Dear Reviewer 1,

We have structured our response in a Q&A form. In each case, your Question/Comment comes first, and then it is followed by our Answer.

1. General remarks

Question/Comment from Reviewer: Paper is in general well structured, and methods are clearly presented. Results of the analyses with proper discussion could be improved.

Authors have done significant work, processing large volumes of the video data manually, and using wavelet and AI methods. Since extensive data preparation and analyses are required for these methods, authors have decided to present the results as bar plots over the sampling points or over the boat transect. This way, majority of the data processing remains hidden and readers have little insight into the detailed functioning of each automated method. The section of the paper dedicated to results and comparison of the methods (4.2.) contains only 4 paragraphs. In my opinion, the results should be discussed in more detail as they are key for validation of your approach.

I suggest that authors additionally present the complete results of the data processing (on an image-by-image scale) in form of frequency distribution curve. This way more metrics about noise in the results would be given to complement the final result.

Overall, presentation of the results (figures and graphs) could be upgraded to match the quality of the conducted research.

Answer: *Thank you for your comment! We followed your suggestion.*

Primarily:

- *We have introduced Fig. 22 and Fig 23. Explanation and reasoning for these new figures are also given: Line 21 – 24 (Abstract), Line 552 – 575; Line 642 – 645.*

Secondarily:

- *To better show the performance of the AI algorithm we have included Fig. 8 (i.e., raw AI results of a survey in one of the sections; before applying moving-average). We have also included explanation on this: Line 400 – 413.*
- *Even though due to the example of Fig. 8, we use the moving-averaged results of the AI when compared to wavelet and physical sieving, we decided to add the AI detection for the specific images taken in the sampling points (basically the AI values from before moving-average in the given point). See: Fig. 14, Fig. 21, Fig. A5, Fig. B5, Fig. C5. We believe these would give extra insight on the current performance and shortcomings of the presented method and help the readers.*

Question: Similarly, authors in their description of the results mention advantages and disadvantages for each method, such as bed armoring effect, sand detection, isolated gravel patches, etc., but this remains described only as a sentence and lacks visual supplement that would provide concrete sources of uncertainties which are crucial for alleviating the shortcomings of the presented approach and maximize its use. This is especially noticeable for wavelet method that has related data in the main manuscript body.

A: *Thank you for your comment! We have introduced visual supplements by inserting Fig. 10, Fig. 12 and Fig. 19. These include bed armouring, different sediment patches and the wavelet method for both cases. We intended to choose the most telling and prominent sample points for each case.*

Q: Authors propose their setup to be moored from the boat on a line and weighted down, assuming that the setup remains horizontal throughout the deployment. However, depending on the line length and drag influence on the setup geometry (which is highly irregular), it is inevitable that setup will be tilted during measurement. Authors haven't presented any details on how this affects the video data, which might be crucial since there are no reference points on the bed.

A: *Thank you for your comment! We have added some explanation in the manuscript, in ch. 4.3. (Line 592 - 606).*

Q: Lastly, authors have addressed implementation challenges in the section 4.3., which will be helpful in further development of this approach. However, challenges they have detected are result of the selected approach, not universal and therefore need to be addressed accordingly. E.g. vessel speed lower than 0.2 m/s is claimed not feasible for straight transect over the river. Since the proposed method aims to cover broader river section and result in 2D maps, why is it important to transect the river? Maybe longitudinal approach would allow for lower velocities while covering the same area? Implementation challenges should be addressed universally and shortcomings of your selected approach (boat type, camera type, illumination, transects, etc.) should be provided with discussion of the data. This is one of the reasons that this section contains repetitive text from earlier, and again some of the findings are repeated in the following section 4.4. Novelty and future work.

