

LAPORAN PBL DEEP LEARNING

Menerapkan Model Pengenalan Gambar VGG-19 untuk Analisis Segmentasi Pemandangan Kota

Armanda Tantra Kumara¹⁾, Taufiqurrahman Syah Effendi²⁾, Siti Nur Azizah³⁾, Hanna Septiani⁴⁾.

Program Studi Sains Data, Jurusan Sains, Institut Teknologi Sumatera

Email : armanda.120450037@student.itera.ac.id,
taufiqurrahman.120450063@student.itera.ac.id, siti.120450063@student.itera.ac.id,
hanna.120450064@student.itera.ac.id

Abstrak

Pengenalan gambar dan analisis pemandangan kota merupakan topik penting dalam pengembangan perkotaan modern yang cerdas dan berkelanjutan. Dalam konteks ini, penerapan Model Pengenalan Gambar VGG-19 telah menarik perhatian sebagai alat yang kuat untuk mengidentifikasi objek, fitur, dan karakteristik dalam pemandangan kota. Model VGG-19, yang dikenal karena kompleksitas dan akurasi tinggi, memungkinkan analisis yang mendalam dan tepat terhadap gambar-gambar kota. Dalam proyek ini kita menerapkan Model Pengenalan Gambar VGG-19 dalam analisis pemandangan kota, mengeksplorasi cara model ini dapat digunakan dalam berbagai aplikasi, seperti pemantauan lalu lintas, identifikasi bangunan, objek, alam, langit, manusia, kendaraan, dan perencanaan perkotaan yang cerdas. Dengan menggabungkan kecerdasan buatan dan visi komputer, penelitian ini bertujuan untuk memahami potensi model VGG-19 dalam meningkatkan pemahaman dan pengelolaan lingkungan perkotaan.

Kata kunci: VGG-19, pengenalan gambar, analisis pemandangan kota.

BAB I

Pendahuluan

1.1 Latar Belakang Masalah

Pengenalan gambar dan analisis perkotaan menjadi salah satu bidang penelitian yang semakin berkembang dengan cepat dalam era digital saat ini. Segmentasi gambar menjadi pilihan yang tepat untuk melakukan analisis dalam suatu data gambar. Segmentasi merupakan sebuah proses identifikasi objek dan pemisahan area tertentu dalam gambar menjadi bagian atau kelas yang lebih jelas. Segmentasi tersebut bertujuan untuk melihat dan mengidentifikasi area yang relevan dalam gambar. Mengingat peran teknologi yang semakin berkembang, teknologi dapat membantu dalam analisis pengembangan perkotaan sehingga dapat memahami aspek dan dinamika urban secara lebih baik. Analisis pemandangan kota memiliki relevansi yang semakin meningkat dalam konteks perencanaan perkotaan yang berkelanjutan. Salah satu terobosan teknologi yang menonjol adalah penggunaan model pengenalan gambar menggunakan VGG-19, yang dimana model ini memiliki kemampuan dalam mengenali unsur-unsur visual. Menggunakan teknologi pengenalan gambar seperti VGG-19 memungkinkan kita untuk secara otomatis mengidentifikasi dan mengklasifikasikan elemen-elemen suatu gambar dengan tingkat akurasi yang tinggi, yang selanjutnya dapat digunakan dalam berbagai aplikasi.

Penerapan teknologi seperti model VGG-19 dalam analisis pemandangan kota membawa potensi besar untuk menghasilkan data berharga yang dapat digunakan oleh pihak yang membutuhkan seperti peneliti, arsitek dan lain-lain. VGG-19 yang dikembangkan oleh tim Visual Geometry Group (VGG) di Universitas Oxford, telah berhasil mendemonstrasikan kemampuannya dalam pengenalan gambar dan pengklasifikasian dengan tingkat akurasi yang sangat tinggi. Model ini terdiri dari 19 lapisan konvolusi dan mampu memahami ciri-ciri kompleks dalam gambar, yang memungkinkan kita untuk melakukan analisis pemandangan kota dengan lebih mendalam dan tepat. Dalam penelitian proyek yang kami lakukan, akan dibahas langkah-langkah teknis yang diperlukan untuk menerapkan model VGG-19 sesuai dengan dataset yang digunakan yaitu pemandangan kota (Cityscape), diawali dengan pengumpulan gambar, preprocessing, hingga ekstraksi fitur.

