Research article

Automated riverbed material analysis using Deep Learning on underwater images

Alexander A. Ermilov[1], Gergely Benkő[1] and Sándor Baranya[1]

[1]Department of Hydraulic and Water Resources Engineering, Budapest University of Technology and Economics, Budapest, 1111, Hungary

Correspondence to: Alexander A. Ermilov (ermilov.alexander@emk.bme.hu)

Abstract. The sediment of alluvial riverbeds plays a significant role in river systems both in engineering and natural processes. However, the sediment composition can show great spatial and temporal heterogeneity, even on river reach scale, making it difficult to representatively sample and assess. Indeed, conventional sampling methods in such cases cannot describe well the variability of the bed surface texture due to the amount of energy and time they would require. In this manuscript, an attempt is made to overcome this issue introducing a novel image-based, Deep Learning algorithm and related field measurement methodology with potential for becoming a complementary technique for bed material samplings and significantly reducing the necessary resources. The algorithm was trained to recognise main sediment classes in videos that were taken underwater in a large river with mixed bed sediments, along cross-sections, using semantic segmentation. Videos were taken on 3 different sites in the Upper Section of the Hungarian Danube. One served for training the AI algorithm, while the other two were for validation. The introduced method is fast, i.e., the videos of 300-400-meter-long sections can be analysed within minutes, with very dense spatial sampling distribution. The goodness of the trained algorithm was evaluated mathematically and via intercomparison with other direct and indirect methods, focusing on the percentages of the detected sediment fractions. For the final evaluation, the sieving analysis of collected physical samples were considered as the ground truth. The results of the AI algorithm were promising, in 64% of the compared sampling points the difference were ≤ 10% from the sieved physical samples, while for the rest of the points it also did not exceed 20%. Besides, the spatial trend in the fraction changes was also well captured along the cross-sections, based upon the visual evaluation of the footages. Suggestions for performing proper field measurements are also given, furthermore, possibilities for combining the algorithm with other techniques are highlighted, briefly showcasing the multi-purpose of underwater videos for hydromorphological assessment.

Keywords: riverbed texture, underwater mapping, sediment classes, Artificial Intelligence, Deep Learning, image-based

1 Introduction

The physical composition of a riverbed plays a crucial role in fluvial hydromorphological processes, as a sort of boundary condition in the interaction mechanisms between the flow and the solid bed. Within these processes, the grains on the riverbed are responsible for multiple phenomena, such as flow resistance (Vanoni and Hwang, 1967; Zhou et al., 2021), stability of the riverbed (Staudt et al., 2018; Obodovskyi et al., 2020), development of bed armour (Rákóczi, 1987; Ferdowsi et al., 2017), sediment clogging (Rákóczi, 1997; Fetzer et al., 2017), fish shelter (Scheder et al., 2015), etc. Through these physical processes, the bed material composition has a determining effect on numerous river-uses, e.g., possibilities of inland waterway transport, drinking water supply through bank filtration, the quality of riverine habitats, etc. Knowledge of riverbed morphology and sediment composition is therefore of major importance in river hydromorphology. In order to gain information about riverbed sediments, in situ field sampling methodologies are implemented.

Traditionally, bed material sampling methods are intrusive (i.e., sediment is physically extracted from the bed for follow-up analysis) and carried out via collecting the sediment grains one-by-one (areal, grid-by-number and pebble count methods, see e.g., Bunte and Abt, 2001; Guerit et al., 2018) or in a larger amount by a variety of grab samplers (volumetric methods, such as WMO, 1981; Singer, 2008). This is then followed by measuring their sizes individually on-site or transporting them to a laboratory for mass-sieving analysis (Fehr, 1987; Diplas, 1988; Bunte and Abt, 2001). These sampling procedures are time- and energy consuming, especially in large gravel and mixed bed rivers, where characteristic grain sizes can strongly vary both in time and space (Wolcott and Church, 1991; USDA, 2007), requiring a dense sampling point allocation. The same goes for critical river reaches, where significant human impact led to severe changes in the morphological state of the rivers (e.g., the Upper section of the Hungarian Danube; Török and Baranya, 2017). When assessing bed material composition on a river reach scale, experts usually try to extrapolate from the samples, and describe larger regions of the bed (even several thousand m^2) by data gathered in a few, several dozen points (see e.g., USDA, 2007; Haddadchi et al., 2018; Baranya et al., 2018; Sun et al., 2021). Gaining a representative amount of the sediment samples is also a critical issue. For instance, following statistical criteria such as those of Kellerhals and Bray (1971) or Adams (1979), a representative sample should weigh ten-to-hundred kg. Additionally, physical bed material sampling methods are unable to directly quantify important, hydromorphological features such as roughness or bedforms (Graham et al., 2005). Due to these constraints, surrogate approaches have recently been intensively tested to analyse the riverbed (see Chapter 2). Unlike the conventional methods, these techniques are non-intrusive and rely on computers and other instrumentation to decrease the need of human intervention and speed up the analyses.

One group of the surrogate approaches is the acoustic methods, where an acoustic wave source (e.g., an Acoustic Doppler Current Profiler; ADCP) is pointed towards the riverbed from a moving vessel, emitting a signal. The strength and frequency of this signal is measured while it passes through the water column, reflecting back to the receiver from the sediment transported by the river, and finally from the riverbed itself. This approach is fast and larger areas can be covered relatively quickly (Grams et al., 2013). While it has already become widely used for describing sediment movement (i.e., suspended sediment, Guerrero et al., 2016; bedload, Muste et al., 2016; and indirectly flow velocity; Shields and Rigby, 2005) and channel shape (Zhang et al., 2008), it has not reached similar breakthrough for riverbed material analysis. Researchers found that it is necessary to apply instrument specific coefficients to convert the signal strength, and these coefficients can only be derived by first validating each instrument using collected sediment samples with corresponding ADCP data. Moreover, the method is sensitive to the bulk density of the sediment and to bedforms (Shields, 2010), while it is also not possible to measure individual grains this way (Buscombe et al., 2014a; 2014b). Hence, the separation of surface roughness from the effects of bedforms is also not possible. Clay and silt patches could be separated with the acoustic approach, but gravel could not be distinguished strongly from sand.

Another group of the surrogate approaches is the application of photography (Adams, 1979; Ibbekken and Schleyer, 1986) and later computer vision or image-processing techniques. During the last two decades, two major subgroups emerged: one uses object- and edge detection (by finding abrupt changes in intensity and brightness of the picture, segmenting objects from each other; Sime and Ferguson, 2003; Detert and Weitbrecht, 2013), and the other one analyses the textural properties of the whole image, using autocorrelation and semi-variance methods to define empirical relationship between image texture and the grain sizes of the photographed sediments (Rubin, 2004; Verdú et al., 2005). The above-mentioned image processing approaches were very time consuming and required mostly site-specific manual settings, however, a few transferable and more automated techniques have also been developed recently (e.g., Graham et al., 2005; Buscombe, 2013). Even though there is a continuous improvement in the applied image-based bed sediment analysis methods, there are still major limitations the users face with, such as:

- Most of the studies (all the ones listed above) focuses on gravel bed rivers, and only a few exceptions can be found in the literature where sand is also accounted for (texture-based methods;e.g.: Buscombe, 2013).
- The adaptation environment was typically non-submerged sediment, instead of underwater conditions (a few exceptions: Chezar and Rubin, 2004; Warrick et al., 2009).
- The computational demand of the image processing is high (e.g., one to ten minutes per image; Detert and Weitbrecht, 2013).
- The analysis requires operator expertise (higher than in case of any conventional method).
- There is an inherent pixel- and image resolution limit (Buscombe and Masselink, 2008 Cheng, 2015; Purinton and Bookhagen, 2019). The finer the sediment, the higher resolution of the images should be (higher calculation time), or they must be taken from a closer position (smaller area and sample per image).

Nowadays, with the rising popularity of Artificial Intelligence (AI), several Machine Learning (ML) techniques have been implemented in image recognition as well. The main approaches of segmentation contra textural analysis still remain; however, an AI defines the empirical relationship between the object sizes (Igathinatane et al., 2009; Kim et al., 2020) or texture types (Buscombe and Ritchie, 2018) in the images and their real sizes. In the field of river sedimentology a few examples can already be found, where ML (e.g., Deep Learning; DL) was implemented. For instance, Rozniak et al. (2019) developed an algorithm for gravel-bed rivers, performing textural analysis. With this approach, information is not gained on individual grains (e.g., their individual shape and position), but rather the general grain size distribution (GSD) of the whole images. At certain points of the studied river basins, conventional physical samplings (pebble count) were performed to provide real GSD information. Using this data, the algorithm was trained (with ~1000 images) to estimate GSD for the rest of the study site, based on the images. The method worked for areas where grain diameters were larger than 5 mm, and the sediment was well-sorted. The developed method showed sensitivity to sand coverage, blurs, reduced illuminations (e.g., shadows) and white pixels. Soloy et al. (2020) presented an algorithm which used object detection on gravel- and cobble covered beaches to calculate individual grain sizes and shapes. 46 images were used for the model training, however, the number of images were multiplied with data augmentation (rotating, cropping, blurring the images; see Perez and Wang, 2017) to enhance the learning session and increase the input data. The method was able to reach a limited execution speed of a few seconds per m^2 and adequately measured the sizes of gravels. Ren et al. (2020) applied an ensemble bagging-based Machine Learning (ML) algorithm to estimate GSD along the 70 km long Hanford Reach of the Columbia River. Due to its economic importance, a large amount of measurement data has been accumulated for this study site over the years, making it ideal for using ML. By the time of the study, 13,372 scaled images (i.e., their millimetre/pixel ratio was known) were taken both underwater and in the dry zones, covering approx. 1 m^2 area each. The distance between the image-sampling points was generally between 50-70 m. An expert defined the GSD (8 sediment classes) of each image by using a special, visual evaluation-classification methodology (Delong and Brusven, 1991; Geist et al., 2000). This dataset was fed to a ML algorithm along with their corresponding bathymetric attributes and hydrodynamic properties, simulated with a 2D hydrodynamic model. Then, it was tested to predict the sediment classes based on the hydrodynamic parameters only. The algorithm performed with a mean accuracy of 53%. Even though this method was not image-based (only indirectly, via the origin of the GSD data), it highlighted the possibilities of an AI for a predictive model, using a high-dimensional dataset. Having such a large data of grain size information can be considered exceptional and takes a huge amount of time to gather, even with the visual classification approach they adapted. Moreover, this was still considered spatially sparse information (point-like measurements, 1 m^2 covered area/image dozens of meters away from each other). Buscombe (2020) used a set of 400 scaled images to train a AI algorithm on image texture properties, using another image-processing method (Barnard et al., 2007) for validation. The algorithm reached a good result for not only gravel, but sand GSD calculation as well, outperforming an earlier, but promising, texture-based method (wavelet analysis; Buscombe, 2013). In addition, the method required fewer calibration parameters than the wavelet image-processing approach. The study also foresaw the possibility to train an AI which estimates the real sizes of the grains, without knowing the scale of one pixel (mm/pixel ratio) if the training is done properly. The AI might learn unknown relationships between the texture and sizes if it is provided with a wide variety (images of several sediment classes) and scale (mm/pixel ratio)) of dataset (however, it is also prone to learn unwanted biases). Recently, Takechi et al. (2021) further elaborated on the importance of shadow- detection and removal, using a dataset of 500 pictures for training a texture-based AI, with the help of an object-detecting image-processing technique (Basegrain; Detert and Weitbrecht, 2013). The previously presented studies, applying ML and DL techniques, significantly contributed to the development and improvement of surrogate sampling methods, incorporating the great potential in AI. However, there are still several shortcomings to these procedures. Firstly, none of the image-based AI studies used underwater recordings, even though the underwater environment offers completely different challenges. Secondly, the training images were always scaled, i.e., the sizes of the grains could be easily reconstructed, which is again complicated to accomplish in a river. Lastly, they were not adapted for continuous (i.e., spatially dense) measurement, but rather focused on a sparse grid-like approach.

The goal of this manuscript is to further investigate the applicability of image processing as a surrogate method, and attempt to break through the above mentioned shortcomings of the AI-based approaches. Hence, we introduce a riverbed material analysing, Deep Learning technique and field measurement methodology, along with our first set of results. The introduced technique aims to eventually become a tool for exploratory mapping of the riverbed, by detecting sedimentation features (e.g., deposition zones of fine sediment, bed armour) and helping decision making for river sedimentation management. Also, the long-term hypothesis of the authors includes the creation of an image-based measurement methodology, where underwater videos of the riverbed could serve multiple sediment related purposes simultaneously. Part of which is the current approach for mapping the riverbed material texture and composition. Others include measuring the surface roughness of the bed (Ermilov et al., 2020) and detecting bedload movement (Ermilov et al., 2022).

Compared to the earlier studies introduced earlier, the main novelty of our manuscript is that both the training and analysed videos are recorded underwater, continuously along cross-sections of a large river. Furthermore, the training is unscaled, i.e., the camera-riverbed distance could vary while recording the videos, without considering image-scale. Moreover, compared to the relatively low number of training images in most of the above referred studies, we used a very large dataset (~15000) of sediment images for the texture-based AI, containing mostly sand, gravel, cobble, and to a smaller extent: bedrock together with some other, non-sediment related objects.

In this current manuscript, first, a literature review is given to better understand the current state of surrogate bedmaterial sampling approaches and their research, gradually leading up to the method of this manuscript and highlighting its relevance. In the third chapter the case studies and the methodology are introduced in detail. The third chapter presents the results and their evaluation, followed by a discussion about the challenges, the novelty and possible continuations of the method. A brief discussion is also given on how the method can support traditional methods and what kind of additional hydromorphological parameters can be provided by such videos, uniquely improving the toolkits of sedimentation engineering. Finally, the main conclusions of the manuscript are drawn.

3 Methods

3.1 Case studies

The results presented in this study are based on riverbed videos taken during three measurement campaigns, in sections of the Danube River, Hungary. The first one was at Site A, Ercsi settlement (~ 1606 rkm), the second one was at Site B, Gönyű settlement (~ 1791 rkm), and the third was at Site C, near to Göd settlement (~ 1667 rkm) (Fig. 1).

Figure 1: The location of the riverbed videos, where the underwater recordings took place. All sites were located in Hungary, Central Europe. The surveys were carried out on the Danube River, Hungary's largest river.

The training of the AI was done using the video images of Site C and a portion of A (test set; see later in Chapter 3.3), while Site B and the rest of the images from A served for validation. The measurements were carried out during daytime, at mid-water regime (Q = 1900 m^3/s) in case of Site A, and low water regime (Q = 1350 m^3/s) at Site B, similarly to Site C (Q = 700 m^3/s). This latter site served only for increasing the training image dataset (i.e., conventional samplings were not carried out at the time of recording the videos), thus we do not go into further details with it for the rest of the manuscript, but the main characteristics are listed in Table 1.

	Site A	Site B	Site C
Q_{survey} [m^3/s]	1900	1350	700
B_{survey} [m]	300 – 450		
$H_{mean, survey}$ [m]	3.5 - 4.5		
S_{survey} [cm/km]	15		
SSC_{survey} [mg/l]	25	20	14
Characteristic riverbed sediment	gravel, sandy gravel	gravel, gravelly sand	gravel, sandy gravel
$Q_{annual,mean}$ [m^3/s]	2000	2200	1400
$Q_{1\%}$ [m^3/s]	5300	5500	4700

Table 1: Main hydromorphological parameters of the measurement sites. Q_{survey}: discharge during survey; B_{survey}: river width during survey; $H_{mean,survey}$: mean water depth during the survey; S_{survey}: riverbed slope during survey; SSC_{survey}: mean suspended sediment concentration during the survey; $Q_{annual, mean.}$: annual-mean of the discharge at the site; $Q_{1\%}$: discharge of 1% probability.

As underwater visibility conditions are influenced by the suspended sediment, the characteristics of this sediment transport is also included in Table 1 (SSC_{survey} – susp. sed. concentration). The highest water depths were around 6-7 m in all cases. In Site A, measurements included mapping of the riverbed with a camera along three separate transects (Fig. 2). At Site B, two transects were recorded.

Figure 2: Bathymetry of Site A and B. The measurement cross-sections are also marked. The vessel moved along these lines from one bank to the other, while carrying out ADCP measurement and recording riverbed videos. Physical bed material samples were also collected in certain points of these sections.

3.2 Field data collection

Fig. 3 presents a sketch of the measurement process with the equipment and a close-up of the underwater instrumentation. During the field measurements, the camera was attached to a streamlined weight (originally used as an isokinetic suspended sediment sampler) and lowered into the water from the vessel by an electric reel. The camera was positioned perpendicularly to the water and the riverbed, in front of the nose of the weight. Next to the camera, two diving lights worked as underwater light sources, focusing into the camera's field of view (FoV). In addition, four laser pointers were also equipped in hand-made isolation cases to provide possible scales for secondary measurements. They were also perpendicular to the bottom, projecting their points onto the underwater camera field of view. Their purpose was to ensure a visible scale (mm/pixel ratio) in the video footages for validation. During the measurement procedure, a vessel crossed the river slowly through river transects, while the position of the above detailed equipment was constantly adjusted by the reel. Simultaneously, ADCP and RTK GPS measurement were carried out by the same vessel, providing water depth, riverbed geometry, flow velocity, ship velocity and position data. Based on this information and by constantly checking the camera's live footage on deck, the camera was lowered or lifted to keep the bed in camera sight and avoid colliding with it. The sufficient camera – riverbed distance depended on the suspended sediment concentration near the bed and the used illumination. The reel was equipped with a register, with its zero adjusted to the water surface. This register was showing the length of cable already released under the water, effectively the rough distance between the water surface and the camera (i.e., the end of the cable). Of course, due to the drag force this distance was not vertical, but this value could be continuously compared to the water depth measured by the ADCP. Differencing these two values, an approximation for the camera – riverbed distance was given all time. The sufficient difference could be established by monitoring the camera footage while lowering the device towards the bed. This value was then to be maintained with smaller corrections during the survey of the given cross-section, always supported by observing the camera recording, and adjusting to environmental changes. The vessel's speed was also adjusted based on the video and slowed down if the video was blurry or the camera got too far away from the bed (see later in Chapter 4.3). The measurements required three personnel to i) drive the vessel, ii) handle the reel, adjust the equipment position, and monitor the camera footage, iii) monitor the ADCP data, while communicating with the other personnel (see Fig. 3).

