SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika

https://ejurnal.itats.ac.id/snestik dan https://snestik.itats.ac.id

Informasi Pelaksanaan :
SNESTIK V - Surabaya, 26 April 2025
Fakultas Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:
DOI : 10.31284/p.snestik.2025.6874

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Klasifikasi Bangunan secara Otomatis Menggunakan Pembelajaran Mendalam dari Gambar Street-View

Ryan Gading A[1], Mahameru A[2], Anggita E[3], M. Andhika K[4], Dian Puspita H[5]

Institut Teknologi Adhi Tama Surabaya[1,2,3,4,5]

e-mail: ryangabdullah99@gmail.com

ABSTRACT

Urban population density mapping or urban utility planning requires a classification map based on individual buildings that are considered much more informative. The goal of this research is to determine how to extract the fine-grained boundaries of individual buildings from a street-view dataset. This paper proposes a general framework for classifying individual building functionality using a deep learning approach. The proposed method is based on a Convolutional Neural Network (CNN) that classifies facade structures from street view images, such as Street-View images. From the experiments conducted, the CNN classifier with the ResNet architecture was able to classify the Street-View data group with an accuracy value of 86.79%. Research shows that overfitting still occurs, necessitating large training data and an increased epoch value during dataset training. We construct a dataset to train and evaluate the CNN classifier. Furthermore, the method is applied to generate a building classification map at the urban area scale.

Keywords: *Deep Learning; CNN; ResNet; Street View Images.*

ABSTRAK

Pemetaan kepadatan populasi perkotaan atau perencanaan utilitas perkotaan memerlukan peta klasifikasi berdasarkan bangunan individual yang dianggap jauh lebih informatif. Tujuan penelitian ini adalah untuk menentukan cara mengekstrak batas-batas terperinci bangunan individual dari kumpulan data tampilan jalan. Makalah ini mengusulkan kerangka kerja umum untuk mengklasifikasikan fungsionalitas bangunan individual menggunakan pendekatan pembelajaran mendalam. Metode yang diusulkan didasarkan pada Jaringan Syaraf Tiruan Konvolusional (CNN) yang mengklasifikasikan struktur fasad dari gambar tampilan jalan, seperti gambar StreetView. Dalam penelitian ini membangun kumpulan data untuk melatih dan mengevaluasi pengklasifikasi CNN. Dari percobaan yang dilakukan pengklasifikasi CNN dengan arsitektur ResNet mampu mengklasifikasikan kelompok data Street-View dengan nilai akurasi 86,79%. Dalam

penelitian yang telah dilakukan masih terjadi overfitting, maka dibutuhkan training data dalam jumlah yang besar serta meningkatkan nilai epoch pada kegiatan traning kelompok data. Lebih jauh, metode ini diterapkan untuk menghasilkan peta klasifikasi bangunan pada skala wilayah perkotaan.

Kata kunci: Pembelajaran Mendalam; CNN; ResNet; Gambar Fasad Jalan.

PENDAHULUAN

Munculnya pembelajaran mendalam (*deep learning*) telah memicu transformasi besar dalam bidang *computer vision*, yang memungkinkan analisis gambar yang sangat akurat dan sepenuhnya otomatis di berbagai aplikasi. Di garis depan kemajuan ini adalah Jaringan Syaraf Tiruan Konvolusional, *Convolutional Neural Networks* (CNN), yang telah memperkuat perannya sebagai arsitektur dasar untuk berbagai tugas penting. Implementasinya tidak terbatas pada klasifikasi gambar, deteksi objek, dan segmentasi semantik saja. Di antara model-model inovatif CNN, arsitektur ResNet, atau Jaringan Residual pada CNN mampu membedakan dirinya secara efektif untuk mengatasi masalah-masalah yang penting seperti masalah gradien yang menghilang (*the vanishing gradient problem*). Kemampuan ini memungkinkan algoritma CNN melakukan kegiatan pelatihan jaringan jauh lebih dalam, yang pada akhirnya kinerja model pengklasifikasi menjadi lebih baik dalam berbagai domain yang menantang [1][2][3].