Eksplorasi studi kasus menunjukkan model VGG-19 telah digunakan dalam menganalisis pemandangan kota di beberapa kota besar. Hasil analisis ini diharapkan dapat memberikan

wawasan dan pemahaman secara mendalam tentang penerapan teknologi pengenalan gambar dalam konteks perkotaan yang berkelanjutan.

1.2 Rumusan masalah

Rumusan masalah adalah langkah penting dalam penelitian yang membantu merinci permasalahan yang akan dipecahkan melalui penelitian tersebut. Berikut adalah beberapa rumusan masalah yang mungkin relevan untuk disertakan dalam konteks penelitian kami:

1. Bagaimana menerapkan model pengenalan gambar VGG-19 untuk menganalisis pemandangan kota secara efisien?
2. Bagaimana variasi cahaya, perspektif, dan kompleksitas struktur perkotaan mempengaruhi performa model VGG-19 dalam menganalisis pemandangan kota?
3. Apakah terdapat kendala atau batasan dalam penerapan model VGG-19 untuk analisis pemandangan kota, dan bagaimana dapat mengatasi kendala tersebut

1.3 Tujuan Penelitian

Berdasarkan perumusan masalah penelitian yang telah dijelaskan di atas, tujuan penelitian dan manfaat penelitian disusun sebagai berikut:

1. Mengimplementasikan Model Pengenalan Gambar VGG-19
2. Menganalisis Performa Model VGG-19 dalam Konteks Pemandangan Kota
3. Menganalisis Kendala dan Batasan dalam Implementasi Model
4. Memberikan Rekomendasi untuk Pengembangan Lebih Lanjut

1.4 Manfaat Penelitian

Manfaat yang diharapkan penelitian dalam pengenalan gambar dalam analisis pemandangan kota menggunakan VGG-19 sebagai berikut:

1. Pengembangan Metode Analisis Pemandangan Kota yang Efisien
2. Peningkatan Akurasi dalam Pengenalan dan Interpretasi Pemandangan
3. Mendorong Inovasi dalam Analisis Visual untuk Pengembangan Kota

BAB II

Tinjauan Pustaka

2.1 Transfer Learning

Pada tahun 2010, Pan et al mengusulkan konsep belajar pengetahuan yang tidak diketahui melalui pengetahuan yang ada yang disebut Transfer Learning. Konsep inti dari pembelajaran ini adalah untuk menemukan kesamaan antara pengetahuan yang ada dan pengetahuan yang belum diketahui. Beberapa domain pengetahuan terlalu abstrak untuk dipelajari, menghasilkan biaya belajar keseluruhan yang tinggi. Oleh karena itu, menggunakan pengetahuan yang ada untuk membantu pembelajaran menjadi penting [10].

Konsep inti dari transfer learning adalah bagaimana menemukan relevansi antara yang diketahui dan yang tidak diketahui dan mempelajari pengetahuan baru. Dalam transfer learning, pengetahuan yang ada biasanya disebut Domain Sumber, dan pengetahuan yang tidak diketahui disebut Domain Target. Pembelajaran ini terutama mempelajari bagaimana memigrasikan pengetahuan dari Domain Sumber ke Domain Target. Di bidang pembelajaran mesin, transfer learning berfokus pada penerapan pengetahuan yang ada ke yang tidak diketahui melalui model yang ditetapkan.