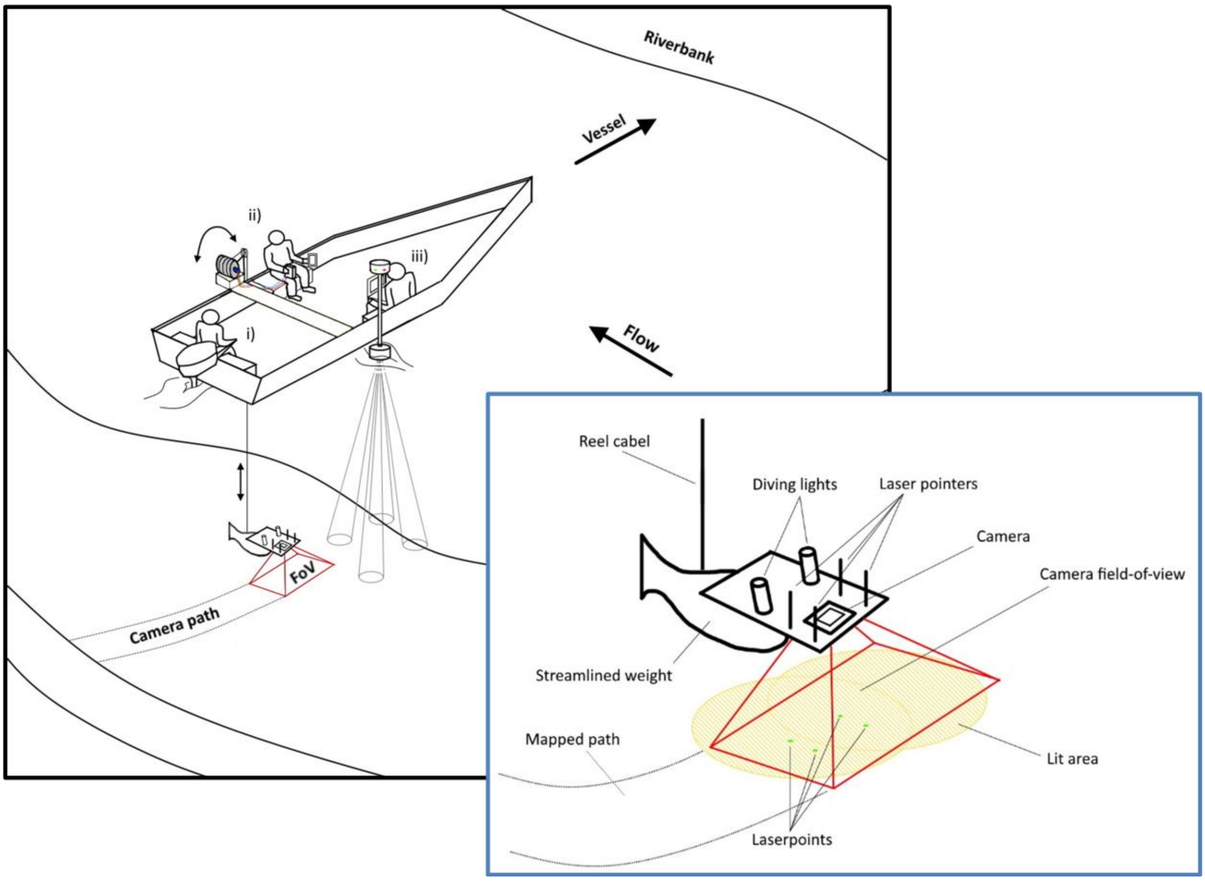

Figure 3: Left: sketch of the measurement process. The vessel was moving perpendicular to the riverbank along a cross-section (i). A reel was used to lower a camera close to the riverbed (ii). Simultaneously, the bed topography and water depth were measured by an ADCP (iii). Right: Close-up sketch of the underwater instrumentation.

The video recordings were made with a GOPRO Hero 7 and a Hero 4 commercial action cameras. Image resolutions were set to 2704x2028 (2.7K) with 60 frame per second (fps) and 1920x1080 (1080p) with 48 fps, respectively. Other parameters were left at their default (see GOPRO 2014; 2018), resulting in slightly different quality of produced images between the two cameras. We found that a 0.2-0.45 m/s vessel speed with 60 fps recording frequency was ideal to retrieve satisfactory images in a range of 0.4-1.6 m camera-bed distances. This meant approximately 15 minutes long measurements per transects. Further attention needed to be paid to the reel and its cable during the crossing when the equipment was on the upstream side of the boat. If the flow velocities are relatively high (compared to the total submerged weight of the underwater equipment), the cable can be pressed against the vessel-body due to the force from the flow itself, causing the reel cable to jump to the side and leave its guide. This results in the equipment falling to the riverbed and the measurement must be stopped to reinstall the cable. For illumination, a diving light with 1500 lumen brightness and 75° beam divergence, and one with 1800 lumen and 8° were used. The four lasers for scaling had 450-520 nm (purple and green) wavelength and 1-5 mW nominal power. Power supply was ensured with batteries for all instruments.

At Site A and Site B, conventional bed material (physical) samplings were also carried out by a grabbing (bucket) sampler along the analysed transects. At each cross-section 4-5 samples were taken, with one exception where we had 10. The measured GSDs were used to validate results of the AI algorithm. Separately, a visual evaluation of the videos was also carried out, where a person divided the transects into subsections based on their dominant sediment classes, after watching the footages.

3.3 Image analysis: Artificial Intelligence and the wavelet method

In this study, we built on the former experiences of the authors, using Benkő et al., 2020 as a proof-of-concept, where the developed algorithm was applied for analysing drone videos of a dry riverbed. The same architecture was used in this manuscript, which is based on the widely used Google's DeeplabV3+ Mobilnet, in which many novel and state-of-the-art solutions are implemented (e.g., Atrous Spatial Pyramid Pooling; Chen et al., 2018). The model was implemented with Pytorch, exploiting its handy API and backward compatibility. The main goal was to build a deep neural network model which can recognise and categorise (via semantic segmentation; Chen et al., 2018) at least three main sediment size classes, i.e., sand, gravel and cobble, in the images, while being quickly deployable. The benefit of the introduced method compared to conventional imagery methods lies in the potential of automation and increased speed. If the annotation and training is carried out thoroughly, analysing further videos can run effortlessly, while the computation time can be scaled down either vertically (using stronger GPUs) or horizontally (increasing the number of GPUs; if parallel analysis of images is desired). In this study a TESLA K80 24GB GDDR5 348bit GPU, an Intel Skylake Intel® Xeon® Gold 6144 Processor (24.75M Cache, 3.50 GHz) CPU with 13GB RAM was used. Also, contrary to other novel image-processing approaches in riverine sediment research (Buscombe, 2013; Detert and Weitbrecht, 2013), the deep convolutional neural network is much less limited by image resolution and mm/pixel ratios, because it does not rely on precise pixel count. This is an important advantage to be exploited here, as we perform non-scaled training and measurements with the AI, i.e., camera-bed distance constantly changed, and size-reference was not used in the images by the AI.

Fig. 4 presents the flowchart of our AI-based image processing methodology. The first step after capturing the videos was to cut them into frames, during which the videos were exploded into sequential images. Brightening and sharpening filters were applied on the images to improve their quality. Next, the ones with clearest outlines and best visibility were chosen. This selection process was necessary because this way the delineation process (learning the prominent characteristics of each class) can be executed accurately, without the presence of misleading or confusing images, e.g., blurry or dark pictures where the features are hard to recognise. For training purposes, we chose three footages from different sections each being ~15 minutes long with 60 fps and 48 fps, resulting in 129 600 frames. In fact, no such large dataset was needed due to the strong similarity of the consecutive

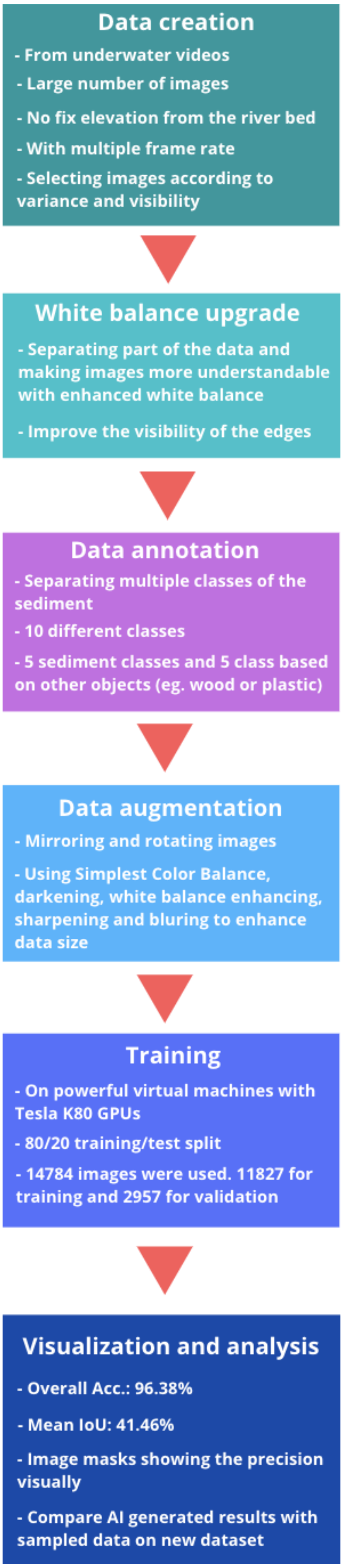

Figure 4: Flowchart of the applied methodology.

frames. The number of images to be annotated and augmented were therefore decreased to ~2000. We also performed a white balance correction on some of the images to improve visibility, making it even easier to later define the sediment class boundaries. We used an additional algorithm to generate more data, with the so-called Simplest Colour Balance method (Limare et al., 2011). It is a simple, but powerful histogram equalisation algorithm which helps to equalise the roughness in pixel distribution.

These steps were followed by the annotation, where we distinguished ten classes. Annotation was carried out by a trained personnel, not by the authors, and performed with the help of an open-source software called PixelAnnotationTool (Breheret, 2017), which enables the user to colour mask large parts of an image based on colour change derivatives (i.e., colour masking part of the images which belong to the same class, e.g., purple/red – sand, green – gravel, yellow – cobble, etc.). The masks and outlines were drawn manually, together with the so-called watershed annotation. That is, when a line was drawn, the algorithm checked for similar pixels in the vicinity and automatically annotated them with the same class. The annotation was followed by a data augmentation step where beside mirroring, cropping, rotating the images (to decrease the chance of overfitting), we also convolved them with different filters. These filters added normally distributed noise to the photos to influence the watershed algorithm and applied sharpening, blurring, darkening, and white balance enhancement. Thus, at the data level, we tried to ensure that any changes in water purity, light, and transparency, as well as colour changes, were adequately represented during training. Images were uniformly converted to 960x540 resolution, scaling them down to make them more usable to fit in the GPU's memory. The next step was to convert all the images from RGB (Red-Green-Blue) based colour to grayscale. This is important because colour images have 3-channels, so that they contain a red, a green, and a blue layer, while grayscale images' pixel can only take one value between 0 and 255. With this colour conversion we obtained a threefold increase in computational speed. In total, a dataset of 14,784 images was prepared (from the ~2000 images of the4 training videos). The next step was to separate this into training and validation sets. In this study, 80% of the data was used for training the Artificial Intelligence, while 20% was to validate the training. It was important to mix the images so that the algorithm selects batches in a pseudorandom manner during training, thus preventing the model from being overfitted.

Finally, after several changes in the hyperparameters (i.e., tuning), the evaluation and visualisation of the training results were performed. Tuning is a general task to do when building Deep Learning Networks, as these hyperparameters determine the structure of the network and the training process itself. Learning rate, for example, describes how fast the network refreshes, updates itself during the training. If this parameter is set too high, the training process finishes quickly, but convergence may not be reached. If it is too low, the process is going to be slow, but it converges. For this reason, nowadays the learning rate decay technique is used, where one starts out with a large learning rate, then slowly reduces it. The technique generally improves optimization and generalization of the Deep Learning Networks (You et al., 2019). In our case, learning rate was initialised to 0.01, with 30000 iteration steps, and the learning rate was reset after every 5000 iterations with a decay of 0.1. Another important parameter was the batch size, which sets the number of samples fed to the network before it updates itself. Theoretical and empirical evidence suggest that learning rate and batch size are highly important for the generalization ability of a network (He et al., 2019). In our study, a batch size of 16 was used (other general values in the literature are 32, 64, 128, 256). We used a cross-entropy loss function.

As previously mentioned, the training of the AI was managed without scaling, without the need for equipped lasers. However, we intended to use the laser pointers to provide a spatial scale for the recorded videos, as a secondary validation. As the lasers were not functioning as we originally hoped, we could not use them constantly during the cross-sectional surveys and could not aim for transactional scaling and validation this way. Instead, we diverted to validation in the points of the physical samplings as we could use the lasers in a few, selected points only. We used a textural image-processing method to analyse the video images of these sampling spots. For this, the already mentioned, transferable wavelet-based signal- and image-processing method was chosen. The method enables to calculate the image-based grain size distribution of the selected pictures. The grey-scale intensity is analysed through pixel-rows and -columns of the image and handled as individual signals. Then, instead of Fourier-transform, the less-constrained wavelet-transform is applied to decompose them. Finally, calculating the power spectra and the sizes (from pixel to millimetre, using the scale) of the wavelet components (each wavelet describes an individual grain) produces the grain size distribution for the given image. Beforehand, this method was proved to be the most efficient, non-AI image-processing method for mixed sediments (Buscombe, 2013; 2020) and was already tested for underwater circumstances in an earlier study by the authors of present manuscript (Ermilov et al., 2020).

4 Results and discussion

4.1 Evaluation of the training

To evaluate the training process, the image series used for the training was analysed by the developed Deep Learning algorithm. Fig. 5 shows results of original images (from the validation set), their ground truth (annotation by the training personnel), as well as the AI prediction (result of the model). The overlays of the original and the predicted images are also shown for better visualization. Calculating the over-all pixel accuracy (i.e., the percent of pixels that were correctly classified) returned a satisfactory result with an average 96% match. As this parameter in object detection and Deep Learning is not a stand-alone parameter (i.e., it can still be high even if the model performs poorly), the mean IoU (intersection-over-union or Jaccard index) was also assessed, indicating the overlap of ground truth area and prediction area, divided by their union (Rahman and Wang, 2016). This parameter showed a much slighter agreement of 41.46%. Interestingly, there were cases, where the model gave better result than the annotating personnel. While this highlighted the importance of thorough and precise annotation work, it also showcased that the number of poor annotations was relatively low, so that the algorithm could still carry out correct learning process and later detections, while not being severely affected by the mistake of the training personnel. Fig. 6 showcases an example for this: the correct appearance of cobble (yellow) in the prediction, even though the user (ground truth) did not define it during the training. As a matter of fact, these positive errors also decrease the IoU evaluation parameter, even though they increase the performance of the AI on the long term. Hence, this shows that pure mathematical evaluation may not describe the model performance entirely. Considering that others also reported similar experience with Deep Learning (Lu et al., 2018) and the fact that 40% and 50% are generally accepted IoU threshold values (Yang et al., 2018; Cheng et al., 2018; Padilla et al., 2020), we considered the 41.46% acceptable. The general quality of our underwater images may have also played a role in lowering the IoU result.

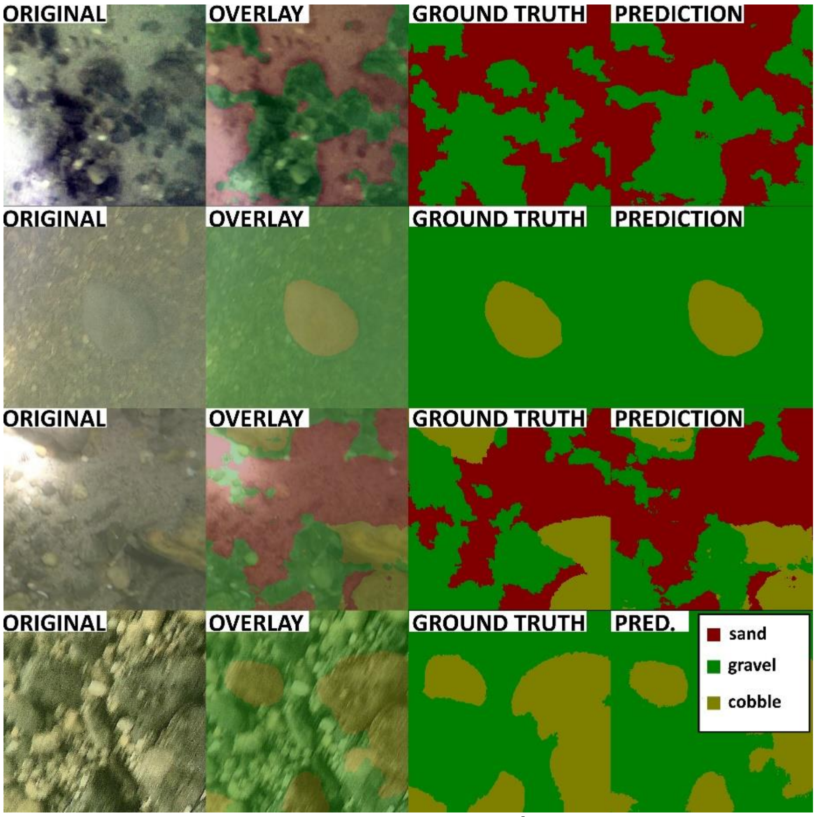

Figure 5: Example comparisons of ground truth (taught pattern, 3rd column) and AI predicted (learnt pattern, 4th column) sediment classes from the training videos showing satisfactory results. The 1st column shows raw images, while the 2nd column overlays the result of the AI detection on the raw image for better visual context.

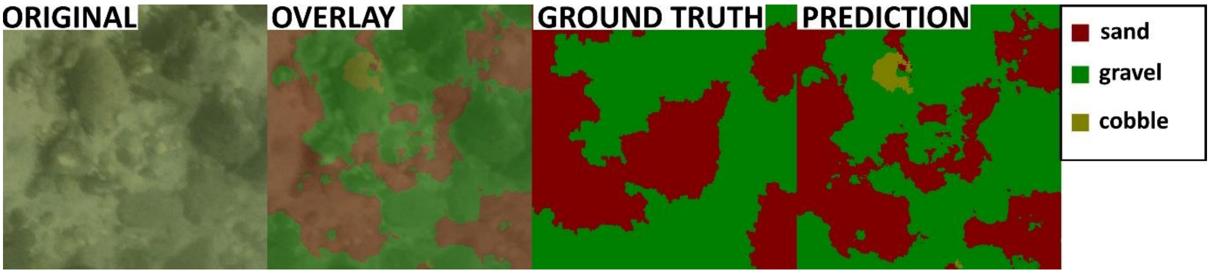

Figure 6: Example comparison of ground truth (taught) and AI predicted (learnt) sediment classes from the training, in case of training personnel mistake during the annotation.

4.2 Intercomparison of methods

In each masked image, the occurring percentage of the given class (i.e., the percentage of the pixels belonging to that class/colour mask, compared to the total number of pixels in the image) was calculated and used as the fraction percentage in that given sampling point. These sediment classes reconstructed by the AI were then compared to three alternative results: i) visual estimation, ii) GSD resulted from conventional grab sampling, iii) wavelet-based image-processing. In the followings, results from two cross-sections will be shown, one from Site A, the video used for the training, and one from Site B, being new for the AI. An averaging window of 15 m was applied on each cross-sectional AI result to smoothen and despike the dataset.

In Fig. 7, the path of the vessel can be seen in Section A - II, at Site A. The path was coloured based on the visual evaluation of the riverbed images. The different colours represent the dominant sediment type seen at the given point of the bed. The locations of the physical bed material samplings are also shown (see yellow markers). Fig. 8 presents the raw (i.e., before moving-average) results of the AI detection of each analysed image along Section A - II. Currently, our approach is sensitive and large spikes, differences can occur in the AI detection between consecutive, slightly displaced video frames. Due to this, and the fact that there is uncertainty in the coordinates of the underwater photos and their corresponding physical samples, it is not recommended to carry out comparisons by selecting certain image and its AI detection. Instead, we applied a moving average-based smoothing for each raw, cross-sectional AI detection, with a window-size corresponding to 15 m. These moving-averages are the ones being compared later in the sampling points to the physical sampling and the wavelet method. For illustration purposes, we provided the raw AI detections of all the sampling point images, even though their result may not be representative of their corresponding moving-average values. Fig. 9 shows the cross-sectional visual classification compared to the AI-detected sediment fractions in percentage after applying moving-average (i.e., the smoothed version of Fig. 8). The noises are mostly caused by sudden changes in lighting conditions. It happens either from losing visual on the riverbed momentarily due to sudden topography changes or from increased suspended sediment concentration.

