

BAB II KAJIAN LITERATUR

2.1 Tinjauan Pustaka

Kemacetan lalu lintas merupakan isu penting yang telah banyak dikaji dalam berbagai penelitian, khususnya dengan pendekatan teknologi berbasis kecerdasan buatan. *Mask R-CNN* yang diperkenalkan oleh Kaiming He, bersama dengan Georgia Gkioxari, Piotr Dollár, dan Ross Girshick, telah menjadi metode unggulan dalam segmentasi objek, termasuk kendaraan di lingkungan lalu lintas yang kompleks [1]. Model ini kemudian dikembangkan lebih lanjut dengan menggunakan arsitektur *backbone ResNet* untuk meningkatkan akurasi dan efisiensi [2].

Di sisi lain, prediksi lalu lintas sering kali melibatkan analisis deret waktu yang memerlukan pemodelan pola *temporal*. Model *LSTM* yang diperkenalkan oleh Hochreiter dan Schmidhuber mampu menangani ketergantungan jangka panjang dalam data sekuensial [3]. Penelitian lain menunjukkan bahwa *LSTM* mampu memprediksi volume lalu lintas dan kecepatan kendaraan dengan performa yang baik jika dibandingkan dengan metode regresi klasik [12], [13].

Penelitian yang menggabungkan teknik deteksi dan prediksi juga telah banyak dilakukan. Misalnya, sistem berbasis *YOLO* yang dikombinasikan dengan metode prediktif seperti *ARIMA* atau *SVR* telah diterapkan untuk mendeteksi kendaraan dan memprediksi kepadatan lalu lintas [14]. Namun, hasil prediksi dari pendekatan ini masih terbatas pada efisiensi waktu tanpa mempertimbangkan akurasi *spasial* yang lebih mendalam.

Pendekatan *Hybrid Mask R-CNN* dan *LSTM* mulai mendapat perhatian karena mampu menggabungkan kekuatan segmentasi *spasial* dan analisis *temporal* [15]. Studi terbaru menunjukkan bahwa kombinasi dua model ini mampu mendeteksi kendaraan secara akurat sekaligus memprediksi kepadatan lalu lintas di waktu mendatang [16]. Namun demikian, tantangan masih ada pada sisi optimasi arsitektur dan parameter model untuk mencapai performa maksimal.

2.1.1 Mask RCNN

Mask R-CNN adalah salah satu algoritma deteksi dan segmentasi objek yang paling canggih dan populer dalam *domain* visi komputer. Algoritma ini dikembangkan oleh Kaiming He, Georgia Gkioxari, Piotr Dollár, dan Ross Girshick sebagai pengembangan dari *Faster R-CNN* [1].

Mask R-CNN menambahkan cabang ketiga pada arsitektur *Faster R-CNN* yang digunakan untuk melakukan segmentasi instansi (instance segmentation), yaitu membedakan objek-objek tidak hanya dari kelasnya tetapi juga secara individu dalam bentuk piksel. Secara umum, *Mask R-CNN* memiliki tiga cabang utama:

1. **Klasifikasi** – untuk mengenali kelas dari setiap objek,
2. **Regresi Bounding box** – untuk menentukan koordinat posisi objek,
3. **Mask** – untuk menghasilkan segmentasi piksel dari objek tersebut.

Mask R-CNN adalah pengembangan dari *Faster R-CNN* yang menambahkan cabang untuk segmentasi objek (*Mask*). Arsitekturnya terdiri dari:

- *Backbone: ResNet* (Residual Network) sebagai ekstraksi fitur
- Region Proposal Network (RPN): Menghasilkan proposal kandidat objek
- RoIAlign: Menyelaraskan proposal untuk ekstraksi fitur yang akurat
- Branch Tambahan: Untuk memprediksi *Mask* objek

a. Arsitektur *Mask R-CNN*

Mask R-CNN menggunakan *backbone* seperti *ResNet* atau *ResNeXt* yang dilengkapi dengan Feature Pyramid Network (FPN) untuk ekstraksi fitur yang lebih kaya dari berbagai skala. Setelah proses ekstraksi fitur, bagian **Region Proposal Network (RPN)** menghasilkan candidate boxes (proposal) yang kemudian disempurnakan oleh proses **RoIAlign**. RoIAlign adalah pembaruan dari metode RoIPool yang mampu menyelaraskan fitur *spasial* dengan lebih akurat, sehingga segmentasi menjadi lebih presisi.