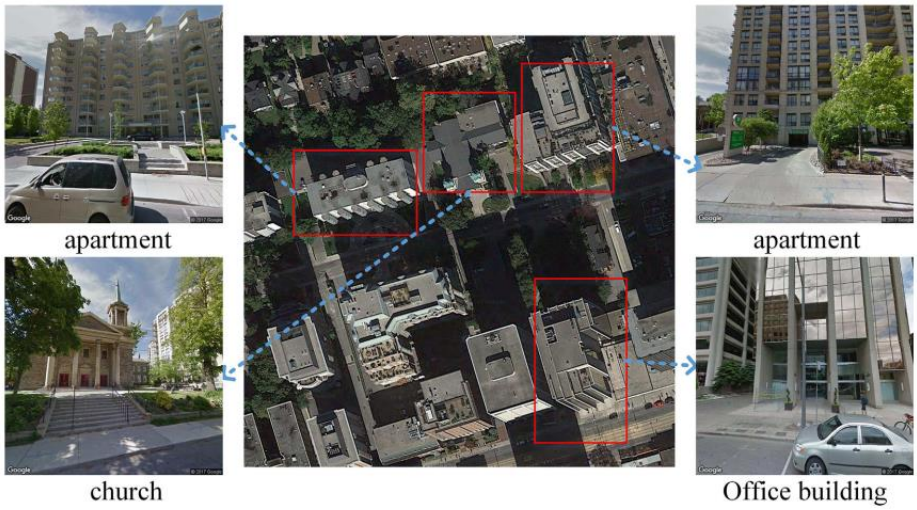
CNN telah banyak diimplementasikan untuk analisis data perkotaan, kegiatan penelitian yang mengimplementasikan CNN didorong oleh kebutuhan mendesak akan perencanaan kota yang lebih efektif, navigasi yang lebih baik, dan penilaian real estat yang terinformasi. Sumber daya yang sangat berharga dalam konteks ini adalah citra tampilan jalan (data citra street-views), yang menawarkan banyak data citra perkotaan beresolusi tinggi, yang memungkinkan para peneliti dan perencana kota untuk memperoleh wawasan penting tentang berbagai tugas, seperti klasifikasi bangunan. Tugas ini melibatkan identifikasi dan pengelompokan berbagai struktur berdasarkan karakteristik visual dan arsitekturalnya. Meskipun demikian, kelompok data street-view menghadirkan tantangan unik, termasuk variabilitas sudut, kondisi pencahayaan yang tidak konsisten, dan penghalang fisik, yang semuanya mempersulit metode klasifikasi citra tradisional [4].

Kelompok data citra tampilan jalan mengungkap lebih banyak detail dari berbagai jenis bangunan daripada bidang atap yang dihasilkan dari peta GIS. Menggunakan kumpulan data citra tampilan jalan dengan bangunan untuk melatih algoritma pembelajaran mendalam Jaringan Syaraf Konvolusional (CNN) pada kegiatan klasifikasi di area yang luas, merupakan kehandalan CNN yang telah menunjukkan kemampuannya dalam tugas serupa. Kegiatan klasifikasi penggunaan lahan pada tingkat bangunan individual bukanlah tugas yang mudah. Peta klasifikasi seperti itu hanya dapat diperoleh melalui basis data kadaster kota, tidak dapat diakses atau terkadang bahkan tidak ada. Memperbarui basis data tersebut tanpa metode otomatis dapat sangat padat karya. Oleh karena itu, klasifikasi bangunan secara otomatis diperlukan dan dapat bermanfaat untuk aplikasi yang terkait dengan perencanaan kota [5].

Makalah ini membahas pemanfaatan CNN berarsitektur ResNet khususnya untuk tugas klasifikasi bangunan menggunakan citra street-view. Dengan memanfaatkan kemampuan ekstraksi fitur canggih yang melekat pada ResNet, studi ini bertujuan untuk meningkatkan akurasi dan generalisasi model yang dirancang untuk klasifikasi bangunan. Kontribusi utama dari penelitian ini meliputi pembuatan kerangka kerja berbasis ResNet yang dioptimalkan untuk kompleksitas data street-view yang memiliki derau dan bersifat heterogen. Lebih jauh, studi ini melakukan evaluasi menyeluruh terhadap kinerja kerangka kerja di berbagai kumpulan data street-view. Implikasi dari temuan ini sangat besar, menawarkan potensi kemajuan dalam strategi pembangunan perkotaan, peningkatan dalam sistem navigasi, dan otomatisasi analisis data arsitektur [6][7].

