This dataset is produced by the National Land Survey of Finland in 2022 and is a collection of images that can be used as training data for semantic segmentation of buildings. The dataset contains 103 true orthophotos and 103 corresponding labels, both in png format. All of the true orthophotos and labels are 2000 x 2000 pixels in size. The images are from five different areas, which can be identified from the name of the images. The true orthophotos are named in the following order: A + area number\_img + image number in the area. The corresponding labels are named in a similar matter; A + area number\_label + image number in the area. The pictures starting with A1 are from Äänekoski, with A2 are from Heinola, with A3 are from Nurmes, with A4 are from Savonlinna, and with A5 are from Toholampi. The pictures from area A1 are 0.4 m in resolution, while the remaining are 0.3 m. The true orthophotos are in RGBA, where the alpha channel is transparent, but can be viewed with GIS software or with image editing software. The labels contain the value 1 for buildings, and no data for areas without buildings.