b. Fungsi Loss

Fungsi Loss *Mask R-CNN*:

$$L = L_{cls} + L_{box} + L_{mask} \quad (2.1)$$

- L_{cls} : Cross-entropy loss untuk klasifikasi.
- L_{box} : Smooth L1 loss untuk regresi bounding box.
- L_{mask} : Binary cross-entropy loss yang dihitung piksel-per-piksel untuk segmentasi mask.

c. Keunggulan *Mask R-CNN*

Mask R-CNN unggul dalam segmentasi objek secara detail, terutama dalam situasi di mana objek saling berdekatan atau tumpang tindih. Ini sangat berguna dalam konteks deteksi kendaraan di jalan raya.

Dalam penelitian oleh Kumar, Sharma, dan Chauhan, *Mask R-CNN* diterapkan untuk mendeteksi kendaraan secara *real-time* dari *video* lalu lintas, dan hasilnya menunjukkan akurasi lebih tinggi dibanding YOLO dan SSD, terutama dalam kondisi padat [2]. Sementara itu, studi oleh Rezaei dan Klette menunjukkan bahwa *Mask R-CNN* tetap menunjukkan performa deteksi yang baik dalam kondisi pencahayaan rendah, menjadikannya ideal untuk aplikasi lalu lintas kota yang kompleks [3].

2.1.2 ResNet (Residual Network)

Pada arsitektur Mask R-CNN, ResNet digunakan sebagai **backbone convolutional neural network** yang berfungsi untuk mengekstraksi fitur visual dari citra lalu lintas. Penggunaan ResNet memungkinkan model memperoleh representasi fitur yang lebih dalam sehingga meningkatkan akurasi deteksi objek kendaraan.

ResNet, atau Residual Network, adalah arsitektur jaringan konvolusional yang diperkenalkan oleh Kaiming He, Xiangyu Zhang, Shaoqing Ren, dan Jian Sun pada tahun 2015. Arsitektur ini menjadi tonggak penting dalam pengembangan *deep learning* karena mampu melatih jaringan yang sangat dalam hingga ratusan lapisan tanpa mengalami *degradasi akurasi* atau *Vanishing gradient* [1].

a. Masalah Jaringan Dalam dan Solusi *ResNet*

Sebelum kemunculan *ResNet*, peningkatan jumlah lapisan dalam jaringan neural sering menyebabkan menurunnya performa karena gradien menjadi sangat kecil atau bahkan hilang saat backpropagation. *ResNet* menyelesaikan masalah ini dengan memperkenalkan skip connection atau shortcut connection, yaitu mekanisme yang memungkinkan input asli untuk dilompati langsung ke lapisan yang lebih jauh tanpa melalui transformasi non-linear penuh.

b. Blok Residual

Unit dasar dari *ResNet* adalah residual block, yang menggunakan shortcut identity mapping:

$$y = F(x, W_i) + x \quad (2.2)$$

- $F(x)$ adalah fungsi transformasi dari layer konvolusi, biasanya terdiri dari dua atau tiga layer berturut-turut.
- x adalah input awal yang ditambahkan langsung ke output layer, menghasilkan output y .

Mekanisme ini memungkinkan jaringan untuk belajar fungsi residual, yaitu selisih antara output yang diinginkan dan input awal, bukan belajar transformasi absolut.

c. Arsitektur *ResNet*

Versi paling umum dari *ResNet* adalah:

- *ResNet-18* dan *ResNet-34*: menggunakan blok dasar (2-layer convolution).
- *ResNet-50*, 101, dan 152: menggunakan *bottleneck block*, yakni tiga layer (1x1, 3x3, 1x1).