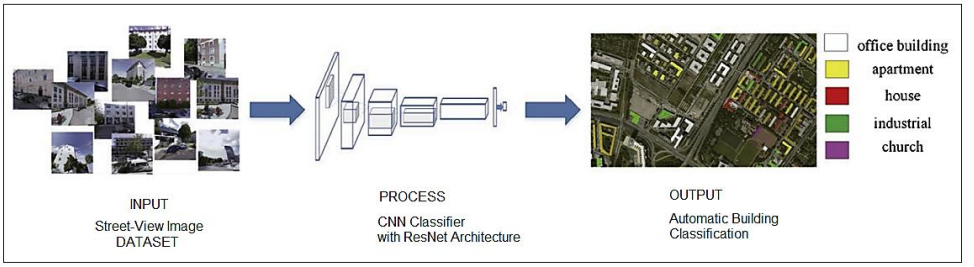
METODE

Kelompok data yang digunakan pada penelitian ini merupakan kelompok data bertipe gambar. Data yang digunakan untuk proses penelitian adalah data terkait Street-View atau "street data" yang memiliki beberapa macam kategori serta dapat diakses pada alamat URL berikut ini: https://www.kaggle.com/datasets/mikhailma/house-rooms-streets-image-dataset/data. Dalam kelompok data tersebut terdapat sembilan kategori yang terdiri dari "apartment", "church", "garage", "house", "industrial", "office building", "retail", dan "roofs" seperti tampak pada Gambar 1 dibawah ini. Kelompok data tersebut sejumlah data gambar jalan yang terdiri dari 17.154 record data gambar. Adanya outlier dalam kelompok data yang digunakan sangat memengaruhi hasil klasifikasi. Maka dilakukan seleksi gambar street view terlebih dahulu yang sesuai dengan sembilan kategori yaitu "apartment", "church", "garage", "house", "industrial", "office building", "retail", dan "roofs". Hanya gambar yang termasuk dalam kategori yang disebutkan di atas yang disimpan untuk klasifikasi lanjutan.



Gambar 1. Kelompok data street view dengan sembilan kategori.

Selanjutnya dilakukan kegiatan melatih pengklasifikasi bangunan CNN-ResNet, pertama-tama diawali dengan membangun kelompok data tolok ukur tampilan jalan yang sesuai, yang berisi total 17.154 data gambar dari sembilan kelas, yaitu "apartment", "church", "garage", "house", "industrial", "office building", "retail", dan "roofs". Pada penelitian ini alur proses automasi kegiatan klasifikasi bangunan dapat diilustrasikan pada Gamber 2. Diawali dengan masukan kelompok data, kemudian berlanjut pada proses klasifikasi. Umumnya CNN yang telah dilatih sebelumnya pada kumpulan data besar dapat disesuaikan dengan baik untuk tugas baru lainnya dengan kumpulan data skala kecil, karena fitur tingkat rendah seperti sudut dan tepi yang dihasilkan oleh lapisan CNN sebelumnya bersifat umum dalam gambar yang berbeda. Representasi gambar tingkat tinggi yang diekstraksi oleh lapisan posterior bergantung pada tugas yang berbeda. Oleh karena itu, menyempurnakan lapisan CNN yang telah dilatih sebelumnya dengan kumpulan data baru telah terbukti menjadi cara yang efisien untuk adaptasi CNN ke tugas pelatihan baru. Untuk lebih meningkatkan kinerja klasifikasi, gambar tampilan jalan untuk setiap contoh bangunan diklasifikasikan, dan kelas bangunan dapat diperoleh dalam tingkat keputusan.

