ABSTRACT

Building is one of the essential features of a topographic map. However, extracting building features using manual digitization or stereo plotting is labor-intensive and time-consuming. Moreover, the condition of settlements in Indonesia is very complex. This problem makes the automation of topographic mapping important, including in the building extraction. One approach that must be optimized for mapping automation is to use computer vision. This research tries to apply Mask R-CNN for building extraction in large-scale topographic mapping. The primary data used in large-scale topographic mapping is the aerial photo, but the very high-resolution satellite image is also used for acceleration. Therefore, Mask R-CNN was applied to the aerial photo and very high-resolution satellite image to understand the output of this algorithm on both data.

As a deep learning network, Mask R-CNN requires at least three stages: data preparation, training and prediction, and accuracy assessment. In training data preparation, the tile size value is 256, and the stride value is 128. Then, in the training process, the parameters follow the default parameters, such as batch size = 1, learning rate = 0.001, weight decay = 0.0001, momentum = 0.9, and trained in 100 epochs. Predictions were carried out on six test images with a minimum detection confidence value of 0.3. Then, after the vectorization process to produce building segments in shapefile format, accuracy assessments were carried out using four evaluation metrics: IoU, precision, recall, and F1 score. Finally, the regularization process was carried out to produce the final output. These processes were carried out on a laptop equipped with an NVIDIA GeForce MX230 GPU with a memory of 4.0 GB. The software used are Global Mapper v18.0, ArcGIS Pro 2.8.0, Anaconda Navigator, Google Colaboratory, and ENVI 5.1.

Based on the training loss graph, it can be concluded that the aerial photo is better to use as a training dataset than the satellite image. Then, in the histogram analysis, inconsistencies occur in the pixel values identified as true positive. Some false positive objects are also segmented because they have clear boundaries, or the case of false negative objects because of the low contrast between the building and its surroundings. In addition, almost all true positive samples have a higher confidence value than false positive objects. The Mask R-CNN can separate adjacent buildings in vectorization results, and the results are close to the ground truth. However, with higher complexity in residential areas, the quality of the segmentation results is lower and further away from the ground truth. Based on the accuracy assessment, the aerial photo gets better accuracy than the satellite image because the true positive, false positive, and false negative values in the aerial photo are better than the satellite image. Finally, the regularization results on the aerial photo are better because they can distinguish small details and match the actual building boundaries.

Key Words: aerial photo, very high-resolution satellite image, building extraction, deep learning, Mask R-CNN

INTISARI

Bangunan adalah salah satu fitur penting dalam peta rupabumi. Meski demikian, ekstraksi fitur bangunan menggunakan digitasi manual atau *stereo plotting* sangat membutuhkan banyak sumberdaya manusia dan waktu. Apalagi, kondisi permukiman di Indonesia sangat kompleks. Masalah ini membuat otomasi pemetaan rupabumi menjadi penting, termasuk dalam ekstraksi bangunan. Salah satu pendekatan yang harus dioptimalkan untuk otomasi pemetaan adalah menggunakan *computer vision*. Penelitian ini mencoba menerapkan Mask R-CNN untuk ekstraksi bangunan dalam pemetaan rupabumi skala besar. Data utama yang digunakan dalam pemetaan rupabumi skala besar adalah foto udara, namun citra satelit resolusi sangat tinggi juga digunakan untuk percepatan. Oleh karena itu, Mask R-CNN diterapkan pada foto udara dan citra satelit resolusi sangat tinggi untuk mengetahui luaran algoritma ini pada kedua data tersebut.

Seperti pada jaringan *deep learning*, Mask R-CNN juga memerlukan paling tidak tiga tahapan: persiapan data, pelatihan dan prediksi, dan uji akurasi. Pada persiapan data latih, nilai *tile size*-nya adalah 256, dan nilai *stride*-nya adalah 128. Kemudian, pada proses pelatihan, parameternya mengikuti parameter *default*, seperti *batch size* = 1, *learning rate* = 0,001, *weight decay* = 0,0001, *momentum* = 0,9, dan dilatih dalam 100 epoch. Prediksi dilakukan pada enam *image* test dengan nilai *detection minimum confidence* adalah 0,3. Kemudian, setelah proses vektorisasi untuk menghasilkan segmen bangunan dalam format shapefile, uji akurasi dilakukan menggunakan empat metrik evaluasi: IoU, *precision*, *recall*, dan *F1 score*. Terakhir, proses regularisasi dilakukan untuk menghasilkan luaran final. Proses-proses tersebut dilakukan pada laptop yang dilengkapi dengan GPU NVIDIA GeForce MX230 dengan memori 4,0 GB. Perangkat lunak yang digunakan diantaranya Global Mapper v18.0, ArcGIS Pro 2.8.0, Anaconda Navigator, Google Colaboratory, dan ENVI 5.1.

Berdasarkan grafik *training loss*, dapat disimpulkan foto udara lebih baik untuk digunakan sebagai dataset latih dibandingkan citra satelit. Kemudian, pada analisis histogram, inkonsistensi terjadi pada nilai piksel yang diidentifikasi sebagai *true positive*. Beberapa objek *false positive* juga tersegmentasi karena memiliki batas yang tegas, atau kasus objek *false negative* karena kontras yang rendah antara bangunan dengan sekitarnya. Selain itu, hampir semua sampel *true positive* memiliki nilai *confidence* yang lebih tinggi dibandingkan objek-objek *false positive*. Pada hasil vektorisasi, Mask R-CNN dapat memisahkan bangunan yang berdekatan dan hasilnya mendekati *ground truth*. Namun dengan kompleksitas yang lebih tinggi pada area permukiman, kualitas hasil segmentasi menjadi lebih rendah dan semakin jauh dari *ground truth*. Berdasarkan uji akurasi, foto udara mendapatkan akurasi yang lebih baik dibandingkan citra satelit karena pada umumnya nilai *true positive*, *false positive*, dan *false negative* pada foto udara lebih baik dibandingkan citra satelit. Terakhir, hasil regularisasi pada foto udara lebih baik karena dapat membedakan detail yang kecil dan sesuai dengan batas bangunan yang sebenarnya.

Kata Kunci: foto udara, citra satelit resolusi sangat tinggi, ekstraksi bangunan, *deep learning*, *Mask R-CNN*