Keunggulan *bottleneck block* adalah efisiensi dalam komputasi tanpa mengurangi kedalaman jaringan.

d. Kelebihan *ResNet*

1. Mampu melatih jaringan yang sangat dalam tanpa mengalami *Vanishing gradient*.
2. Efisien secara komputasi, terutama pada arsitektur dengan *bottleneck block*.
3. Sangat fleksibel dan dapat digunakan sebagai *backbone* berbagai arsitektur seperti *Mask R-CNN*, *Faster R-CNN*, *RetinaNet*, dan sebagainya.
4. Performanya unggul dalam benchmark seperti ImageNet dan COCO dengan akurasi yang lebih tinggi dibanding model konvolusional lainnya seperti VGG.

e. Penerapan *ResNet* dalam Visi Komputer

ResNet telah menjadi *backbone* standar dalam berbagai aplikasi visi komputer modern. Dalam konteks penelitian ini, *ResNet* digunakan sebagai extractor fitur visual untuk *Mask R-CNN*, karena kemampuannya menghasilkan representasi *spasial* yang kuat dan dalam. Seperti yang dijelaskan oleh He et al. [1], *ResNet-50* dan *ResNet-101* terbukti efektif untuk tugas segmentasi objek dalam *Dataset COCO* dan *PASCAL VOC*.

Studi oleh Zoph dan Le [2] menunjukkan bahwa bahkan dalam arsitektur hasil penCarian otomatis (*neural architecture search*), *ResNet* tetap menjadi baseline yang kuat dan kompetitif.

2.1.3 Teori *Long Short-Term Memory (LSTM)*

Dalam penelitian ini, model LSTM digunakan untuk mempelajari pola perubahan kepadatan lalu lintas berdasarkan data deret waktu yang dihasilkan dari proses deteksi kendaraan. Variabel **vehicle_count** dan **area_ratio** pada setiap frame video disusun sebagai sekuens data temporal yang kemudian diproses oleh LSTM untuk memprediksi kondisi kepadatan lalu lintas pada waktu berikutnya.

Long Short-Term Memory (LSTM) adalah jenis Recurrent Neural Network (*RNN*) yang dirancang untuk mengatasi keterbatasan *RNN* konvensional dalam mengingat informasi jangka panjang. *LSTM* dikembangkan oleh Sepp Hochreiter dan Jürgen Schmidhuber pada tahun 1997 [1]. Arsitektur ini sangat efektif dalam menangani data sekuensial seperti deret waktu, teks, dan *video*.

a. Masalah pada RNN Konvensional

RNN tradisional cenderung gagal dalam mempelajari dependensi jangka panjang karena mengalami masalah **Vanishing gradient** atau **exploding gradient** selama proses backpropagation. Akibatnya, model hanya mampu mengingat informasi jangka pendek dan gagal memahami konteks sekuensial yang lebih panjang.

b. Struktur Arsitektur LSTM

LSTM mengatasi masalah ini dengan memperkenalkan memori internal yang disebut **cell state** dan tiga gerbang utama yang mengatur informasi yang disimpan dan dibuang:

– **Forget Gate:**

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.3)$$

Gerbang ini menentukan informasi mana dari cell state lama yang akan dibuang.

– **Input Gate:**

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.4)$$

$$\hat{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2.5)$$

Menentukan informasi baru yang akan disimpan ke cell state.

– **Update Cell State:**

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \quad (2.6)$$

– **Output Gate:**

$$h_t = o_t * \tanh(C_t) \quad (2.7)$$

Menentukan output tersembunyi (hidden state) untuk waktu t .

c. Keunggulan LSTM

1. **Kemampuan mengingat jangka panjang** melalui cell state yang dipertahankan selama banyak langkah waktu.
2. **Efektif untuk data sekuensial dan deret waktu**, termasuk prediksi lalu lintas, pemrosesan bahasa alami, dan pengenalan aktivitas *video*.
3. **Mengatasi kelemahan RNN konvensional**, seperti overfitting dan kegagalan memahami konteks jangka panjang.

d. Penerapan LSTM dalam Prediksi Lalu Lintas

Dalam konteks sistem transportasi, LSTM banyak digunakan untuk memprediksi volume kendaraan, kecepatan lalu lintas, dan kemacetan berdasarkan data historis. Penelitian

oleh Xiaolei Ma, Zhiyong Tao, dan Yu Wang menunjukkan bahwa *LSTM* mengungguli model-model statistik konvensional seperti ARIMA dan SVR dalam prediksi lalu lintas jangka panjang [2].