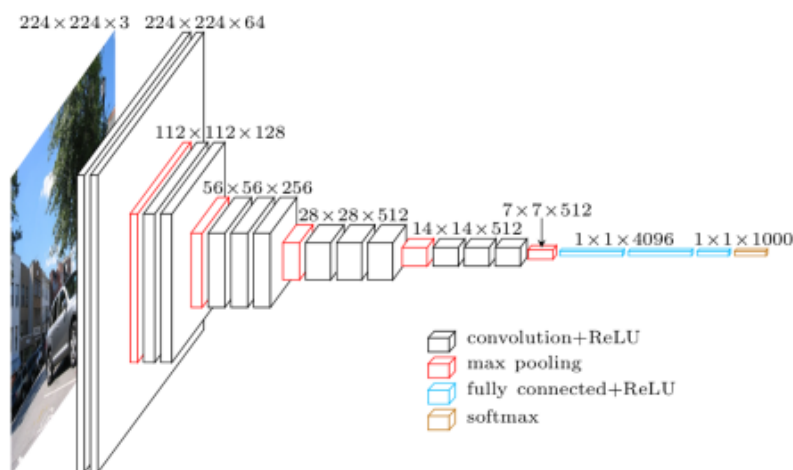
2.2 Deep Learning

Merupakan bagian dari metode pembelajaran mesin berdasarkan jaringan saraf tiruan dengan pembelajaran representasi. Kata sifat "dalam" mengacu pada penggunaan banyak lapisan dalam jaringan. Metode yang digunakan dapat berupa pengawasan, semi-pengawasan, atau tanpa pengawasan[11]. arsitektur seperti jaringan saraf dalam, jaringan kepercayaan mendalam, jaringan saraf berulang, jaringan saraf konvolusional dan transformator telah diterapkan pada berbagai bidang termasuk visi komputer, pengenalan suara, pemrosesan bahasa alami, terjemahan mesin, bioinformatika, desain obat, analisis citra medis, ilmu iklim, inspeksi material dan program permainan papan, yang menghasilkan hasil yang sebanding dan dalam beberapa kasus melampaui kinerja ahli manusia [12][13][14]. Dalam deep learning, metode CNN atau Convolutional Neural Network sangatlah bagus dalam menemukan fitur yang baik pada citra ke lapisan berikutnya untuk membentuk hipotesis nonlinier yang dapat meningkatkan kompleksitas sebuah model. Model yang kompleks tentunya akan membutuhkan waktu

pelatihan yang lama sehingga di dunia deep learning penggunaan GPU sudah sangatlah umum [15].

2.3 VGG

VGG adalah singkatan dari Visual Geometry Group; ini adalah arsitektur Convolutional Neural Network (CNN) standar dalam dengan banyak lapisan. Yang “dalam” mengacu pada jumlah lapisan dengan VGG-16 atau VGG-19 yang terdiri dari 16 dan 19 lapisan konvolusional. Arsitektur VGG adalah dasar dari model pengenalan objek yang inovatif. Dikembangkan sebagai jaringan neural dalam, VGGNet juga melampaui garis dasar dalam banyak tugas dan kumpulan data di luar ImageNet. Selain itu, saat ini arsitektur ini masih menjadi salah satu arsitektur pengenalan gambar yang paling populer.



Gambar 1. Visual Geometry Group

- **VGG-16**

Model VGG16 mencapai hampir 92,7% akurasi pengujian top-5 di ImageNet. ImageNet adalah kumpulan data yang terdiri dari lebih dari 14 juta gambar yang termasuk dalam hampir 1000 kelas. Selain itu, ini adalah salah satu model terpopuler yang dikirimkan ke ILSVRC-2014. Ini menggantikan filter berukuran kernel besar dengan beberapa filter berukuran kernel 3x3 satu demi satu, sehingga membuat peningkatan signifikan dibandingkan AlexNet. Model VGG16 dilatih menggunakan GPU Nvidia Titan Black selama beberapa minggu. Seperti disebutkan di atas, VGGNet-16 mendukung 16 lapisan

dan dapat mengklasifikasikan gambar ke dalam 1000 kategori objek, termasuk keyboard, hewan, pensil, mouse, dll. Selain itu, model ini memiliki ukuran input gambar 224 x 224.

