High-speed Rail (HSR) Line Building Dataset Introduction

Zhengzhou-Xi'an HSR line and Beijing-Shanghai HSR line are representative high-speed rail lines in east-west and north-south directions of China, operating mileage in line of about 484 kilometers and 1318 kilometers, respectively. The Zhengzhou-Xi'an HSR line is the main road connecting the central and western China, which cut the shortest running time between the two cities from former more than six hours to less than two hours. It will meet the growing demand of passenger and cargo transportation in central and western China, and help promote local development. The Beijing-Shanghai HSR line, which connects two major economic zones in China, i.e., Bohai Economic Rim and the Yangtze River Delta, has drawn great attain in the world. This railway is considered to be one of the busiest railway lines in China.

We manually edited and created a high-speed rail line building dataset and named it HSR line building dataset. As depicted in Figure 1 and Figure 2, this dataset contains a total of 336 2000 × 2000 images across all HSR line areas, each of which is composed of R (red), G (green) and B (blue) channels, with 0.5m spatial resolution. There are 221 2000 × 2000 images uniformly collected from Zhengzhou-Xi'an HSR line area and 115 2000 × 2000 images uniformly collected from Beijing-Shanghai HSR line area (the area of yellow color, 1km around the HSR line). The original images covering the Zhengzhou-Xi'an HSR line area and Beijing-Shanghai HSR line area were obtained from Google Earth, which were taken and released by Google Earth on January 21, 2019 and covered 3600 km^2. As shown in Figure 1 and Figure 2, we collected image samples of our dataset uniformly along the HSR line area images to ensure that these samples can cover every typical area as much as possible in HSR lines and manually interpreted for the training, verifying (the blue dots) and testing purpose (the red dots). In our experiment, 22 image tiles were randomly selected as test set and the remaining 199 tiles were used for training and validation for Zhengzhou-Xi'an HSR line building dataset. 80 image tiles were randomly selected as test set and the remaining 35 tiles were used for training and validation for Beijing-Shanghai

HSR line building dataset. Where the validation data were randomly selected from training sets with a ratio of 20%.

For the production of image sample labels, we set up a professional team to fully annotated each image, identifying and providing a polygon footprint for each building using ENVI and ArcGIS software to produce high-quality building vector map. Any visible building footprints were labeled to represent the shape of buildings and adjacent buildings were marked separately as far as possible in our annotation. As the nature of manually interpretation, some small errors are inevitable especially for complex scenes. Consequently, we carried out triple cross checking carefully after the initial annotation to minimize the risk of false judgment. Finally, the corresponding raster label maps, as shown in section of data display, are derived from the building footprint vector maps as ground truth label data, and the building area and non-building area is represented by blue color and white color, respectively.

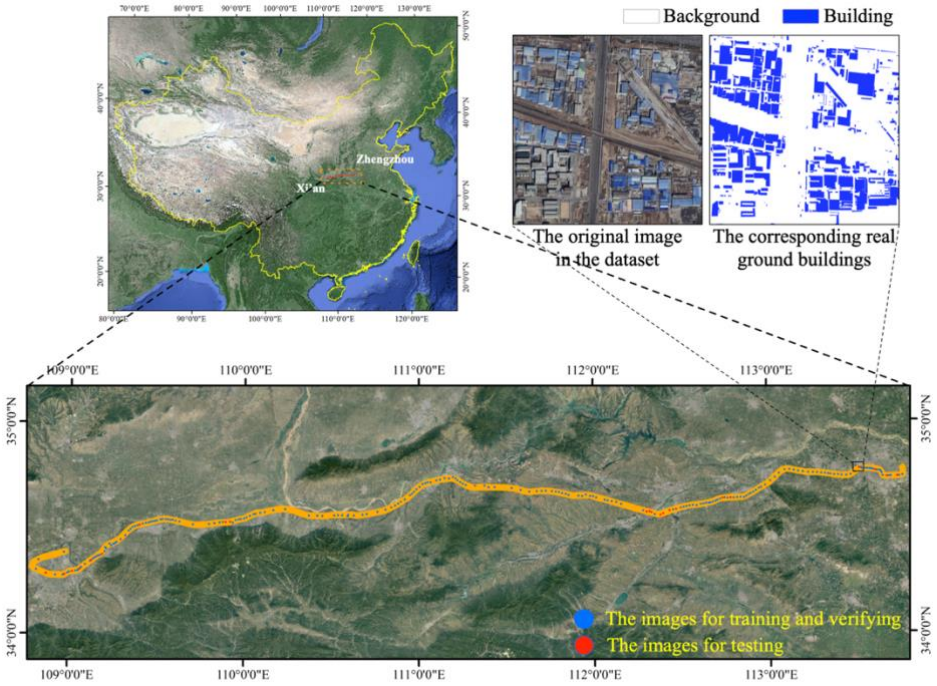

Figure 1: Area covered and the position of sample images by the Zhengzhou-Xi'an HSR line

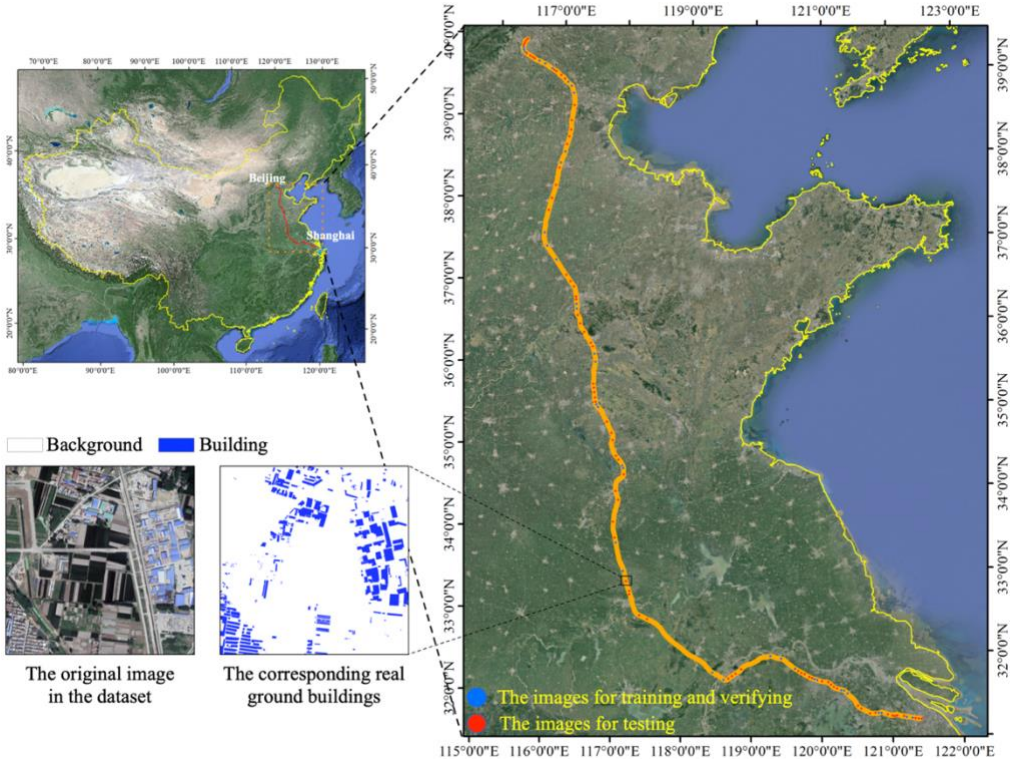

Figure 2: Area covered and the position of sample images by the Beijing-Shanghai HSR line