A: *Thank you for your comment! Relating to the vessel speed and longitudinal survey paths, we have added some explanation in the manuscript, in ch. 4.3. (Line 580 – 592). Regarding the repetition of illumination, we have removed this part from ch. 4.3. and inserted it at ch. 4.2, where the problem of illumination first came up and is visually supplemented (Line 517 - 523). We did our best to remove/separate the specific challenges from ch. 4.3. and added them to the section where the setup is introduced, ch. 3.2. (Line 221 – 231).*

2. Specific remarks

Q: Keywords are relatively general and do not offer information of specific contents of the paper. I suggest to drop the keywords "rivers", "sedimentology", "underwater" and "mapping", and use the following keywords for distinction instead: "riverbed texture" and "underwater mapping" (or something along these lines).

A*: Thank you for your comment! We complied with your suggestion and modified the keywords to:*

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

Q: There are several issues with referencing approach used in the paper:

Reference "Török et al., 2017" is not appropriate to be representative of bed armour development since the paper focuses on CFD modelling.

A: *Thank you for your comment! We have replaced it with Ferdowski et al., 2017. (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)*

Q: On ln43 several references from the same author are used where one would be sufficient (Church et al., 1987; Wolcott and Church, 1991; Rice and Church, 1998; USDA, 2007)

A: *Thank you for your comment! We have removed Church et al., 1987 and Rice and Church, 1998.*

Q: On more occasions there are several references grouped together, masking specific relevance for each of them - see ln48, ln73, ln83, ln86, ln88, etc. I suggest that in cases where all references support the same statement select the ones most representative, and in cases where more statements are given you connect the reference with each of them (e.g. ln86)

A: *Thank you for your comment! We have removed several references and left the most representative ones.*

Q: Reference "Kellerhals and Bray, 1971" is used 3 times in the paper, always grouped with several others and twice with Adams. I suggest that you use it where it is most appropriate (ln50) and drop it for general statements (ln40 and ln82)

A: *Thank you for your comment! We have followed your suggestions.*

Q: In the goal of the paper contribution is highlighted as "…through improved (continuous, quick, covering larger areas) data collection.". I would suggest that you rephrase this into "…through more extensive data collection." Since it is hard to argue that method is:

…continuous (data is collected over single in situ survey)

…quick (single point collection is quicker, both for collection and processing)

…covering large area (what is large area, and does this conflict with two previous advantages)

I suggest that these advantages authors address in the discussion and drop from the goal itself since they are not straightforward.

A: *Thank you for your comment! We have followed your suggestions.*

Q: Considering the size of the paper and volume of the work conducted, the goal of the paper is a bit short. I suggest that you expand the goal of the paper to reflect the specific contribution to the field.

A: *Thank you for your comment! We have followed your suggestion and extended the description of our goal in ch. 1. (Line 152 – 161).*

Q: Methodology section briefly describes the three locations of Danube River where data was collected, introducing flow rate data, SSC, etc. This data might be useful for someone familiar with the river, which most of the readers won't be. Please put the presented data in context – e.g. provide complementary duration data, long-term average data, etc. which would allow estimation of conditions under which surveys were performed (low flow, average flow, flood).

A: Thank you for your suggestion! We have followed your suggestion and added $Q_{1\%}$. We also attempted to clarify the Table.

Q: Part of the Methodology focusing on the equipment lacks information that would help understand the data quality, in the context of the maintaining the setup distance from the bed. Please expand the current description with the data about the desired height above the bed and what does it depend on (supposably illumination of the FOV). Similarly, you initially state that size-reference wasn't used in the images (ln253), and after offer contrary statement that laser pointers were used to provide scale (ln294). Probably laser pointers do not offer constant distance due to the bed irregularities, but clarification of the way they were used would be helpful.

A: *Thank you for your comment! We have expanded the camera - riverbed distance related description: (Line 580 – 592).*

In ln173, it was originally said the training was unscaled for the AI. Ln371 - 384 was about the wavelet method, which inherently requires scaling. The wavelet method was only applied in the physical sampling points, so the laser were only used in these points. As we had some problems with them, we did not use the lasers constantly, during the transversal surveys, only when we reached over a given physical sampling point. We have clarified these statements and updated the manuscript (Line 341 – 346).