Penelitian lain oleh Yisheng Lv, Yanjie Duan, dan Wenwen Wang menunjukkan bahwa *LSTM* mampu belajar representasi non-linear dan kompleks dari aliran lalu lintas dengan akurasi tinggi [3].

2.1.4 Evaluasi Model Deep Learning

Evaluasi kinerja model dalam penelitian ini menggunakan beberapa metrik kuantitatif, yaitu:

- *Mean Average Precision (mAP)* dan *Intersection over Union (IoU)* untuk mengukur akurasi deteksi kendaraan.
- *Root Mean Square Error (RMSE)* dan *Mean Absolute Error (MAE)* untuk mengukur tingkat kesalahan prediksi kemacetan.

Metrik ini digunakan untuk menilai seberapa efektif model *Hybrid Mask R-CNN-ResNet* dan *LSTM* dalam mendeteksi serta memprediksi kemacetan secara *real-time*.

Dalam konteks sistem lalu lintas perkotaan, evaluasi model deep learning tidak hanya berfungsi untuk mengukur performa matematis, tetapi juga merepresentasikan kemampuan model dalam menangkap karakteristik *spasial* dan *temporal* lalu lintas. Akurasi deteksi kendaraan yang tinggi belum tentu menghasilkan prediksi kemacetan yang akurat apabila informasi *spasial* tersebut tidak direpresentasikan secara tepat dalam dimensi waktu. Oleh karena itu, diperlukan mekanisme integrasi *spasial-temporal*, di mana hasil deteksi kendaraan seperti jumlah kendaraan per *frame*, distribusi posisi, dan tingkat kepadatan lalu lintas diolah menjadi fitur sekuensial yang kemudian dipelajari oleh model *Long Short-Term Memory (LSTM)*. Dengan demikian, metrik mAP dan IoU digunakan untuk mengevaluasi kualitas analisis *spasial*, sedangkan RMSE dan MAE digunakan untuk mengevaluasi ketepatan analisis *temporal* dalam satu kesatuan sistem hybrid.

Perkembangan terbaru dalam penelitian prediksi lalu lintas menunjukkan munculnya berbagai model spatio-temporal yang lebih kompleks, seperti **Convolutional LSTM (ConvLSTM)**, **Graph Neural Network (GNN)**, dan **Transformer berbasis attention mechanism**. ConvLSTM mampu menangkap hubungan spasial dan temporal secara simultan melalui operasi konvolusi pada struktur sel memori. Sementara itu, Graph Neural Network digunakan untuk merepresentasikan jaringan jalan sebagai graf sehingga hubungan antar ruas jalan dapat dimodelkan secara eksplisit. Model Transformer juga mulai digunakan

dalam prediksi lalu lintas karena kemampuannya mempelajari dependensi jangka panjang secara lebih efisien melalui mekanisme attention.

Meskipun model-model tersebut menunjukkan performa tinggi pada dataset skala besar, kompleksitas komputasinya relatif lebih tinggi. Oleh karena itu penelitian ini menggunakan pendekatan pipeline antara deteksi objek berbasis CNN dan pemodelan temporal berbasis LSTM yang lebih sederhana namun tetap memungkinkan analisis eksperimen secara terkontrol.

2.2 Penelitian Terdahulu

1. Xu, K. C., Wang, G., Yan, S., Yu, J., Zhang, B., Dai, S., Li, Y., & Xu, L., “Fast Vehicle and Pedestrian Detection Using Improved *Mask R-CNN*,” *Mathematical Problems in Engineering*, 2020. Penelitian ini memperkenalkan modifikasi pada arsitektur *Mask R-CNN* dengan mengganti *backbone* standar dengan *ResNet-86* untuk meningkatkan kecepatan inferensi serta akurasi deteksi kendaraan dan pejalan kaki pada kondisi nyata. *Dataset* yang digunakan berbasis Microsoft COCO dan kondisi jalan perkotaan, dengan hasil menunjukkan pengurangan waktu proses dan peningkatan akurasi sekitar 5%. Penulis juga membahas tantangan deteksi objek kecil dan tumpang-tindih yang kerap terjadi dalam lingkungan lalu lintas padat.

