

Intisari

Tutupan lahan merupakan salah satu informasi geospasial yang penting dalam mendukung perencanaan tata ruang, khususnya pada skala menengah hingga besar seperti skala 1:25.000 hingga 1:5.000, dimana keberadaan objek detail seperti jalan dan permukiman menjadi elemen penting dalam proses perencanaan dan pengambilan keputusan. Perubahan tutupan lahan yang berlangsung cepat menuntut data yang memiliki cakupan luas, pembaruan berkala, dan akses terbuka untuk pemantauan yang efisien. Kebutuhan tersebut dapat dipenuhi oleh citra Sentinel-2 yang memiliki resolusi temporal tinggi dan gratis. Di sisi lain, perkembangan metode klasifikasi tutupan lahan saat ini banyak memanfaatkan pendekatan *deep learning* karena kemampuannya dalam mengekstraksi fitur secara otomatis dan menghasilkan akurasi yang lebih baik dibandingkan metode konvensional. Namun, resolusi spasial 10 meter pada citra Sentinel-2 masih menjadi keterbatasan dalam merepresentasikan objek berukuran kecil sehingga detail spasial yang menjadi masukan model *deep learning* menjadi kurang optimal. Oleh karena itu, pendekatan *super-resolution* seperti Real-ESRGAN, DiffuSR, dan CRefDiff dikembangkan untuk meningkatkan detail spasial citra. Penelitian ini bertujuan untuk membandingkan ketiga metode tersebut serta mengevaluasi pengaruhnya terhadap performa klasifikasi tutupan lahan menggunakan CNN 2D berbasis U-Net dengan mekanisme Swin Transformer.

Penelitian dilakukan melalui 3 tahapan utama, yaitu *preprocessing* citra, peningkatan resolusi spasial menggunakan ketiga model *super-resolution*, serta klasifikasi tujuh kelas tutupan lahan, yaitu permukiman, vegetasi tinggi, vegetasi rendah, sawah, badan air, jalan, dan tanah terbuka. Evaluasi kualitas citra *super-resolution* dilakukan menggunakan metrik *Peak Signal-to-Noise Ratio* (PSNR), *Structural Similarity Index Measure* (SSIM), dan *Learned Perceptual Image Patch Similarity* (LPIPS) dengan citra UAV sebagai referensi. Sementara itu, evaluasi hasil klasifikasi dilakukan menggunakan *confusion matrix* untuk memperoleh nilai *Precision*, *Recall*, *F1-score*, *Intersection over Union* (IoU), dan *Accuracy*. Terdapat tujuh objek yang diklasifikasi, yaitu permukiman, vegetasi tinggi, vegetasi rendah, sawah, badan air, jalan, dan tanah terbuka.

Hasil penelitian menunjukkan bahwa setiap model *super-resolution* memiliki karakteristik berbeda dalam meningkatkan resolusi spasial citra. Secara kuantitatif, Real-ESRGAN menunjukkan akurasi dan efisiensi terbaik dengan PSNR 21,07 dB, SSIM 0,558, dan LPIPS 0,224. DiffuSR menunjukkan performa menengah dengan PSNR 20,68 dB, SSIM 0,398, dan LPIPS 0,338, sedangkan CRefDiff menghasilkan PSNR $18,97 \pm 0,84$ dB, SSIM $0,338 \pm 0,044$, dan LPIPS $0,414 \pm 0,079$. Meskipun metrik kuantitatifnya lebih rendah, CRefDiff menghasilkan detail spasial yang lebih tajam secara visual melalui pendekatan *reference-based diffusion*. Integrasi citra hasil *super-resolution* ke dalam model CNN 2D berbasis U-Net meningkatkan akurasi klasifikasi dibandingkan citra Sentinel-2 resolusi asli 10 m. Akurasi dan efisiensi terbaik diperoleh pada CRefDiff dengan *Accuracy* 0,8182, *F1-Score* 0,7838, dan *Mean IoU* 0,6596, terutama dalam meningkatkan segmentasi objek kecil seperti permukiman dan jalan. DiffuSR menunjukkan performa moderat dengan akurasi 0,7927 dan *Mean IoU* 0,4437, sementara Real-ESRGAN menghasilkan akurasi 0,6671 dan *Mean IoU* 0,1662. Secara keseluruhan, penerapan *super-resolution* terbukti meningkatkan akurasi klasifikasi tutupan lahan berbasis *deep learning* pada citra Sentinel-2 serta mendukung pemutakhiran peta dasar secara lebih detail dan efisien.

Kata kunci: tutupan lahan, *deep learning*, *super-resolution*, Sentinel-2, CNN 2D

Abstract

Land cover is an important geospatial information component for supporting spatial planning, particularly at medium to large map scales such as 1:25,000 to 1:5,000, where detailed objects such as roads and buildings become essential elements in planning and decision-making processes. Rapid land cover changes require data with wide coverage, frequent updates, and open access to enable efficient monitoring. These requirements can be addressed by Sentinel-2 imagery, which provides high temporal resolution and free accessibility. On the other hand, recent developments in land cover classification methods have increasingly utilized deep learning approaches due to their capability to automatically extract features and achieve higher accuracy compared to conventional methods. However, the 10-meter spatial resolution of Sentinel-2 imagery remains a limitation in representing small objects, resulting in less optimal spatial detail as input for deep learning models. Therefore, super-resolution approaches such as Real-ESRGAN, DiffuSR, and CRefDiff have been developed to enhance spatial image details. This study aims to compare these three methods and evaluate their influence on land cover classification performance using a 2D CNN based on U-Net architecture integrated with a Swin Transformer mechanism.

The research was conducted through three main stages: image preprocessing, spatial resolution enhancement using the three super-resolution models, and classification of seven land cover classes, namely built-up areas, high vegetation, low vegetation, paddy fields, water bodies, roads, and bare land. The quality of super-resolution images was evaluated using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS), with UAV imagery used as reference data. Meanwhile, classification performance was evaluated using a confusion matrix to obtain Precision, Recall, F1-score, Intersection over Union (IoU), and overall Accuracy.

The results indicate that each super-resolution model demonstrates different characteristics in enhancing spatial image resolution. Quantitatively, Real-ESRGAN shows the best accuracy and efficiency with a PSNR of 21.07 dB, SSIM of 0.558, and LPIPS of 0.224. DiffuSR demonstrates moderate performance with a PSNR of 20.68 dB, SSIM of 0.398, and LPIPS of 0.338, while CRefDiff produces a PSNR of 18.97 ± 0.84 dB, SSIM of 0.338 ± 0.044 , and LPIPS of 0.414 ± 0.079 . Despite having lower quantitative metrics, CRefDiff produces visually sharper spatial details through its reference-based diffusion approach. The integration of super-resolution imagery into the 2D CNN U-Net model improves classification performance compared to the original 10-meter Sentinel-2 imagery. The best classification accuracy and effectiveness are achieved by CRefDiff with an Accuracy of 0.8182, F1-Score of 0.7838, and Mean IoU of 0.6596, particularly in improving segmentation of small objects such as buildings and roads. DiffuSR shows moderate performance with an accuracy of 0.7927 and Mean IoU of 0.4437, while Real-ESRGAN produces an accuracy of 0.6671 and Mean IoU of 0.1662. Overall, the implementation of super-resolution methods has been proven to improve deep learning-based land cover classification accuracy on Sentinel-2 imagery and supports more detailed and efficient base map updating.

Keywords: land cover, deep learning, super-resolution, Sentinel-2, CNN 2D