- **VGG-19**

Konsep model VGG19 (juga VGGNet-19) sama dengan VGG16 hanya saja model ini mendukung 19 lapisan. Angka “16” dan “19” menunjukkan jumlah lapisan bobot dalam model (lapisan konvolusional). Artinya VGG19 memiliki tiga lapisan konvolusional lebih banyak dibandingkan VGG16. Kami akan membahas lebih lanjut tentang karakteristik jaringan VGG16 dan VGG19 di bagian akhir artikel ini.

2.4 Segmentasi

Segmentasi gambar atau citra merupakan proses pembagian citra dalam ukuran piksel tertentu menjadi beberapa bagian atau objek tertentu. Pemisahan yang dilakukan berdasarkan perbedaan yang mencolok atau kesamaan intensitas piksel dalam satu dimensi gambar. Pendekatan dengan metode klasik seperti thresholding, clustering, dan pemisahan objek berdasarkan tekstur telah memberikan landasan segmentasi. Pendekatan yang lebih modern seperti jaringan saraf konvolusional dalam deep learning mampu membuat segmentasi gambar menjadi presisi dan adaptif. Metode lain yang dapat digunakan berbasis grafis seperti watershed dan region-based berkontribusi dalam pemisahan objek dan memahami struktur citra lebih mendalam.

BAB III

Metode Penelitian

3.1 Teori Dasar

A. Model Pengenalan Gambar VGG-19: [1]

Arsitektur Convolutional Neural Network (CNN): VGG-19 adalah salah satu arsitektur CNN yang mendalam yang digunakan untuk tugas pengenalan gambar. Ini terdiri dari 19 lapisan (hence the name "VGG-19") dan sangat baik dalam mengekstraksi fitur-fitur dari gambar. Layer Konvolusi: Layer konvolusi dalam VGG-19 adalah bagian utama dalam proses ekstraksi fitur gambar. Mereka berfungsi sebagai penghasil fitur penting yang membantu model mengenali pola dan objek dalam citra, yang pada gilirannya berguna dalam tugas analisis pemandangan kota atau tugas pengenalan gambar lainnya.

B. Transfer Learning:[4]

Konsep Transfer Learning: Laporan dapat menjelaskan konsep transfer learning dan mengapa VGG-19 sering digunakan sebagai model dasar untuk tugas pengenalan gambar. Penjelasan dapat melibatkan bagaimana model pre-trained pada dataset besar dapat digunakan sebagai landasan untuk tugas khusus seperti analisis pemandangan kota.

C. Ekstraksi Fitur:[5]

Ekstraksi Fitur pada Layer Tertentu: menjelaskan mengapa layer-layer tertentu dari VGG-19 sering digunakan untuk ekstraksi fitur dalam tugas analisis pemandangan kota.

D. Evaluasi dan Performa:[6]

Metrik Performa: menjelaskan metrik performa yang digunakan untuk mengevaluasi model, seperti akurasi, F1-score, atau metrik khusus lainnya yang sesuai dengan tugas analisis.

3.2 Metode

Pada project ini digunakan data pemandangan kota yang berfokus pada pemahaman semantik pemandangan jalanan perkotaan yang akan diproses dengan CNN. Dataset terdapat 30 class yang dikelompokkan ke dalam 8 kategori (permukaan datar, manusia, kendaraan, konstruksi, objek, alam, langit dan void. ukuran setiap gambar 256 x 512 pixel yang diambil di 50

kota. Setiap file merupakan gabungan dengan foto asli di bagian kiri gambar, di samping gambar berlabel (output segmentasi semantik) di bagian kanan.

A. Pengumpulan Data

Data yang diperlukan untuk melakukan penelitian ini adalah data cityscapes image pairs tampak atas permukaan bumi yang berfokus permukaan datar, manusia, kendaraan, konstruksi, objek, alam, langit dan void. Data cityscapes image pairs kemudian akan dibagi menjadi data train dan data validasi dengan hasil 2780 dan 695 [7].