2. Bhawana Parihar & Poonam Chimmwal, “Traffic Flow Prediction using *LSTM* Networks: A *Deep learning* Approach for *Real-time* Traffic Forecasting,” *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 11, 2022. Penelitian ini menggunakan model *Long Short-Term Memory (LSTM)* untuk meramalkan aliran lalu lintas secara *real-time* berdasarkan data sensor jalan. Penulis menguji berbagai pengaturan hyper-parameter seperti jumlah layer, ukuran batch, dan learning rate, serta membandingkannya dengan metode tradisional seperti ARIMA dan regresi linier. Hasil menunjukkan bahwa *LSTM* berhasil menangani non-linearitas dan variabilitas lalu lintas dengan cukup baik, meningkatkan akurasi prediksi untuk jangka pendek.

3. Zhang, T., Lu, Z., Chen, Q., Zhong, H., Zhang, Y., Xue, J., & Wu, C., “A Spatial-Temporal Short-Term Traffic Flow Prediction Model Based on Dynamical-Learning Graph Convolution Mechanism,” arXiv preprint, May 2022. Penelitian ini menggabungkan mekanisme Graph Convolutional Network (GCN) dengan *LSTM* dalam satu kerangka

spasial-temporal: GCN untuk menangkap hubungan antar ruas jalan (topologi jaringan) dan *LSTM* untuk menangkap pola *temporal* aliran lalu lintas. Dengan menggunakan *Dataset* nyata, model mereka menunjukkan peningkatan akurasi prediksi hingga 7% dibandingkan baseline *LSTM* tunggal. Studi ini relevan terhadap persimpangan jalan karena mengatasi kompleksitas struktur jalan sebagai graf.

4. Qi, Y., & Cheng, Z., “Research on Traffic Congestion Forecast Based on Deep learning,” *Information*, vol. 14, no. 2, Art. 108, Feb. 2023. Penelitian ini fokus pada prediksi kemacetan (bukan hanya aliran kendaraan) menggunakan model *deep learning* yang memanfaatkan variabel seperti kepadatan kendaraan, kecepatan rata-rata, dan waktu tempuh. Metodologi menggabungkan *CNN* untuk ekstraksi fitur *spasial* dan *LSTM* untuk prediksi *temporal*. Hasil menunjukkan akurasi prediksi hingga 93%. Penulis juga membahas kendala data lalu lintas dengan gangguan eksternal (cuaca, insiden) serta menyarankan integrasi data tambahan untuk meningkatkan performa.

5. “Improved Mask R-CNN Multi-Target Detection and Segmentation for Autonomous Driving in Complex Scenes,” *Sensors*, vol. 23, no. 8, Art. 3853, 2023. (Fang, S.; Zhang, B.; Hu, J.). Penelitian ini mengusung peningkatan arsitektur *Mask R-CNN* dengan *backbone* ResNeXt dan modul perhatian “Efficient Channel Attention (ECA)”. Strategi bottom-up path enhancement pada FPN ditambahkan untuk memperkuat fitur *multi-skala*. Hasil eksperimen pada *Dataset CityScapes* menunjukkan mAP deteksi mencapai 62.62% (peningkatan ~4.7%) dan mAP segmentasi mencapai 57.58% (peningkatan ~3.96%). Studi menyoroti tantangan deteksi objek banyak dan tumpang-tindih di lingkungan lalu lintas kompleks.