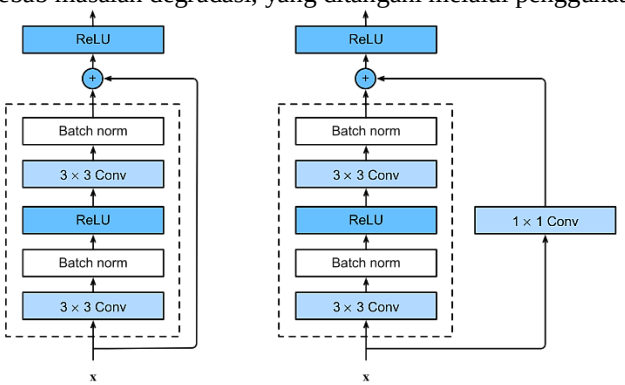


Gambar 2. Alur proses automasi klasifikasi gedung dari gambar Street-View

Koneksi Residual dalam CNN-ResNet, dalam sebuah model jaringan saraf multilayer, dapat diilustrasikan subjaringan dengan sejumlah lapisan bertumpuk. Nyatakan fungsi dasar yang dilakukan oleh subjaringan ini sebagai $H(x)$, dimana x merupakan sebuah masukan dari subjaringan. Pembelajaran residual mengubah parameterisasi dalam subjaringan ini dan membiarkan lapisan parameter mewakili "fungsi residual" $F(x) = H(x) - x$. Luaran y dari sub jaringan tersebut kemudian direpresentasikan sebagai :

$$y = F(x) + x \quad (1)$$

Operasi dari "$+x$" pada persamaan (1) diimplementasikan melalui sebuah "skip connection" yang melakukan pemetaan identitas untuk menghubungkan masukan subnetwork dengan luaran-nya. Koneksi ini disebut sebagai "residual connection". Fungsi $F(x)$ pada persamaan (1) sering kali direpresentasikan oleh perkalian matriks yang diselingi dengan fungsi aktivasi dan operasi normalisasi (misalnya, normalisasi batch atau normalisasi layer). Secara keseluruhan, salah satu subnetwork ini disebut sebagai "blok residual" seperti tampak pada Gambar 3. Blok Diagram ResNet pada penelitian S. Hochreiter, tahun 1998 [3] Sebuah deep residual network dibangun dengan menumpuk blok-blok ini secara sederhana. Rumusan pembelajaran residual memberikan manfaat tambahan untuk mengurangi masalah gradien yang menghilang sampai batas tertentu. Namun, penting untuk mengakui bahwa masalah gradien yang menghilang bukanlah akar penyebab masalah degradasi, yang ditangani melalui penggunaan normalisasi.



Gambar 3. Blok Diagram ResNet

Komponen utama dalam ResNet terdiri dari lima bagian, yang pertama adalah Lapisan Konvolusional. ResNet menggunakan beberapa lapisan konvolusional untuk mengekstraksi fitur dari gambar input. Setiap operasi konvolusional melibatkan penerapan filter (atau kernel) ke data input, menghasilkan peta fitur yang menyoroti aspek penting dari gambar. Proses ini diulang di beberapa lapisan untuk menangkap fitur yang semakin abstrak. Bagian kedua adalah Blok Residual atau inovasi inti ResNet, yang mencakup koneksi lewati yang melewati satu atau beberapa lapisan. Desain ini memungkinkan input ditambahkan langsung ke output blok, memfasilitasi aliran gradien yang lebih baik selama backpropagation dan mengurangi masalah yang terkait dengan

gradien yang menghilang. Setiap blok residual biasanya terdiri dari dua atau tiga lapisan konvolusional diikuti oleh koneksi pintasan yang menambahkan input asli ke output yang ditransformasikan.

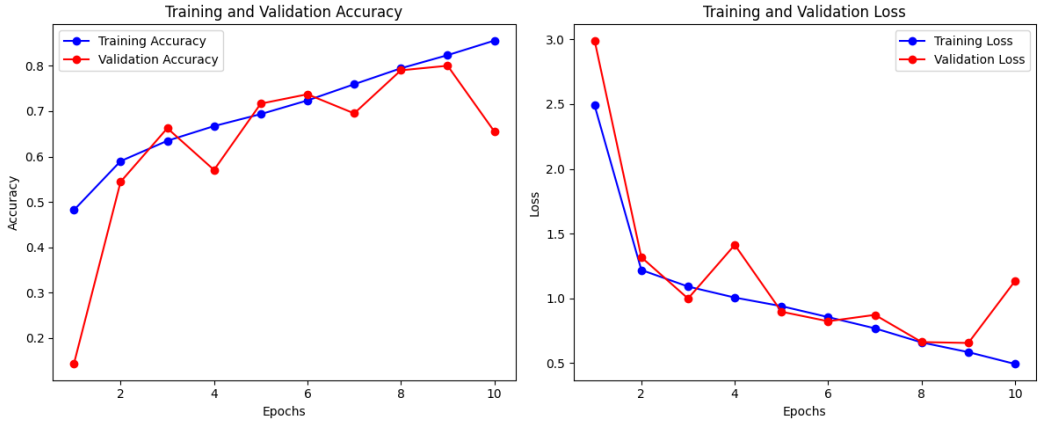
Berikutnya bagian Fungsi Aktivasi, setelah setiap lapisan konvolusional, fungsi aktivasi, umumnya ReLU (Rectified Linear Unit), diterapkan untuk memperkenalkan non-linearitas ke dalam model. Langkah ini penting untuk memungkinkan jaringan mempelajari pola kompleks dalam data. Bagian yang keempat adalah Lapisan Penggabungan yang digunakan untuk mengurangi peta fitur, mengurangi dimensinya sambil mempertahankan informasi penting. Proses ini membantu mengurangi beban komputasi dan mengendalikan overfitting dengan meringkas informasi fitur. Bagian yang terakhir adalah Lapisan yang Terhubung Sepenuhnya di akhir jaringan, lapisan yang terhubung sepenuhnya mengambil keluaran yang diratakan dari lapisan penggabungan terakhir dan menghasilkan klasifikasi akhir berdasarkan fitur yang dipelajari. Tahap ini mengintegrasikan semua fitur yang dipelajari untuk membuat prediksi tentang data gambar masukan.

