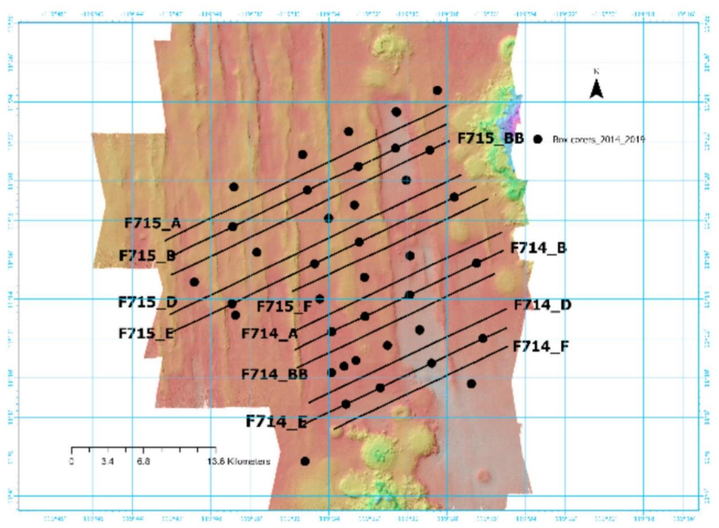
Enhanced estimation of polymetallic nodule resources from seafloor images using Mask R-CNN and YOLO with Segment Anything Model (SAM) annotations

Karol Kabała, Tomasz Kogut, Arkadiusz Tomczak
Maritime University of Szczecin, Wały Chrobrego 1-2, 70-500 Szczecin, Poland,
{k.kabala, t.kogut, a.tomczak}@pm.szczecin.pl

I. GOAL

This study aims to develop and apply accurate methods for estimating the abundance of polymetallic nodules using seafloor imagery. By employing advanced instance segmentation models (Mask R-CNN and YOLO), the research focuses on the automatic detection and measurement of individual nodules. This allows for reliable estimation of both their number and size, which is essential for assessing the volume of available mineral resources. To streamline and enhance the dataset annotation process needed for training the models, the Segment Anything Model (SAM) is also utilized. Ultimately, the goal is to improve the precision of resource estimation in support of the responsible and sustainable development of deep-sea mineral extraction.

II. DATA



- 30 373 bottom images were acquired at a resolution of 5184 × 3456 pixels
- 16 survey lines
  - 12 survey lines were located in the H22NE area
  - 2 lines in the H33 area
  - 2 outside the designated areas.
- 31 bottom samples were taken using a box corer (black points)

(a) Image of polymetallic nodules on bottom,
(b) nodules extracted from bottom in box corer,
(c) sorted nodules on measuring scale

III. METHODS

Two deep learning–based instance segmentation models were developed to automatically detect and delineate individual polymetallic nodules in seabed imagery.

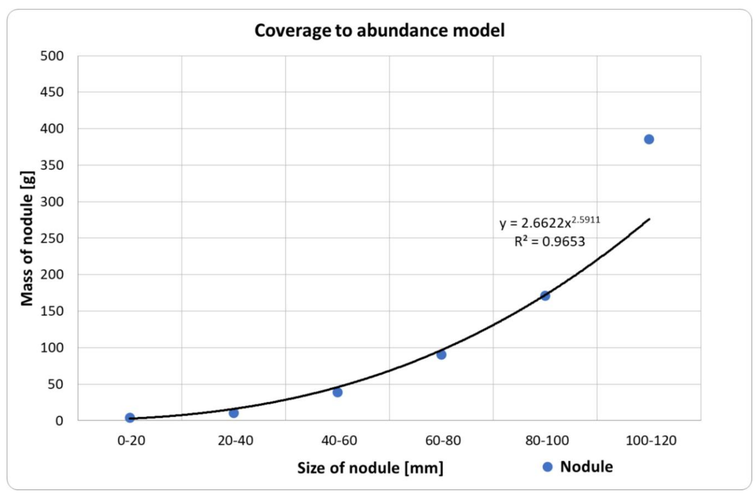
1. **Mask R-CNN**
2. **YOLOv5**

The model development process relied on transfer learning, allowing the use of pre-trained weights as a starting point, and was followed by targeted fine-tuning of hyperparameters to adapt the models to the specific visual characteristics of underwater environments.

TRAINING DATA PREPARATION

We used the Segment Anything Model (SAM) to create initial segmentation masks for the training set. While SAM proved effective in handling most image regions, it occasionally misclassified boundaries or merged overlapping objects. In these cases, we manually corrected the masks to ensure the dataset remained consistent and accurately annotated. These adjustments were necessary to prevent noise in the training process and to maintain a high standard of label quality. The annotation process was designed to guide the neural network by indicating the expected locations of nodules within the images. It was particularly important to include nodules that were partially obscured by sediment, as this ensures a more accurate representation of their true occurrence.

