Information Modelling and Knowledge Bases XXXIII
M. Tropmann-Frick et al. (Eds.)

doi:10.3233/FAIA210476

Deep Learning for Knowledge Extraction from UAV Images[1]

S. Brezani [a], R. Hrasko [a], D. Vanco [a], J. Vojtas [a], P. Vojtas [a,b 2]
[a] *Globesy ltd. Zilina, Slovakia*
[b] *Dpt. Software Engineering, Charles University, Prague, Czechia*

Abstract. We study possibilities and ways to increase automation, efficiency, and digitization of industrial processes by integrating knowledge gained from UAV (unmanned aerial vehicle) images with systems to support managerial decision-making. Here we present our results in the secondary wood processing industry. First, we present a deployed solution for repeated area and volume estimated calculations of wood stock areas from our UAV images in the customer's warehouse. Processing with the commercial software we use is time-consuming and requires annotation by humans (each time aerial images are processed). Second, we present a partial solution where for computing areas of woodpiles, the only human activity is annotating training images for deep neural networks' supervised learning (only once in a while). Third, we discuss a multicriterial evaluation of possible improvements concerning the precision, frequency, and processing time. The method uses UAVs to take images of woodpiles, deep neural networks for semantic segmentation, and an algorithm to improve results. (semantic segmentation as image classification at a pixel level). Our experiments compare several architectures, backbones, and hyperparameters on real-world data. To calculate also volumes, the feasibility of our approach and to verify it will function as envisioned is verified by a proof of concept. The exchange of knowledge with industrial processes is mediated by ontological comparison and translation of OWL into UML. Furthermore, it shows the possibility of establishing communication between knowledge extractors from images taken by UAVs and managerial decision systems.

Keywords. automation of industrial processes; decision support; knowledge and information modeling and discovery; deep neural learning; modeling multimedia information and knowledge; content-based multimedia data management; UAV; photogrammetry; semantic segmentation

1. Introduction

Our long term research project is focused on studying possibilities and ways to increase automation, efficiency, and digitization of production, technological and logistic processes in the automotive industry using autonomously controlled UAV (unmanned aerial vehicles) means combined with ICT equipment for real-time processing and evaluation of acquired data according to Industry 4.0.

[1] This publication was realized with support of the Slovak Operational Programme Integrated Infrastructure in frame of the project: Intelligent systems for UAV real-time operation and data processing, code ITMS2014+: 313011V422 and co-financed by the European Regional Development Fund

[2] Corresponding Author, KSI MFF UK, Malostranske nam. 25, 118 00 Prague1, Czech Republic; E-mail: vojtas@ksi.mff.cuni.cz

There are numerous UAVs applications in managing civil infrastructure assets, such as routine bridge inspections, disaster management, power line surveillance, and traffic monitoring. This article describes our experience with an internally developed and deployed solution that uses a commercial photogrammetric product in the wood processing industry. Furthermore, we design new methods and prototypes in the mentioned above Industry 4.0 direction. That is, we increase automation of all processes, decrease the need for human expert intervention and interconnect our application with a decision support system via an ontology.

Industry 4.0 is the digital transformation of manufacturing and related industries and value creation processes in organizations, including logistics, supply chain, finance, accounting, and human resources. It helps manufacturers with current challenges by becoming more flexible and reacts easier to changes in the market. It can increase the speed of innovation and is very consumer-centered, leading to faster design processes. Implementation of this trend in an organization focuses on creating detailed digital models of reality, optimally real-time. This digital model (digital twin, see [23] for a framework reducing reality to a model) makes it much easier to oversee, control, and actively manage all production and manufacturing processes. A critical prerequisite is the acquisition of detailed data that can be processed and transformed into the knowledge needed for qualified management decisions by enriching the classic high-level data of the ERP system (e.g., orders and deliveries, accounting, plant management) with little-detailed operation data. It is commonly achieved using barcodes, QR codes, and scanners or using different IoT sensors.

Nevertheless, some data cannot be obtained, collected, or measured automatically. Appropriately equipped workers are necessary for manual collection, processing, and visual or sound data transformation. Humans' processing of visual or sound data means high costs and very long data update intervals. For example, inventory of externally stored material, such as containers, coal, wood stockpiles, or freshly made cars, can take several hours and days and often requires more personnel with adequate equipment — measuring equipment, dedicated software, or a protective kit. After an inventory check, the data is entered manually into the basic ERP systems, far from real-time processing. Based on this data, no real-time correction is possible. Only subsequent actions can be performed.

The research project aims to automatically collect outdoor visual data using pre-programmed UAVs and automatically process and transform them into knowledge using advanced computational tools such as machine learning based on deep neural networks. Deploying this solution to a real production facility can bring the capability of automatic data collection and processing of visual data regularly, with direct integration to core ERP systems in the form of alerts or data transfer. This way, the outdoor reality could be manageable almost in real-time. Our ambition is to deploy such a solution in the automotive production plant or its suppliers for logistics, warehousing, security, or maintenance. We start our research with automatic measurements of wood stockpiles in the wood storage facility. Slovakia is the fourth largest forest-covered country in the EU (with about 41% of the area, after Sweden, Finland, and Austria). The wood processing industry characterizes lower profit margins than in other sectors. Therefore, it is necessary to create value-added products in Slovakia and not just export raw wood abroad. For this reason, we consider it essential to bring new procedures and solutions using UAVs and intelligent image processing.

Our company has a research and development department, where we prepare prototypes in knowledge modeling and processing high-quality images from different

sources, such as UAV aerial images. Our starting point for this paper is a deployed solution to calculate wood stockpiles volume in the customer's warehouse using UAV images. The photogrammetry software we use to process UAV images requires annotating the area of interest from an expert. It can be challenging when multiple customers need to be served. Developing a generic solution that makes this annotation automated can be exciting in many practical applications.

2. Use case description

The customer is active in secondary wood processing, also known as value-added wood products manufacturing, generally defined as continuous manufacturing beyond lumber production. This customer needs information about the temporal development in wood stockpiles.

The on-site process begins by setting calibration points, placing them manually, and marking a known length. It is necessary for precise photogrammetry processing. The next step is to set the flight route data manually so the drone (UAV) is ready to fly and take appropriate area pictures. Afterward, pictures are taken and collected to fit automatic processing by a commercial tool, Pix4D, which creates an orthophoto map [17]. The created orthophoto map trained user manually annotates the areas of interest, which took the trained user about 20 minutes (depending on the area shape). Finally, Pix4D can calculate the woodpiles' area and volume.