6. Dai, X., “Research on Urban Traffic Flows Prediction Model Based on Deep learning,” *Applied and Computational Engineering*, vol. 94, Nov. 2024. Penelitian ini memanfaatkan data historis kendaraan, cuaca, dan kondisi jalan untuk melatih model *LSTM* dalam prediksi aliran lalu lintas di jalan antar negara bagian (Interstate-94, AS). Skema training 80% dan testing 20% digunakan, dengan evaluasi menggunakan MAE dan RMSE. Hasil menunjukkan performa model *LSTM* cukup baik dalam menangani dinamika lalu lintas meski tantangan pada data eksternal tetap besar. Penulis menekankan kebutuhan integrasi data *real-time* dan pengujian skala besar untuk penerapan smart city.

7. “Network Level Spatial Temporal Traffic Forecasting with Hierarchical-Attention-LSTM,” Dec 2024. Penelitian ini propose model hierarkis *LSTM* dengan mekanisme attention pooling untuk meramalkan kondisi lalu lintas network-wide, bukan hanya satu

tautan jalan. Model mampu menangani pola waktu yang berbeda skala dan memprediksi kemacetan anomali dengan lebih akurat. Studi ini menunjukkan bahwa penggunaan attention dan struktur hierarkis meningkatkan akurasi prediksi dan dapat menangani data luar kebiasaan (outlier).

8. Slimane, J. B., & Ammar, M. B., “A Novel Approach of Traffic Congestion and Anomaly Detection with Prediction Using *Deep learning*,” *Journal of Electrical Systems*, vol. 20, no. 3s, 2024. Penelitian ini mengembangkan kerangka *deep learning* yang menggabungkan *CNN* untuk deteksi kemacetan/anomali dan *LSTM* untuk prediksi aliran kendaraan. *Dataset* yang digunakan mencakup berbagai pola kemacetan, termasuk abnormal (insiden, kecelakaan). Hasilnya menunjukkan bahwa model baru lebih unggul dibanding metode konvensional dalam akurasi, presisi, dan efisiensi komputasi. Penulis juga menekankan pentingnya sistem *real-time* untuk “smart traffic management”.

9. “Traffic Flow Prediction Using *Deep learning* Techniques in Urban Road Networks,” *Applied Mathematics and Nonlinear Sciences*, vol. 10, no. 1, Jan 2025. (Mar 17 2025). Penelitian ini mengevaluasi berbagai teknik *deep learning* (*LSTM*, Transformer, GNN) dalam prediksi aliran lalu lintas di jaringan jalan perkotaan. Model *temporal-spasial* dieksplorasi dan data dibagi kategori: “regular flow”, “episodic event”, dan “meteorological data”. Hasil mengindikasikan bahwa kombinasi model (ensemble) lebih unggul dibanding tunggal dan memperbesar akurasi serta robustitas terhadap kondisi ekstrim.

10. Pasha, M. K., Atmadja, A. R., & Firdaus, M. D., “Intelligent Traffic Management System Using *Mask Regions-Convolutional Neural Network* (MR-*CNN*),” *KINETIK*, vol. 10, no. 3, pp. 329-338, Aug 2025. Penelitian lokal di Indonesia ini menerapkan varian *Mask R-CNN* dengan *backbone ResNet-50* pada rekaman CCTV persimpangan Pasteur-Pasopati Bandung. Hasil menunjukkan akurasi deteksi kendaraan 80.7%, presisi makro 0.89, recall 0.83, dan F1-score 0.82. Sistem juga mampu menyesuaikan durasi lampu hijau berdasarkan tingkat kepadatan kendaraan secara *real-time*, sehingga memiliki implikasi praktis bagi manajemen lalu lintas perkotaan di Indonesia.

Berdasarkan kajian terhadap penelitian terdahulu, dapat disimpulkan bahwa sebagian besar penelitian masih memisahkan proses deteksi kendaraan dan prediksi kemacetan sebagai dua tahapan yang berdiri sendiri. Penelitian berbasis *Mask R-CNN* umumnya berfokus pada peningkatan akurasi deteksi kendaraan secara *spasial*, sementara penelitian prediksi lalu lintas berbasis *LSTM* lebih banyak memanfaatkan data numerik atau sensor