Figure 7: The path of the vessel and camera in Section A - II, Site A. The polyline is coloured based on the sediment features seen during visual evaluation of the video. Yellow markers are the locations of physical bed material samplings. (Map created with Google Earth Pro)

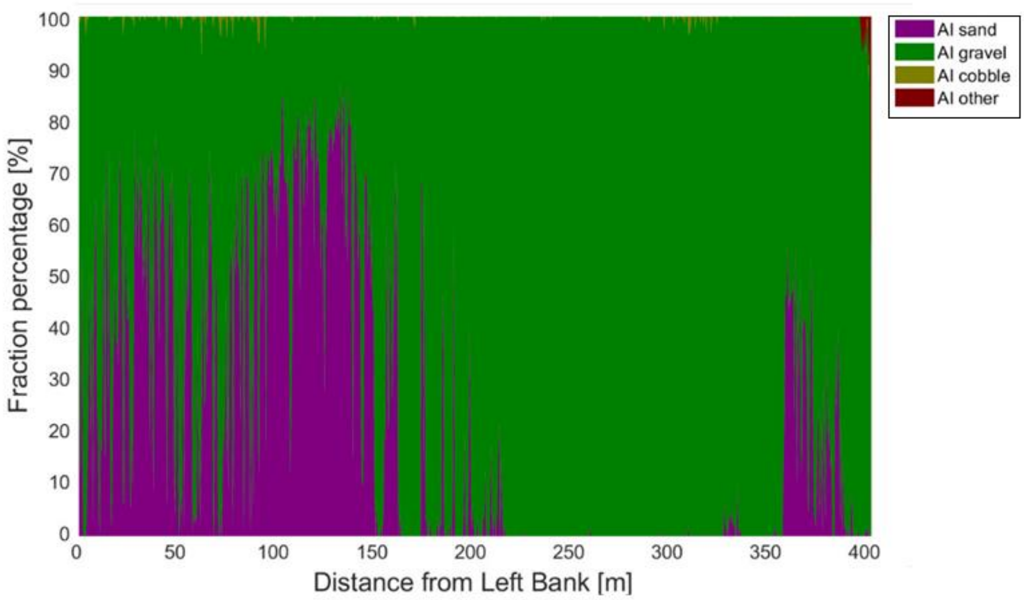

Figure 8: The sediment fraction percentage results of every image, analysed by the AI along Section A - II. While the trends are apparent, the sensitivity of the method at its current state can be observed. AI result before applying moving-averaging.

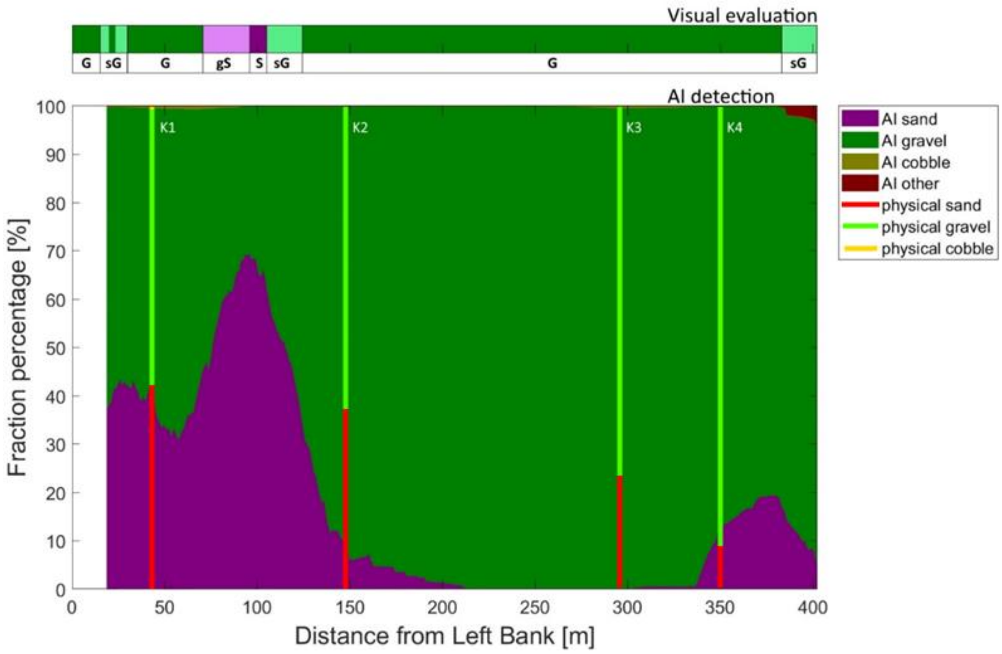

Figure 9: Section A - II. The visual evaluation of the dominant sediment features in the video (top) compared to sediment fraction percentage, recognised by the AI (bottom). AI result after applying moving-averaging. The visual evaluation included four classes: gravel – G, sandy gravel – sG, gravelly sand – gS, sand – S,). The fractions of the physical samples are shown as verticals.

The AI result shows satisfactory match with the human evaluation. For example, around 100 m from the left bank, between K1 and K2 sampling points, the AI peaks with around 70% sand an 30% gravel correctly. Furthermore, on the two side of this peak a steep transition to gravel and decreasing sand occurs, similarly to the eye observation, marked as sandy gravel and gravelly sand. Mixed sediment zones were also correctly identified by the AI at both riverbanks.

Next, the AI estimated sediment classes were compared with both the physical samples and the wavelet method at each sampling locations (Fig. 7). Overall, the AI based classification agrees well with the physical samples, however, at sample K3 the ~20% sand content was neither reconstructed by the AI, nor could be observed by eye in the images. Considering that the gravel dominates the bed sediments, the absence of sand fraction in the AI results might be explained with the difference between the surface GSD and subsurface GSD. While both the AI and the eye observation-based assessment focus on the bed surface, the physical sampling represents a thicker layer, including the subsurface layer, too. Indeed, the so-called bed armouring phenomenon, taking place in the vicinity of the thalweg in mixed-bed rivers, leads to coarser surface grains and finer subsurface grains (see e.g., Wilcock, 2005). This may also explain the case of K2 as it was located closer to the thalweg.

Fig. 10 presents an image of the collected physical sample in K3 together with its sieving result as well as the underwater image of the riverbed surface in K3, and the results of the two different image processing methods. Bed armouring is indeed present in the sampling point as the finer, sand fraction cannot be observed on the riverbed surface, yet they appear in the collected bed material sample. In Fig. 11., supportive images of bed armouring are provided, taken during our surveys in the Upper section of the Hungarian Danube. We broke the surface armour to showcase the presence of the underlying finer fractions.

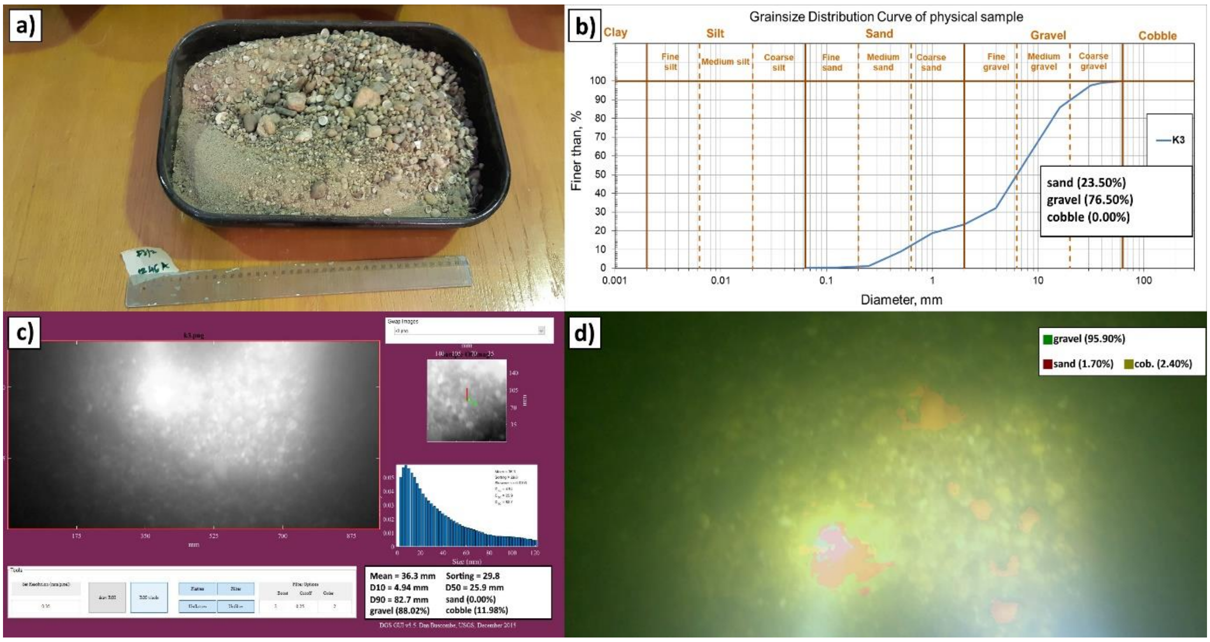

Figure 10: Bed armour in sampling point K3. The measured percentages of fractions are also presented, respectively. a) Image of the collected physical sample, containing both sand and gravel fractions. b) GSD curve of the physical sample, produced with sieving analysis. c) Wavelet analysis result of the image, taken in the sampling point. d) AI detection result in the sampling point.

Figure 11: Images of bed armouring, taken during our surveys in the Upper section of the Hungarian Danube.

As for the wavelet analysis-based imagery technique, an overall slight overestimation of the coarse particles can be observed, and the sand classes are, in fact, not reconstructed correctly. This finding agrees well with the field experiences of Ermilov et al. (2020), where the authors indicated the strong sensitivity of the wavelet technique on the image resolution, and showed that to detect a grain, the diameter must be at least three times larger than the pixel. In this survey, the camera was indeed closer to the riverbed at sampling points K1 and K4 and the wavelet algorithm was able to detect coarse sand, but finer sand was neglected yielding the lower percentages, seen in Fig. 12.

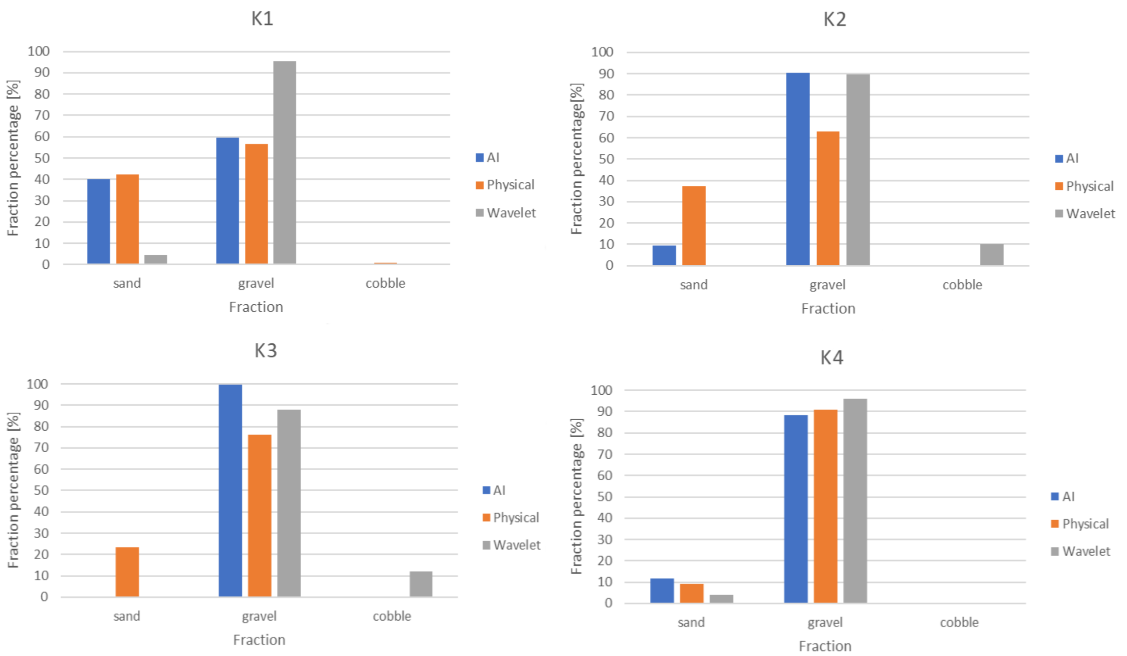

Figure 12: Comparison of sediment fraction % at the sampling locations from the moving-averaged AI detection, conventional sieving and the wavelet-based image processing method. Section A - II.

The images of the bed from the sampling points are shown in Fig. 13, while Fig. 14 presents the AI detection overlayed on these specific images, for visualisation purposes. As it was seen through the example of Fig. 8, a moving-average has to be used for treating occurring noises. Hence, the AI results plotted in the bar- and smoothed sectional plots are not necessarily representative of these instantaneous snapshots, such as Fig. 14.

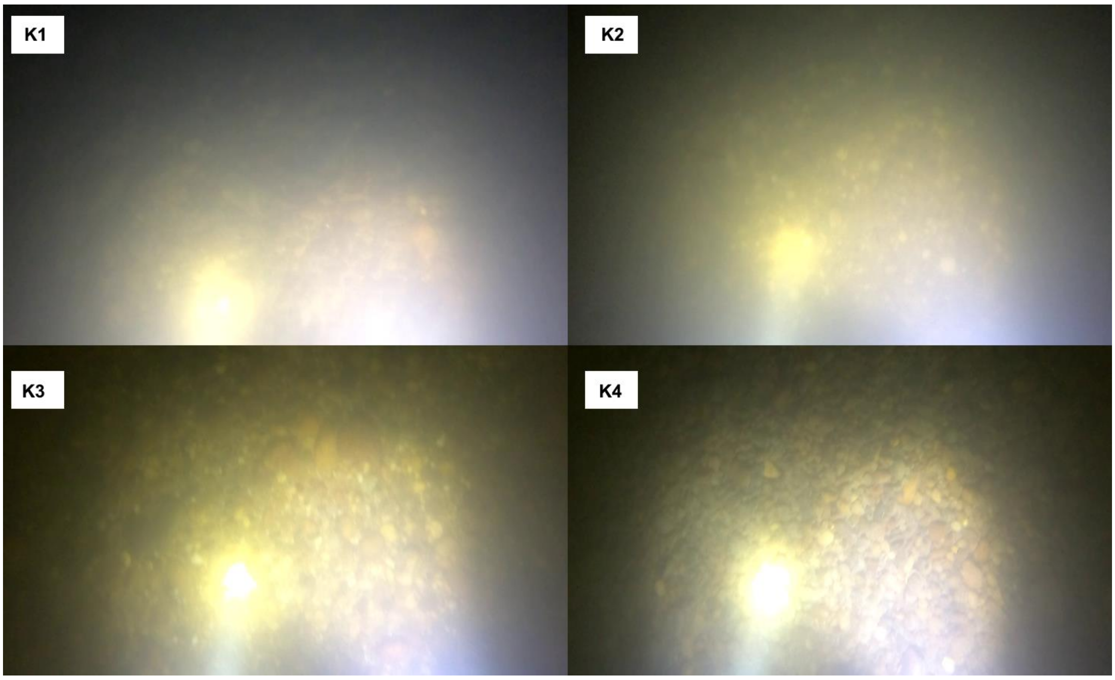

Figure 13: Riverbed video images at the sampling points in Section A - II.

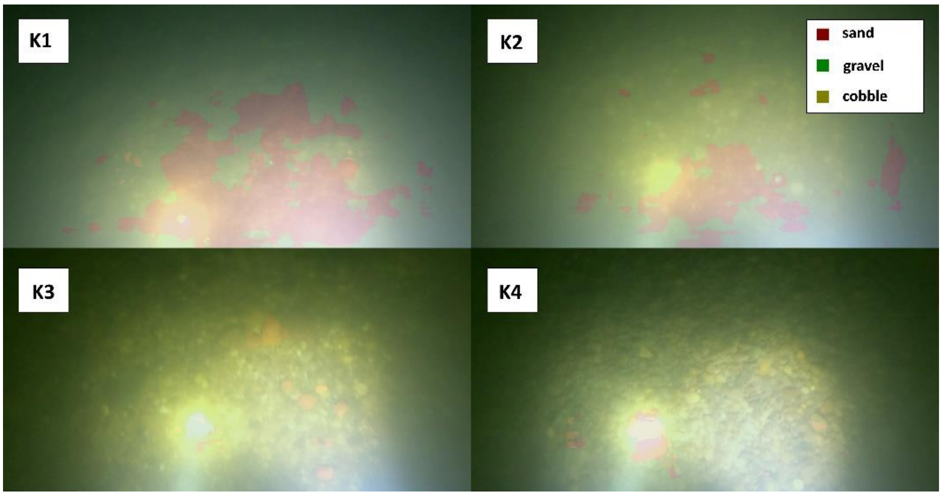

Figure 14: Riverbed video images overlapped with their raw, AI detection result, at the sampling points in Section A - II.

At site B (Fig. 15) the river morphology is more complex compared to Site A as a groyne field is located along the left bank (see again Fig. 2). As such, the low flow regions between the groynes yields the deposition of fine sediments, and much coarser bed composition in the narrowed main stream. Even though no images containing only fine sediments were applied in the training phase of the AI algorithm, the model still managed to successfully distinguish these zones. The fine sediments in the deposition zone at the left bank was adequately estimated and showed a good match with the visual evaluation for the whole cross-section (see Fig. 16).