Q: In the Section 4.1, explaining the training of Deep Learning authors present example of erroneous particle detection of the user. Although this is good and informative example, in my opinion it shouldn't be presented as "ground truth" and rather included as sidenote explaining that training data needs to be carefully selected since you noticed errors in user judgement (n.b., who was the user – one of the authors or trained personnel?)

A: *Thank you for your comment! We have modified the Figure and separated the case of erroneous detection, along with its description. Fig. 5 and Fig. 6, Line 366 – 372.*

N.b. The user was trained personnel. We have now also included this in the text. Line 296.

3. Technical remarks

Q: On ln 31 “fluvial navigation” -> more appropriate would be “fairway placement”, but the connection is loose so I suggest that you replace it with use more relevant to the grain sire instead.

A: *Thank you for your comment! We have followed your suggestion and looked for the most appropriate expression. We have decided with “inland waterway transport”.*

Q: On ln 31 “riverbed structure” -> unclear, does this represent morphology?

A: *Thank you for your comment! We have corrected the sentence to:*

“Knowledge of riverbed (~~structure~~) morphology and (~~grain~~) sediment composition is therefore of major importance in river hydromorphology.”

Q: “Riverbed” and “river bed” are used interchangeably throughout the manuscript. Please proofread the manuscript.

A: *Thank you for your comment! We have corrected it.*

Q: Goal (aim) of the paper is combined with Introduction, making it indistinctive. Please separate the aim of the paper into separate paragraph.

A: *Thank you for your comment! We have followed your suggestion! Line 152 – 161.*

Q: Figure 1 is very simple and lacks details (Danube is not highlighted and therefore indistinctive from other rivers in the figure).

A: *Thank you for your comment! We have updated the figure and highlighted the Hungarian Section of the Danube.*

Q: Can the background orthophoto data be added on the Figures 2 and 3?

A: *Thank you for your comment! We have followed your suggestion! We also merged them into one figure.*

Q: Names of the sections (transects) are very hard to follow since they do not follow any logical order (I suppose they do for the authors, but I suggest that you rename them to achieve clarity for the readers).

A: *Thank you for your comment! We have followed your suggestion!*

Q: “Streamlined weight” -> isokinetic suspended sediment sampler?

A: Thank you for your comment! Yes. We have added the explanation:

> *“ (…) streamlined weight (originally used as an isokinetic suspended sediment sampler) (…)”. Line 211 – 212.*

Q: Methodology section would benefit from added flow chart (process chart) since video decomposition and enhancement is carried out through several steps.

A: *Thank you for your comment! We have added a flow chart in Fig. 4.*

Citation: https://doi.org/10.5194/esurf-2022-56-RC1

Dear Reviewer 2,

We have structured our response in a Q&A form. In each case, your Question/Comment comes first, and then it is followed by our Answer.

Question/comment from Reviewer: In general, the manuscript is well written and structured, although I recommend some small re-structuring at the beginning. The methods are mostly explained in a clear way, however, the way of presenting the results have to be modified to improve the scientific quality of the manuscript before publishing.

Answer: *Thank you for your comment! We followed your suggestion.*

Primarily:

- *We have introduced Fig. 22 and Fig 23. Explanation and reasoning for these new figures are also given: Line 21 – 24 (Abstract), Line 552 – 575; Line 642 – 645.*

Secondarily:

- *To better show the performance of the AI algorithm we have included Fig. 8 (i.e., raw AI results of a survey in one of the sections; before applying moving-average). We have also included explanation on this: Line 400 – 413.*
- *Even though due to the example of Fig. 8, we use the moving-averaged results of the AI when compared to wavelet and physical sieving, we decided to add the AI detection for the specific images taken in the sampling points (basically the AI values from before moving-average in the given point). See: Fig. 14, Fig. 21, Fig. A5, Fig. B5, Fig. C5. We believe these would give extra insight on the current performance and shortcomings of the presented method and help the readers.*

1. General comments:

Q: The abstract contains a brief description of the methods and its advantages compared to conventional measuring procedures and it is mentioned what is included in the manuscript without naming results. Basically, no results are presented in the abstract. E.g., how many sites showed an successful performance etc. One reason might be the missing quantitative evaluation of the performance of the new technique compared to other techniques (see also comment below).

