[**Instructions**: Each data package must populate this readme template. Delete the instructions (marked with brackets throughout) before saving. If using the template in Git, save as .md. If starting with the SFA GitLab template, this template is pre-populated. **For inclusion in your ESS-DIVE data package, save as .pdf.**The readme is not limited to the information or sections described below. No matter what is included in the readme file, the documentation should be written for an audience that has never seen the data package contents before. Descriptions should be written as clearly and concisely as possible. We strongly recommend having someone review the content before finalizing.]

[**Instructions for data packages**: Follow the [step-by-step ‘how to’ instructions](https://pnnl.sharepoint.com/:w:/r/teams/SubsurfaceBiogeochemicalResearchSFA/Shared%20Documents/General/SFA%20Data%20and%20Software%20Management/How-to-Publish-Data-Package.docx?d=wced5db769f7e4d328434ea7ad7350d59&csf=1&web=1&e=Px7pHV).]

|  |
| --- |
| **Title** |

Machine learning photogrammetric analysis of images provides a scalable approach to study riverbed grain size distributions

|  |
| --- |
| **Summary** |

This data package is associated with the publication “Machine learning photogrammetric analysis of images provides a scalable approach to study riverbed grain size distributions” submitted to XXX (Regier et al, XXX) and a GitHub repository (https://github.com/peterregier/d50\_computer\_vision). The distribution of sediment grain size in streams and rivers is often quantified by the median grain size (d50), a key metric for understanding and predicting hydrologic and biogeochemical function of streams and rivers. Manual methods to measure d50 are time-consuming and ignore larger grains, while model-based methods to estimate d50 often over-generalize basin characteristics, and therefore cannot accurately represent site-scale heterogeneity. Here, we apply a machine learning photogrammetry methodology (You Only Look Once, or YOLO) for estimating d50 for grains > 2 mm based on images collected from streams and rivers throughout the Yakima River Basin (YRB). To understand how photogrammetric methods may help bridge the gaps in resolution and accuracy between manual and model-based d50 estimates, we compared YOLO d50 values to manual and model-based estimates across the YRB. We found distinct differences among methods for d50 averages and variability, and relationships between d50 estimates and basin characteristics. We discuss the advantages and limitations of the YOLO algorithm versus current methods, and explore potential future directions to combine d50 methods to better estimate spatiotemporal variation of d50, and improve incorporation into basin-scale models. This data package is associated with

|  |
| --- |
| **Critical Details** |

**[Section Optional]**

[**Instructions**: This section is optional and may be less relevant for models. It may contain a description of a labeling scheme and mapping across samples, replicates, sensors, experiments, sites, or files. It may contain clarifying information about differences among samples, replicates, sensors, experiments, sites, or files. It may contain clarifying information on how to work with the data, analyses, or models or what processing has been done.]

|  |
| --- |
| **Data Package Structure** |

**[Section Required]**

[**Instructions**: SFA data packages are required to follow ESS-DIVE Reporting Formats. Follow the [step-by-step ‘how to’ instructions](https://pnnl.sharepoint.com/:w:/r/teams/SubsurfaceBiogeochemicalResearchSFA/Shared%20Documents/General/SFA%20Data%20and%20Software%20Management/How-to-Publish-Data-Package.docx?d=wced5db769f7e4d328434ea7ad7350d59&csf=1&web=1&e=Px7pHV).]

[**Instruction:** Include the following sentence or something similar in your readme.: “Please see [name of FLMD csv] for a list of all files contained in this data package and descriptions for each. Data dictionaries for csv files have \_dd” appended to the file names.”]

[**OPTIONAL Instruction**: In addition to the required FLMD csv, you can choose to explain specific folders or files in this readme as desired.]

[**OPTIONAL Instruction**: You can choose to include a diagram/map of your data package structure in this readme.

[**OPTIONAL Instruction**: **For PC users:**

1. Type “cmd” in the search to get “Command Prompt” and right click to select “Run as administrator”
2. Under “Command Prompt”, type the path where your data package saved.
   1. If the path is on other disk on your local computer than “C” drive, or on share folder and mapped on your PC as “Y” drive,
      1. Type > Y: \\ ##this command leads to your Y drive
      2. Type > cd “folder name under Y drive”
   2. If the path is on PNNL One-drive folder
      1. Type> c:\Users\renh686\OneDrive – PNNL
3. Type “tree” and hit ‘enter’ to create an image of the directory structure of the path (<https://docs.microsoft.com/en-us/windows-server/administration/windows-commands/tree>).

Note: During the processing, use “Ctrl+C” to abort the current task and regain user control.

1. Copy and paste the command output to the readme file. To render it nicely in markdown, use code block by placing triple backticks before and after the directory tree . E.g., ```/directory tree```

[**OPTIONAL Instruction**: **For Mac/Linux:**

1. Open a terminal and go to the root directory of your data package. E.g. `cd /path/to/data\_package`

2. Type `tree -d -L 2` to show the directory tree with a maximum display depth of 2. See full examples [here]( <https://www.howtoforge.com/linux-tree-command/>)

3. Copy and paste the command output to the readme file. To render it nicely in markdown, use code block by placing triple backticks before and after the directory tree. E.g., ```/directory tree```

|  |
| --- |
| **Citations and Acknowledgements** |

**[Section Required]**

[**Instruction include the following text for SFA packages or replace with the text from your manuscript acknolwedgments:** This research was supported by the U.S. Department of Energy (DOE) Biological and Environmental Research (BER) Environmental System Science (ESS) program (<https://ess.science.energy.gov/>) through the Pacific Northwest National Laboratory River Corridor Science Focus Area (SFA). PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DE-AC05-76RL01830.]

[**Instruction when publishing to ESSDIVE and there is an associated manuscript, include the statement:** “Cite this data package with the appropriate DOI. Cite the associated manuscript in any work that that uses analyses or conclusions presented in the manuscript. To cite the paper: [**Instruction**: Insert full paper citation]]

[**Instruction** **for software development, include the following statement verbatim:** “This material was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the United States Department of Energy, nor Battelle, nor any of their employees, nor any jurisdiction or organization that has cooperated in the development of these materials, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness or any information, apparatus, product, software, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. PACIFIC NORTHWEST NATIONAL LABORATORY operated by BATTELLE for the UNITED STATES DEPARTMENT OF ENERGY under Contract DE-AC05-76RL01830.”]

[**Instruction** **for WHONDRS data packages, include the following and modify as needed: “**- Please acknowledge the U.S. Department of Energy (DOE) Biological and Environmental Research (BER) Environmental System Science (ESS) program (<https://ess.science.energy.gov/>) — which generously provides funding to the WHONDRS — in work that uses this dataset. If using FTICR-MS, NMR, LC-MS, or GC-MS data, please also acknowledge the U.S. DOE Environmental Molecular Sciences Laboratory (EMSL; grid.436923.9).”]

|  |
| --- |
| **Contact** |

**[Section Required]**

[**Instruction**: Insert Contact name, Contact email]

|  |
| --- |
| **Change History** |

**[Section Required When on ESS-DIVE]**

**[Instruction**: **NOTE: this section only needs to be included in the ESS-DIVE data package readme. It does not need to be included in the GitLab/GitHub readme. The purpose is to indicate if you have revised a published data package after it has been published. Populate row 1 during first publication. Populate row 2 and beyond as needed if revisions are done.]**

|  |  |  |
| --- | --- | --- |
| Version 1 | [Insert Date] | Original data package publication |
|  |  |  |