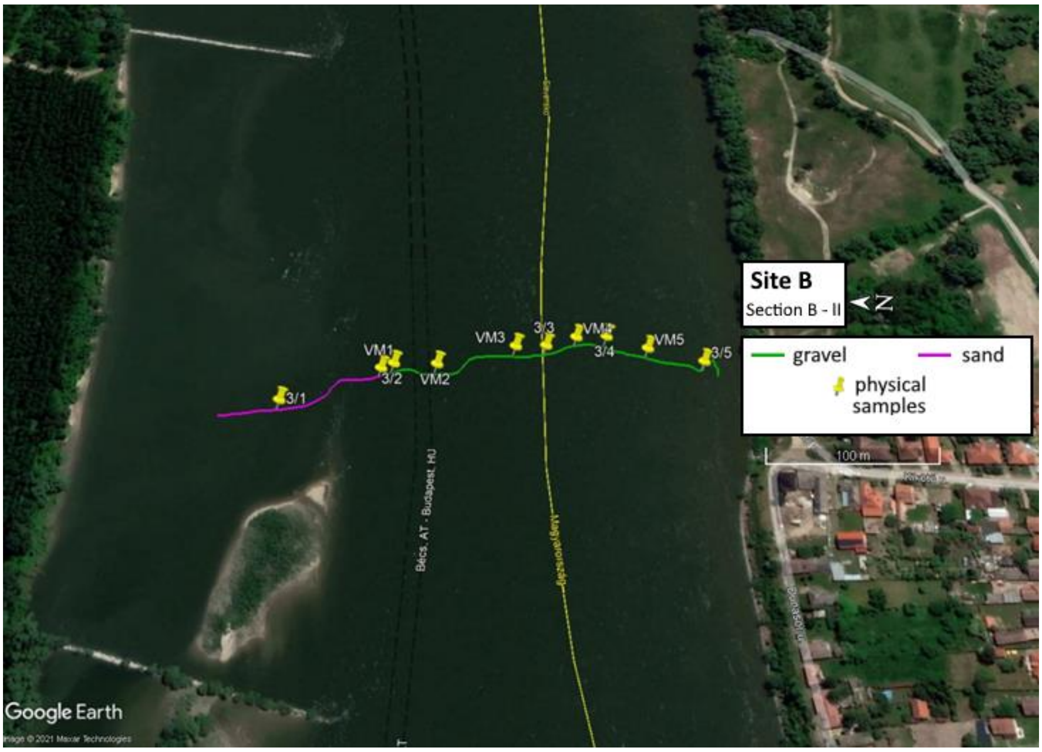

Figure 15: The path of the vessel and camera in Section B - II, Site B. The polyline is coloured based on the sediment seen during visual evaluation of the video. Yellow markers are the locations of physical bed material samplings. (Map created with Google Earth Pro)

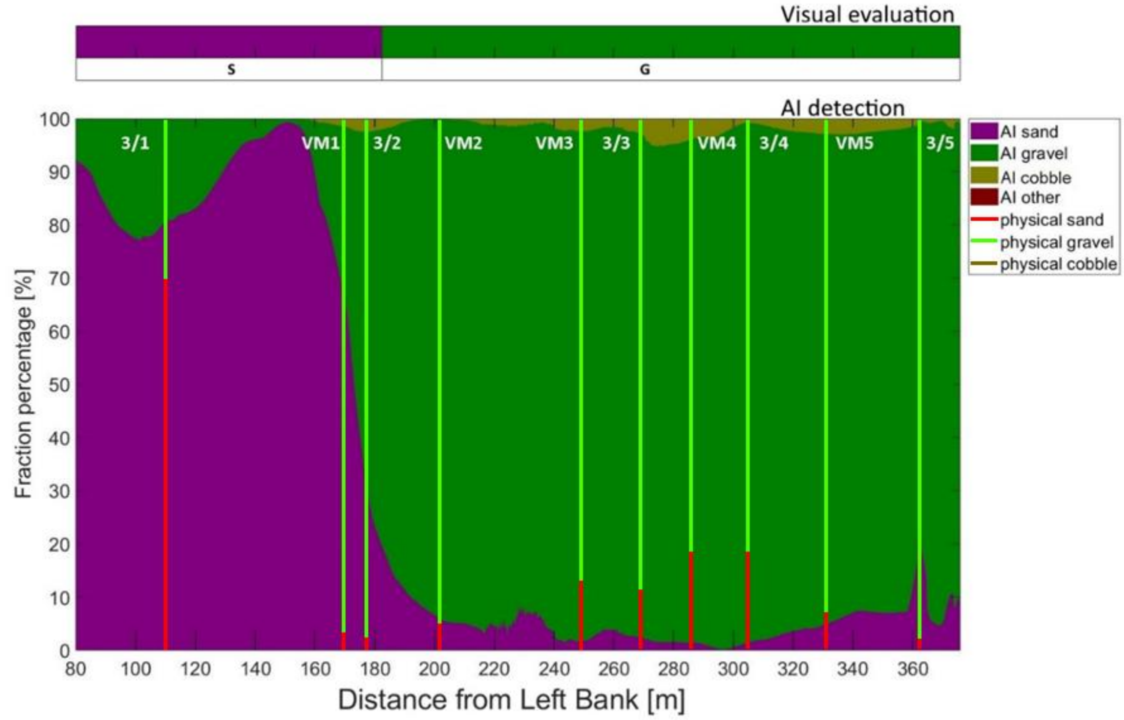

Figure 16: Sediment fraction percentages in Section B - II, recognised by the AI. The visual evaluation included two classes: gravel – G, sand – S). The fractions of the physical samples are shown as verticals.

When comparing the AI results with the physical samples, the match is acceptable for most of the samples, such as 3/1, VM2, VM3, 3/3, VM4, 3/4, VM5, respectively, with a highest difference of 10%. The significant disagreements arose at sampling points VM1 and 3/2. Indeed, these points are located around the border of the sediment deposition zone, showing steeply decreasing amount of sand moving away from the left bank (see the variation from point 3/1 towards 3/2 in Fig. 17). This trend is successfully calculated by the AI algorithm, but the physical samples for points VM1 and 3/2 show strong gravel dominance with negligible amount of sand (see Fig. 18). Resembling the findings at the other study site, this difference can likely be explained with the disturbed physical samples, which contain the sediments of the subsurface layer, too. In this case, however, the fine sediment layer is accumulated on the gravel particles. It is also worth noting that the physical samples are analysed by weighing the different sediment size classes, resulting in weight distribution. On the other hand, the imagery methods provide surface distributions, hence having a thin layer of fine sediments on the top can strongly bias the resulted composition (Bunte and Abt, 2001; Sime and Ferguson, 2003; Rubin et al., 2007). In Fig. 19/a, the grainsize frequency diagram (blue) showcases how the wavelet method incorrectly detected a gravel and a larger cobble mode and it did not manage to identify sand. Again, this was due to not achieving the sufficient image resolution required by the wavelet method. Fig. 19/b on the other hand presents the AI detection for the very same image, with satisfactory results. The algorithm managed to segment the gravels and the sand patches.

At sampling point 3/5 a weaker, but still satisfactory agreement was found. Here, the AI indicated 20% sand in contrast with the physical samples. Analysis of the raw videos may indicate that the suspended sediment concentration was higher in this region and the transported fine particles frequently became visible passing through the light beams, eventually causing disturbance in the AI analysis. Another issue in the AI algorithm was associated with the illumination. Using a diving light with small beam divergence proved counterproductive. The high intensity, focused light occasionally caused overexposed zones (white pixels) in the bed image, misleading the AI and resulting in detecting incorrect classes there. The use of wide beam divergence lamps is recommended instead, with uniform light. Examples for these problems are illustrated in Fig. 20 (a: overexposure, b: moving suspended sediment). Fig. 21 presents the AI detection overlayed on the images taken exactly in the sampling points.

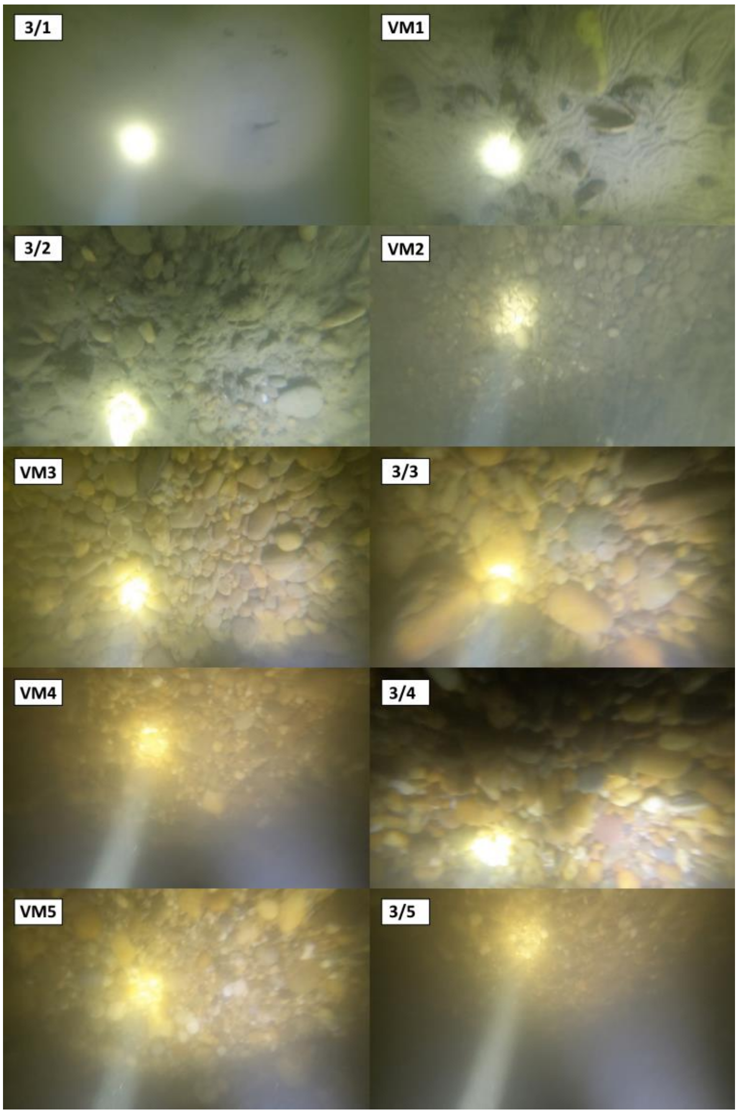

Figure 17: Riverbed video images at the sampling points in Section B - II.

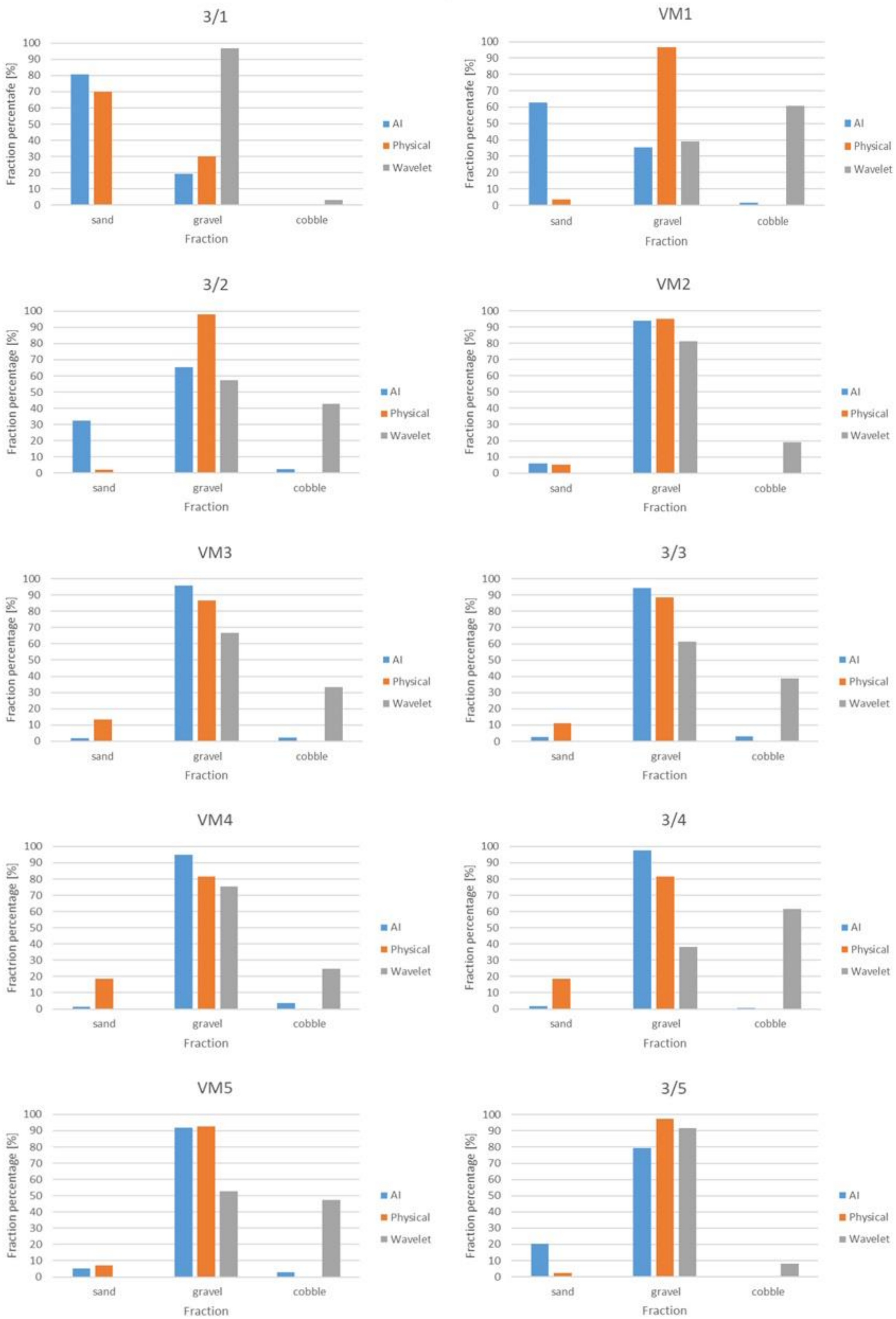

Figure 18: Comparison of sediment fraction % at the sampling locations from the moving-averaged AI detection, conventional sieving and the wavelet-based image processing method. Section B - II.

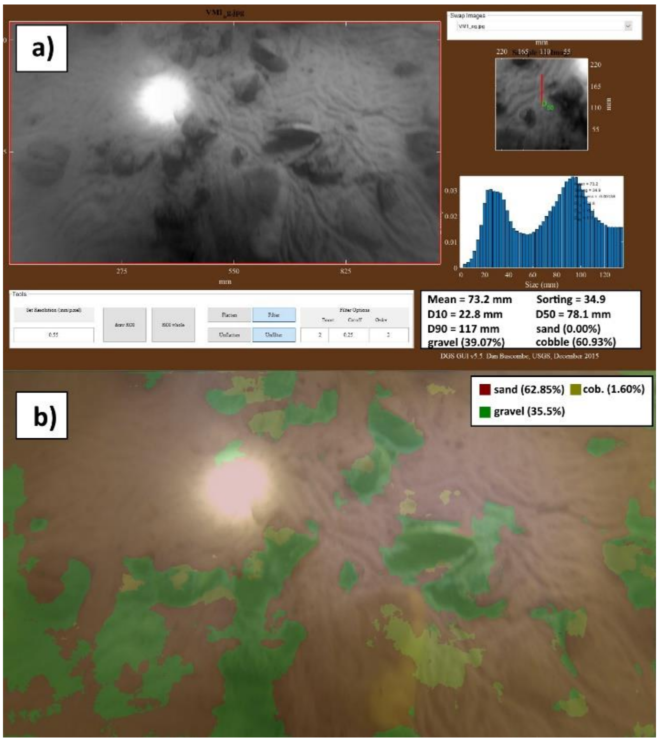

Figure 19: a) Wavelet analysis result of the underwater image in VM1. b) AI detection result of the same image.

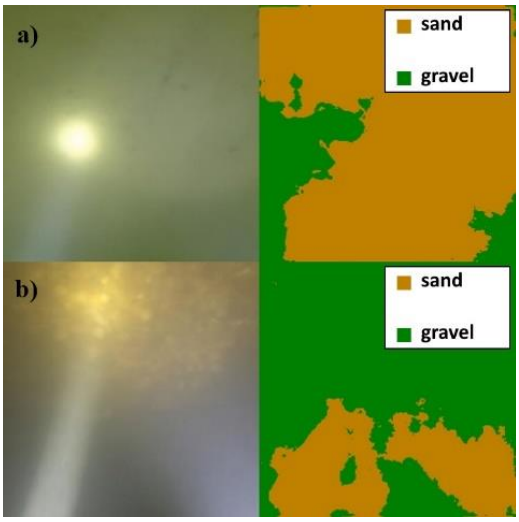

Figure 20: The effect of strong diving light on the AI algorithm in: a) purely sand covered zone and b) darker zone with higher SSC. The original images are on the left, while the AI detections can be found on the right.

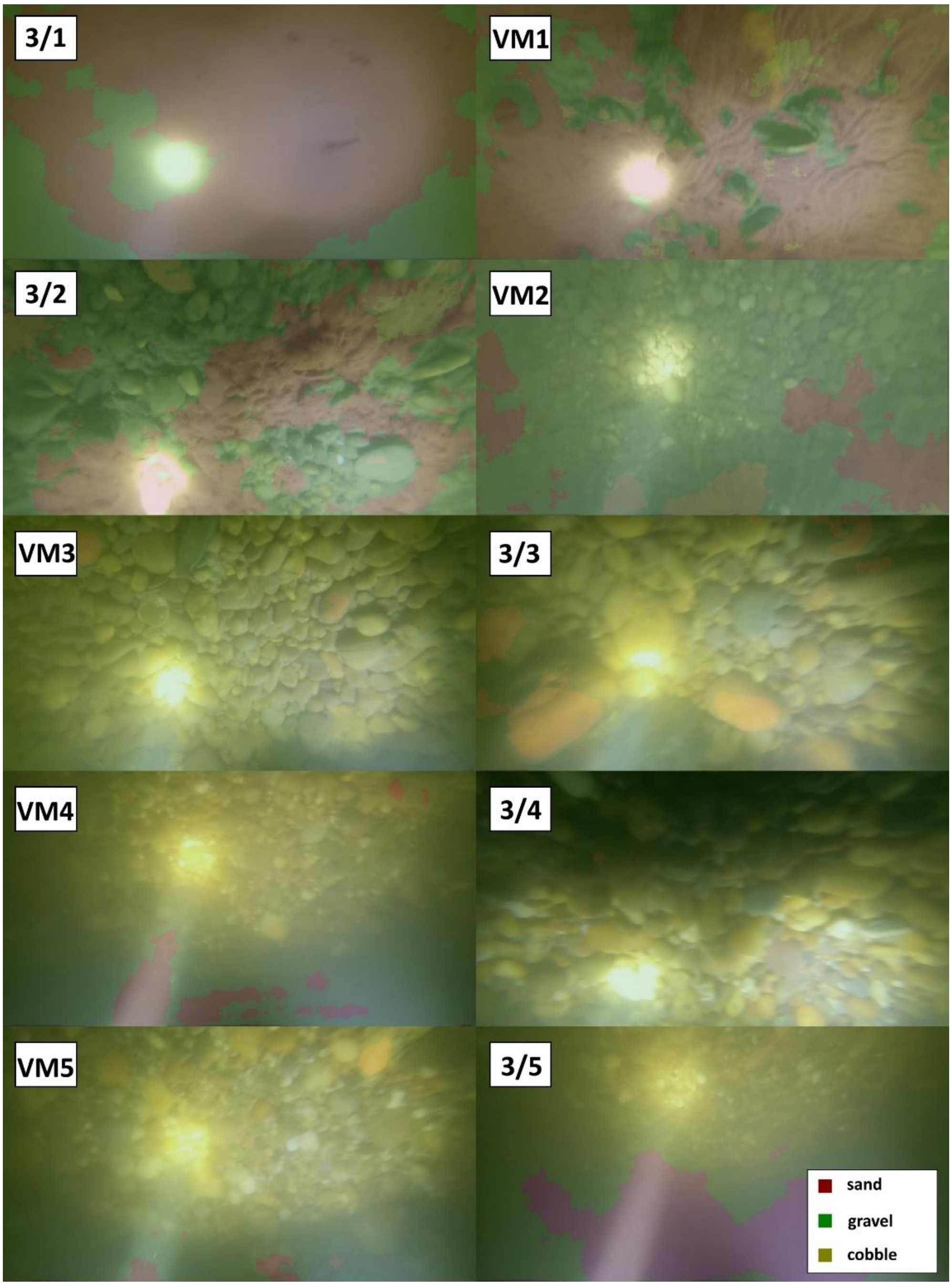

Figure 21: Riverbed video images overlapped with their raw, AI detection result, at the sampling points in Section B - II.

Results of the other measurements can be found in the Appendix. Fig. A2, B2 and C2 show that the trend of riverbed composition from the visual evaluation is well-captured by the AI in the other cross-sections of the study as well (see sampling points F/3, F/4, 1/1, 1/2 1/3, respectively). The resolution limit of the wavelet approach was further noticeable (see Fig. A3, B3 and C3) as it was not able to detect sand, similarly to the sampling points presented earlier. In Section A - III, traces of possible bed armouring were found as neither the eye observation nor the AI detected sand class in the images (see Fig. B4) of F/1 and F/2, even though the physical samples contained this fraction. In sampling point 1/4, the AI correctly detected the mixing of sand and gravel, but the physical sample showed the dominance of sand rather than the gravel fraction (see Fig. C3). The possible explanation behind these differences resembles what was the case for the K3 sampling point, introduced earlier: there was an additional finer fraction in the subsurface layer, hidden from the camera.