Our goal is to eliminate manual processing steps to have a generic solution (with an API) that automated this process. Thus, our solution could be deployed for more customers without the need for a trained human expert annotator. We hope this can be very interesting in many practical applications also beyond the wood industry.

The solution we present here works using neural network-trained solutions for semantic segmentation and domain ontology in multimedia/spatial environments. Our main contributions are:

- Experiences and data from a deployed system using UAVs and the professional/commercial photogrammetry software.
- A new system integration solution interconnecting UAV aerial images from a wood log warehouse with the decision support system mediated by an ontology and customers' requirements. Depending on the application need, we can tune our solution up along several axes (e.g., precision, execution time, amount of human expert activity, frequency).
- An experimental prototype of the UAV aerial image processing system, based on several alternative deep neural network architectures and several pre-trained backbones. We improve semantic segmentation with a new algorithm.
- Experiments with calculations of the area of the wood logs pile base with real-world data from several UAV flights during 9 months and their comparison with the results from Pix4D.
- The extracted knowledge can be sent to the decision support system using a general external ontology equipped with the respective domain extension.

Other extensions of this use case can classify the type or quality of wood or work in places where manual calibration point settings are impossible. For example, the idea is to use a car catalog with known car dimensions. An alternative task may be to decide

only whether the warehouse wood reserve has increased or decreased. Another option would be to estimate the amount of wood delivered for a given time compared to the declared invoice for delivery (combined with vehicle weighting and license plate reader).

3. First deployed application and experiences

The UAV systems can quickly and efficiently collect data and capture vast areas of the globe from the surface in various spectra. The most commonly used spectrum for imaging is the orthophoto layer (geometrically corrected ("orthorectified") such that the scale is uniform). The data captured from the UAV also contains field data, which in conjunction with the orthophoto layer, we can process photogrammetrically post-process and thus provide data with higher added value. In addition, photogrammetry provides optical and mathematical methods and tools for calculating spatial/dimensional coordinates based on digital photography from the scanned area.

The main issue is how to use these data to get optimal information for specific tasks. First, let us focus on calculating estimates of wood stockpiles volumes.

- Several parameters affect the measurement results:
- Ability to obtain the most accurate information for 3D processing
- Processing time (higher accuracy takes longer)
- Degree of automation (how much manual work of an expert is required)

At first sight, it is clear that these parameters conflict with current technology standards in the field. However, we already successfully deployed our first solution in the customer warehouse, where a large amount of wood was processed. The primary task is to estimate the volume of regular wood stockpiles and their changes in time. As we mentioned above, the necessary process of collecting aerial pictures using drones is in place, so we already have calibrated system using drones, which collects the aerial area images.

In further processing, Pix4Dmapper[17] transforms the geodetic coordinates of the images' common points into a single 3D model of the scanned area into a point cloud. Such an approach can achieve high accuracy, but it is time-consuming for processing. The whole process displays a red arrow procedure in Figure 1. The rising demands on precision require more processing time in the range of hours to days.

To address processing time the automation is necessary. The automated data processing process follows the green arrow procedure in Figure 1. That shows how our new solution works. However, the only difference is which steps are manual and which are automated.

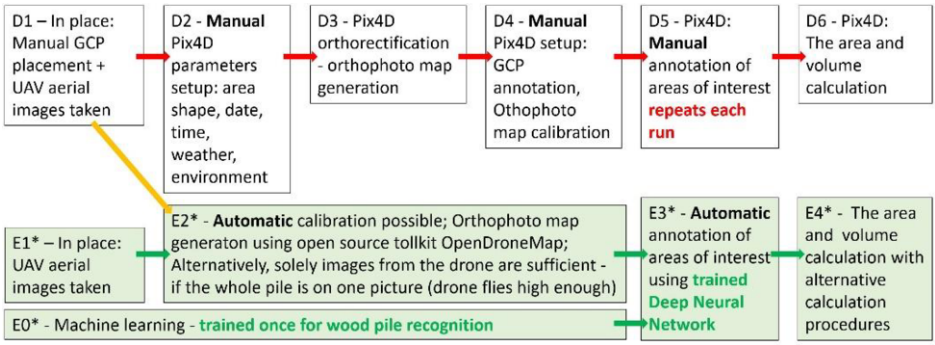

Figure 1. Procedure for processing the already deployed solution (red arrows) using GCP (ground control points) and our new approach (green arrows) with an alternative without GCP and functional area calculation and proof of concept of volume calculation

The already mentioned product process (red arrows procedure in Figure 1) takes place almost automatically, except for the three manual steps:

1. The first manual step (D2 in Figure1) is manual program (Pix4D) processing settings. The program has several attributes that are necessary to select before the process according to the parameters of the monitored environment and the subject of measurements, such as environment type, measured object shape, surrounding environment, the quality of the collected data (e.g., the influence of current weather on local brightness), required quality and precision of the results and other specifications.
2. The second manual step (D4 in Figure 1) is calibration using ground control points (GCPs) marked and measured at the beginning of the scan. The calibration itself consists of entering these parameters into the program based on their identification from the evaluated area's images. Again, we use state-of-the-art satellite technology with a deviation at the centimeter level to target GCP reference points. This step is performed on-site just before the drone takes off and is a manual annotation in the program itself (Figure 2).

Figure 2. Ground Control Points (GCPs) annotation in Pix4D

3. The third manual step (D5 in Figure 1) is to annotate the edges precisely to demarcate and define the wood base's shape (in all directions). This step is critical for accurate area and volume calculation. It consists of manually marking edges of a convex area directly in the orthophoto map, between which we want to measure the exact value of dimensions in space (see Figure 3). By the way, it is the most time-consuming step, and the prerequisite is trained professionals.

Figure 3. Manual annotation of the area of interest

After these three manual steps, the program (Pix4D) can calculate the area of annotated surface and volume of the corresponding wood stockpiles.

4. Model training and data annotation

The Pix4D photogrammetric method has proved to be valid and gives satisfactory results for our customers. However, in this method's background, we still find many activities requiring the manual work of highly trained users. Moreover, it would be necessary for ontology engineering and connection to a decision support system to have complete control over the application (e.g., using API).

