) and Slicing Aided Hyper Inference (SAHI) introduced originally in [2] and further studied in author's publication [A1]. Evaluation takes place on the custom Tomato360 dataset, well-suited due to its high-resolution images and the distinct growth pattern of the tomatoes.

The methodologies were implemented in Python using PyTorch, and MMDetection [5]. The original images were split into training, validation, and test parts and processed using either fixed-size patches or artificial-size slicing.

Two object detection models: Faster R-CNN [20] and Tood [7], were compared, with training utilizing Stochastic Gradient Descent and adjustable learning rates. A mask identifying object group placements in test images during the ASSAHI technique were generated using a basic FCN-Unet [21] architecture. For comprehensive details on the methodologies, data preprocessing, model architecture, training, and inference, refer to the full dissertation text.

The semantic segmentation model, FCN-Unet, which is critical for ASSAHI, was able to segment and identify foreground objects from the background with a mean accuracy (mACC) of 79.69% and a mean Intersection over Union (mIOU) of 77.28%.

Comparing the results of models trained using SAFT and ASSAFT with test results produced by SAHI and ASSAHI, it was found that ASSAFT and ASSAHI substantially increased all evaluated metrics compared to the SAFT and SAHI methodologies. However, using ASSAHI alone with a model fine-tuned by SAFT did not yield better performance; see Tab. 5.1.

On top of that, the ASSAHI methodology is combined either with FCN-Unet segmentation mask prediction or with masks obtained from ground true an-

Tab. 5.1 Evaluation of proposed ASSAFT and ASSAHI methodologies on a Tomato360 dataset from Faster-RCNN DCN and Tood DCN models. Each row presents results from two different fine-tuning methods, SAFT and ASSAFT, combined with different inference techniques, SAHI and ASSAHI. For ASSAHI, the segmentation input is from either an FCN-Unet or a ground truth mask (gtmask) is presented.

model & method		mAP	mAP50	Precision	Recall	F1-score
faster-rcnn DCN						
SAFT	SAHI	0.398	0.674	0.67	0.75	0.71
	ASSAHI FCN-Unet	0.397	0.669	0.67	0.75	0.71
ASSAFT	ASSAHI FCN-Unet	0.431	0.713	0.71	0.80	0.75
	ASSAHI gtmask	0.455	0.753	0.78	0.85	0.82
Tood DCN						
SAFT	SAHI	0.367	0.645	0.65	0.72	0.68
	ASSAHI FCN-Unet	0.365	0.640	0.64	0.72	0.68
ASSAFT	ASSAHI FCN-Unet	0.432	0.703	0.70	0.77	0.74
	ASSAHI gtmask	0.461	0.747	0.75	0.82	0.78

SAFT - Slicing Aided Fine Tuning
ASAFT - Artificial Size Slicing Aided Fine Tuning
SAHI - Slicing Aided Hyper Inference
ASSAHI - Artificial Size Slicing Aided Hyper Inference

notation (gtmask) to showcase the full potential of ASSAHI, i.e., how much we can improve the results by utilizing a better semantic segmentation model. We can see that the difference between ASSAHI with FCN-Unet and gtmasks is noticeable and provide room for improvement.

Despite the improvements in object detection precision, the usage of ASSAFT and ASSAHI significantly increased the demand for computational resources, particularly during the training phase. This resulted in approximately five times more training data samples and an extended training process. However, the additional computational needs of ASSAHI compared to SAHI during the inference phase were relatively minimal. Since the inference time and the object detection accuracy are the most important in practical usage, we can conclude that the enhancements in object detection precision make the additional computational requirements worthwhile.