Finally, to quantify the efficiency of the introduced Deep Learning algorithm, we have selected the points which were determined to be the most comparable between the applied analysis methods. The decision was based on how well the given physical sample and riverbed image corresponded to each other. In the case of bad correspondence, the point was removed from the final comparison. These removed points belonged to one of the two scenarios that was discussed earlier: i) bed armouring, or ii) surface accumulation of a relatively thin sand layer, covering the gravel particles underneath. As a result, 11 points were kept and used for the final evaluation. These were: 1/1 from section B-I; 3/1, VM2, VM5, 3/5 from section B-II; F3, F4 from section A-I; K1, K2, K4 from section A-II; and A5 from section A-III. Fig. 22 presents the AI-measured sediment fraction percentages against the data from their corresponding, sieved physical samples (i.e., the bar plots values, respectively). The differences between the AI and the physical samples for most of the points (7 out of 11) were less or equal to 10%, while in the remaining 4 they did not exceed 20%.

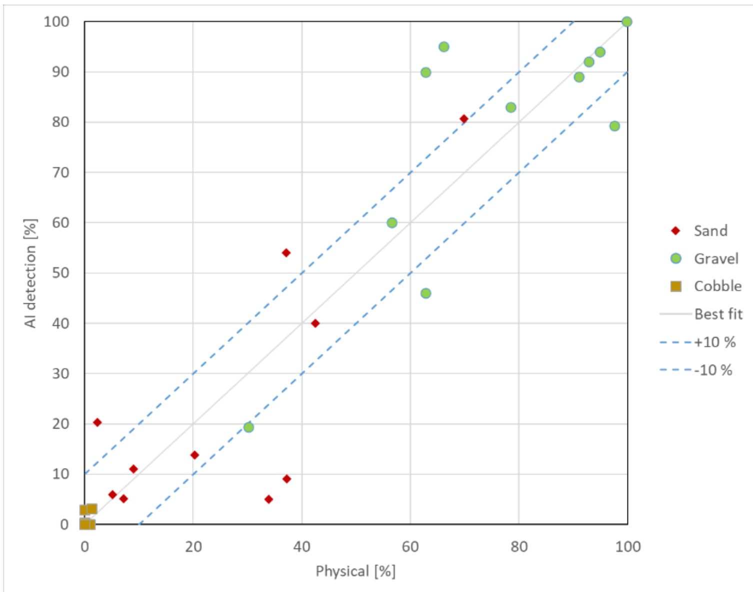

Figure 22: Comparison of sediment fractions between the AI detection and physical samples, for the selected sampling points.

Fig. 23 on the other hand compares the performance of the AI to the wavelet method. In this case, the discrepancies are greater, but this is due to the limitations of the wavelet approach, discussed earlier. For instance, the wavelet detected large amounts of cobbles in 4 points, while neither the AI, nor the physical samples (Fig. 22) did so. Furthermore, it was unable to recognise the sand fraction almost completely.

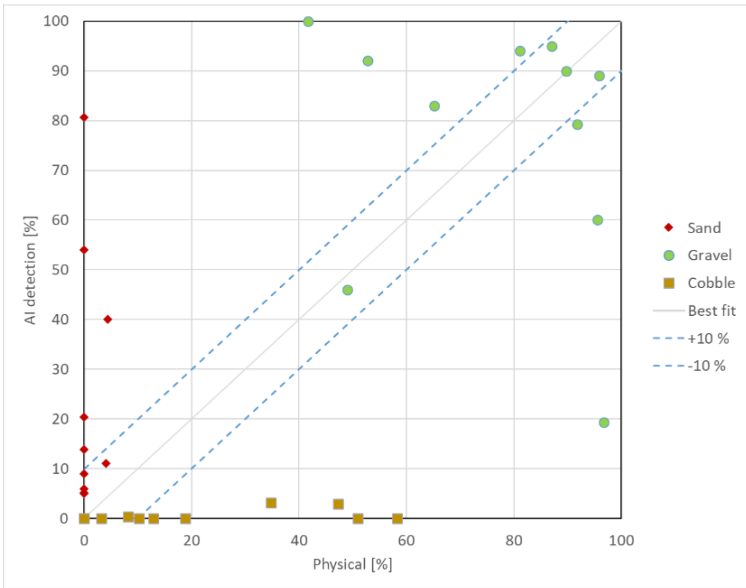

Figure 23: Comparison of sediment fractions between the AI detection and the wavelet approach, for the selected sampling points.

Based on the results presented in this manuscript, it could be established that the AI managed to recognise the main features of the riverbed material composition from underwater videos with satisfactory accuracy.

4.3 Implementation challenges

The power supply for the entire imaging infrastructure, i.e., for the camera, the diving lights and lasers, was ensured by batteries. However, due to the low temperature at the river bottom, the battery level decreased extremely fast, compared to normal circumstances. Providing the power supply directly from the motorboat engines can overcome this issue. To keep the camera in the adequate height also caused difficulties, since getting too close to the bed can harm the devices, lifting too high, on the other hand, will result in poor image quality. The measured instantaneous ADCP flow depth data was used therefore to keep the bed in camera sight, while maintaining proper boat velocity to avoid blurry images. Choosing a higher recording frequency, however, can be beneficial and alter this limitation, when provided. Lower velocities could not be maintained as the river would have moved the vessel out of the section. An alternative solution can be to move on longitudinal (streamline) paths instead of transects. This would allow for lower vessel speed. This would increase the time of the measurement, which still could be profitable if the images are of higher quality. However, the conventional way for river bathymetry surveys is to move on transversal, cross-sectional paths, due to the river bathymetry having a lower spatial variation along streamlines, compared to the changes that occur in the transversal direction (Benjankar et al., 2015; Kinsman, 2015). As such, it may require carrying out a relatively dense set of longitudinal paths to gain proper information, further increasing the time demand. Thus, for this alternative, higher attention needs to be paid towards choosing these paths and the interpolation method. Another challenge can be the influence of drag force on the measurement setup. In our case, even though the main body itself was a streamlined weight, equipping the other tools on it turned the setup geometry irregular. Additionally, we found that our setup was a bit nose-heavy. Due to this, and the drag force combined, the camera tilted forward during the measurements. As a result, the lower parts of the raw images were sometimes too dark, as the camera was looking over the riverbed, and not at the lit part of the bed. Examples for this could be seen in Fig. 17 (VM4, VM5, 3/5). Hence, in this manuscript we decided to crop out the lower 25% of all raw images generally, before processing them to avoid this element of uncertainty. On the long term however, this effect could be reduced by building a streamlined container (e.g., 3D-printed body, or a body similar to unmanned underwater vehicles') with slots in it for each device, and also by improving the weight distribution. Furthermore, we hypothesize that by using lasers (as originally planned in this study) during the measurements, the known structure (i.e., the position and distances) of the laser points when the setup is perpendicular to the bed, can help to orthorectify the images. This will decrease the effect of occasional tilting when one wishes to carry out size analysis on the images. In our case, we presented how the wavelet method had inherently bigger issues (i.e., image resolution limit), which could not be caused by the camera tilting since those would be in a significantly lower magnitude of error.

As for the training of the AI algorithm with the underwater images, the illumination is indeed a more crucial aspect, compared to normal imagery methods. In many cases only the centre areas of the images were clearly visible, whereas the remaining parts were rather dark and shady. Determining the boundaries between distinct sediment classes for these images was challenging even for experienced eyes. This quality issue certainly generated some incorrect annotations. To overcome this issue, manually varying the white balance thus enhancing the visibility of the sediment could improve the training to some extent. It is worth noting that when Deep Learning methods are to be used, most of the problems arise from the data side (Yu et al., 2007), whereas issues related to the applied algorithms and hardware are rare. This is because data is more important from an accuracy perspective than the actual technical infrastructure (Chen et al., 2020). The time demand of image annotation (data preparation) is relatively high, i.e., a trained person could analyse roughly 10 images per hour. On the other hand, as introduced earlier, a great advantage of using AI is the capability of improving the quality of training itself, often yielding better agreement with reality, compared to the manual annotation. Similar results have been reported by Lu et al., (2018). This at the same time proves that with the introduced approach, there is no need for very precise manual training, thus a fast and effective training process can eventually be achieved.

The validation of the Deep Learning algorithm is far from straightforward. In this study, four approaches were adapted, i.e., a mathematical approach, and comparison with three other measurement methods, respectively. The mathematical approach was based on calculating pixel accuracy and the Intersection-over-union parameter, as it is usually done in case of Deep Learning methods to describe their efficiency (e.g., Rahman and Wang, 2016). However, the latter parameter was shown to be decreased even when the model improved. Consequently, using only the mathematical evaluation in this study could not describe adequately the model performance. Hence, the results were compared to those of three other methods: i) visual evaluation of the image series, ii) a wavelet-based image-processing method (using the method of Buscombe, 2013) and iii) riverbed composition data from physical samples. Considering the features of the applied methods, the first one, i.e., the visual observation, is expected to be the most suitable for the model validation. Indeed, when assessing the bed surface composition by eye, the same patterns are sought, i.e., both methods focus on the uppermost sediment layer. On the other hand, the physical sampling procedure inherently represents subsurface sediment layers, leading to different grain size distributions in many cases. For instance, as shown above, if bed armour develops in the riverbed and the sampler breaks-up this layer, the resulted sample can contain the finer particles from the subsurface layer. On the contrary, in zones where a fine sediment layer is deposited on coarse grains, i.e., a sand layer on the top of a gravel bed, the physical samples represent the coarse material too, moreover, considering that the sieving provides weight distribution this sort of bias will even enhance the proportion of the coarse particles. Attempts were made to involve a third, wavelet-based method for model validation. However, this method failed when finer particles, i.e., sand, characterized the bed. This is an inherent limitation of these type of methods, as discussed earlier, i.e., the pixel size, is simply not fine enough to reconstruct the small grain diameters in the range below fine gravel. Lastly, the most suitable sample points were selected to quantify the performance of the AI. Holding the sieved physical samples as ground truth, the AI showed promising results. In 64% of the points, the difference between AI-detected and physical sediment percentages was less than or equal to 10%. In the rest, it did nut surpassed 20%.

As it is known, the ML and DL models can learn unknown relationships in datasets, but unwanted biases as well. With our current dataset, in our opinion, these biases would be the darker tones of visible grain texture and the lack of larger grain sizes. This way our model in its current state is only applicable effectively in the chosen study site, until the dataset is not expanded with additional images from other rivers or regions. However, the purpose of the manuscript was to introduce the methodology itself and its potential in general and not to create a universal algorithm.

4.4 Novelty and future work

The introduced image-based Deep Learning algorithm offers novel features in the field of sedimentation engineering. First, to the authors' knowledge, underwater images of the bed of a large river have not yet been analysed by AI. Second, the herein introduced method enables extensive (and still relatively quick) mapping of the riverbed, in contrast to most of the earlier approaches, where only several points or shorter sections were assessed with imagery methods. Third, the method is much faster compared to conventional samplings or non-AI based image-processing techniques. The field survey of a 400 m long transect took ~15 minutes, while the AI analysis took 4 minutes (approx. 7 image/s). The speed range of 0.2-0.45 m/s of the measurement vessel and the 15 minutes per transect complies with the operating protocol of general ADCP surveys on rivers (e.g., RD Instruments, 1999; Simpson, 2002; Mueller and Wagner, 2013). Hence, the developed image-based measurement can be carried out together with the conventional boat-mounted ADCP measurements, further highlighting its time efficiency. Indeed, the method is a great alternative approach for assessing riverbed material on-the-go, in underwater circumstances. As an extensive and quick mapping tool, it can support other types of bed material samplings in choosing the sampling locations and their optimal number. Furthermore, it can be used for quickly detecting areas of sedimentation and their extent, as it was shown in Ch. 4.2. (Fig. 16). This way, it may support decisions regarding the maintenance of the channel or the bank-infiltrated drinking water production (detecting colmation and colmated zones). Fourth, a novel approach was used for the imaging and model training. As the camera-bed distance were constantly changing, the mm/pixel ratio also varied. Hence, no scale was defined for the algorithm beforehand. Earlier Deep Learning methods for sediment analysis all applied fixed camera heights and/or provided scaling for the AI. It should be noted that these were airborne measurements, mapping the dry zone of the rivers. In an underwater manner, it is extremely challenging to keep a fixed, constant camera height due to the spatially varying riverbed elevations. Hence, it is of major importance that this manuscript introduces a methodology and a Deep Learning algorithm which neglect the need for scaling. This way, the method is faster and easier to build, but also simpler to use. Of course, as a trade-off, the method, as of now, cannot reconstruct detailed grainsize distributions. Indeed, the purpose was rather to provide a uniquely fast bed material mapping tool, additionally with a much denser spatial resolution than the conventional methods, saving up significant resources.

Originally, beside the three main sediment grain classes introduced in the manuscript (sand, gravel, cobble), others were also defined during annotation (e.g., bedrock, clams), but due to class imbalance (i.e., dominance of the three sediment classes), these were not adapted successfully. There is a good potential in improving the method through transfer learning (see Zamir et al., 2018) using broader dataset, involving other sediment types. Another possible way to counter imbalance is the use of so-called weighted cross entropy (see Lu et al., 2019) on the current dataset, which will also be investigated in our case.

Since the introduced method offers a quick way to provide extensive, spatially dense bed material information of its composition, it may be used to boost the training dataset of predictive, ensemble bagging-based Machine Learning techniques (e.g., Ren et al., 2020) and improve their accuracy. Furthermore, the method can support the implementation of other imagery techniques. For instance, using one of the training videos of this study the authors managed to reconstruct the grain-scale 3D model of a riverbed section with the Structure-from-Motion technique (Ermilov et al., 2020), enabling the quantitative estimation of surface roughness. Underwater field cameras can also be used for monitoring and estimating bedload transport rate (Ermilov et al., 2022) by adapting LS-PIV and the Statistical Background Model approach. This latter videography technique may also be used with moving cameras (e.g., Hayman and Ekhlund, 2003), which enables its adaptation into our method by e.g., detecting bedload movement in the cross-section.

The statistical representativity of the introduced method, as a surface sampling technique, needs to be also addressed in future work. Following and building upon the experience of conventional, surface sampling procedures (e.g., grid sampling; Diplas, 1988) may prove to be beneficial, where they provided the exact number of gravel particles needed to be included (Wolman, 1954) to satisfy the representativity criteria. Then, using edge- and blob-detection would enable to calculate and compare the number of gravel particles in the images to this value. Furthermore, we intend to apply 2 cameras, with overlapping FOVs for increasing the covered area (and the representativity) during surveys. Besides, it would also improve the accuracy of the Structure-from-Motion technique mentioned earlier.

5 Conclusion

A novel, artificial intelligence-based riverbed sediment analysis method has been introduced in this manuscript, which uses underwater images to reconstruct the spatial variation of the characteristic sediment classes. The method was trained and validated with a reasonably high number (~15000) of images, collected in a large river, in the Hungarian section of the Danube. The main novelties of the developed Deep Learning based procedure are the followings: i) underwater images are used, ii) the method enables mapping of the riverbed along the measurement vessel's route with very dense spatial allocation, iii) cost-efficient, iv) works without scaling, i.e., the distance between the camera and the riverbed can vary. Consequently, in contrast with conventional pointwise bed sediment analysis methods, this technique is robust and capable of providing continuous sediment composition data covering whole river reaches, eventually providing the possibility to set up 2D bed material maps. In this way, river reach scale hydromorphological assessments can be supported, where the composition of bed surface is of interest, e.g., when performing habitat studies, parameterising 2D and 3D computational hydrodynamic and morphodynamic models, or assessing the impact of restoration measures.

Financial support. The first author acknowledges the support of the ÚNKP-21-3 New National Excellence Programme of the Ministry for Innovation and Technology, and the National Research, Development and Innovation Fund, Hungary.

Code availability. The code written and used in this manuscript is available here: https://bmeedu-my.sharepoint.com/:f:/g/personal/ermilov_alexander_emk_bme_hu/EjI2neM4AOZGsBkYgKReViEBBzRFRFoYyLlmo6SzTB_qDQ?e=AqpqHI

Data availability. The dataset and results can be accessed using the following link: https://bmeedu-my.sharepoint.com/:f:/g/personal/ermilov_alexander_emk_bme_hu/EhoGx64sP1tFnj8Z1OdMZAsBZWd5gDYzPyodSUDdWFjeiw?e=hKIXjq

Author contributions. GB developed the code and carried out the training process. AAE carried out the fieldwork, evaluated the results, did the laboratory analysis, and collaborated with GB in improving the images. SB oversaw and directed the project, while managing the financial- and equipment background.

Competing interest. The contact author has declared that none of the authors has any competing interest.

Acknowledgements. The authors would like to thank our students Dávid Koós, Gergely Tikász, Schrott Márton and our technicians István Galgóczy, István Pozsgai, Károly Tóth and András Rehák for fieldwork support.

References

Adams, J.: Gravel Size Analysis from Photographs. J. Hydraul. Div., 1979, 105, 1247–1255. doi/10.1061/JYCEAJ.0005283, 1979.

Anglin, D. R., Haeseker, S. L., Skalicky, J. J., Schaller, H., Tiffan, K. F., Hatten, J. R., et al.: Effects of Hydropower Operations on Spawning Habitat, Rearing Habitat, and Standing/Entrapment Mortality of Fall Chinook Salmon in the Hanford Reach of the Columbia River. US Fish and Wildlife Service, final Report. Available at: https://pubs.er.usgs.gov/publication/70179516, 2006.

Baranya, S., Fleit, G., Józsa, J., Szalóky, Z., Tóth, B., Czeglédi, I. and Erős, T.: Habitat mapping of riverine fish by means of hydromorphological tools. Ecohydrology, Volume 11, Issue 7 e2009. Available at: https://doi.org/10.1002/eco.2009 , 2018.

Barnard, P., Rubin, D., Harney, J. and Mustain, N.: Field test comparison of an autocorrelation technique for determining grain size using a digital beachball camera versus traditional methods. Sedimentary Geology, 201(1–2): 180–195., 2007.

Benjankar, R., Tonina, D., Mckean, J.: One-dimensional and two-dimensional hydrodynamic modelling derived flow properties: Impacts on aquatic habitat quality predictions. Earth Surf. Process. Landf. 2015, 40, 340–356.

Benkő, G., Baranya, S., Török, T. G., and Molnár, B.: Folyami mederanyag szemösszetételének vizsgálata Mély Tanulás eljárással drónfelvételek alapján (in English: Analysis of composition of riverbed material with Deep Learning based on drone video footages). Hidrológiai Közlöny, 100, 61–69., 2020.Manuscript

Breheret, A.: Pixel Annotation Tool. Av. at: https://github.com/abreheret/PixelAnnotationTool, 2017.

Bunte, K. and Abt, S. R.: Sampling Surface and Subsurface Particle-Size Distributions in Wadable Gravel- and Cobble-Bed Streams for Analyses in Sediment Transport, Hydraulics, and Streambed Monitoring; General Technical Report (GTR), U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: Fort Collins, CO, USA, 2001.

Buscombe, D. and Masselink, G.: Grain size information from the statistical properties of digital images of sediment. Sedimentology, 56, 421–438. doi/10.1111/j.1365-3091.2008.00977.x, 2008.

Buscombe, D.: Transferable wavelet method for grain-size distribution from images of sediment surfaces and thin sections, and other natural granular patterns. Sedimentology, 60 1709–1732., 2013.

Buscombe, D., Grams, P. and Kaplinski, M.: Characterizing riverbed sediment using high-frequency acoustics: 1. Spectral properties of scattering. Journal of Geophysical Research: Earth Surface, doi: 10.1002/2014JF003189, 119:12, (2674-2691), 2014a.

Buscombe, D., Grams, P. and Kaplinski, M.: Characterizing riverbed sediment using high-frequency acoustics: 2. Scattering signatures of Colorado Riverbed sediment in Marble and Grand Canyons. Journal of Geophysical Research: Earth Surface, doi/full/10.1002/2014JF003191, 119:12, (2674-2691), 2014b.

Buscombe, D. and Ritchie, A. C.: Landscape Classification with Deep Neural Networks. Geosciences, 8, 244. Available at: https://doi.org/10.3390/geosciences8070244 , 2018.