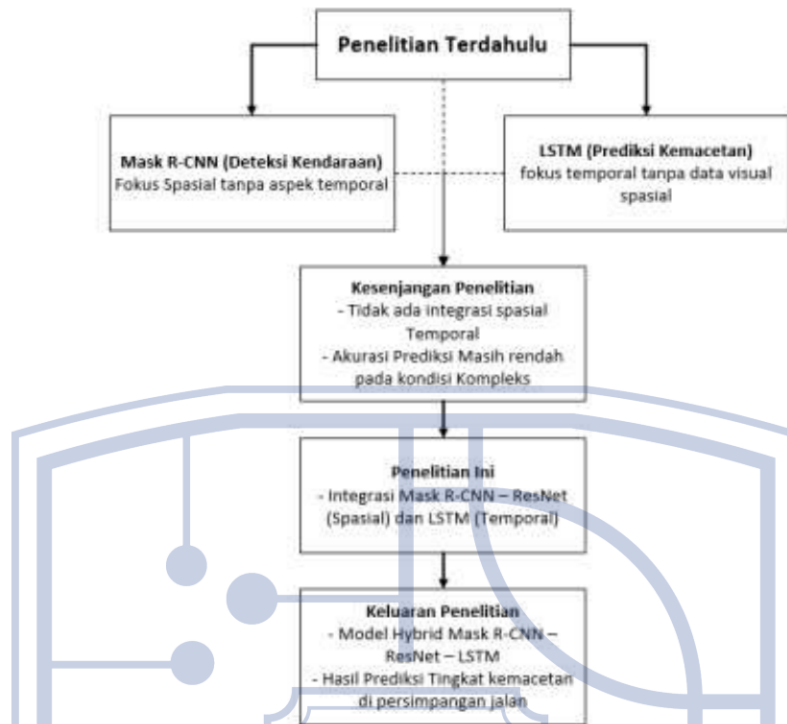
tanpa melibatkan informasi visual secara langsung. Kondisi ini menunjukkan adanya kesenjangan penelitian dalam pengintegrasian hasil deteksi visual berbasis segmentasi instansi dengan pemodelan *temporal*. Kesenjangan tersebut menjadi dasar pengembangan model *hybrid Mask R-CNN–ResNet* dan *LSTM* yang diusulkan dalam penelitian ini.

2.3 Kerangka Konseptual

Kerangka konseptual penelitian ini dibangun berdasarkan hasil sintesis dari teori dan penelitian terdahulu pada bidang deteksi kendaraan dan prediksi lalu lintas berbasis *deep learning*. Dari hasil kajian, terlihat bahwa penelitian terdahulu cenderung terbagi ke dalam dua arus besar. Arus pertama berfokus pada analisis spasial melalui deteksi kendaraan menggunakan model seperti Mask R-CNN. Arus kedua berfokus pada analisis temporal melalui prediksi lalu lintas menggunakan model seperti LSTM.

Penelitian oleh Xu Liang, Chen Ming, dan Wang Hui [5] serta Fang Yuxiang, Zhao Ke, dan Liu Rong [6] menunjukkan bahwa algoritma *Mask R-CNN* mampu mendeteksi kendaraan dengan tingkat akurasi tinggi pada berbagai kondisi lalu lintas. Namun, kedua penelitian tersebut hanya menekankan pada kemampuan model dalam menganalisis citra secara *spasial* dan belum mengintegrasikan aspek waktu yang sangat penting dalam konteks perubahan kondisi lalu lintas. Sementara itu, penelitian oleh Bhawana Kumari dan Deepak Chimmwal [7] serta Qi Chenglong dan Cheng Weiyu [8] telah membuktikan efektivitas model *LSTM* dalam melakukan prediksi arus kendaraan berbasis data sekuensial. Akan tetapi, model *LSTM* yang mereka gunakan tidak memiliki mekanisme untuk memahami konteks visual atau hubungan *spasial* antar objek kendaraan dalam *video*.