Suppose we created a system that can proceed automatically without intervention. In that case, we could streamline the entire process of regular daily inventory measurements and at the same time effectively evaluate and monitor the movements (increase and decrease in the area) of material in the warehouse. We would potentially gain an overview of materials' movement over large areas of one or more warehouses of different customers. It would find justification in several industries by extending today's limited capabilities to almost unlimited use with automated UAVs for regular inventories. The idea to create an automated system for identifying the content (storage space occupancy) of stored wood in the warehouses of a wood processing company has been our quest.

For localizing the woodpiles in the images, we used deep neural networks (DNN). DNNs are used to solve several types of tasks such as image classification, object detection, etc. In our case, we use DNNs to solve an image segmentation type task with two classes (woodpile, background on pixel level). Since training models from scratch is very computationally and data-intensive, we used a transfer learning method to train our models. In this method, an existing model - which has been trained on a different dataset (e.g., an image segmentation model with a backbone trained to discriminate 80 object types) is used, and this model is then trained on a new task and data sample - identifying woodpiles. The advantage of this approach lies in the fact that the original model with a backbone was already able to identify basic shapes and their combinations. Thus, subsequent training just adapted this model to the new task and data.

The image segmentation task is of supervised machine learning type and hence needs annotated data (ground truth) for training. The annotated image represents the image itself and its metadata, defined as the relevant objects' location in the image. We

used the *Label Studio tool*[3] to annotate the images. We used it to mark the wood stockpiles on a sample of pictures. For annotation, we used original images from 3 flights in 4K resolution. These images have been scaled down and split into smaller 480x480px images. We manually annotated 1000 images. Subsequently, we split the data images into a training dataset (800 images) and a validation dataset (200 images).

We used augmentation within the training cycle to prevent overfitting in neural networks to increase the size and diversity of annotated data. The *imgaug library*[4] provides a wide range of transformations in order to transform both image and segmentation data. In our case, we used affine transformations (rotation, shift, zoom), contrast adjustment, noise generation, and the like.

As a baseline implementation of image segmentation models, we used the *Segmentation Models library*[5]. This library implements 4 model architectures for binary and multi-class classification (U-Net, PSPNet, Linknet, and FPN). In addition, the library uses the transfer learning method and allows using one of the 25 pre-trained networks (trained to classify the *2012 ILSVRC ImageNet dataset*[6]) as a backbone for the semantic segmentation architecture. This method makes it possible to use a trained neural network, or part of it, for another (related) category of tasks. In our experiment, we used 3 architectures (U-Net, PSPNet, and FPN) and 2 backbones (ResNet-18 and VGG-16). These models were trained using our annotated training and validation dataset.

Several factors affect the performance and accuracy of the trained model and the speed of training and interference during deployment in the production environment. These include:

- The chosen architecture - determines the model's performance, stability, the time required for its training and interference, and other aspects of neural network models. The development of neural network architectures for semantic segmentation is an area of intensive scientific research and development. An overview of current architectures can be found on the link7.
- The selected type of pre-trained neural network backbone - is essential when the transfer learning method is applied. As with architecture selection, the backbone neural network selection affects model performance, stability, training time, and interference.

[3] https://github.com/heartexlabs/label-studio

[4] https://github.com/aleju/imgaug

[5] https://github.com/qubvel/segmentation_models

[6] http://image-net.org/challenges/LSVRC/2012/

[7] https://paperswithcode.com/methods/category/segmentation-models

5. Methods, tools, and experiments

Figure 4 shows the progression of our experiment, which consists of multiple phases: First phase <*1. ODM*> shows the generation of an orthophoto map from the input images using the OpenDroneMap[8] tool.

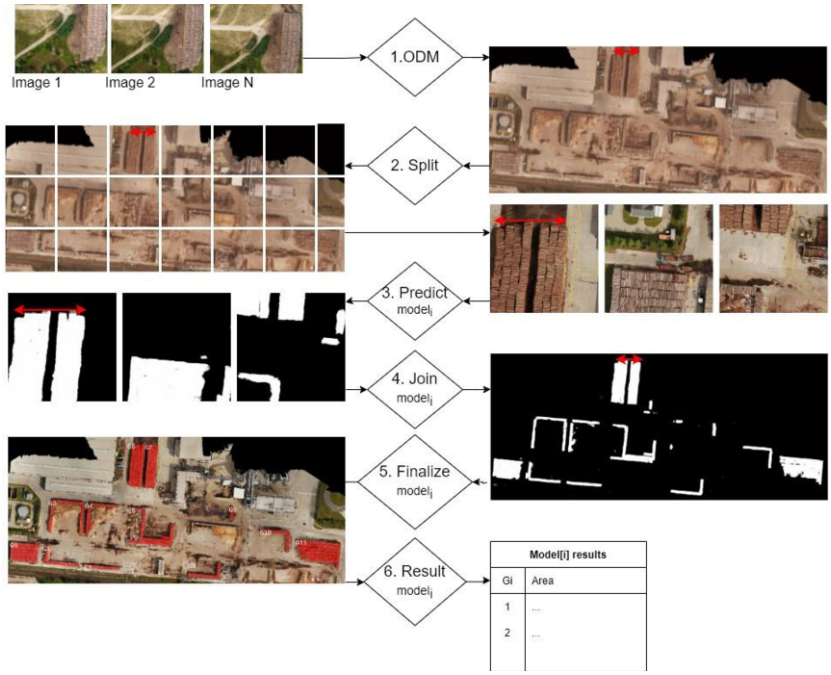

Figure 4. The progression of our experiment. See details below in the following text.

These images are in 4K quality, and their number is in the order of hundreds. This step is necessary because it is not always possible to have the whole woodpile on a single image (Figure 5).

The generated orthophoto map shows the entire monitored area in a single image. However, the orthophoto map is too large for further direct processing (~ 18000x12000 px) and therefore needs to be processed. Thus, the next phase <*2. Split*> splits the input orthophoto map into N images with a resolution of 480x480 px.

Figure 5. The input image shows three different parts of different woodpiles, none of which is visible as a whole.