Tab. 5.2 Comparison of crop estimates made by an agronomist and by the proposed model.

harvest	current day		in 7 days	
row number	27	29	27	29
actual crop yield [kg]	30.40	38.20	92.20	99.60
agronomist's estimate [kg]	50.00	50.00	110.00	110.00
agronomist's error [%]	-64.50	-30.89	19.35	10.44
model's estimate [kg]	29.07	30.45	90.19	98.20
model's error [%]	4.38	20.29	2.18	1.41

5.4 Practical Applications in Tomato Greenhouse

This section moves beyond theoretical concepts to focus on tangible, real-world applications within a tomato farming environment. The first application addresses the detection of whiteflies on yellow sticky tags, which are commonly employed for pest monitoring in greenhouses. The proposed system utilizes SAFT and SAHI techniques and replaces repetitive human work. The study shows that the model's precision is competitive with human work, achieving a similar F1-score as human operators versus professional phytopathologists.

The next application focuses on tomato fruit detection in a greenhouse. A final proposed solution utilizes the Tood model enhanced with novel processing methods ASSAFT and ASSAHI, as described in the previous section 5.3. To evaluate the system's performance on the entire dataset, the data were divided into five folds, and the final results were aggregated from all folds test predictions. The resulting scores were solid, with a **precision** of **0.85**, a **recall** of **0.93**, and an **F1-score** of **0.89**.

To practically assess the methodology, two rows of Belioso tomatoes were captured at Bezdinek Greenhouse, with an agronomist estimating crop yield from each row this day and the following week. These estimates were compared with the model's predictions, which were calculated using an average tomato weight of 38.5g. The model consistently provided more accurate predictions. Such precision indicates reliable, practical usability in a tomato greenhouse scenario, where currently available crop estimates can vary substantially, and $\pm 20\%$ is a tolerable error from a commercial point of view.

6 FULFILLMENT OF THE DOCTORAL THESIS AIMS

This section summarizes the efforts undertaken to achieve the objectives of this dissertation, which were initially defined as follows:

1. **Appraise the current state of the research area:** The area of deep learning applied in computer vision is a rapidly evolving research field marked by frequent incremental advancements rather than groundbreaking discoveries. Despite this, diligent efforts were made to stay abreast of the latest studies published in reputable journals and conferences, with only the most relevant ones to the dissertation topic being selected. A thorough overview of these select studies can be found in the Literature Review of the doctoral thesis.
2. **Develop and curate a custom dataset in a Tomato greenhouse:** In partnership with NWT, a tech-oriented company, and Bezdinek, a tomato farming enterprise, the Tomato360 dataset was created. The dataset is introduced in the author's paper [A1]. The process of data acquisition, image production, and labeling is detailed in the results section 5.1 and further in the dissertation. This section also includes basic statistics and identifies major challenges presented by the dataset.
3. **Investigate and compare the effectiveness of attention mechanisms**: Section 5.2 presents two case studies of attention mechanism integration into deep CNNs. The first study, published in the impacted journal [A2], documents a successful implementation of attention gates in medical image segmentation. In the second case, an ablation study of spatial attention incorporation into different object detection models was conducted and tested on the Tomato360 dataset.
4. **Develop an enhanced deep learning pipeline**: The effectiveness of deep learning methodologies depends significantly on a robust deep learning pipeline. All the fundamental components of such a pipeline

are outlined in the Methodology part of the doctoral thesis. The practical knowledge needed to establish such a comprehensive methodology overview was gathered over the whole doctoral studies of the author, and its correctness is confirmed in a successful computer vision application in different application fields published by the author of this thesis: [A1, A2, A4, A5, A6, A7, A8, A9]. Moreover, the novel techniques, Artificial Size Slicing Aided Fine Tuning (ASSAFT) and Artificial Size Slicing Aided Hyper Inference (ASSAHI) are introduced in sections 4.1.1 and 4.1.2, respectively. Those techniques are specifically tailored to handle high-resolution images with small objects within to be detected.