Buscombe, D.: SediNet: a configurable deep learning model for mixed qualitative and quantitative optical granulometry optical granulometry. Earth Surface Processes and Landforms, 45, 638-651. DOI: 10.1002/esp.4760, 2020.

Chandler, J., Lane, S. N. and Ashmore, P.: Measuring river-bed and flume morphology and parameterising bed roughness with a KODAK DCS460 digital camera. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Part B7., 2000.

Chen, C., Zhang, P., Zhang, H., Dai, J., Yi, Y., Zhang, H. and Zhang, Y.: Deep Learning on Computational-Resource-Limited Platforms: A Survey. Volume 2020, Article ID 8454327. Available at: https://doi.org/10.1155/2020/8454327, 2020.

Chen, L., Zhu, Y., Isola, P., Papandreou, G., Schroff, F. and Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Proceedings of the European conference on computer vision (ECCV) (pp. 801-818). https://arxiv.org/abs/1802.02611., 2018.

Cheng, D., Li, X., Li, W. H., Lu, C., Li, F., Zhao, H. and Zheng, W. S.: Large-Scale Visible Watermark Detection and Removal with Deep Convolutional Networks. In book: Pattern Recognition and Computer Vision. First Chinese Conference, PRCV, Guangzhou, China, Proceedings, Part III. DOI: 10.1007/978-3-030-03338-5_3, 2018.

Cheng, Z., and Liu, H.: Digital grain-size analysis based on autocorrelation algorithm. Sedimentary Geology, 327, 21-31. Available at: https://doi.org/10.1016/j.sedgeo.2015.07.008, 2015.

Chezar, H. and Rubin, D. M.: Underwater Microscope System. United States Patent Office, The United States of America as represented by the Secretary of the Interior, US Patent No. 6,680,795 B2., 2004.

Church, M. A., McLean, D. G., and Wolcott, J. F.: Sediments transport in Gravel Bed Rivers. Chap.: Riverbed Gravels: Sampling and Analysis. John Wiley and Sons, New York, 43–88, 1987.

Delong, M. D. and Brusven, M. A.: Classification and spatial mapping of riparian habitat with applications toward management of streams impacted by nonpoint source pollution. Environmental Management, 15:565-571. DOI: 10.1007/BF02394745, 1991.

Detert, M. and Weitbrecht, V.: User guide to gravelometric image analysis by BASEGRAIN. In Advances in Science and Research; Fukuoka, S., Nakagawa, H., Sumi, T., Zhang, H., Eds.; Taylor and Francis Group: London, UK, 2013; pp. 1789–1795. ISBN 978-1-138-00062-9., 2013.

Diplas, P.: Sampling Techniques for Gravel Sized Sediments. Journal of Hydraulic Engineering. DOI: 10.1061/(ASCE)0733-9429(1988)114:5(484), 1988.

Ermilov, A.A., Baranya, S. and Török, G.T.: Image-Based Bed Material Mapping of a Large River. Water, 12, 916. Available at: https://doi.org/10.3390/w12030916, 2020.

Ermilov, A. A., Fleit, G., Conevski, S., Guerrero, M., Baranya, S., & Rüther, N.: Bedload transport analysis using image processing techniques. ACTA GEOPHYSICA, 1895-6572 1895-7455. http://doi.org/10.1007/ s11600-022-00791-x, 2022.

Fehr, R.: Einfache Bestimmung der Korngrös¬senverteilung von Geschiebematerial mit Hilfe der Linienzahlanalyse (In English: Simple detection of grain size distribution of sediment material using line-count analysis). Schweizer Ing. und Archit., 105, 1104–1109., 1987.

Ferdowsi, B., Ortiz, C. P., Houssais, M., & Jerolmack, D. J. (2017). Riverbed armouring as a granular segregation phenomenon. Nature Communications 2017 8:1, 8(1), 1–10. https://doi.org/10.1038/s41467-017-01681-3

Fetzer, J., Holzner, M., Plötze, M. and Furrer, G.: Clogging of an Alpine streambed by silt-sized particles – Insights from laboratory and field experiments. Water Research, Volume 126, Pages 60-69. https://doi.org/10.1016/j.watres.2017.09.015, 2017.

Geist, D. R., Jones, J., Murray, C. J. and Dauble, D. D.: Suitability criteria analyzed at the spatial scale of redd clusters improved estimates of fall chinook salmon (Oncorhynchus tshawytscha) spawning habitat use in the Hanford Reach, Columbia River. Canadian Journal of Fisheries and Aquatic Sciences, 57: 1636-1646., 2000.

Gilcher, M. and Udelhoven, T.: Field Geometry and the Spatial and Temporal Generalization of Crop Classification Algorithms—A Randomized Approach to Compare Pixel Based and Convolution Based Methods. Remote Sens., 13, 775., 2021.

GOPRO Hero 4 Silver: User Manual. Available at: https://gopro.com/content/dam/help/hero4-silver/manuals/UM_H4Silver_ENG_REVA_WEB.pdf, 2014.

GOPRO Hero 7 Black: User Manual. Available at: https://gopro.com/content/dam/help/hero7-black/manuals/HERO7Black_UM_ENG_REVA.pdf, 2018.

Graham, D. J., Reid, I. and Rice, S. P.: Automated sizing of coarse-grained sediments: image-processing procedures. Mathematical Geology, 37, 1–28. https://doi.org/10.1007/s11004-005-8745-x, 2005.

Graham, D. J. Rollet, A.J., Piégay, H. and Rice, S. P.: Maximizing the accuracy of image-based surface sediment sampling techniques. Water Resour. Res., 46, W02508. https://doi.org/ 10.1029/2008WR006940, 2010.

Grams, P. E., Topping, D. J., Schmidt, J. C., Hazel, J. E. and Kaplinski, M.: Linking morphody-namic response with sediment mass balance on the Colorado River in Marble Canyon: Issues of scale, geomorphic setting, and sampling design, J. Geophys. Res. Earth Surf., 118, 361–381, doi:10.1002/jgrf.20050., 2013.

Guerit, L., Barrier, L., Liu, Y., Narteau, C., Lajeunesse, E., Gayer, E., Métivier, F.: Uniform grain-size distribution in the active layer of a shallow, gravel-bedded, braided river (the Urumqi River, China) and implications for paleo-hydrology. Earth Surface Dynamics. 6. 1011-1021. DOI: 10.5194/esurf-6-1011-2018., 2018.

Guerrero, M., Rüther, N., Szupiany, R., Haun, S., Baranya, S. and Latosinski, F.: The Acoustic Properties of Suspended Sediment in Large Rivers: Consequences on ADCP Methods Applicability. Water, 8, 13; doi:10.3390/w8010013, 2016.

Haddadchi, A., Booker, D.J. and Measures, R.J.: Predicting riverbed substrate cover proportions across New Zealand. Catena, Volume 163, pp. 130-146. Available at: https://doi.org/10.1016/j.catena.2017.12.014, 2018.

Hayman, E., Eklundh, J.: Statistical Background Subtraction for a Mobile Observer. Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set 0–7695–1950–4/03, 2003.

He, F., Liu, T., Tao, D.: Control batch size and learning rate to generalize well: theoretical and empirical evidence. Neural Information Processing Systems, 2019.

Ibbeken, H., and Schleyer, R.: Photo-sieving: A method for grain-size analysis of coarse-grained, unconsolidated bedding surfaces. Earth Surf. Process. Landforms, 11, 59–77. Available at: https://doi.org/10.1002/esp.3290110108, 1986.

Igathinathane, C., Melin, S., Sokhansanj, S., Bi, X., Lim, C .J., Pordesimo, L. O. and Columbus, E. P.: Machine vision based particle size and size distribution determination of airborne dust particles of wood and bark pellets. Powder Technol., 196, 202–212. Available at: https://doi.org/10.1016/j.powtec.2009.07.024, 2009.

Kellerhals, R. and Bray, D. I.: Sampling Procedures for Coarse Fluvial Sediments. J. Hydraul. Div., 97, 1165–1180., 1971.

Kim, H., Han, J. and Han, T. Y.: Machine vision-driven automatic recognition of particle size and morphology in SEM images. Nanoscale, 12, 19461–19469. Available at: https://doi.org/10.1039/D0NR04140H, 2020.

Kinsman, N.: Single-Beam Bathymetry Data Collected in Shallow-Water Areas near Gambell, Golovin, Hooper Bay, Savoonga, Shishmaref, and Wales, Alaska 2012–2013; Department of Natural Resources. Division of Geological & Geophysical Surveys: Fairbanks, AK, USA, 2015.

Le, Q. V.: Building high-level features using large scale unsupervised learning. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, pp. 8595–8598., 2013.

Leopold, L. B.: An Improved Method for Size Distribution of Stream Bed Gravel. Water Resour. Res., 6, 1357–1366. https://doi.org/10.1029/WR006i005p01357, 1970.

Limare, A., Tal, M., Reitz, M. D., Lajeunesse, E., and Métivier, F.: Optical method for measuring bed topography and flow depth in an experimental flume. Solid Earth, 2, 143–154, https://doi.org/10.5194/se-2-143-2011., 2011.

Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 60, pages 91–110, 2004.

Lu, S., Gao, F., Piao, Ch. and Ma, Y.: Dynamic Weighted Cross Entropy for Semantic Segmentation with Extremely Imbalanced Data. 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM). doi: 10.1109/AIAM48774.2019.00053, 2019.

Mueller D. S., Wagner, Ch. R.: Measuring Discharge with Acoustic Doppler Current Profilers from a Moving Boat. USGS, Chapter 22 of Book 3, Section A. https://pubs.usgs.gov/tm/3a22/, 2009.

Mueller D. S., Wagner, Ch. R.: Measuring discharge with acoustic Doppler current profilers from a moving boat, version 2.0. https://www.researchgate.net/publication/284587353_Measuring_discharge_with_acoustic_-Doppler_current_profilers_from_a_moving_boat, 2013.

Muste, M., Baranya, S., Tsubaki, R., Kim, D., Ho, H., Tsai, H. and Law, D.: Acoustic mapping velocimetry. Water Resour. Res., 52, 4132–4150, doi:10.1002/2015WR018354., 2016.

Obodovskyi, O., Habel, M., Szatten, D., Rozlach, Z., Babiński, Z., Maerker, M.: Assessment of the Dnieper Alluvial Riverbed Stability Affected by Intervention Discharge Downstream of Kaniv Dam. Water, 12(4):1104. https://doi.org/10.3390/w12041104, 2020.

Padilla, R., Netto, S. M. and da Silva, E. A. B.: A Survey on Performance Metrics for Object-Detection Algorithms. Conference: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP). DOI: 10.1109/IWSSIP48289.2020, 2020.

Perez, L. and Wang, J.: The Effectiveness of Data Augmentation in Image Classification using Deep Learning. arXiv preprint arXiv:1712.04621. Av. at: https://arxiv.org/abs/1712.04621. 2017., 2017.

Purinton, B. and Bookhagen, B.: Introducing PebbleCounts: A grain-sizing tool for photo surveys of dynamic gravel-bed rivers. Earth Surf. Dyn., 7, 859–877. https://doi.org/10.5194/esurf-7-859-2019, 2019.

Rákóczi, L.: Selective erosion of noncohesive bed materials. Geografiska Annaler. Series A, Physical Geography, Vol. 69, No. 1, pp. 29-35. https://doi.org/10.2307/521364, 1987.

Rákóczi, L.: Identification of river channel areas inclined for colmation, based on the analysis of bed material. Vízügyi Közlemények, LXXIX., chapter 3., 1997.

Rahman, M. A. and Wang, Y.: Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation. In: Bebis G. et al. (eds) Advances in Visual Computing. ISVC 2016. Lecture Notes in Computer Science, vol 10072. Springer, Cham. https://doi.org/10.1007/978-3-319-50835-1_22, 2016.

RD Instruments – Acoustic Doppler Current Profilers – Appliction Note: https://www.comm-tec.com/library/Technical_Manuscripts/RDI/FSA-004%20Mode1.pdf, 1999.

Rice, S. and Church, M.: Grain size along two gravel-bed rivers: Statistical variation, spatial pattern and sedimentary links. Earth Surf. Process. Landf., 23, 345–363., 1998.

Ren, H., Hou, Z., Duan, Z., Song, X., Perkins, WA., Richmond, M. C., Arntzen, E. V. and Scheibe, T. D.: Spatial Mapping of Riverbed Grain-Size Distribution Using Machine Learning. Front. Water, 2:551627. doi: 10.3389/frwa.2020.551627, 2020.

Rozniak, A., Schindler, K., Wegner, J. D. and Lang, N.: Drone images and Deep Learning for river monitoring in Switzerland. Semester project. Institute of Geodesy and Photogrammetry, Project, Swiss Federal Institute of Technology (ETH) Zurich, 2019.

Rubin, D. M.: A simple autocorrelation algorithm for determining grain-size from digital images of sediment. J. Sed. Res., 74, 160–165., 2004.

Rubin, D. M., Chezar, H., Harney, J. N., Topping, D. J., Melis, T. S., Sherwood, C. R.: Underwater microscope for measuring spatial and temporal changes in bed-sediment grain size. Sedimentary Geology, Volume 202, Issue 3, Pages 402-408, https://doi.org/10.1016/j.sedgeo.2007.03.020, 2007.

Scheder, C., Lerchegger, B., Flödl, P., Csar, D., Gumpinger, C. and Hauer, C.: Riverbed stability versus clogged interstitial: Depth-dependent accumulation of substances in freshwater pearl mussel (Margaritifera margaritifera L.) habitats in Austrian streams as a function of hydromorphological parameters. Limnologica, Volume 50, January 2015, Pages 29-39. https://doi.org/10.1016/j.limno.2014.08.003, 2015.

Shields, F. D., Jr. and Rigby, J. R.: River habitat quality from river velocities measured using acoustic Doppler current profiler. Environ. Manage.; 36(4):565-75. doi: 10.1007/s00267-004-0292-6., 2005.

Shields, F. D. Jr.: Aquatic Habitat Bottom Classification Using ADCP. Journal of Hydraulic Engineering, Vol. 136, Issue 5, 2010.

Sime, L. C. and Ferguson, R. I.: Information on grain-sizes in gravel bed rivers by automated image analysis. J. Sed. Res., 73, 630–636., 2003.

Simpson, M. R.: Discharge Measurements Using a Broad-Band Acoustic Doppler Current Profiler. USGS, Open-File Report 01-1, https://pubs.usgs.gov/of/2001/ofr0101/, 2002.

Singer, M. B.: A new sampler for extracting bed material sediment from sand and gravel beds in navigable rivers. Earth Surface Processes and Landforms 33(14):2277 – 2284 DOI: 10.1002/esp.1661, 2008.

Soloy, A., Turki, I., Fournier, M., Costa, S., Peuziat, B. and Lecoq, N.: A Deep Learning-Based Method for Quantifying and Mapping the Grain Size on Pebble Beaches. Remote Sens., 12, 3659, doi:10.3390/rs12213659, 2020.

Staudt, F., Mullarney, J. C, Pilditch, C. A. and Huhn, K.: Effects of grain-size distribution and shape on sediment bed stability, near-bed flow and bed microstructure. Earth Surface Processes and Landforms, 44(5). DOI: 10.1002/esp.4559, 2018.

Sun, Z., Zheng, H. and Sun, L.: Analysis on the Characteristics of Bed Materials in the Jinghong Reservoir on the Lancang River. Sustainability, 13, 6874. https://doi.org/10.3390/su13126874, 2021.

Takechi, H., Aragaki, S. and Irie, M.: Differentiation of River Sediments Fractions in UAV Aerial Images by Convolution Neural Network. Remote Sens., 13, 3188. https://doi.org/10.3390/rs13163188, 2021.

Taravat, A., Wagner, M. P., Bonifacio, R. and Petit, D.: Advanced Fully Convolutional Networks for Agricultural Field Boundary Detection. Remote Sens., 13, 722., 2021.

Török, G. T., Baranya, S. (2017) Morphological Investigation of a Critical Reach of the Upper Hungarian Danube. Periodica Polytechnica Civil Engineering. 61(4), pp. 752–761. https://doi.org/10.3311/PPci.10530, 2017.

USDA: Guidelines for Sampling Bed Material. Technical Supplement 13A, 2007.

Vanoni, V. A. and Hwang, L. S.: Relation between Bed Forms and Friction in Streams. J. Hydraulics Division., 93 (3), 121–144. doi:10.1061/JYCEAJ.0001607, 1967.

Verdú, J. M., Batalla, R. J. and Martinez-Casanovas, J. A.: High-resolution grain-size characterisation of gravel bars using imagery analysis and geo-statistics. Geomorphology, 72, 73–93., 2005.

Warrick, J. A., Rubin, D. M., Ruggiero, P., Harney, J. N., Draut, A. E. and Buscombe, D.: Cobble cam: Grain-size measurements of sand to boulder from digital photographs and autocorrelation analyses. Earth Surf. Process Landf., 34, 1811–1821. https://doi.org/10.1002/esp.1877, 2009.

Wilcock, P. R.: Persistance of armor layers in gravel-bed streams. Hydrology and Land Surface Studies. https://doi.org/10.1029/2004GL021772, 2005.

Wolcott, J. F., Church, M.: Strategies for sampling spatially heterogeneous phenomena: The example of river gravels. Journal of Sedimentary Research. v. 61, no. 4, p. 534–543, 1991.

Wolman, M. G.: Method of sampling coarse river bed material. Trans. Am., Geophysical Union, 35(6), 951-956.

WMO: Measurement of river sediments: prepared by the Rapporteur on Sediment Transport of the Commission for Hydrology. Report, World Meteorological Organization - No. 561, Operational hydrology report (OHR)- No. 16, 1981.

Yang, F., Yi, M., Cai, Y., Blasch, E., Sullivan, N., Sheaff, C., Chen, G. and Ling, H.: Multitask Assessment of Roads and Vehicles Network (MARVN). Proceedings Volume 10641, Sensors and Systems for Space Applications XI, 106410D, https://doi.org/10.1117/12.2305972, 2018.

You, K., Long, M., Wang, J., Jordan M.I.: How Does Learning Rate Decay Help Modern Neural Networks? https://doi.org/10.48550/arXiv.1908.01878, 2019.

Yu, L., Wang, S. and Lai, K.K.: Data Preparation in Neural Network Data Analysis. In book: Foreign-Exchange-Rate Forecasting with Artificial Neural Networks. DOI: 10.1007/978-0-387-71720-3_3, 2007.

Zamir, A. R., Sax, A., Shen, W., Guibas, L., Malik, J. and Savarese, S.: Taskonomy: Disentangling Task Transfer Learning. In Proceedings of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 3712–3722. doi: 10.1109/CVPR.2018.00391, 2018.

Zhang, Q., Shi, Y., Chen, Z. and Jiang, T.: ADCP measured flow current of the middle-lower Changjiang River channel. Front. Earth Sci., China 2, 1–9. https://doi.org/10.1007/s11707-008-0016-y, 2008.

Zhou, Y., Lu, J., Jin, Z., Li, Y., Gao, Y., Liu, Y. and Chen, P.: Experimental Study on the Riverbed Coarsening Process and Changes in the Flow Structure and Resistance in the Gravel Riverbed Downstream of Dams. Front. Environ. Sci., https://doi.org/10.3389/fenvs.2021.611668, 2021.

Zhu, J., Park, T., Isola, P. and Efros, A. A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. arxiv, https://arxiv.org/abs/1703.10593, 2020.