B. Analisa Data

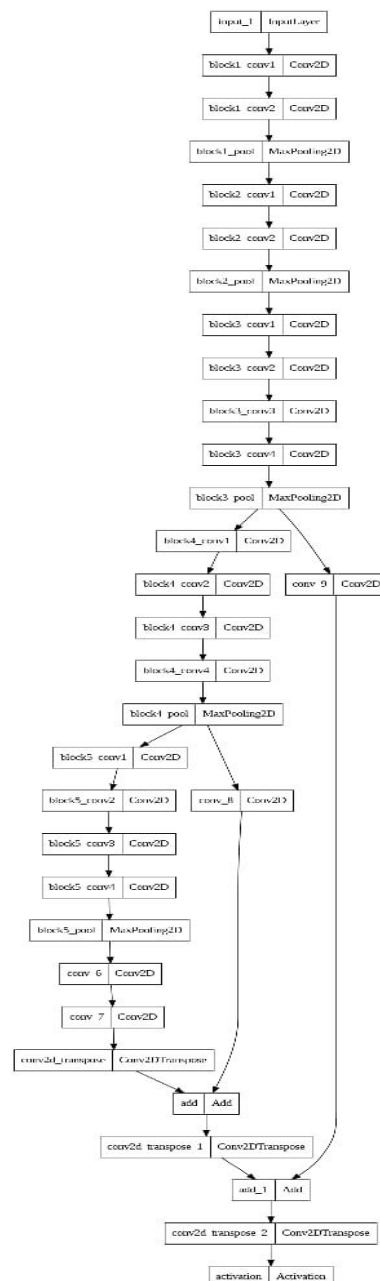
Data latih dan data uji input yang berupa citra satelit tampak atas bumi Objek yang diberi label di latar depan tidak boleh memiliki lubang, yaitu jika ada latar belakang yang terlihat 'melalui' beberapa objek di latar depan, maka itu dianggap sebagai bagian dari latar depan. Hal ini juga berlaku untuk wilayah yang sangat tercampur dengan dua kelas atau lebih: wilayah tersebut diberi label kelas latar depan. Contoh: daun pohon di depan rumah atau langit (semua pohon), jendela mobil transparan (semua mobil) [8].

C. Pengenalan VGG-19

Kenapa kami memilih VGG19 karena pertama, Pemandangan kota bisa sangat kompleks dengan berbagai elemen, seperti bangunan, jalan raya, taman, kendaraan, dan banyak lagi. Model yang lebih dalam seperti VGG-19 mungkin lebih mampu menangani keragaman dan kompleksitas ini karena memiliki lebih banyak lapisan dan kapasitas untuk memahami fitur-fitur yang rumit dalam gambar. Kedua, Karena VGG-19 memiliki lebih banyak lapisan, biasanya memiliki kemampuan yang lebih baik dalam mempelajari representasi fitur yang lebih dalam dan kompleks. Ini dapat menghasilkan akurasi yang lebih tinggi dalam tugas. Ketiga, dalam menggunakan transfer learning, model VGG-19 yang telah dilatih sebelumnya mungkin sudah memiliki representasi fitur yang kuat untuk objek-objek yang umumnya ada dalam pemandangan kota. Ini dapat menghemat waktu dan sumber daya pelatihan.

Dibawah ini terdapat **Gambar 3.** yang merupakan plot dari model vgg19 yang telah dibuat. Dari plot tersebut menunjukkan beberapa lapisan konvolusi (Conv2D) dan lapisan pengelompokan (MaxPooling2D) yang digunakan saat melatih model. Setiap lapisan konvolusi bertanggung jawab untuk mengekstraksi fitur-fitur kompleks dari gambar, sedangkan pada

lapisan pengelompokan (MaxPooling2D) memiliki tujuan untuk mereduksi dimensi gambar dan mengurangi jumlah parameter dengan masih mempertahankan informasi penting yang terdapat dalam gambar. Arsitektur model VGG19 memiliki beberapa block konvolusi yang bertujuan untuk mengekstraksi fitur dari gambar yang dimasukkan dalam model transfer learning yang digunakan. Masing-masing block terdiri dari beberapa lapisan konvolusi dan lapisan pengelompok, yang dimana dalam setiap lapisan tersebut memiliki parameter tertentu yang akan diatur atau disesuaikan selama pelatihan model. Sehingga model tersebut dapat mempelajari representasi fitur yang tepat dari data yang digunakan.



