

Earth Surf. Dynam. Discuss., author comment AC2  
<https://doi.org/10.5194/esurf-2021-67-AC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.



## Reply on RC1

Xingyu Chen et al.

---

Author comment on "Convolutional neural networks for image-based sediment detection applied to a large terrestrial and airborne dataset" by Xingyu Chen et al., Earth Surf. Dynam. Discuss., <https://doi.org/10.5194/esurf-2021-67-AC2>, 2022

---

1. Response to referee #1, Byungho Kang

### **Byungho Kang:**

The paper explores the applicability of CNN-based image segmentation on determining the grain size, which could eventually substitute the pre-existing methods for grain detection. The authors verify the reliability of the proposed grain sizing method by comparing it with hand-measured labels and manual sieving.

Promising as it is, the paper needs a small number of revisions.

### **Response:**

Thank you for the nice summary of our work.

### **Byungho Kang:**

It seems necessary to elaborate the general training process for the U-net (e.g., hyperparameters values, number of epochs for training, what kind of optimization method were used, how the tiles were being selected for training, et Cetra). The inclusion of a section or a table could help.

### **Response:**

We added a paragraph (L201-L209) to elaborate upon the U-net implementation.

The U-net was implemented based on a python library pytorch (Paszke et al., 2019). The cross entropy loss function and the stochastic gradient descent were used for model

optimization. Model hyperparameters were tuned based on grid searching optimization and 5-fold random cross validation (Goodfellow et al., 2016). The training speed for U-net is influenced by the number of images in the training datasets, the batch size and the number of training epoch. Given a fixed training datasets, the hyperparameter number of training epoch was tuned first, followed by the learning rate. The maximum batch size depends on GPU memory and we preferred larger batch size for faster training speed when several batch size values result in a similar error during the cross-validation. The optimized model hyperparameters are: (1) number of training epoch = 150; (2) learning rate = 0.005; (3) batch size = 96; and (4) image tile size = 512. The optimum image tile size was determined based on the analysis in section 5.2. '

**Byungho Kang:**

Likewise, it would be better to provide more information on manual labeling. Even labeled by two observers, labeling 128,461 grains would take significant hours, and I assume it was based on an auto/semi-automatic algorithm that captures the grains from images.

**Response:**

To better train and evaluate our method, the 128,461 grains manual labeling dataset were all prepared by hand. The two observers cross-checked their results with each other, such that all the manually labeled images were from the consensus of two observers.

We added more details in the preparation of the manual labeling process in the manuscript (L137-L142).

**Byungho Kang:**

It would be better to include information about the number of grains in test images (both for manual labeling and U-net based prediction). This could improve the overall credibility (of error calculation) than merely providing the percentile-based information.

**Response:**

Accepted. We added a column 'Average # of Grains in each image' in Table 1 for manual labeling. Since manual labeling was used as the baseline method in Section 4.2, 4.3 and 4.4, we think that the number of grains for manual labeling is sufficient. Meanwhile, the percentile-based error is based on the calculation of grain weight and can provide different information from the number of grains in each image.

**Byungho Kang:**

The comparison between the sieving dataset and other image-based methods(Fig.6a) needs more clarification (e.g., conversion of physical scale (mm to pixels) should be mentioned in section 4, not 5)

**Response:**

Revised. We added a paragraph (L248-L251) to better explain the comparison between sieving and other methods (e.g., the conversion of physical scale and the correction of

image photo distortion). We also added a section to introduce how a predicted image was transferred to grain size information (L187-L192).

We implemented a systematic analysis on how image resolution influences the predictive ability of Grain ID in section 5.3, the topic is different with the comparison of sieving, we think the section should be in the discussion part not in the result part.

**Byungho Kang:**

Some minor typos at line 244: Fig. 4b should be replaced with Fig. 5b.

**Response:**

Revised.