Appendix

Appendix A Site A - Section A - I

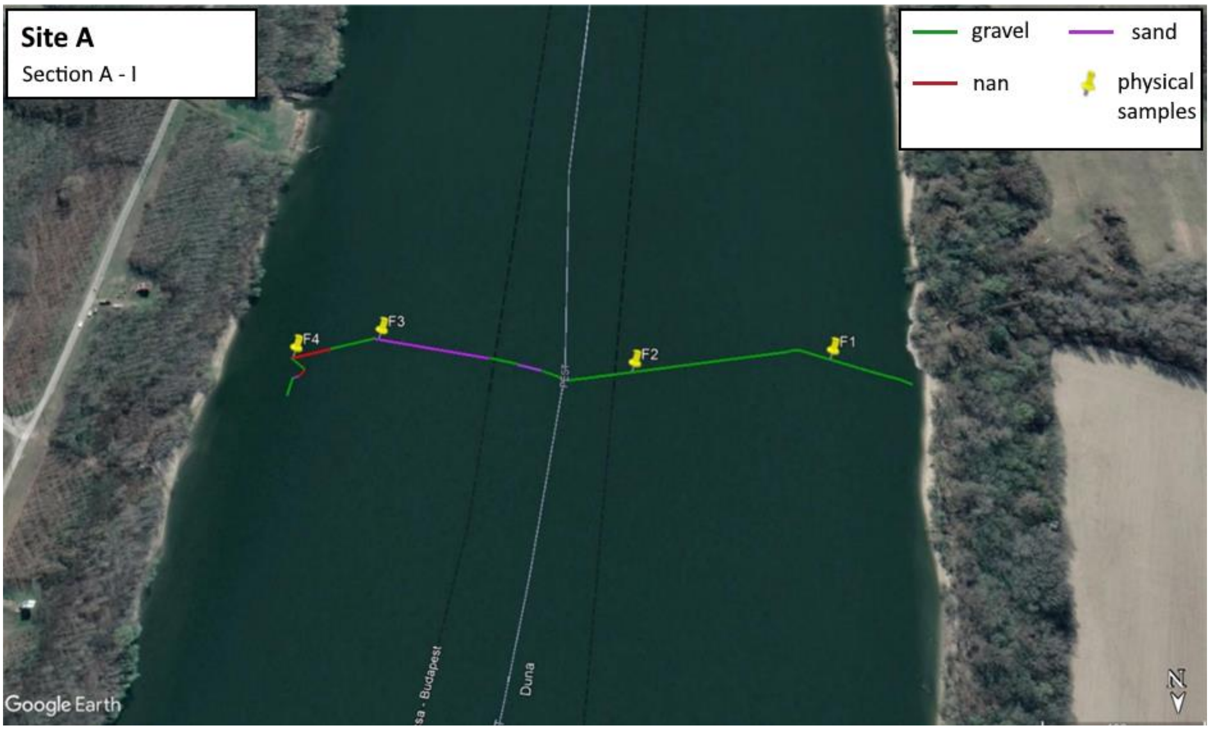

Figure A1: The path of the vessel and camera in Section A - I, Site A. The polyline is coloured based on the sediment seen during visual evaluation of the video. Yellow markers are the locations of physical bed material samplings. (Map created with Google Earth Pro)

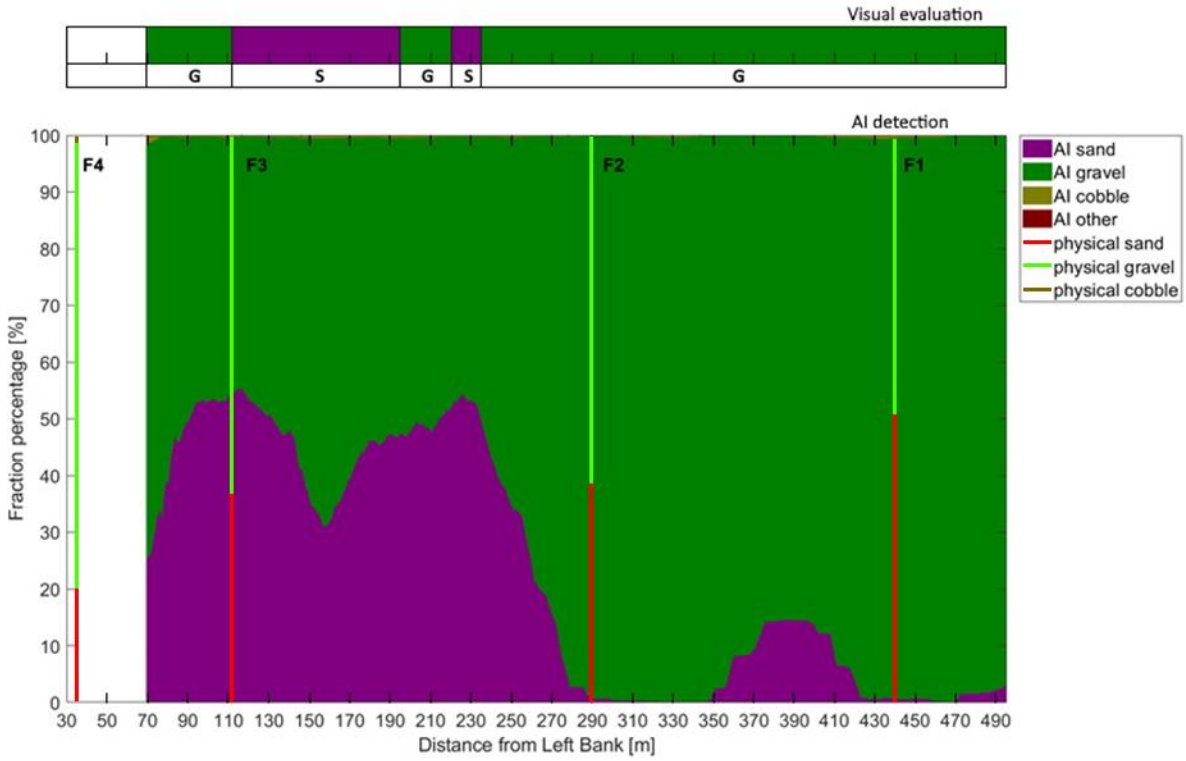

Figure A2: Sediment fraction percentages in Section A - I, recognised by the AI. The visual evaluation included two classes: gravel – G, sand – S). The fractions of the physical samples are shown as verticals.

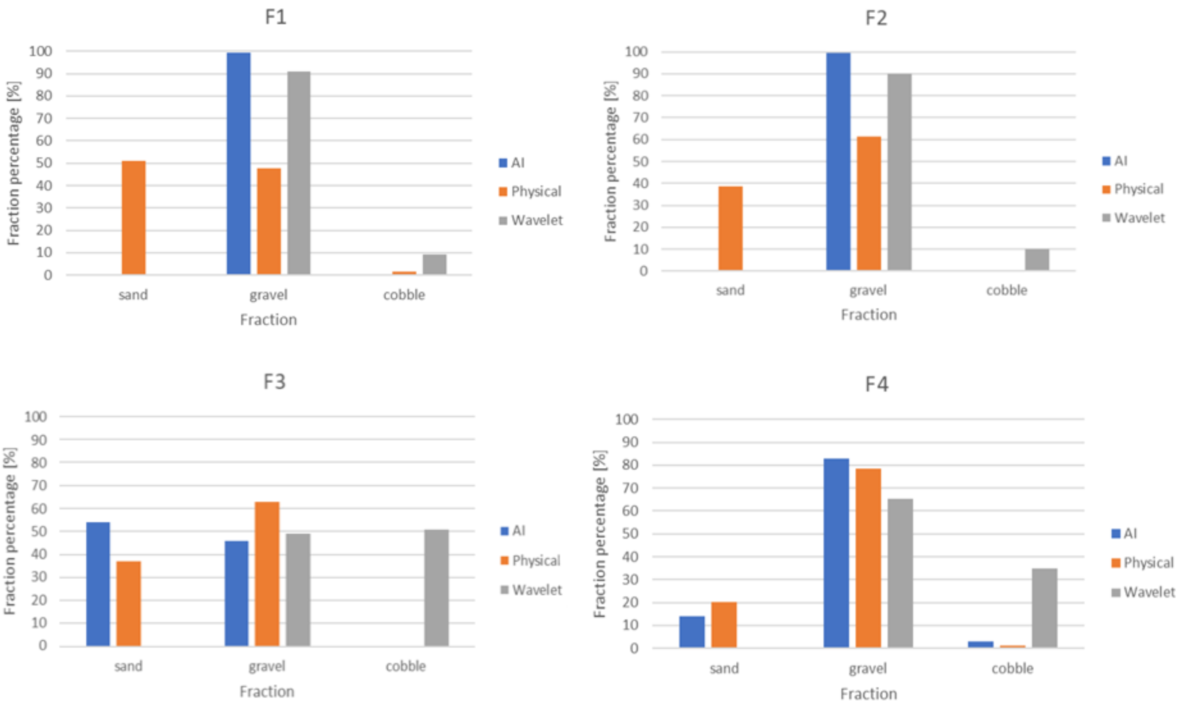

Figure A3: Comparison of sediment fraction % at the sampling locations from the moving-averaged AI detection, conventional sieving and the wavelet-based image processing method. Section A - I.

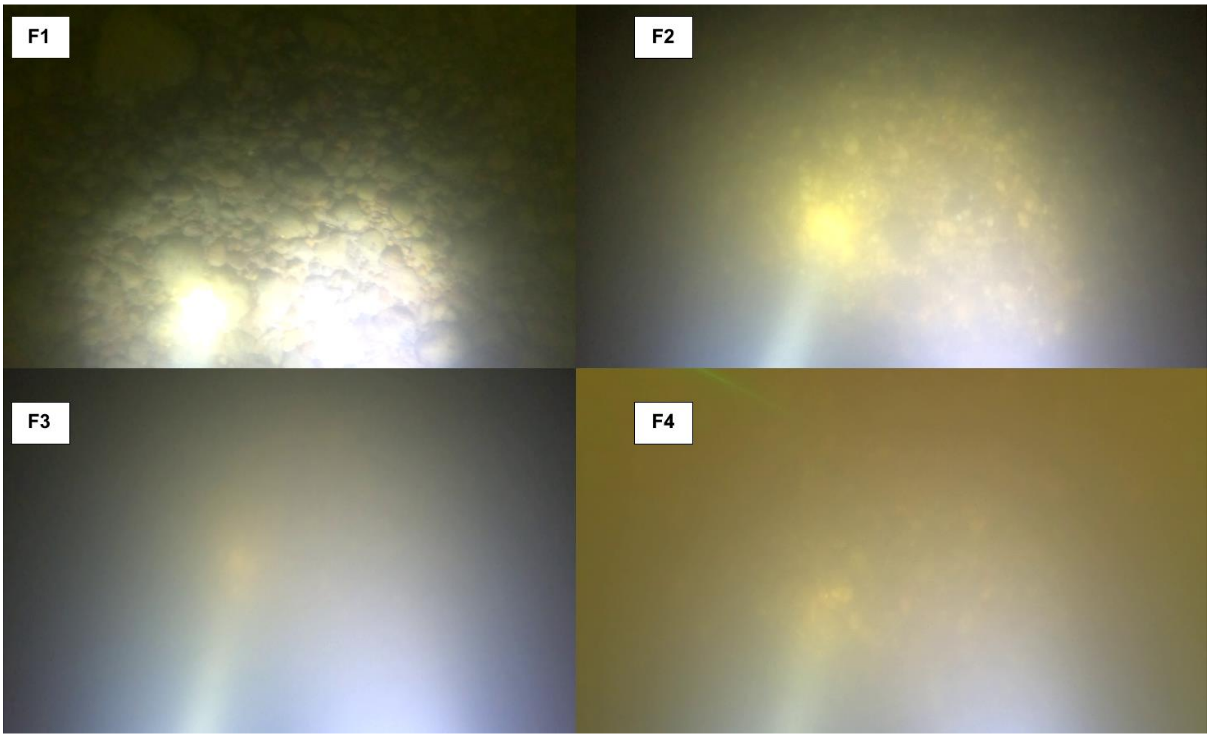

Figure A4: Riverbed video images at the sampling points in Section A - I.

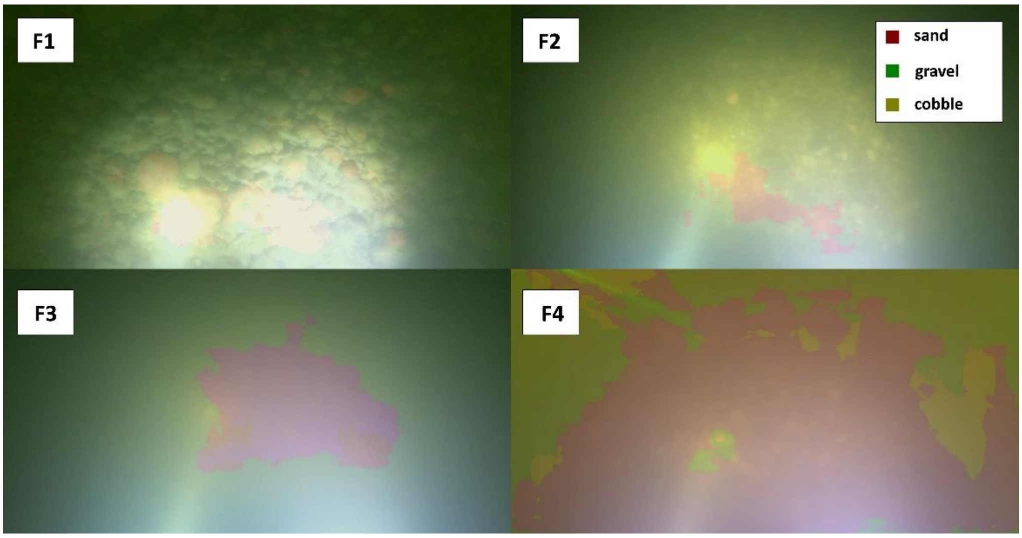

Figure A5: Riverbed video images overlapped with their raw, AI detection result, at the sampling points in Section A - I.

Appendix B Site A – Section A - III

Figure B1: The path of the vessel and camera in Section A - III, Site A. The polyline is coloured based on the sediment seen during visual evaluation of the video. Yellow markers are the locations of physical bed material samplings. (Map created with Google Earth Pro)

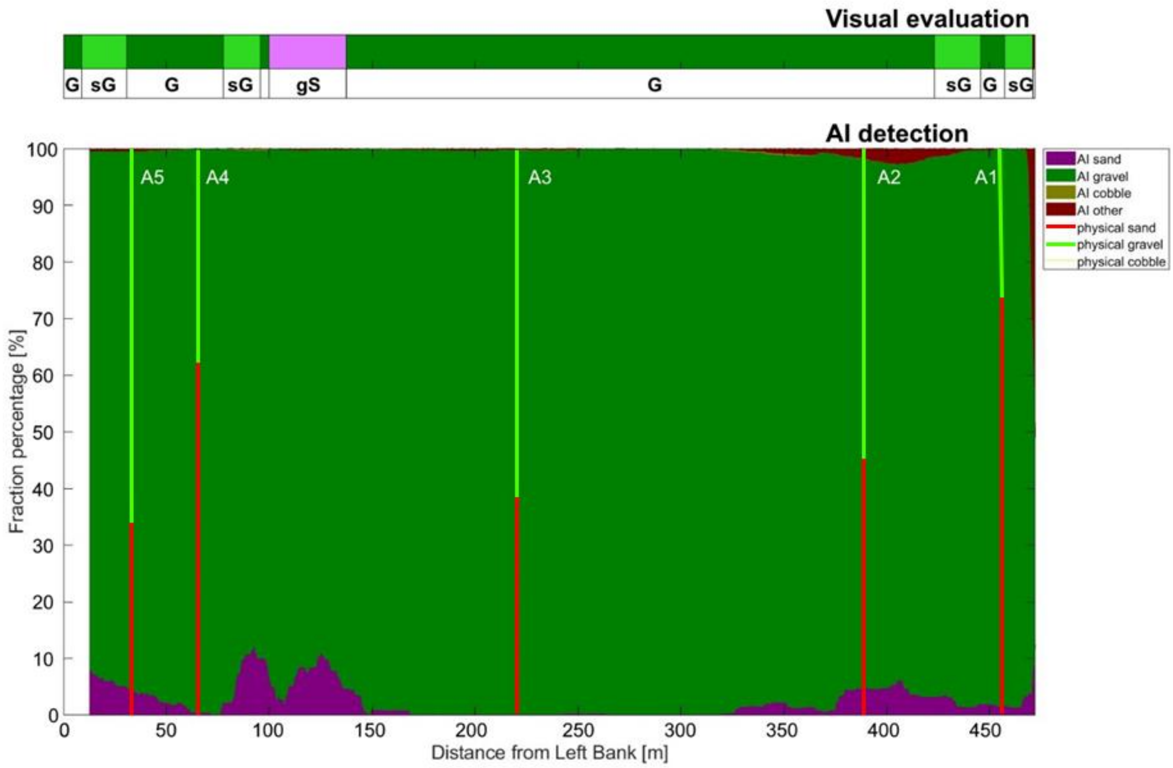

Figure B2: Sediment fraction percentages in Section A - III, recognised by the AI. The visual evaluation included three classes: gravel – G, sandy gravel – sG, gravelly sand - gS). The fractions of the physical samples are shown as verticals.

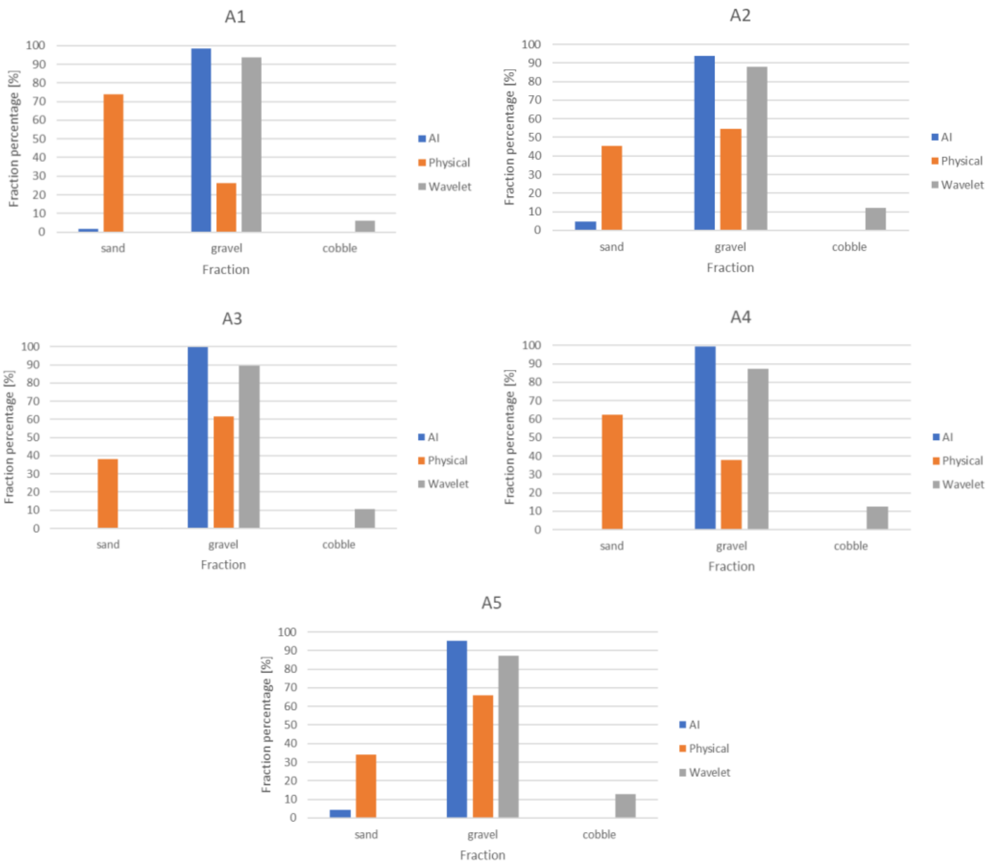

Figure B3: Comparison of sediment fraction % at the sampling locations from the moving-averaged AI detection, conventional sieving and the wavelet-based image processing method. Section A - III.