Gambar 3. Visualisasi Arsitektur Model

Dalam preprocessing terdapat perubahan skala yang dilakukan, tujuannya untuk memudahkan perangkat lunak jaringan saraf dalam pengolahan data. Selain itu mengubah rotasi gambar untuk dapat memahami objek dalam berbagai orientasi. Membuat generator data untuk melatih dan menguji model jaringan saraf. Secara dinamis generator mengambil gambar dari direktori yang digunakan selama pelatihan. Penggunaan generator berguna untuk menghindari memuat seluruh dataset gambar dalam memori secara bersamaan.

BAB IV

Pembahasan

4.1 Hasil dan Pembahasan

Cityscape dataset merupakan dataset yang berisi potongan foto dari sebuah video yang diambil dari kendaraan berjalan di Jerman. Gambar-gambar yang termuat dalam dataset berupa gambar asli dan label segmentasi semantik. Label segmentasi semantik adalah pengolahan citra dan penglihatan komputer yang bertujuan untuk mengkategorikan pixel dalam gambar ke suatu objek atau area yang berbeda. Segmentasi semantik menghasilkan pemetaan pixel-pixel yang memberikan detail objek dan elemen gambar. Adapun contoh gambar mentah dari dataset Cityscape sebagai berikut.

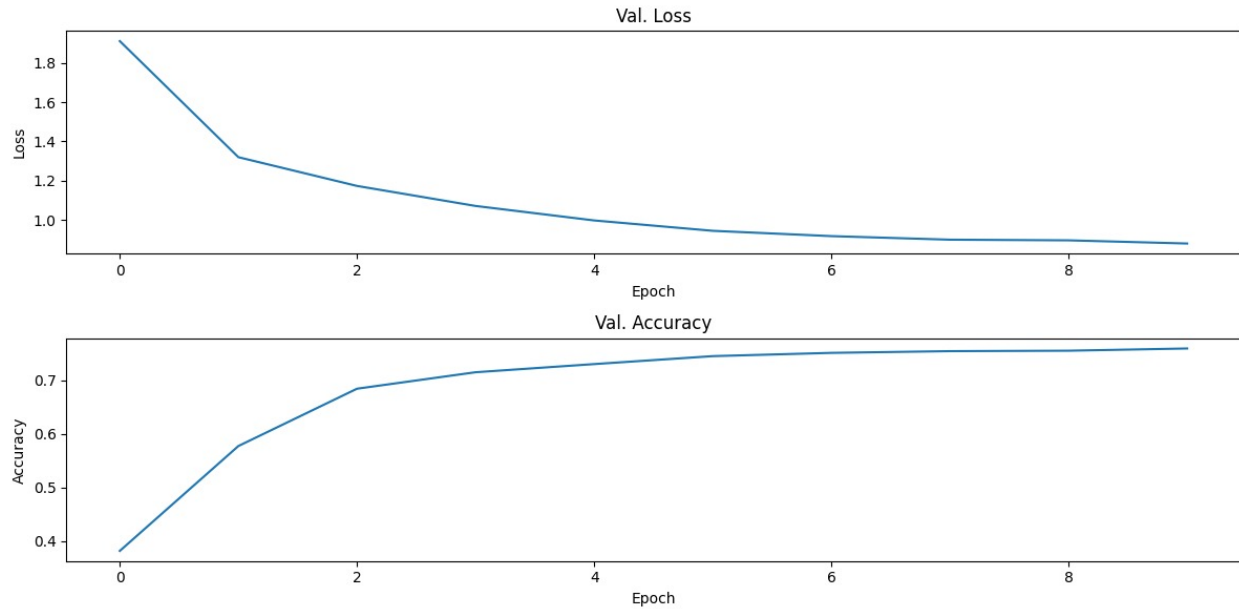


Gambar 2. Gambar asli



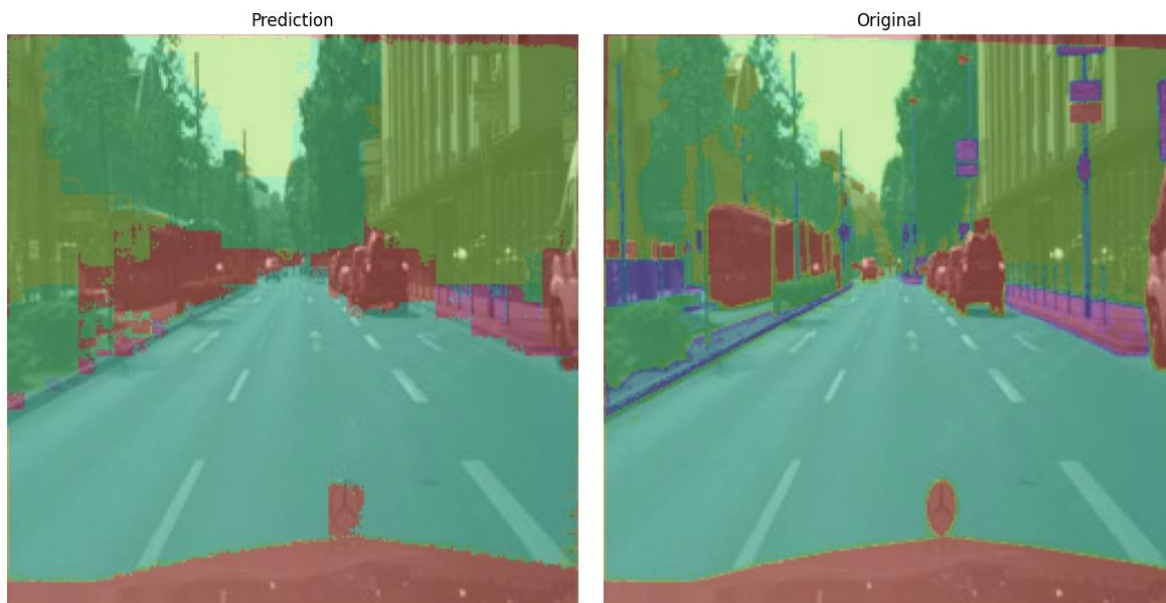
Gambar 3. Segmentasi semantik