5. **Evaluate the proposed pipeline on the custom dataset**: The newly proposed methods of Artificial Size Slicing Aided Fine Tuning (ASSAFT) and Artificial Size Slicing Aided Hyper Inference (ASSAHI) are successfully applied to a custom-made Tomato360 dataset. The methodologies were tested using two different object detection model architectures. The effects of each methodology component were analyzed in the results, in section 5.3, along with the demands on time and computational resources.

6. **Analyze the impact and practicality of the proposed methods**: This dissertation thoroughly examines the implications and feasibility of incorporating attention mechanisms and the newly developed processing pipeline for small object detection in high-resolution images. Despite promising outcomes from the application of attention mechanisms in medical image data [A2], a similar mechanism did not yield significant improvements for the Tomato360 dataset as discussed in section 5.2. In contrast, the novel image slicing techniques - ASSAFT and ASSAHI, showcased substantial improvements in the final tomato detection performance. This reinforces the importance of a methodically constructed deep learning pipeline as a critical determinant of successful real-world applications of deep learning models. Moreover, the final proposed solution proved to be a reliable basis for predicting tomato crop yields in a real-world setting, as is documented in section 5.4.

7 IMPACT OF WORK ON SCIENCE AND PRACTICE

Deep learning techniques, and especially deep convolutional neural networks, occupy the field of computer vision nowadays, outperforming other techniques substantially. Despite the success of deep CNN techniques, there are difficulties inherent to their applicability. First, large datasets are needed for the successful training of deep CNN models, which requires a considerable amount of resources. Aside from problems due to the cost of acquisition, labeling, and data anonymization techniques, the methodology of processing and dealing with the data strongly influence the final method's success rate. This work seeks to establish an overview of the current standard techniques and best practices to set up the logical, consistent pipeline applicable to different computer vision tasks.

The practical knowledge needed to establish such a comprehensive methodology overview was gathered over the whole doctoral studies, and its correctness is confirmed in successful computer vision applications in different fields published by the author of this thesis: [A1, A2, A4, A5, A6, A7, A8, A9]. This thesis, moreover, documents the application of deep convolutional neural networks in a practical, real-world application from a commercial farming environment. From the first problem definition to a final solution.

The assignment of tomato fruit counting appears repeatedly throughout this work and creates the connection between theoretical research and practical application. In this practical example, this work documents the complex and non-linear process of creating a real-world dataset, shedding light on the unique challenges that arise in specific application contexts and proposing solutions to address them. Those challenges are not included in common large-scale datasets utilized by the computer vision community and therefore needed a slightly different approach. By providing a detailed account of how these challenges were identified and addressed, this dissertation underscores the need for flexibility and innovation in the application of deep learning techniques to real-world problems. Furthermore, it highlights the potential value that custom datasets can bring in furthering our understanding of how deep learning techniques behave in varying contexts and how they can be adapted and optimized for it.

A substantial effort is made to document the decision process of development and employ extensive analysis to empower the decision with comprehensive information. While the importance of architecture changes and extensions proved not to be very significant in the problem of tomato fruit detection, the significance of a well-structured deep learning pipeline was reinforced through the execution of the newly proposed methodologies: Artificial Size Slicing Aided Fine Tuning (ASSAFT) and Artificial Size Slicing Aided Hyper Inference (ASSAHI). These methods provide an innovative approach to data processing, specifically aimed at small object detection groups in high-resolution images. The results from the application of ASSAFT and ASSAHI on the Tomato360 dataset yielded notable enhancements in the success rates of object detection. These outcomes further underline the crucial role that data management plays in effectively implementing deep learning techniques into practice.

Finally, this doctoral thesis significantly contributes to both the scientific community and practical applications by highlighting the importance of a robust deep learning pipeline, introducing innovative methodologies for enhancing object detection in high-resolution images, and demonstrating the process and value of creating and using custom, real-world datasets. It is a stepping stone that bridges the gap between theory and practice in the field of deep learning, shedding light on the path for future research and applications.