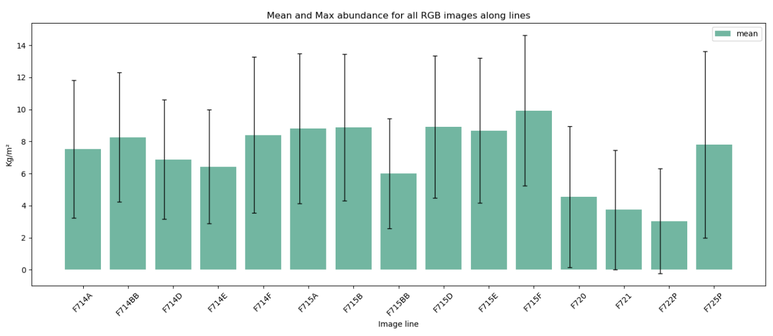
VOLUME ESTIMATION

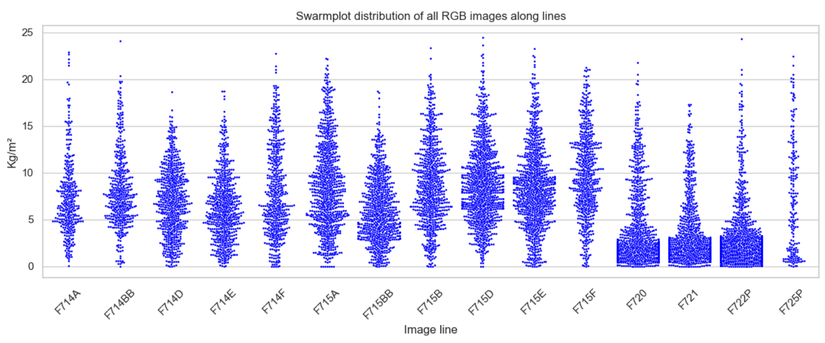


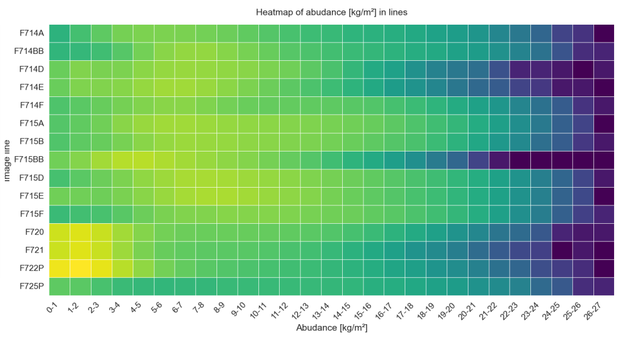
We developed an empirical model to translate the observed surface coverage of polymetallic nodules into estimates of their subsurface volume. This approach was grounded in direct measurements from collected physical samples, which provided a reliable link between image-based coverage estimates and actual volumetric abundance.

IV. RESULT

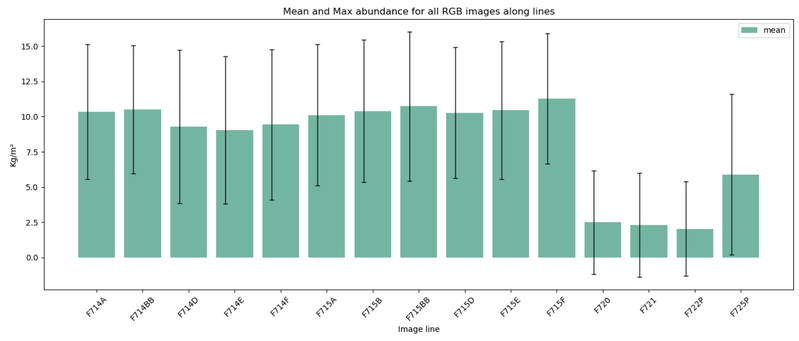
ANALYSIS OF ABUNDANCE VARIABILITY USING MASK R-CNN

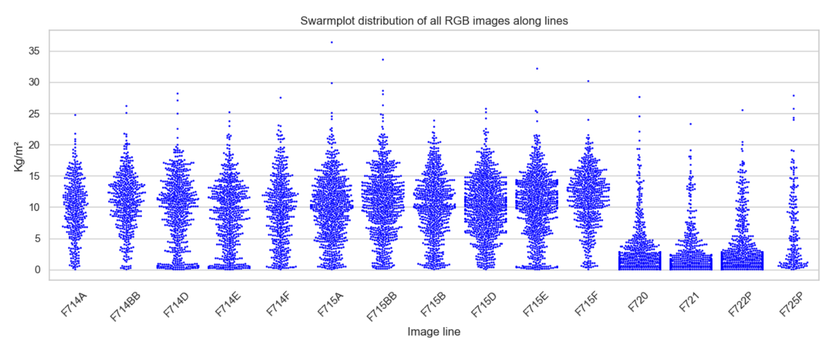


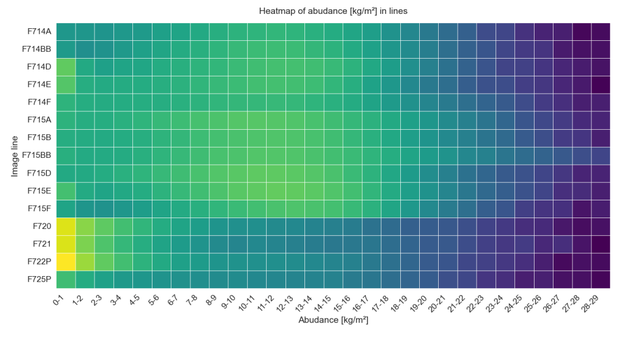




ANALYSIS OF ABUNDANCE VARIABILITY USING YOLOv5







COMPARISON MASK R-CNN AND YOLOv5 IN THE H22NE AREA