

Although all the points have been fulfilled, there are several weak spots. Namely:

- Section 3.1: 'A convolutional neural network (CNN) is a shift-invariant.' That is not valid, it is shift-equivariant.
- Fig. 4.2: a nice pipeline. But when crops are created and the object detection network is trained, why not to use existing network, such as tomatoDet or annotated datasets? At least two are publicly available.
- Section 5.3.1: why use such an obsolete backbone as ResNet50?
- Section 5.3.1: the task of segmentation consists of two highly imbalanced classes, foreground vs. background. Using standard cross-entropy, the class with a smaller occurrence (foreground) will be suppressed. This can be seen in Table 5.6 where the mIOU is 77.0 and the IOU of the foreground is only 55.5. Focal loss or Dice loss (or combination of both, ideally) must be used. The most surprising thing is that the combination of Dice + CE is mentioned in Section 5.2.1.
- Section 5.3.1: training a network that includes batch normalization needs a large enough batch size to produce stable results. Here, the batch size of 2 is not adequate. In such a case, decreasing size of images even more than authors propose is the correct way as same as using mixed precision for training the network.
- The work does not include graphs illustrating training/validation losses to see how big the overfitting is and to estimate if the augmentations were chosen correctly.

These remarks do not decrease the overall quality of the work and may be answered during the defense of the thesis.

I acknowledge Ing. Alžběta Turečková proved with the thesis her qualification and skills and I recommend her for the degree Doctor of Philosophy, Ph.D.

Ostrava, 1.9.2023

doc. Mgr. Petr Hurtik, Ph.D.

University of Ostrava



Referee's report on the PhD thesis

Alžběta Turečková: Deep Learning Methods Applied in Computer Vision

The described topic is actual and reflects the current trend of bringing deep learning techniques into industry to increase the efficiency of manufacturing, or planting tomatoes in this case.

All the goals defined in the work were fulfilled, but some of them deserve to be extended. That is the case mainly of the first goal. Here, I am missing information on global pooling, regularization layers, or modern activation functions in Section 3.1; Section 3.2 ends with year 2018 and thus misses information on current methods such as DETR, the whole family of YOLO, CenterNet, DetectorRS, and many others approaches. The subpart of the fifth goal, namely 'Measure its performance against existing standard techniques used for small object detection.', has not been fulfilled; there is a lack of comparison with the standard way based on crops.

I consider the way to solve the defined tomato problem to be generally correct. The main emphasis is given to preprocessing and postprocessing of data for object detection network and to design own custom training pipeline consisting of two networks in total. The idea ASSAFT-ASSAHI, which seems to be the main contribution of the author, must be pointed out because it is great, smart, and simple. It also opens a direction for future work, namely integrating the two networks (segmentation + detection) into one solution that will be able to be trained in an end-to-end manner.

The author showed that the idea of ASSAFT-ASSAHI has also application in another task, whitefly counting. Also, thanks to the good results from tomato detection, the author can make a prediction of the future yield of tomato crops. The connection between research and industry is perfect and may serve as a starting point for further collaboration and improvement of the system.

The formal level is suitable, and the work does not include major language issues. The negative point is that the style of bibliography is not consistent. Regarding the fluency of the text, there are two issues. Firstly, Section 5.4.1 is almost identical to Section 5.3.1 - declaring that the training scheme is nearly the same and pointing out the differences only may increase the readability of the work. Secondly, the work deals with tomatoes, then accidentally jumps to tumor segmentation in Section 5.2 and again continues with tomatoes. I consider Section 5.2 to be confusing.

Parts of the work have been published in journals and conference proceedings, where Alžběta Turečková (Vlachynská) is also the first author in most of the publications. The publication record is appropriate and enhances the fact that the thesis has an adequate level of quality.

implementaci mechanismů v medicínské segmentaci obrazů, zatímco druhá zkoumá nasazení navrženého algoritmu v různých modelech detekce objektů na datasetu Tomato360.

Práce se důkladně věnuje návrhu a vývoji technik, které byly speciálně navrženy pro detekci malých objektů v obrázcích s vysokým rozlišením. Autorka detailně popisuje všechny klíčové komponenty tohoto přístupu, včetně nových technik nazvaných Artificial Size Slicing Aided Fine Tuning (ASSAFT) a Artificial Size Slicing Aided Hyper Inference (ASSAHI).

Autorka aplikovala a testovala navržené algoritmy na vlastním datasetu Tomato360 a porovnávala jejich výkon s existujícími standardními technikami pro detekci malých objektů. Výsledky byly podrobně analyzovány, zahrnující přesnost, robustnost a efektivitu navržených postupů. Autorka se zabývala výpočetními nároky, škálovatelností a obecností těchto navržených metod. Z výsledků je patrné, že nové navržené metody zpracování dat, ASSAFT a ASSAHI, dosáhly významných vylepšení v úspěšnosti detekce rajčat.

Celkově práce přináší důkladný přehled o výzkumu v oblasti hlubokého učení a jeho praktickém využití. Autorka zdůrazňuje důležitost inovativních metod pro úspěšnou aplikaci hlubokého učení v reálných scénářích. Práce také dokumentuje hodnotu vlastních datasetů a jejich využití při zkoumání chování metod hlubokého učení v různých kontextech.

Jedinou výtku, kterou mám na dizertantku, je ohledně její publikační činnosti. Přestože má autorka již 12 publikačních výstupů, tak se domnívám, že zaměření a potenciál její práce měl být uplatněn v prestižnějších časopisech (Springer, Elsevier).