[8] https://www.opendronemap.org/

The next phase <*3. Predict*> is the actual prediction/localization of the woodpiles. We used in parallel several trained deep neural network models ($model_i$) for the image segmentation task (described in section 4. Model training and data annotation). The input for such a model is a 3-channel (RGB) image with a resolution of 480x480px. The output is a 1-channel (gray-scale) image with the same resolution, where each pixel determines the probability with which a woodpile is or is not present at a given location. Thus, the output of the whole phase is N gray-scale images with a resolution of 480x480px. Since we chose several image segmentation models within the experiment, section *5.2 Experiments* presents the results of each model.

For further processing, for each $model_i$, these N predictions need to be combined. These images are merged in the next phase <*4. Join*>. For the join, an analogous procedure as in the phase <*2. Split*> is used.

After the predictions join, an image analogous to the orthophoto map containing the predicted positions of the woodpiles is produced. Since the result from the prediction is only a probabilistic map, these results still need to be processed. The final processing takes place in phase <*5. Finalize*>. For a more detailed description of this processing, see the next section, *5.1 Finalize*.

5.1. Finalize

The output from the model (each $model_i$) is a segmentation mask. The segmentation mask represents the probability that each pixel of the input image is part of the wood stockpile. The post-processing task is to convert such a segmentation mask of probabilities into a set of contours. Subsequently, it is possible to transform these regions into a set of polygons. This procedure is shown in Figure 6.

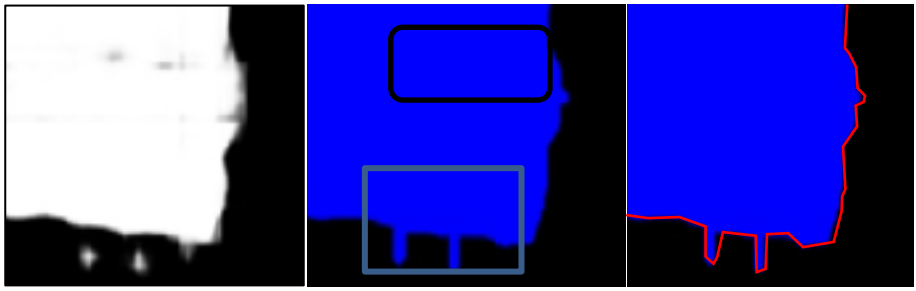

Figure 6. Segmentation mask with probabilities (black and white image with shades of gray), detected area by our algorithm (blue-black image), and projected detected area approximated by polygons (in red)

The left image shows the probability segmentation mask contains empty spaces (places without woodpiles), protrusions, and areas with different probabilities (places where the algorithm located woodpiles). The middle image represents the mask processed by our algorithm. It clearly defines where the stored timber is and where it is not.

The conversion of a segmentation mask with probabilities to a set of polygons directly impacts the prediction quality. Several hyperparameters can influence the conversion algorithm, which can significantly impact the final accuracy of the model prediction. We searched for the values of these parameters manually when solving

subtasks. The algorithm uses the OpenCV[9] library. Initially, the segmentation mask is converted to binary using the threshold parameter. Subsequently, the algorithm performs the close morphological operation on the binary mask using a kernel with the shape *kernel_shape* and size (*kernel_size*, *kernel_size*). The morphological operation serves to close the "holes" which are visible in Figure 6, left, for example. First, contours (contiguous areas of similar color and intensity) are searched for on such a modified binary mask. Next, the algorithm iterates the contours. If the contour area is smaller than the *min_area*, the algorithm excludes such a contour from further processing. Otherwise, the algorithm tries to approximate the contour using a polygon (*epsilon* parameter).Code converting a segmentation mask to a set of polygons:

```
INPUT: mask, threshold, epsilon, kernel_size
SET result TO []
SET mask_bin TO mask > threshold
CALL cv2.getStructuringElement WITH kernel_size, cv2.MORPH_ELLIPSE RETURNING
kernel
CALL cv2.morphologyEx WITH mask_bin, kernel, cv2.MORPH_CLOSE RETURNING mask_mod
CALL cv2.findContours WITH mask_mod RETURNING contours
FOR EACH contour IN countors
    CALL cv2.contourArea WITH contour RETURNING area
    IF area < min_area THEN
        CONTINUE
    END IF
    CALL cv2.arcLength WITH contour RETURNING arcl
    SET econtour TO arcl * epsilon
    CALL cv2.approxPolyDP WITH contour, econtour RETURNING cpoly
    APPEND cpoly TO result
END FOR
OUTPUT: result
```

Figure 7. Areas with stored wood that were the survey subject, marked in red, with the group identifier as a recognized area

5.2. Experiments

Figure 7 shows the area where we performed our experiments. The woodpiles are annotated (marked in red) by our algorithm. There are 12 woodpiles labeled G0 to G11 in the figure. To evaluate the performance of the models, we also need to know the actual area of the woodpiles - the ground truth (GT). We obtained these using the Pix4D tools. The ground truth of the stack area ranges from 98 m^2 (G9) to 1794 m^2 (G11).

[9] https://opencv.org/

The following table (Table 1) shows the results of the model predictions and GT areas of G0 - G11.

Table 1. Comparison of the results of our predictions of models for surfaces G0 to G11 with the ground truth (GT).

	Model prediction m²						GT m²
	FPN ResNet18	FPN VGG16	PSPNet ResNet18	PSPNet VGG16	UNet ResNet18	UNet VGG16	Pix4D
G0	1084	1080	1096	1078	1087	1083	1104
G1	334	314	344	317	344	311	325
G2	536	534	540	528	518	523	597
G3	218	214	207	214	214	211	234
G4	431	350	440	512	467	383	368
G5	915	867	898	890	929	909	948
G6	1147	1149	1156	1177	1150	1155	1138
G7	886	890	903	894	886	891	887
G8	293	281	283	292	290	282	298
G9	87	86	97	83	88	86	98
G10	281	278	272	263	281	276	305
G11	1685	1765	1708	1720	1755	1752	1794

Let M be the set of models, G be the set of areas {G0, …, G11}. A_{mg} is the model prediction for $m \in M$ and area $g \in G$, and GT_g is the ground truth area for $g \in G$. We evaluated the performance of the models using the MAE, CAE, and MAPE.