A: *Thank you for your comment! We have introduced Fig. 22 and Fig 23. Explanation and reasoning for these new figures are also given: Line 21 – 24 (Abstract), Line 552 – 575; Line 642 – 645.*

Q: Regarding the structuring the authors choose to write introduction and literature review separately leading to the issue that the shortcomings and restrictions of alternative measuring concepts are included in the chapter of the literature review but the objective and goals of this manuscript is included in the introduction. This does not read straightforward and I recommend to rethink the structuring in this context.

A: *Thank you for your comment! We agree so we have updated and restructured Ch. 1. accordingly.*

Q: The authors present a new AI-approach and used the River Danube as study site. Did the authors make a proof-of-concept or some experiments to test the AI approach under easier boundary conditions? From my perspective, the new AI approach should be first evaluated for easy conditions (maybe shallow and clear water, laboratory conditions, manually selected surface compositions), before applying it at much more challenging conditions.

A: *Thank you for your comment! Yes, we have updated the manuscript and now shortly mention Benkő et al., 2020 as our "proof-of-concept" study, in ch. 3.3., Line 260 – 263.*

Note for the Reviewer: In Benkő et al., 2020 the same AI architecture was applied for analysing drone videos of a dry riverbed. The drone was easily maintaining constant distance from the bed during its flight, enabling scaling of the images. Furthermore, the visibility conditions were far more friendly (compared to underwater conditions), as the measurements were carried out during clear and sunny weather.

Q: Another question that can/should have been discussed is about the representativity of the evaluated areas. Similar to the weight of a sediment sample an area can be specified to be sure that it is representative for the river. Well, I see that this manuscript is mainly to present the method but the representativity aspect should at least be discussed and be mentioned in the outlook.

A: *Thank you for your comment! Indeed, the question of representativity needs to be discussed in future work. In this manuscript, as you mention, the exploration of field application and performance of the AI was our main goal. As such, we included some thoughts on this as outlook, in Ch. 4.4, Line 697 – 704.*

Q: My main concern is the presentation of the results. Although I like the way of showing the results (Figure 5-7), it is too less from a scientific point of view. The manuscript include only a fraction of the verification and validation process. I recommend that authors think about a way to include all results in the paper in a quantitative way. Although the mathematical tools (pixel accuracy, IoU) might not be perfect instruments for the validation, they provide valuable information that can be calculated for the entire datasets and results can be presented e.g. in box plots. This would include the entire work that was done and provide much more information compared to the currently presented results. I also recommend for Figures 8,13 to calculate the differences between the AI approach and the others (for all datasets) and again…the differences can be plotted e.g. as box plots. Overall, this means that a more quantitative way of the verification should be achieved.

A: *Thank you for your comment! We have introduced Fig. 22 and Fig 23. Explanation and reasoning for these new figures are also given: Line 21 – 24 (Abstract), Line 552 – 575; Line 642 – 645.*

2. Specific comments:

Q: Abstract Line 21-22:

The last sentence of the abstract represents the aim/goal/purpose of this manuscript? I suggest shifting the purpose to a more appropriate place.

A: *Thank you for your comment! We have removed the last sentence.*

Q: Keywords Line 23:

From my perspective the selected keywords are not descriptive enough for the presented work (too general words/terms)

A: A: *Thank you for your comment! We have modified and changed the keywords.*

Q: Introduction Line 55-58:

These lines describe the goal of the paper. However, the goal is not described precisely. The authors mention to "introduce a Deep Learning-based technique" without writing for what. My recommendation is to rephrase the goal/objective of the manuscript e.g. by formulating a hypothesis.

A: *Thank you for your comment! We have followed your suggestion and extended the description of our goal in ch. 1. (Line 152 – 161).*

Q: Introduction Line 56, 64, 183 ff

I am not sure, if the term "paper" is sufficient or the correct term for a scientific article (maybe too colloquial?), of if "manuscript" would be the better term (I am not a native speaker...).