8 CONCLUSION

This dissertation presents a thorough exploration of small object detection within high-resolution images, focusing on applications across various domains. An in-depth study on attention mechanisms was conducted, where the successful implementation of a U-Net model with attention gates led to improved detection of abdominal organs and tumors in CT images. This achievement shed new light on the importance of attention in complex medical imaging. However, the research also discovered that similar attention mechanisms did not prove to be beneficial in the specific case of tomato detection, providing practical insights into the domain-specific nature of these techniques.

In the agricultural context, the work introduced a comprehensive framework for tomato fruit detection, demonstrating a multi-faceted approach that synergizes various cutting-edge techniques. Leveraging the Tood object detection model, novel methods: Artificial Size Slicing Fine Tuning (ASSFT), and Artificial Size Slicing Hyper Inference (ASSAHI), were developed, resulting in a solid F1-score of 0.89. These innovative techniques allowed for accurate yield predictions in a real-world setting, outperforming common agronomist estimates and providing an economically advantageous solution.

In conclusion, the research detailed in this dissertation contributes substantially to both the field of computer vision and practical applications within the medical and agricultural sectors. By advancing the understanding of attention mechanisms, innovating in small object detection, and demonstrating real-world applicability in applications from Tomato greenhouse, this work establishes a robust and reliable approach to high-resolution image analysis. The insights and methodologies developed throughout this research provide a robust foundation for future exploration, setting the stage for further refinement and expansion into diverse applications and challenges within object detection and beyond.

REFERENCES

[1] Deep Image: Scaling up Image Recognition. *CoRR*. 2015, abs/1501.02876. Dostupné z: `<http://arxiv.org/abs/1501.02876>`. Withdrawn.

[2] Akyon, F. C., Cengiz, C., Altinuc, S. O., Cavusoglu, D., Sahin, K. and Eryuksel, O. SAHI: A lightweight vision library for performing large scale object detection and instance segmentation, November 2021.

[3] Bilic, P. et al. The Liver Tumor Segmentation Benchmark (LiTS). *CoRR*. 2019, abs/1901.04056. Dostupné z: `<http://arxiv.org/abs/1901.04056>`.

[4] Chen, I.-T. and Lin, H.-Y. Detection, Counting and Maturity Assessment of Cherry Tomatoes using Multi-spectral Images and Machine Learning Techniques. In *VISIGRAPP (5: VISAPP)*, pp. 759–766, 2020.

[5] Chen, K. et al. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv preprint arXiv:1906.07155*. 2019.

[6] Ding, J. et al. Object detection in aerial images: A large-scale benchmark and challenges. *IEEE transactions on pattern analysis and machine intelligence*. 2021, 44, 11, pp. 7778–7796.

[7] Feng, C., Zhong, Y., Gao, Y., Scott, M. R. and Huang, W. Tood: Task-aligned one-stage object detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3490–3499. IEEE Computer Society, 2021.

[8] Fuglie, K. The growing role of the private sector in agricultural research and development world-wide. *Global Food Security*. 2016, 10, pp. 29–38. ISSN 2211-9124. doi: https://doi.org/10.1016/j.gfs.2016.07.005.

[9] Girshick, R., Donahue, J., Darrell, T. and Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*. 2015, 38, 1, pp. 142–158.

[10] Gongal, A., Amatya, S., Karkee, M., Zhang, Q. and Lewis, K. Sensors and systems for fruit detection and localization: A review. *Computers and Electronics in Agriculture*. 2015, 116, pp. 8–19.

[11] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*. 2012.

[12] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. and Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*. 2021, 18, 2, pp. 203–211.

[13] Kirillov, A., He, K., Girshick, R., Rother, C. and Dollár, P. Panoptic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9404–9413, 2019.

[14] Liu, G., Nouaze, J. C., Touko Mbouembe, P. L. and Kim, J. H. YOLO-Tomato: A Robust Algorithm for Tomato Detection Based on YOLOv3. *Sensors*. 2020, 20, 7. ISSN 1424-8220. doi: 10.3390/s20072145.