Na dizertantku mám následující otázky:

1. Jak byla aplikována a ověřena navržená řešení na reálných scénářích, zejména na farmě Bezdíněk?
2. Jaký byl dopad této disertační práce na vědeckou komunitu a praxi? Jak byly tyto výsledky přijaty a jak mohou být využity ve vědeckém a praktickém kontextu?
3. Co bylo v průběhu Vaší vědecké práce nejtěžší v kontextu cílů práce?

Závěrečné doporučení

Přes výše uvedené připomínky přináší dizertační práce Ing. Alžběty Turečkové přinejmenším zajímavé poznatky a splňuje kritéria na dizertační práci kladené, a proto ji doporučuji k obhajobě.

Jestliže uchazeč uspokojivě zodpoví uvedené otázky a bude vhodně reagovat v rámci veřejné rozpravy, doporučuji mu udělit vědecko-akademickou hodnost "Philosophiae Doctor" (zkratka Ph.D.).

V Ostravě, dne 10.9.2023

Martin Kotyrba

POSUDEK DISERTAČNÍ PRÁCE

Uchazeč: Ing. Alžběta Turečková
Studijní program: Inženýrská informatika
Fakulta aplikované informatiky
UTB Zlín

Téma práce: Využití metod hlubokého učení v počítačovém vidění

Oponent: doc. RNDr. Martin Kotyrba, Ph.D.
Přírodovědecká fakulta
Katedra informatiky a počítačů
Ostravská univerzita

Předložená dizertační práce Ing. Alžběty Turečkové se zabývá problematikou zpracování dat při praktickém použití technik hlubokého učení pro detekci objektů v obrazech s vysokým rozlišením. Práce přináší seznámení se s novými technikami ASSAFT a ASSAHI a tyto techniky byly v praxi použity pro odhad sklizně rajčat ve skleníku, proto má tato práce velký potenciál a nachází významné praktické uplatnění. Téma práce je vysoce aktuální. Práce je dělena na 8 kapitol včetně úvodu a závěru a má 118 stran.

Po formální stránce je práce kvalitní, obsahuje analýzu současného stavu, cíle práce, korektně použitou citační normu. Práce je logicky strukturovaná a psaná přehledně. Práce je strukturována tak, že postupně plní stanovené cíle. Níže uvádím podrobnější informace o každém z těchto cílů a způsob, jak byly dosaženy:

Zhodnocení aktuálního stavu výzkumné oblasti

Autorka práce provedla podrobnou literární rešerši, kde sledovala vývoj výzkumné oblasti v oblasti hlubokého učení v počítačovém vidění. Tato rešerše zahrnovala studium relevantních článků, publikací a konferenčních příspěvků. Autorka vybrala ty nejdůležitější a aktuální zdroje a integrovala je do práce, poskytující tak komplexní přehled o stavu výzkumu v této oblasti.

Vytvoření a spravování vlastního datasetu

Autorka spolupracovala s firmou NWT a farmou Bezdinek na vytvoření vlastního datasetu nazvaného Tomato360. V této části práce jsou detailně popsány postupy pro akvizici dat, tvorbu obrázků a jejich označení. Autorka také identifikovala hlavní výzvy spojené s vytvářením reálného datasetu, které se liší od standardních velkých datasetů používaných v běžných testovacích formách pro obor počítačového vidění.

Výzkum

V této části práce jsou popsány dvě případové studie, které se zaměřily na implementaci mechanismů do různých architektur konvolučních neuronových sítí (CNN). První studie dokumentuje úspěšnou

Review of the dissertation thesis submitted by Alžběta Turečková

Title: Deep Learning Methods Applied in Computer Vision

The thesis is, besides the introduction, composed of three main parts. The first part contains a thorough state of the art about applying deep learning methods in image processing, including object detection, semantic segmentation, and attention mechanism used in images. The next part summarizes the research's full methodology, including data processing and preparation, model construction, and training and testing approaches. Finally, validation metrics are introduced and described. The last part contains achieved results on two types of data - medical and tomato with the application of suggested approaches using an attention mechanism and compared with standard techniques. Together, the work is well-structured and easy to read.

The idea is well-defined, and the two application areas are well-selected. The topic is very actual because the attention mechanism is under deep investigation in many areas of data processing. Its application in images is natural and brings an entirely new system for image processing with added context. The author also emphasizes using the attention mechanism in high-resolution photos that adds complexity to further processing due to the huge number of objects in a single picture and reduces the sizes of the detected objects. The results achieved and presented in the thesis showed the value of the suggested approach in the described areas. Therefore, the novel approaches presented in work, attention mechanism in medical images, and added slicing tuning and hyper inferencing, are valuable and worth further analysis and application in different areas.

The language and structure of the thesis are on a perfect level. The text is well written, the English quality is excellent, and I found no mistakes or typos. The sources and references are deep and actual, and all related sources are mentioned.

What I found disturbing is that the lack of highlighting of best results in the result tables complicates evaluating the methods used. Also, the number of evaluated models and compared approaches is relatively small and should be improved with other Vision-transformer-based models to assess the quality of the suggested approach.

The author's publication includes twelve works, three journals, and conference papers. The number of results is above the average and more than sufficient for the Ph.D. defense. As a result of the Ph.D. study, it is good to list publications and present a good view of the community's critical acceptance of the proposed methods.

When I consider all aspects of the presented work, achieved results, and the publications, I think that the work done by the students is fascinating and promising for future research and

I recommend the thesis for defense and award of the Ph.D. title.

Ostrava, 08.09.2023

prof. Ing.
Jan Platoš,
Ph.D.

Digitally signed by
prof. Ing. Jan Platoš,
Ph.D.
Date: 2023.09.08
11:52:58 +02'00'

prof. Ing. Jan Platoš, Ph.D.

Department of Computer Science, FEECS

VSB-Technical University of Ostrava