Pada **Gambar 2.** merupakan gambar asli dari data cityscape, gambar tersebut menunjukkan hasil potongan gambar suatu video yang terdapat dalam kamera kendaraan. Kemudian pada **Gambar 3.** merupakan segmentasi semantik dari gambar atau foto asli sebelumnya, pada segmentasi tersebut beberapa objek dibedakan dengan warna yang berbeda-beda. Dari data tersebut yang selanjutnya diproses untuk analisis model dan prediksi keakuratan model terhadap data gambar cityscape. Berdasarkan metode yang telah dilakukan dengan model VGG19 terdapat perubahan val_loss dan val_accuracy dalam setiap epoch pelatihan. Dalam membuat hasil prediksi yang akurat dan baik dilakukan variasi learning rate, batch, dan epoch. Dari variasi model yang telah dilakukan didapatkan hasil prediksi pada gambar dibawah ini.



Gambar 4. Visualisasi Hasil Prediksi

Berdasarkan **Gambar 4** didapatkan hasil akurasi sebesar 78% pada epoch 10, dimana jika akurasi semakin naik pada saat epoch bertambah dan loss semakin turun saat epoch bertambah berarti menunjukkan bahwa model tersebut berjalan cukup baik. Hasil prediksi model visualisasi segmentasi ditampilkan untuk memperlihatkan kualitas segmentasi yang telah dilakukan, dari visualisasi tersebut terlihat bahwa grafik akurasi semakin meningkat seiring dengan bertambahnya epoch dan sebaliknya grafik loss menurun seiring dengan bertambahnya epoch. Visualisasi tersebut menunjukkan hasil prediksi dari data cityscape yang telah diuji berdasarkan model dan parameter yang digunakan.