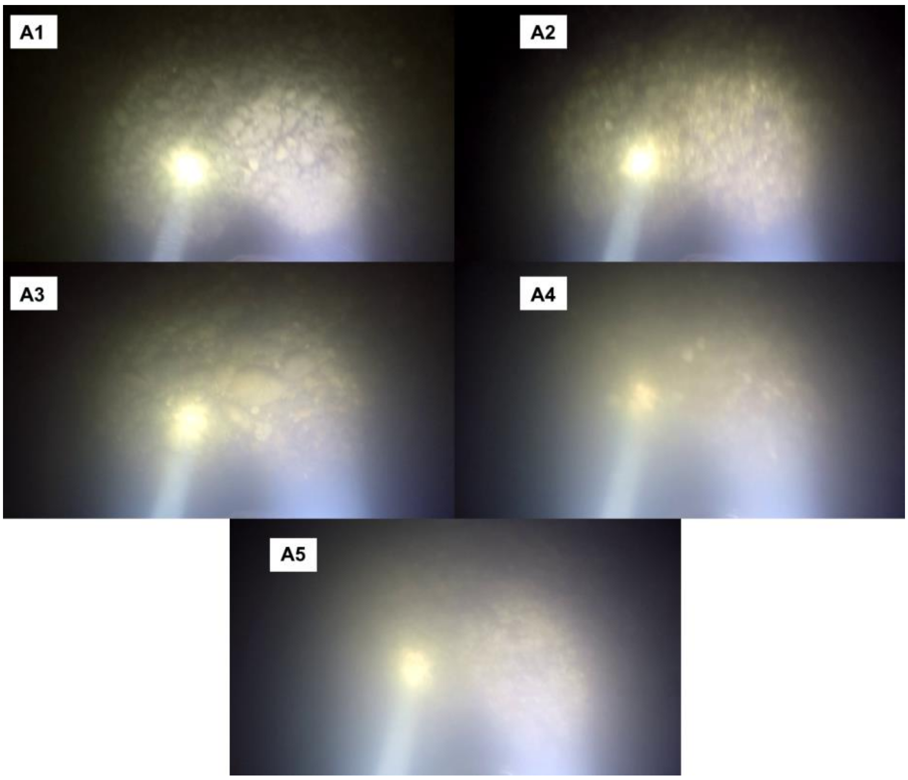

Figure B4: Riverbed video images at the sampling points in Section A - III.

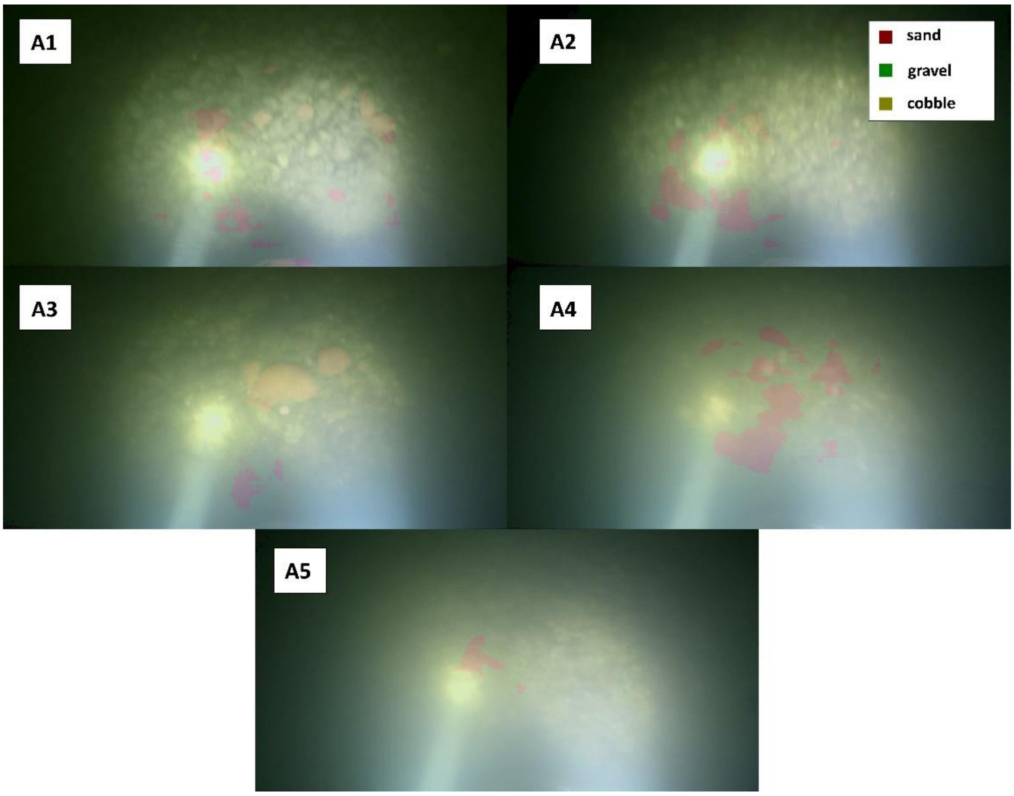

Figure B5: Riverbed video images overlapped with their raw, AI detection result, at the sampling points in Section A - III.

Appendix C Site B – Section B - I

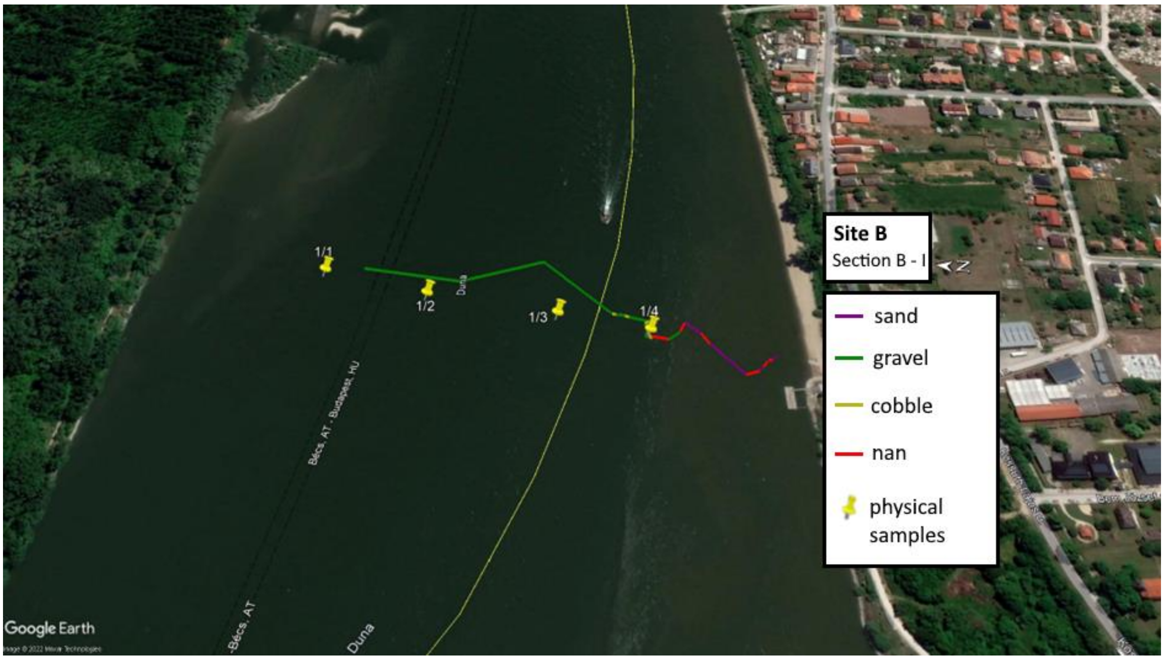

Figure C1: The path of the vessel and camera in Section B - I, Site B. The polyline is coloured based on the sediment seen during visual evaluation of the video. Yellow markers are the locations of physical bed material samplings. (Map created with Google Earth Pro)

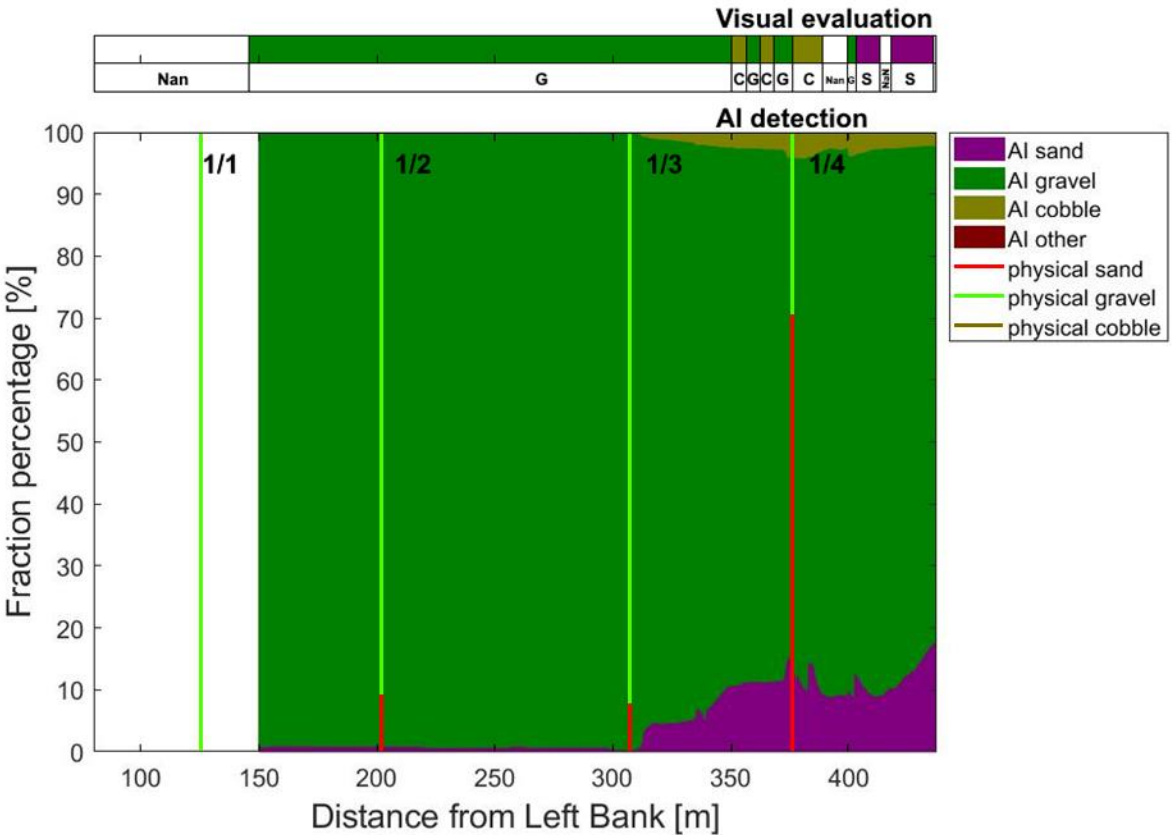

Figure C2: Sediment fraction percentages in Section B - I, recognised by the AI. The visual evaluation included two classes: gravel – G, sand – S). The fractions of the physical samples are shown as verticals.

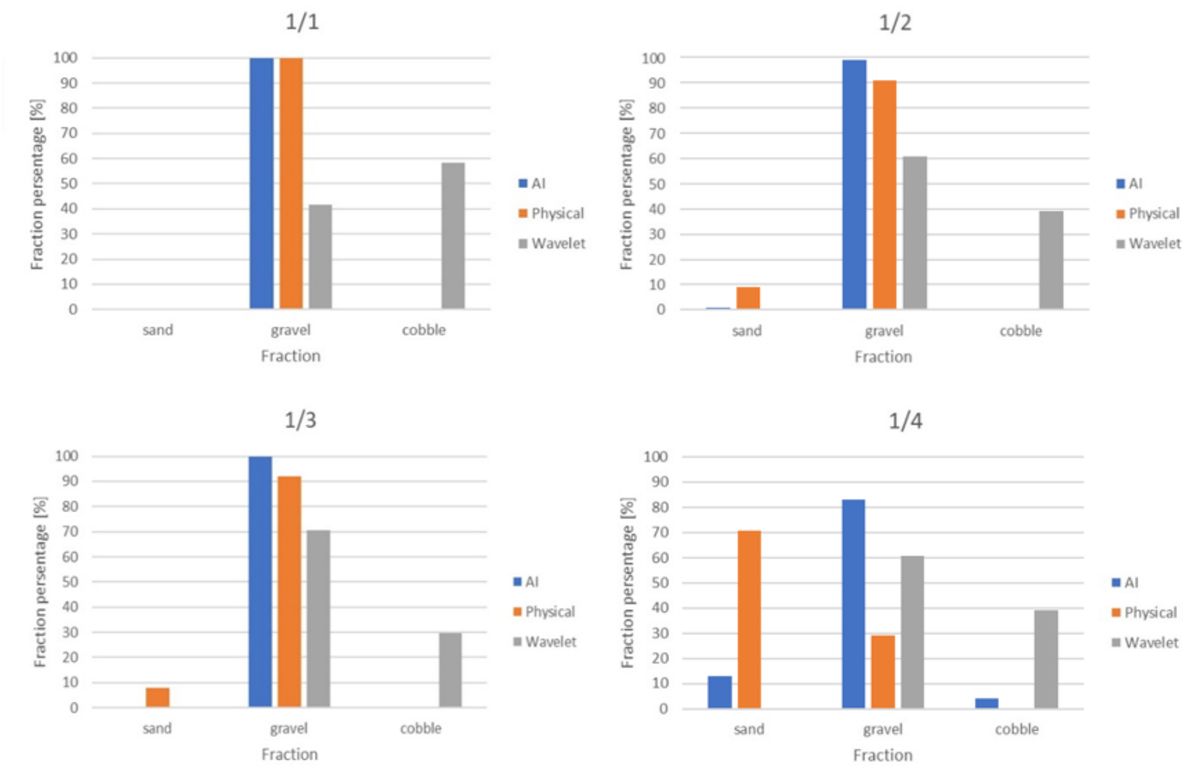

Figure C3: Comparison of sediment fraction % at the sampling locations from the moving-averaged AI detection, conventional sieving and the wavelet-based image processing method. Section B - I.

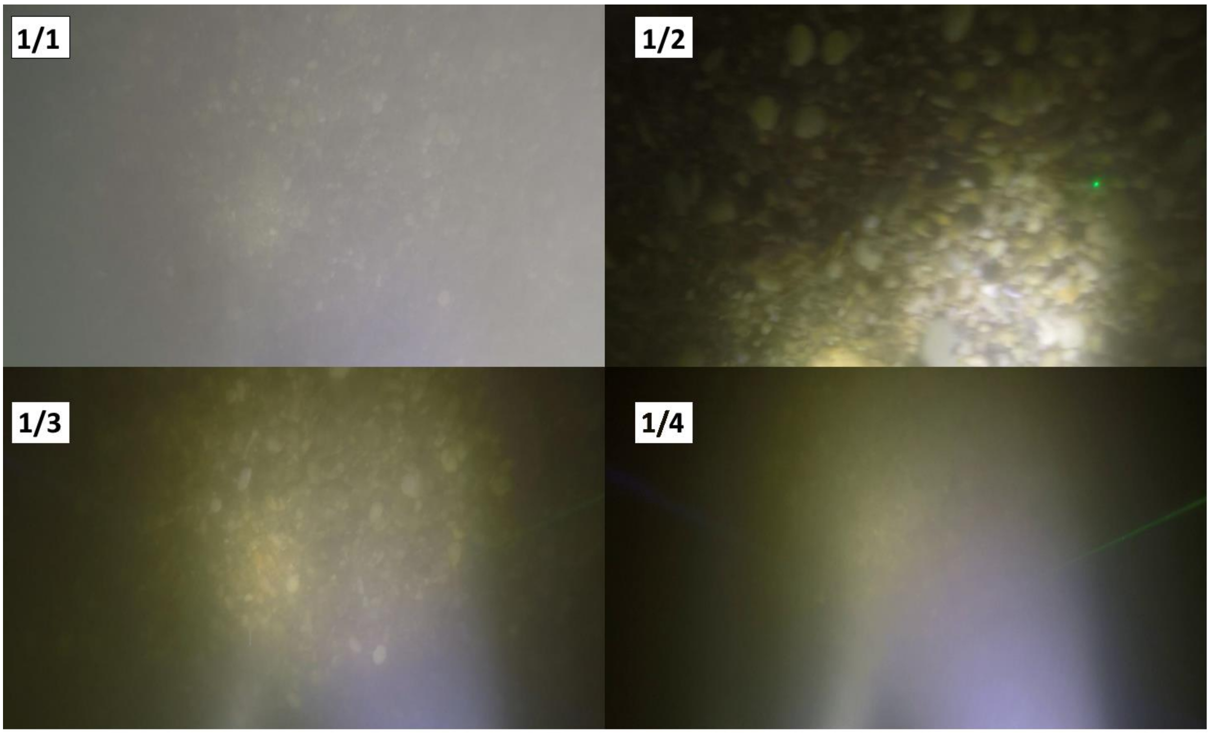

Figure C4: Riverbed video images at the sampling points in Section B - I.

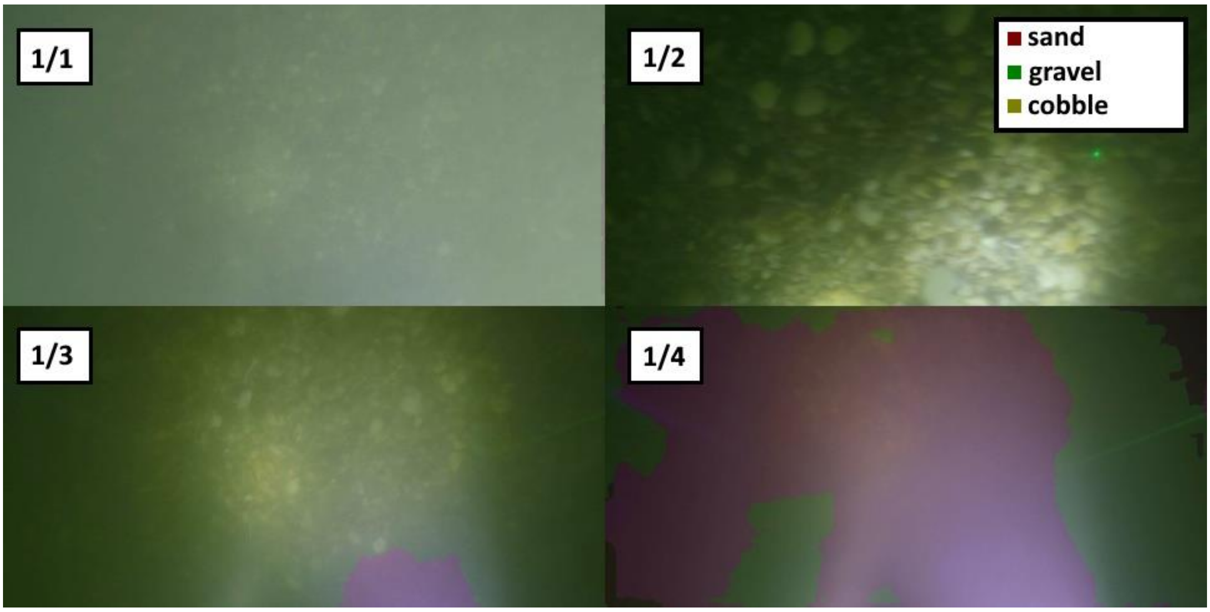

Figure C5: Riverbed video images overlapped with their raw, AI detection result, at the sampling points in Section B - I.