A: *Thank you for your comment! We have replaced the word with "manuscript".*

Q: Literature Review, Line 67

"One group of the surrogate approaches are acoustic methods" instead of "One group of the surrogate approaches is the acoustic methods"

A: *Thank you for your comment! We have corrected it.*

Q: Literature Review, Line 106-107

This last sentence does not fit into the listing because it describes the effect of the limitation. However, I do not agree with this last sentence in general. Although image-based analyses might be time-consuming in post-processing, they are very helpful and quick in the field, e.g. to collect sediment data for several kilometres of a river that can also be seen as quick and continuous measurement. From my perspective, the question of "quick" depends on the goal of a survey…

A: *Thank you for your comment! We have removed the last sentence.*

Q: Literature Review, Line 116-117

How can get information about a GSD without detecting the size of particles? Please clarify or rephrase.

A: *Thank you for your comment! We rephrased it to: "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". Line 107 - 109.*

Q: Literature Review, Line 162-168

The article structure using an introduction (with specification of goals), then an literature review with shortcoming of previous method and benefits of the presented new method is a bit strange. I recommend to include the literature review into the introduction (most common way) and then objectives/aims/goals can be specified based on shortcoming of previous methods and/or novelty/benefits of the new method.

A: *Thank you for your comment! We agree so we have updated and restructured Ch. 1. accordingly.*

Q: Figure 1

The formulation of the caption could be improved.

A: *Thank you for your comment! We have modified the caption.*

Q: Figure 2, Figure 3:

The captions should include that the figure is showing the bathymetry of Site A with three measurement transects…I also recommend to combine them to one figure and distinguish between A and B (e.g. Figure 2A à Site A and Figure 2B à Site B), because the information is identical in both.

A: *Thank you for your comment! We have followed your suggestion!*

Q: Figure 4:

please check upper and lower case after colon.

A: *Thank you for your comment! We have corrected it.*

Q: Method, Line 232

"At each cross-section had 4-5 samples were taken" à remove "had"

A: *Thank you for your comment! We have corrected it.*

Q: Methods, Line 286-287

I recommend not to use the word “approximately” here but use the correct percentages for training and validation. Otherwise, the reader gets the impressions that the authors are not sure about the percentage allocation.

A: *Thank you for your comment! We have corrected it. Even though the exact number of these images (46) was published, but we could not find information*

Q: Method, Line 289-292

What are these hyperparameters and what do they change, why do they to be changed? What are the numbers related learning rate 0.01, what is a decay of 0.1 and what a batch size of 15. What is the cross-entropy loss function doing? Well, I am not an expert of AI or any other kind of learning algorithms but a few words explaining the numbers and expressions would significantly increase the comprehensiveness.

A: *Thank you for your comment! We have updated the manuscript and added some explanation to the hyperparameters. Line 326 – 338.*

Q: Results and Discussion, Line 308, 316

Remove the “.” After “Figure 5”… actually, sometimes it is written Fig. and sometimes Figure, please check the entire manuscript for consistency (and if necessary author guidelines).

A: *Thank you for your comment! We have corrected and updated the manuscript accordingly.*

Q: Results and Discussion, Line 310

Why is the expression “pixel accuracy” written in italic?

A: *Thank you for your comment! It was a mistake. We have corrected and updated the manuscript accordingly.*

Q. Figure 5

The caption should describe the 4 single columns. Not necessary to include an evaluation into the caption.

A: *Thank you for your comment! We have corrected and updated the manuscript accordingly.*

Q: Figure 7, 11

Maybe better: “The fractions of the physical samples are shown as verticals”. I recommend to increase the thickness of the verticals for better visualisation.