[15] Long, J., Shelhamer, E. and Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[16] Mu, G., Lin, Z., Han, M., Yao, G. and Gao1, Y. Segmentation of kidney tumor by multi-resolution VB-nets. Technical report, Shanghai United Imaging Intelligence Inc., Shanghai, China, 2019.

[17] Mu, Y., Chen, T.-S., Ninomiya, S. and Guo, W. Intact Detection of Highly Occluded Immature Tomatoes on Plants Using Deep Learning Techniques. *Sensors*. 2020, 20, 10. ISSN 1424-8220. doi: 10.3390/s20102984.

[18] Oktay, O. et al. Attention U-Net: Learning Where to Look for the Pancreas. *CoRR*. 2018, abs/1804.03999. Dostupné z: <`http://arxiv.org/abs/1804.03999`>.

[19] Parico, A. I. B. and Ahamed, T. Real Time Pear Fruit Detection and Counting Using YOLOv4 Models and Deep SORT. *Sensors*. 2021, 21, 14. ISSN 1424-8220. doi: 10.3390/s21144803.

[20] Ren, S., He, K., Girshick, R. and Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pp. 91–99, 2015.

[21] Ronneberger, O., Fischer, P. and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

[22] Simpson, A. L. et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *CoRR*. 2019, abs/1902.09063. Dostupné z: `<http://arxiv.org/abs/1902.09063>`.

[23] Xu, Z.-F., Jia, R.-S., Liu, Y.-B., Zhao, C.-Y. and Sun, H.-M. Fast Method of Detecting Tomatoes in a Complex Scene for Picking Robots. *IEEE Access*. 2020, 8, pp. 55289–55299. doi: 10.1109/ACCESS.2020.2981823.

[24] Zhao, Y., Gong, L., Huang, Y. and Liu, C. A review of key techniques of vision-based control for harvesting robot. *Computers and Electronics in Agriculture*. 2016, 127, pp. 311–323. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2016.06.022.

[25] Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q. and Ling, H. Detection and Tracking Meet Drones Challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021, pp. 1–1. doi: 10.1109/TPAMI.2021.3119563.

[26] Zhu, X., Cheng, D., Zhang, Z., Lin, S. and Dai, J. An empirical study of spatial attention mechanisms in deep networks. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6688–6697, 2019.

AUTHOR'S PUBLICATIONS

Journal Publications with Impact Factor

[A1] Turečková, A., Tureček, T., Janků, P., Vařacha, P., Šenkeřík, R., Jašek, R., Psota, V., Štěpánek, V., Komínková Oplatková, Z., Slicing aided large scale tomato fruit detection and counting in 360-degree video data from a greenhouse. In *Measurement*, vol. 204, pp 111977, 2022, Elsevier. ISSN 0263-2241. DOI: `10.1016/j.measurement.2022.111977`

[A2] Tureckova, A., Turecek, T., Kominkova Oplatkova, Z., Rodríguez-Sánchez, A. J. Improving CT Image Tumor Segmentation Through Deep Supervision and Attentional Gates. In *Frontiers in Robotics and AI*, vol. 7, pp 106, 2020, Frontiers. ISSN 2296-9144. DOI: `10.3389/frobt.2020.00106`

[A3] Li, H., Zhang, H., Xu, Y., Tureckova, A.„ Zahradník, P., Chang, H., Neuzil, P. Versatile digital polymerase chain reaction chip design, fabrication, and image processing, In *Sensors and Actuators B: Chemical*, vol. 283, pp. 677-684, 2019. Elsevier B.V. ISSN 0925-4005. DOI: `10.1016/j.snb.2018.12.072`.