Gambar 5. Hasil Prediksi Segmentasi

Diatas, **Gambar 5.** merupakan visualisasi dari prediksi model yang sebesar 78% terhadap data yang digunakan. Dari visualisasi tersebut menunjukkan beberapa objek belum tercover oleh warna-warna segmentasi gambar. Hal ini sesuai dengan akurasi nilai prediksi model segmentasi gambar.

BAB V

Penutup

5.1 Kesimpulan

Dalam percobaan yang telah kami lakukan, Cityscape dataset yang berisi gambar dari video yang diambil dari kendaraan di 50 kota termasuk gambar asli dan segmentasi semantik. Dengan proses preprocessing yang mencakup normalisasi intensitas pixel, rotasi gambar, dan penggunaan generator data kami berhasil melatih dan menguji model jaringan saraf. Namun dalam pelatihan model yang telah dilakukan didapatkan hasil yang kurang efisien karena hanya menginput 2 epoch. Dari 2 epoch yang diinputkan diperoleh akurasi sebesar 78%. Dan visualisasi arsitektur yang diberikan memberikan gambaran kepada pembaca bagaimana arsitektur model berjalan.

5.2 Saran

Model yang telah dilatih memberikan hasil yang baik untuk digunakan pada tugas segmentasi gambar berikutnya. Dengan mempertimbangkan hasil pelatihan dan visualisasi segmentasi, model VGG19 secara efektif dapat digunakan untuk tugas segmentasi citra dengan dataset cityscape image pairs. Hasil penilaian lanjutan diperlukan untuk mengevaluasi performa model secara mendalam dan implementasi yang lebih luas.

REFERENSI

- [1] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.
- [2] Yang, X., Luo, P., Loy, C. C., & Tang, X. (2015). WIDER FACE: A Face Detection Benchmark. Proceedings of the IEEE International Conference on Computer Vision (ICCV).
- [3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is All You Need. Advances in Neural Information Processing Systems (NeurIPS).
- [4] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In Advances in neural information processing systems (pp. 3320-3328).
- [5] Arsanjani, J. J., Sayyad, G., Sarai, A., & Wakefield, S. (2013). Integrating space syntax into GIS for evaluating the impact of urban form on pedestrians' walking behavior: The case of Siena, Italy. *Environment and Planning B: Planning and Design*, 40(3), 505-527.
- [6] Evaluasi dan Performa: [Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... & Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the ACM International Conference on Multimedia (MM) (pp. 675-678).
- [7] <https://www.jtiulm.ti.ft.ulm.ac.id/index.php/jtiulm/article/download/99/64>
- [8] <https://www.cityscapes-dataset.com/dataset-overview/>
- [10] J. C. Hung, K. C. Lin, and N. X. Lai, "Recognizing learning emotion based on convolutional neural networks and transfer learning," *Applied Soft Computing Journal*, vol. 84, p. 105724, 2019, doi: 10.1016/j.asoc.2019.105724.
- [11] LeCun, Yann; Bengio, Yoshua; Hinton, Geoffrey (2015). "Deep Learning". *Nature*. **521** (7553): 436–444.
- [12] Ciresan, D.; Meier, U.; Schmidhuber, J. (2012). "Multi-column deep neural networks for image classification". *2012 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3642–3649.
- [13] Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey (2012).
- [14] *echCrunch*. 25 May 2017.from the original on 17 June 2018. Retrieved 17 June 2018.
- [15] Danukusumo, Kefin Pudi. 2017. *Implementasi Deep Learning Menggunakan Convolutional Neural Network Untuk Klasifikasi Citra Candi Berbasis GPU*.