We can define the MAE (*mean absolute error*) metric as:

$$MAE(m) = \frac{\sum_{g \in G} |A_{mg} - GT_g|}{|G|} \quad (1)$$

The CAE (*cumulative absolute error*) metric is defined as:

$$CAE(m) = \sum_{g \in G} |A_{mg} - GT_g| \quad (2)$$

The MAPE (*mean absolute percentage error*) metric is defined as

$$MAPE(m) = \frac{100}{|G|} \sum_{g \in G} \left| \frac{A_{mg} - GT_g}{GT_g} \right| \quad (3)$$

Table 2. Comparison of individual models predictions using MAP (*mean absolute error*), CAE (*cumulative absolute error*), and MAPE (*mean absolute percentage error*) metrics

	MAE / m²	CAE / m²	MAPE / %
UNet VGG16	**25.39**	**304.72**	**5.63**
FPN VGG16	26.41	316.93	5.64
UNet ResNet18	28.98	347.74	6.88
FPN ResNet18	30.24	362.87	5.87
PSPNet ResNet18	33.55	402.63	6.48
PSPNet VGG16	42.48	509.71	9.19

The table shows that the model with U-Net architecture and VGG16 backbone is the best in all metrics. The MAE metric expresses the average absolute error of the area prediction. In the case of the best model, this value is ±25.39 m^2. Because the areas of the measured stacks are diametrically different in size 98 m^2 vs. 1794 m^2 (G9 / G11), the MAPE metric, which abstracts the area size, is also of interest. This metric expresses the average absolute percentage deviation from GT. Again, the best model has this value ± 5.63%.

The last CAE metric expresses the total absolute deviation of all G areas from GT. The CAE of the best model is 304.72 m^2. If we compare this value with the total area of woodpiles in the area $GT_{all} = \sum_{g \in G} GT_g = 8096.141\ m^2$, we get a deviation of 3.76 % (304.72 ÷ 80.96141). The error rate from the point of view of individual areas g, $g \in G$ shows the graph in Figure 8.

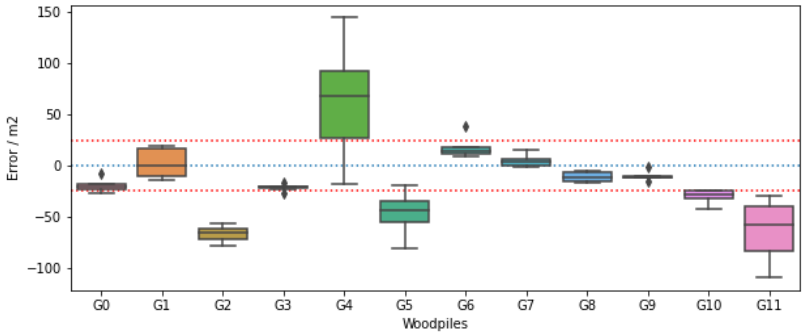

Figure 8. The graph shows the error rate of model predictions for each woodpile.

The graph shows G2, G4, G5, G10, and G11 areas, for which all models show a high error rate. Therefore, these areas can be described as "difficult". Similarly, the G4 area in which most models showed a high error rate. We assume that these areas contain patterns/structures of wood that the trained models cannot recognize. Figure 9 shows

images of the areas G4 and G11 with the marked problem parts. An ongoing analysis of these areas can therefore help increase the accuracy of the solution.

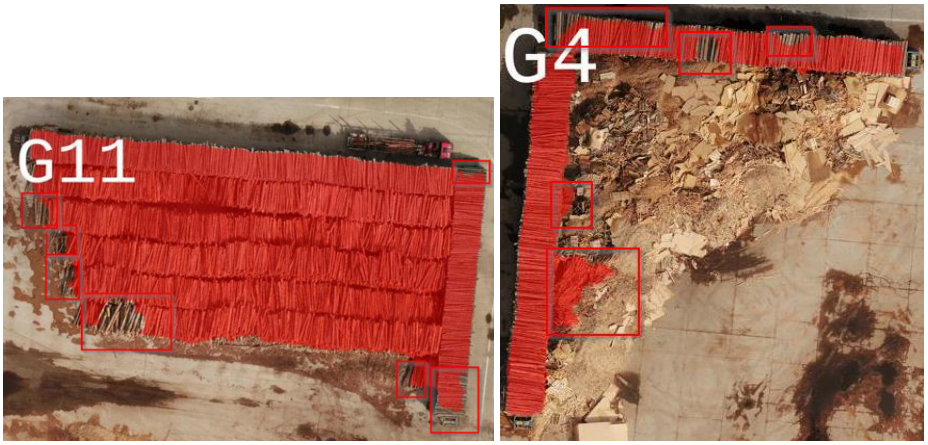

Figure 9. The red squares are the problematic parts in areas G4 and G11, where the models systematically fail.

6. Possible future development – proof of concept

Next, we describe our experience with further developments. These were neither fully implemented in our experimental tool nor fully evaluated so that we can describe their status only as proof of concept.

6.1. The backbone retrain

Our following motivation is to increase the accuracy of a neural network by the backbone improvement and its convergence to error loss elimination and a progression towards a network state where the network has learned to appropriately respond to a set of training patterns within some margin of error.

The backbone in the neural network serves as a features extractor. It is often trained on different tasks such as image classification and different data than we use. An interesting approach may be to retrain the backbone on a task similar to the required objective. For example, we need more annotated training data to retrain the backbone to image classification. However, data annotation is time-consuming and economically demanding; therefore, it is not worth using this exercise.

A self-supervised learning task, known as Contrastive Learning, does not require annotated data, maybe a suitable approach. An example is a SimCLR method[2]. Using this method, we can train the backbone on a large set of real data (in the order of 10,000 images) and thus adapt it to our needs and then use it in our semantic segmentation models.

6.2. Increasing the diversity of the training dataset

We aim to incrementally improve the following research training dataset by manually annotating the images where the model showed the highest error rate (see Figure 10).

Figure 10. Example of images with incorrect prediction - candidates for training set extension

6.3. Data input enrichment

In our research, we mainly focused on identifying wood stockpiles from UAV aerial images. Implemented models currently use these images as a prediction input, with each image being 3-channel (RGB).

Since the wood stockpiles are spatial objects, extending the models' inputs by the 4th channel is an elevation map (see Figure 11).