A: *Thank you for your comment! We have corrected and updated the manuscript accordingly.*

Q: Results and Discussion, Line 358

“The images of the bed from the sampling points are in shown in Figure 9” à Remove the first “in”

A: *Thank you for your comment! We have corrected and updated the manuscript accordingly.*

Q: Results and Discussion, Line 367

please correct “thalweq”

A: *Thank you for your comment! We have corrected and updated the manuscript accordingly.*

Q: Results and Discussion, Line 400-401

please correct “of the of the sediment”

A: *Thank you for your comment! We have corrected and updated the manuscript accordingly.*

Q: Results and Discussion, Line 406-407

I guess the authors mean that the weighting of the physical samples resulted in a mass distribution!?

A: *Thank you for your comment! The manuscript mentioned it as: “The collected samples were analysed in laboratory by drying, sieving, and weighing to provide local grain size distribution.” in ch. 3.2.*

To avoid confusion, we have removed “volumetric distribution” and replaced it with “weight distribution”. Line 505 and Line 636.

Q: Results and Discussion, Line 430-432

In this sentence, the authors describe the results of the other sections in the appendix and find a “well-captured” performance of the AI. What is the base for a well-captured AI? Why are not the values mentioned in the “Method”-Section published?

A: *Thank you for your comment! We have introduced Fig. 22 and Fig 23. Explanation and reasoning for these new figures are also given: Line 21 – 24 (Abstract), Line 552 – 575; Line 642 – 645.*

Q: Results and Discussion, Line 439-442

This statement need to be proven by data. How many successful detection did the AI (percentage)…what means successful in this way? I strongly recommend that the authors underpin such statements with data.

A: *Thank you for your comment! We have introduced Fig. 22 and Fig 23. Explanation and reasoning for these new figures are also given: Line 21 – 24 (Abstract), Line 552 – 575; Line 642 – 645.*

Q: Results and Discussion, Line 454-456

Did the authors investigated the effect of lighting? E.g. experiments in shallow clear rivers without lighting compared to the dark deeper investigations? What are the limitations for the AI regarding the lighting beam? Would it be feasible that the AI learn what grey colour value lighting is? It would be interesting to get some more details on this aspect! Especially what can be done to avoid it…

A: *Thank you for your comment! As of now, only plans exists for further investigating the question of lighting. Unfortunately, the scope of the present manuscript does not include further analysis for this. As we mention in the manuscript, we think one practical solution would be to use lamps with uniform light, ones which are less focused. Other option could be to teach the AI to recognise the light beam and torching effect and simply remove the detected area from calculation. This of course would decrease the effective analysed area and the representativeness.*

Q: Results and Discussion, Line 482-483

I do not agree that the mathematical values are senseless, although they are of course not a perfect tool for validation purposes. However, they allow for analyses in a certain value range (e.g. 0.4 – 0.6, as mentioned earlier by the authors) and to investigate all data and plot them in a meaningful way (e.g. box plots). This would be highly beneficial…

A: *Thank you for your comment! We have introduced Fig. 22 and Fig 23. Explanation and reasoning for these new figures are also given: Line 21 – 24 (Abstract), Line 552 – 575; Line 642 – 645.*

Q: Results and Discussion, Line 542ff

Structure-from-motion required overlapping images from different perspectives, but the camera provides images of the riverbed from only one side, I cannot not imagine that this can be used to generate accurate enough 3D-models of the riverbed.

A: *Thank you for your comment! As the camera passes over an object (sediment particle) in its path, several consecutive images will be taken of the object along the direction of the camera movement. Of course, it means higher uncertainty in the transverse direction. In this sense however, it is the same as airborne Lidar surveys where the drones are also not sweeping through the area from many perspectives, but rather carry out single-flight measurements. For estimating bed roughness for example, it is enough to have a firm grasp in the direction of the camera movement. Partially for this reason, we chose the directional roughness calculation method in our earlier study (Ermilov et al., 2020.) where we tested underwater Sfm. If one wishes to have better 3D results in the transverse, then applying two cameras simultaneously, with overlapping FOV would be beneficial. This is part our plans,*

partially for this very reason, but also to increase the size of the analysed area and improve the representativity of the method.

Citation: https://doi.org/10.5194/esurf-2022-56-RC2