Journal Publications Indexed in Scopus

[A4] Tureckova, A., Holik, T. Kominkova Oplatkova, Z. Dog Face Detection Using YOLO Network. In *MENDEL*, vol. 26, num. 2, pp 17-22, 2020. DOI: `10.13164/mendel.2020.2.017`

Conference Proceedings

[A5] Tureckova, A., Turecek, T., Kominkova Oplatkova, Z. ICIP 2022 Challenge: PEDCMI, TOOD Enhanced by Slicing-Aided Fine-Tuning and Inference, In *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 4292–4295, Bordeaux, France, 2022. DOI: 10.1109/ICIP46576.2022.9897826.

[A6] Tureček, T., Vařacha, P., Turečková, A., Psota, V., Janků, P., Štěpánek, V., Viktorin, A., Šenkeřík, R., Jašek, R., Chramcov, B., Grivas, I., Komínková Oplatková, Z., Scouting of Whiteflies in Tomato Greenhouse Environment Using Deep Learning. In *Agriculture Digitalization and Organic Production*, pp. 323-335, Singapore, 2022. Springer Singapore. ISBN 978-981-16-3349-2.

[A7] Tureckova, A., Turecek, T., Kominkova Oplatkova, Z., Rodríguez-Sánchez, A. J. KiTS challenge: VNet with attention gates and deep supervision, In *KiTS 2019 challenge*, preprint, 2020. URL http://results.kits-challenge.org/miccai2019/manuscripts/tureckova_2.pdf.

[A8] Tureckova, A., and Rodríguez-Sánchez, A. J. ISLES Challenge: U-Shaped Convolution Neural Network with Dilated Convolution for 3D Stroke Lesion Segmentation, In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 319-327, Cham, 2019. Springer International Publishing. ISBN 978-3-030-11723-8. DOI: 10.1007/978-3-030-11723-8_32.

[A9] Vlachynska, A., , Kominkova Oplatkova, Z. , Turecek, T. Dogface Detection and Localization of Dogface's Landmarks, In *Artificial Intelligence and Algorithms in Intelligent Systems*, pp. 465-476, Cham, 2019. Springer International Publishing. ISBN 978-3-319-91189-2. DOI: 10.1007/978-3-319-91189-2_46.

[A10] Vlachynska, A., Kominkova Oplatkova, Z., Sramka, M. The coordinate system of the eye in cataract surgery: Performance comparison of the circle Hough transform and Daugman's algorithm, In *AIP Conference Proceedings*, vol. 1863. 2017. AIP Publishing. DOI: 10.1063/1.4992259.

[A11] Vlachynska, A., Cerveny, J., Cmiel, V., Turecek, T. Automatic Image-Based Method for Quantitative Analysis of Photosynthetic Cell Cultures, In *Hybrid Artificial Intelligent Systems*, pp. 402-413, Cham, 2016. Springer International Publishing. ISBN 978-3-319-32034-2. DOI: 10.1007/978-3-319-32034-2_34.

[A12] Vlachynska, A., Sramka, M. Artificial Neural Networks Application in Intraocular Lens Power Calculation, In *Conference: 9th EUROSIM Congress on Modelling and SimulationAt: Oulu, Finland*, 2016. DOI: 10.1109/EUROSIM.2016.45.

CURRICULUM VITAE

Name: Alžběta Turečková, maiden name Vlachynská
E-mail: tureckova@utb.cz
Birth: 27th May 1991
Nationality: Czech

Education and Training

2015 - now **Engineering Informatics (Doctoral studies)**
Tomas Bata University in Zlin, Faculty of Applied Informatics
Aim of dissertation: Soft Computing Methods in Computer Vision

- Researcher (application of research, grants, projects)
- Artificial Intelligence Laboratory member (ailab.fai.cz)
- Lecturrer

2013 - 2015 **Biomedical Engineering and Bioinformatics** (Master's degree)
Brno University of Technology, FEEC
Thesis: Photosynthetic cell suspension cultures quantitative image data processing
Overall study results: passed with honor.