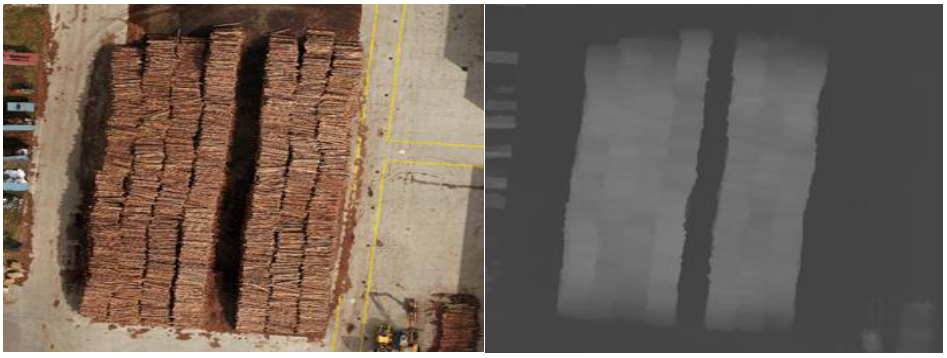

Figure 11. Example of wood stockpiles with the corresponding elevation map

An elevation map can be obtained directly from a LIDAR device as an elevation map or as a side output of the orthophoto map generation. Furthermore, the elevation map (see Figure 12) could extend the input of the prediction model by another channel - the elevation map. Thus, the input can be a 4-channel image (RGB + elevation map) for a composite model that improves semantic segmentation prediction. In the future, using an elevation map could be an important step in volume estimation. Our proof of concept shows it is feasible as envisioned.

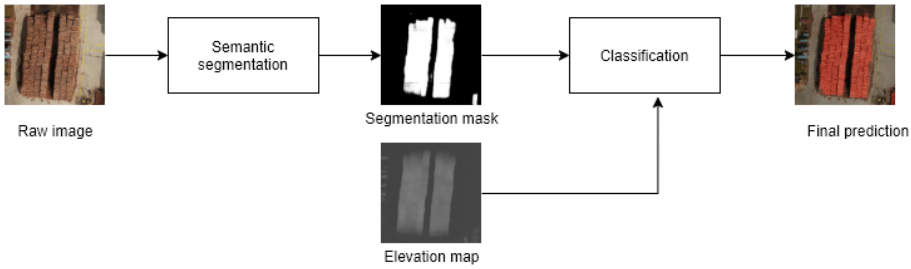

Figure 12. Schematic use of elevation map for improving semantic segmentation

Even under the assumption that the use of elevation maps can be a significant contribution, there are situations where its contribution can be debatable or misleading, for example, when the timbers merge with the surroundings, as in Figure 13 below.

Figure 13. Example of an elevation map where the wood mass overlaps with the environment (beams vs. sawdust or wood chips)

6.4. Other ideas

In paper [14], the authors describe the usage of car catalogs for object recognition. UAV aerial image is compared to one in the catalog with a similarity measure. In this way, it would be possible to estimate real dimensions in the future without the need for ground calibration (known vehicles have known dimensions). Some other objects can have known dimensions, e.g., track gauges. However, similar considerations are, so far, only a future work.

7. Decision support system enriched by knowledge extracted from UAV aerial images

We briefly describe how knowledge extracted from UAV aerial images can be sent to a decision support system. Start with a flat general ontology (e.g., DBPedia[10], Schema[11]),

[10] http://dbpedia.org/ontology/
[11] https://schema.org/

then we extend it with a Spatio-temporal model [16], and finally with a domain ontology (e.g., here [15]). The connection between OWL and UML can be made by [1] (or by owl2uml, which is a Protege plugin[12])

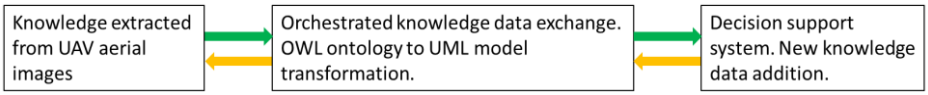

Figure 14. Knowledge exchange

Communication between our knowledge extractor and managerial decision system (inner company processes) goes both ways, as indicated in Figure 14. The right-to-left direction is first dedicated to automated communication of user requirements. There are several conflicting possible user preferences. We mention three axes here: the time required for one flight processing, required accuracy, and required repetition frequency.

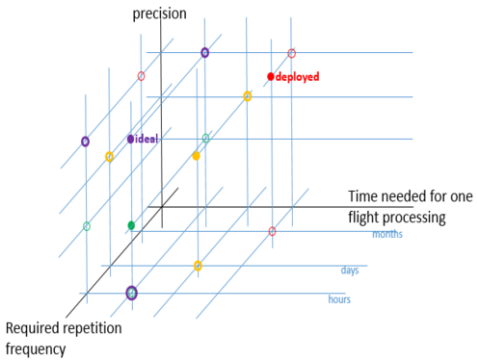

Figure 15. Different optimization strategies

Figure 15 illustrates interrelations between several alternatives: in **red,** our deployed solution with high precision, the considerable time needed for processing, and required repetition in months (solid bullet is the 3D position and bullets with no shape fill are respective projections to 2D planes). **Orange** and **green** are alternatives we consider in our experiments. **Violet** is a position of an ideal point for a user that wants all criteria of maximal benefit. This is a clear multicriterial situation, and we use our learning of aggregation function (to have an FLN-class preference model), see [11]. Many architectures, backbones, and other hyperparameters allow us to move almost continuously along with coordinates within a reasonable range. Trained preference models can then find an optimum in the area.

Our next plans are devoted to more general specifications of customer user requirements in natural language. Using our NLP techniques (see [4]), we can parse sentences into dependency trees, and after automated annotation, we can learn, e.g., a new domain of interest, which can be an entry into web search. In our example, these Datalog rules are pretty simple. In more complicated domains, this learning can be more involved. Communication is based on semantic models on both sides. Figure 16 shows an application diagram for data acquisition and subsequent knowledge extraction.