Untuk kegiatan berikutnya dilakukan pengukuran kinerja algoritma pengklasifikasi CNN-ResNet menggunakan matriks confusion. Matriks confusion merupakan tabel yang menunjukkan seberapa baik kinerja algoritme klasifikasi dengan membandingkan nilai prediksi dengan nilai aktual. Ini adalah metode visualisasi yang membantu mengidentifikasi kelas data mana yang paling sering salah diklasifikasikan. Selanjutnya menunjukkan matriks confusion yang dinormalkan dari semua jaringan terlatih yang dievaluasi oleh data uji, dan nilai Accuracy terkait dari delapan kelas bangunan, masing-masing.

HASIL DAN PEMBAHASAN

Memahami training loss dan validation loss sangat penting untuk mengevaluasi pengklasifkasi CNN selama proses pelatihan. Kedua metrik tersebut memberikan informasi tentang seberapa baik model belajar dari data pelatihan dan menggeneralisasikannya ke data yang tidak terlihat. Training loss merupakan alat yang mengukur kesalahan model pada kelompok data latih. Pengukuran ini dihitung setelah setiap batch selama pelatihan kelompok data dan mencerminkan seberapa baik model tersebut sesuai dengan data pelatihan yang digunakan. Saat pelatihan berlangsung, kerugian pelatihan biasanya berkurang pada saat model belajar untuk meminimalkan kesalahan pada kelompok data latih. Selain training loss pengukuran yang lain dalah validation loss merupakan pengukur kesalahan pada kelompok data validasi terpisah yang belum terlihat oleh model selama kegiatan pelatihan. Pengukur ini dihitung setelah setiap periode (yaitu, setelah satu lintasan lengkap melalui data pelatihan). Tujuan pelacakan kerugian validasi adalah untuk menilai seberapa baik model dapat menggeneralisasi ke data baru yang tidak terlihat. Pada penelitian yang telah dilakukan ini mengunakan training loss dan validation loss untuk melihat kinerja pengklasifikasi CNN.

Pada Gambar 4(a) dapat dilihat akurasi validasi lebih tinggi daripada akurasi pelatihan, ini bukanlah perilaku ideal untuk pengklasifikasi CNN. Hal ini dapat disebabkan oleh model tidak dilatih dan diuji dengan data yang tepat. Sehingga indikasi overfitting terjadi pada percobaan yang dilakukan. Jika akurasi validasi jauh tertinggal dari akurasi pelatihan, hal ini menunjukkan overfitting. Mirip dengan akurasi validasi, kesenjangan yang signifikan dalam akurasi pelatihan dapat mengindikasikan overfitting. Jika akurasi pelatihan jauh lebih tinggi daripada akurasi validasi, hal ini menyiratkan overfitting. Akurasi adalah metrik yang digunakan untuk mengevaluasi kinerja model prediktif. Akurasi ditentukan dengan mengambil jumlah total prediksi yang benar yang dibuat oleh model dan membagi angka tersebut dengan jumlah keseluruhan prediksi yang dicoba. Perhitungan ini memberikan rasio yang jelas yang mencerminkan seberapa sering prediksi model selaras dengan hasil aktual, yang memberikan wawasan tentang efektivitasnya.