2010 - 2013 **Biomedical Technology and Bioinformatics** (Bachelor's degree)
Brno University of Technology, FEEC
Thesis: Searching adenine and guanine rich regions
Overall study results: passed with honor.

Study Internships

04 - 08/2018 **University of Innsbruck, Austria**
The deep learning models for image classification and segmentation
Internship focused on understanding and modifying CNN-based image classification and segmentation models. A medical data segmentation model was created for the ISLES 2019 challenge, published and presented at MICCAI 2019.

05 - 07/2017 **University of Oviedo, Oviedo, the Kingdom of Spain**
Image-based analysis of root development in Arabidopsis
Developed a program for automatic measurement of root length and root hair count in Arabidopsis.

01 - 04/2017 **Northwestern Polytechnical University, Xi'an, P. R. China**
Automatic image-based analysis of qPCR on the chip
Worked on quantifying DNA detection on a chip, based on the detection of positive/negative sub-reaction cells in fluorescent microscope images. The research findings were later published in journal.

Work Experience

10/2019-now	**Tomas Bata University in Zlin, Faculty of Applied Informatics** Lecturer of courses Software Project Development Tools, Computer Science Basic and Neural Networks. Consultant and tutor of bachelor and diploma theses. Junior researcher in A. I. Lab (ailab.fai.cz). Team member of projects from the Technology Agency of the Czech Republic. Her main research interest and professional focus lay on artificial intelligence methods applied in computer vision, deep learning, convolutional neural networks, and medical image processing.
2015-2017	**GEMINI oční klinika a.s.** Biomedical engineer (Biomedical engineering, Medical devices) Manager of internal IS. She did requirements analysis of medical work-flow and their implementation into clinical system. Alpha / beta -testing of clinical system. Definition and management of clinical reports.

Publication Activity

Orcid ID: 0000-0002-5566-7393
Nr. publications: 12 (3 IF journals, 1 Scopus journal, 8 conference proceedings)
Nr. citations: 85 (excluding self-citations)
H-index: 5

Personal Skills and Competences

Mother tongue: Czech
Other languages: English (C1)
Computer skills and competences: Python, C++, Matlab, PyTorch, OpenCV
Social skills and competences: Friendly and communicative, she is capable of working independently and also thrive in team settings, including those with an international composition. Her adaptability has been honed through multiple study internships and participation in project teams under the Technology Agency of the Czech Republic (TACR). These experiences have fortified her interpersonal skills and ability to adjust to diverse work environments.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ASSAFT	Artificial Size Slicing Aided Fine Tuning
ASSAHI	Artificial Size Slicing Aided Hyper Inference
bbox	bounding box
COCO	Common Objects in Context, a large-scale dataset.
CV	Computer Vision
FLOP	Floating-Point Operation
FLOPS	Floating-Point Operation Per Second
GFLOPS	Giga Floating-Point Operations Per Second
IOU	Intersection Over Union
IOS	Intersection Over Smaller Area
DL	Deep Learning
GREEDYNMM	Greedy Non-maximum merging
M	Mega, Millions, 10^6
NMM	Non-maximum merging
NMS	Greedy non-maximum suppression
SAFT	Slicing Fine Aided Tuning
SAHI	Slicing Hyper Aided Inference
T	Tera, trillions, 10^{12}
YST	Yellow Sticky Tag

Ing. Alžběta Turečková, Ph.D.

Deep Learning Methods Applied in Computer Vision

Využití metod hlubokého učení v počítačovém vidění

Doctoral Thesis Summary

Published by Tomas Bata University in Zlin

Nad Stráněmi 4511,760 05 Zlín, Czech Republic

First edition

Circulation: published electronically

Typography: Ing. Alžběta Turečková, Ph.D.

This publication did not pass through a language nor an editorial revision.

The year of publication: 2023

ISBN 978-80-7678-191-7