[12] https://protegewiki.stanford.edu/wiki/OWL2UML

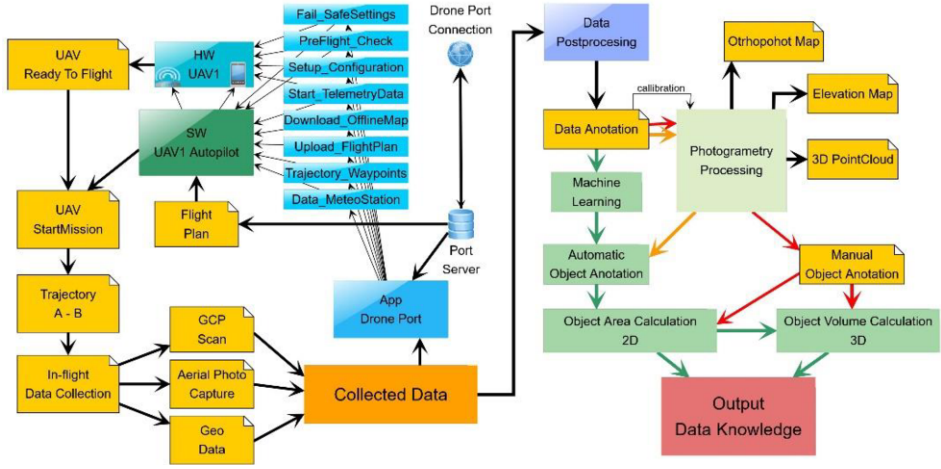

Figure 16. Data collection on the left; Knowledge extraction on the right

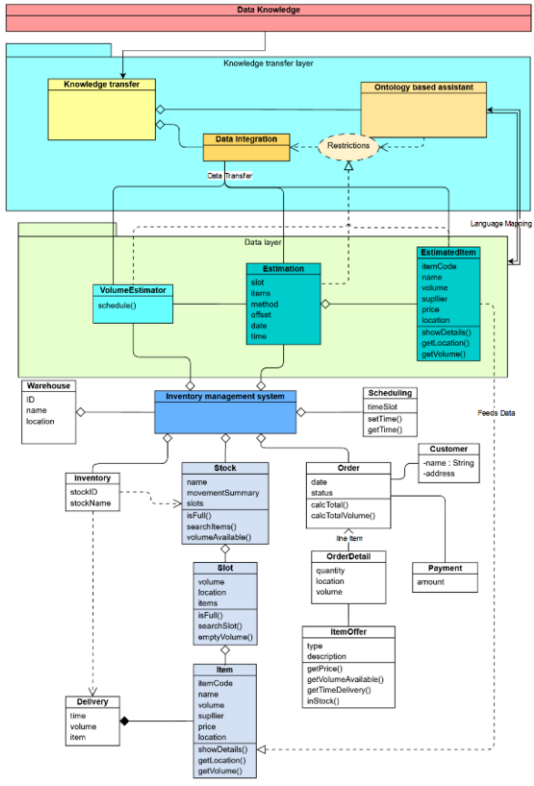

Figure 17. Example of integration with any general inventory management module (illustrated snapshot)

The extracted data knowledge is possible to transfer into any inventory management system negotiated by ontologies and contribute to the decision process. Inventory management modules worldwide provide processes of maintaining the appropriate level of stock in a warehouse. Inventory management activities involve identifying inventory requirements, setting targets, providing replenishment techniques and options, monitoring storage item usages, reconciling the inventory balances, and reporting inventory status. Integrating inventory management modules with other ERP modules (sales, purchase, and finance modules) allows ERP systems to generate vigilant executive-level reports.

One of the most crucial parts of collecting inventory data information nowadays is obtaining and processing vivid stock data. Real-time data provides unique opportunities to increase the capability of production efficiency in manufacturing environments. Real-time data collection is on the rise in every industry using very advanced approaches to data

collection - machine learning (ML), artificial intelligence (AI), deep learning (DL), unmanned aerial vehicle (UAV), and many more.

Using these advanced technologies, we can develop faster and reasonably precise solutions to obtain warehouse wood mass in real-time. Every warehouse dealing with wood mass has procedures (advanced procedures) on measuring volume wood mass in stock. Most of the time, it is based on volume estimation by experienced workers checking the stock regularly.

Our approach based on using drones to monitor and execute volume estimation can provide data knowledge to transfer to any inventory management system with an ontology assistant's help. Of course, several restrictions and rules must be applied to map the local domain model and data. However, such a knowledge management ontology tool can smartly integrate the resources into a coherent corpus of interrelated information as an inventory management data addition. Figure 17 shows a brief example of such a model.

Ontologies offer an alternative way to cope with heterogeneous representations of customer internal inventory management models. The domain model implicit in an ontology can be considered a unifying structure for giving information a common representation and semantics. The idea is that collected and calculated data can transparently transform from the UAV data processing model by introducing the knowledge transfer ontology-based assistant.

8. Related work

There are many commercial products and many more UAV aerial image processing applications: [17], [5], and, e.g., use of GNSS, ArcGIS reported by [13], to mention a few. [13] noticed that ArcGIS was 12-20 times faster than the use of GNSS, with comparable precision. On the other side, we can see that their camera and imaging gave a smaller density of points than ours. The main difference between our method and that of [13] is the increased precision of measurements achieved by a camera with higher resolution and a new UAV hardware generation. Secondary differences are twofold: A) we used the precise targeting of space using GCP, the reality against the virtual model, which created an even more accurate digital projection of the terrain. B) We adapted the data processing process in the Pix4D process settings, intending to achieve maximum accuracy.

Wood or forest-related studies, where objects may be under treetops, often use geographic or geological tools (see, e.g. [3]) to calculate accurate volumes using Lidar data.

Great motivation for us was papers [22] and [7]. Further ideas from [14] and [18] were also very inspiring. A big help was a lot of public domain software – detailed references are in footnotes on the appropriate place in this paper.

There is yet a broader context of our work, namely integrating neural (subsymbolic) and symbolic AI, which can be considered as a fifth dimension (added to 3D + time) as used in [10] and [19]. While machine learning has advanced thanks to deep neural networks rapidly, the trial and error approach it uses is similar to the way humans learn. It is sometimes failing due to the lack of data or context. However, humans developed language and other systems that make it possible to pass on knowledge directly to others who integrate that into their knowledge. [8] looks at the rules which enable knowledge transfer to work best in AI and integrate them with existing machine learning approaches.