(a) (b)

Gambar 4 (a) Training and Validation Accuracy dan (b) Training and Validation Loss

Pada Gambar 4(b) adalah gambar grafik training loss dan validation loss yang dihasilkan dari percobaan yang telah dilakukan. Dari grafik tersebut dapat diartikan bahwa validation loss lebih besar daripada training loss, seperti yang terlihat pada gambar. Ini biasanya menunjukkan bahwa model tersebut overfitting, dan tidak dapat digeneralisasi pada data baru. Secara khusus, model tersebut berkinerja baik pada data pelatihan tetapi buruk pada data baru dalam set validasi. Pada suatu titik, kerugian validasi menurun tetapi mulai meningkat lagi. Alasan penting untuk kejadian ini adalah bahwa model tersebut mungkin terlalu rumit untuk data atau, model tersebut dilatih untuk jangka waktu yang lama. Dalam kasus ini, pelatihan dapat dihentikan saat kerugiannya rendah dan stabil, ini biasanya dikenal sebagai penghentian dini. Penghentian dini adalah salah satu dari banyak pendekatan yang digunakan untuk mencegah overfitting.

Percobaan yang telah dilakukan pengklasifikasi CNN dengan arsitektur ResNet mampu untuk mengklasifikasikan kelompok data gambar Street-View dengan nilai accuracy: 0.8679, kemudian nilai loss: 0.4593, nilai val_accuracy: 0.6554, dan nilai val_loss: 1.1342. Dengan hasil tersebut masih dibutuhkan penyesuaian kembali agar nilai accuracy dari model pengklasifikasi CNN-ResNet dapat ditingkatkan.

KESIMPULAN

Dalam penelitian ini, disajikan kerangka kerja untuk memprediksi klasifikasi bangunan dengan pendekatan deep learning, dapat menghasilkan peta yang lebih informatif untuk skala wilayah dan kota. Dengan pendekatan ini, akurasi yang tinggi dapat dicapai untuk klasifikasi penggunaan lahan masing-masing bangunan. Kegiatan klasifikasi peta bangunan ini akan membantu mendapatkan wawasan tentang wilayah perkotaan, dan memiliki potensi untuk banyak analisis perkotaan yang inovatif, misalnya pemetaan kepadatan populasi perkotaan beresolusi sangat tinggi, pemahaman struktur sosial perkotaan, analisis struktur ekonomi kota, dan perencanaan perkotaan umum. Dalam penelitian yang telah dilakukan masih terjadi overfitting, maka dibutuhkan training data dalam jumlah yang besar serta meningkatkan nilai epoch pada kegiatan traning kelompok data. Untuk pekerjaan di masa mendatang, guna meningkatkan kinerja klasifikasi, dapat dilengkapi dengan fitur deskripsi teks yang terkait dengan gambar media sosial dan informasi teks yang ditampilkan dalam gambar.

DAFTAR PUSTAKA

[1] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J. Big Data*, vol. 8, no. 1, Dec. 2021, doi:

10.1186/s40537-021-00444-8.

[2] H. Alaeddine and M. Jihene, "Deep Residual Network in Network," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–9, 2021, doi: 10.1155/2021/6659083.

[3] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Int. J. Uncertainty, Fuzziness Knowldege-Based Syst.*, vol. 6, no. 2, pp. 107–116, 1998, doi: 10.1142/S0218488598000094.

[4] J. Kang, M. Körner, Y. Wang, H. Taubenböck, and X. X. Zhu, "Building instance classification using street view images," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 44–59, 2018, doi: 10.1016/j.isprsjprs.2018.02.006.

[5] R. Han, X. Fan, and J. Liu, "EUNet: Edge-UNet for Accurate Building Extraction and Edge Emphasis in Gaofen-7 Images," *Remote Sens.*, vol. 16, no. 13, pp. 1–21, 2024, doi: 10.3390/rs16132397.

[6] Y. Zhao, X. Zhang, W. Feng, and J. Xu, "Deep Learning Classification by ResNet-18 Based on the Real Spectral Dataset from Multispectral Remote Sensing Images," 2022.

[7] Y. Wang, *Research on Image Classification Based on ResNet*, no. Iciaai 2023. Atlantis Press International BV. doi: 10.2991/978-94-6463-300-9.

Halaman ini sengaja dikosongkan