[9] aims to bridge between the two paradigms. Authors discuss neural-symbolic integration in relation to the Semantic Web field, focusing on promises and possible benefits for both, and report on some current research on the topic. In [20], the authors address the contemporary problem of learning neural networks from relational data and knowledge representations. As they note: while virtually all standard models are limited to data in the form of fixed-size tensors, the relational data are omnipresent in the interlinked structures of the Internet and relational databases. Likewise, background knowledge in the form of relational logic rules or rich graph-based structures is often available in many domains, yet impossible or very difficult to exploit with the standard deep learning models.

In the case of object detection, the knowledge can be integrated after neural learning using knowledge about object classes, which are usually expressed by nouns.

The particular added value of [10] and [19] is the highly elaborated user interface, which increases intuitiveness.

In our approach on the symbolic AI side, we were motivated by [16], [15], [1], [6], and our former work, e.g. [4] and [21]. We used our [11] approach to fuzzy multicriterial systems based on the Fagin-Lotem-Naor FLN class of models. In one direction, we learn the preference model; in the opposite direction, we use the trained model to send users useful information.

Connections to the automotive industry are our long-term interest. In the paper [12], the authors describe a questionnaire for Spain's industry and its conclusions. Maybe we could go that way as well.

9. Conclusions and future work

Our long-term interest in studying possibilities and ways to increase automation, efficiency, and digitization of industrial processes using autonomously controlled UAV means interconnected with managerial decision support systems (especially in the automotive industry). In this paper, we took the first step to master the necessary methods in a much simpler domain. We consider two-way communication of knowledge and requirements between our system and an industry managerial system.

Here we presented our results in the secondary wood processing industry. First, we presented a deployed solution for calculating woodpiles volume from our UAV flight images during a time frame of 9 months. Processing is based on commercial photogrammetry software with some manual human intervention necessary. Second, we developed alternative automated solutions based on deep neural network learning and experimented with several deep neural network architectures, several backbone variants, and hyperparameters on real-world data.

Future work considers the use case extensions, e.g., concerning wood quality, monitoring the whole process from the forest via sawmills, transportation, various warehouses, and a more significant number of users.

In future work (hopefully in the final version after the conference), we would like to extend our experiments by studying the influence of loss function, training set, learning rate, learning scheduler, and regularization on final results.

References

[1] Brockmans S. et al. Visual Modeling of OWL DL Ontologies Using UML. In – ISWC 2004, McIlraith S.A. et al. eds. LNCS 3298, Springer 2004, 198-213

[2] T. Chen et al. A Simple Framework for Contrastive Learning of Visual Representations. In Proc. 37th International Conference on Machine Learning, PMLR 119:1597-1607, 2020 also https://arxiv.org/abs/2002.05709v3

[3] F. Chudy et al. Identification of Micro-Scale Landforms of Landslides Using Precise Digital Elevation Models, Geosciences 2019, 9, 117; doi:10.3390/geosciences9030117

[4] J. Dedek. Semantic annotations. 2012 Ph.D. thesis, Charles University, Faculty of Mathematics and Physics, Dpt. Software Engineering, advisor P. Vojtas https://dspace.cuni.cz/handle/20.500.11956/41689

[5] Volume Measurement with Drones, https://support.dronedeploy.com/docs/volume-measurement,

[6] J. Euzenat, P. Shvaiko. Ontology Matching, Springer 2013

[7] G. Ghiasi et al. Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation, arXiv:2012.07177v1

[8] G. Gottlob - Knowledge Processing, Logic, and the Future of AI - World Logic Day 2021, Vienna Center for Logic and Algorithms, January 14th, 2021 https://www.youtube.com/watch?v=j-HerlYBJEw

[9] P. Hitzler et al. Neural-symbolic integration and the Semantic Web. Semantic Web 11(2019)1-9

[10] Y. Kiyoki et al. System with SPA-Based Semantic Computing for Integrating and Visualizing Ocean-Phenomena with "5-Dimensional World-Map", Information Modelling and Knowledge Bases XXXII, IOS Press, pp. 76-91, January 2021

[11] Kopecky, M., Vojtas, P.: Visual E-Commerce Values Filtering Framework with Spatial Database metric. Computer Science and Information Systems, 17,3 (2020) 983–1006

[12] A. Martínez Sánchez, M. Pérez Pérez. Supply chain flexibility and firm performance: A conceptual model and empirical study in the automotive industry INTERNATIONAL JOURNAL OF OPERATIONS AND PRODUCTION MANAGEMENT 25,7 (2005) 681-700

[13] M Mokroš et al. Unmanned aerial vehicle use for wood chips pile volume estimation. ISPRS - The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 41 (2016) 953-956

[14] T. Moranduzzo and F. Melgani, "Detecting Cars in UAV Images With a Catalog-Based Approach," in IEEE Transactions on Geoscience and Remote Sensing, 52,10 (2014) 6356-6367

[15] A. Öhgren. Developing an Ontology for Wood-related Industry: An Experience Report. Research Report 04:6, School of Engineering, Jönköping University access December 10th, 2020

[16] Ch. Parent et al. Spatio-temporal conceptual models: Data structure + space + time. in ACM-*GIS'99*

[17] Pix4D: Professional photogrammetry&drone mapping https://www.pix4d.com/

[18] Radovic, M.; Adarkwa, O.; Wang, Q. Object Recognition in Aerial Images Using Convolutional Neural Networks. *J. Imaging* 2017, *3*, 21

[19] S. Sasaki et al. Global & Geographical Mapping and Visualization Method for Personal/Collective Health Data with 5D World Map System, Information Modelling and Knowledge Bases XXXII, IOS Press, pp. 134-149, January 2021

[20] G. Sourek, V. Aschenbrenner, F. Zelezny, S. Schockaert, O. Kuzelka. Lifted relational neural networks: Efficient learning of latent relational structures." In: Journal of Artificial Intelligence Research (JAIR) 62 (2018), pp. 69–100

[21] V. Vanekova, P. Vojtas. Comparison of Scoring and Order Approach in Description Logic EL(D). In SOFSEM 2010, J. van Leeuwen et al. eds. LNCS 5901, pp. 709-720, Springer 2010

[22] Hengshuang Zhao et al. Pyramid Scene Parsing Network, arXiv:1612.01105v2

[23] Brezani, S., Vojtas, P. (2021). Aggregation for flexible Challenge-Response. In FQAS'21, Andreasen, T. et al. eds. LNCS 12871, Springer, Cham, pages 209-- 222