Keterbatasan lain juga ditemukan pada penelitian Zhang Qiang, Liu Fang, dan Wu Jie [9], di mana integrasi *CNN* dan *RNN* yang mereka usulkan masih menghasilkan waktu pemrosesan yang tinggi (*inference latency*) sehingga tidak optimal untuk aplikasi *real-time*. Dalam konteks Indonesia, penelitian oleh Adi Prasetyo dan Rahmawati Dini [10] menggunakan *Mask R-CNN* untuk deteksi kendaraan dengan hasil akurasi tinggi, namun penelitian tersebut berhenti pada tahap deteksi tanpa melanjutkan ke prediksi kemacetan yang bersifat *temporal*. Penelitian lain seperti yang dilakukan oleh Li Jun, Huang Shan, dan Zhou Tong [11] serta Sun Hao, Zhao Lei, dan Tang Yong [13] mencoba menerapkan *attention mechanism* untuk meningkatkan deteksi kendaraan, tetapi masih memiliki keterbatasan pada aspek generalisasi model dan belum menggabungkan analisis *temporal* secara komprehensif.



Gambar 2. 1 Kerangka Konseptual

Gambar di atas hasil adaptasi atau sintesis dari beberapa sumber teori utama yang dijadikan dasar. Proses deteksi dan prediksi kemacetan di area ini memerlukan pendekatan yang mampu menangkap pola *spasial* (posisi dan pergerakan kendaraan dalam ruang) serta pola *temporal* (perubahan kondisi lalu lintas dari waktu ke waktu). Oleh karena itu, konsep yang dikembangkan dalam penelitian ini menggabungkan kekuatan *Convolutional Neural Network (CNN)* untuk representasi *spasial* dan *Recurrent Neural Network (RNN)* tipe *LSTM* untuk analisis *temporal* dalam satu sistem *hybrid*.

Penelitian ini dimulai dengan tahap pengumpulan dan pra-pemrosesan data berupa citra atau *video* lalu lintas yang diambil dari kamera pengawas di area persimpangan. Data ini kemudian melalui proses normalisasi, konversi warna, dan segmentasi awal untuk memastikan kualitas input yang optimal. Tahap pra-pemrosesan ini penting untuk mengurangi noise visual yang dapat menyebabkan kesalahan deteksi pada sistem berbasis citra.

Selanjutnya, sistem melakukan deteksi kendaraan menggunakan algoritma *Mask R-CNN* yang berbasis *ResNet*. Arsitektur *ResNet* berperan sebagai *backbone* untuk ekstraksi fitur mendalam dari citra, sedangkan *Mask R-CNN* menambahkan cabang segmentasi yang menghasilkan *bounding box* dan *Mask instance* bagi tiap kendaraan. Output dari tahap ini bukan hanya jumlah kendaraan, tetapi juga informasi *spasial* seperti posisi dan ukuran objek

di setiap *frame video*. Data *spasial* tersebut kemudian dikonversi menjadi *feature sequence* yang menjadi input untuk tahap prediksi.

Tahap berikutnya adalah prediksi kemacetan menggunakan model *Long Short-Term Memory (LSTM)*. *LSTM* memanfaatkan urutan fitur dari hasil deteksi untuk mengenali pola *temporal*, misalnya peningkatan jumlah kendaraan dalam jangka waktu tertentu atau penurunan kecepatan rata-rata antar *frame*. Dengan mempelajari hubungan antar waktu ini, *LSTM* dapat memprediksi kemungkinan terjadinya kemacetan dalam interval waktu berikutnya. Model ini sangat relevan untuk sistem peringatan dini (*early warning system*) dan pengaturan sinyal lalu lintas adaptif.

Integrasi kedua model ini disebut sebagai *model Hybrid Mask R-CNN–ResNet–LSTM*. Kombinasi ini memungkinkan sistem untuk melakukan deteksi objek kendaraan dengan presisi tinggi sekaligus memahami dinamika perubahan lalu lintas secara *temporal*. Proses integrasi dilakukan melalui mekanisme *feature fusion*, di mana output fitur *Mask R-CNN* yang telah melalui lapisan *ResNet* disusun dalam format sekuensial sebelum dimasukkan ke jaringan *LSTM*. Pendekatan ini diharapkan dapat meningkatkan konsistensi hasil prediksi, mengurangi kesalahan akibat deteksi yang tidak stabil, serta memberikan hasil yang lebih akurat dalam memproyeksikan kondisi kemacetan.

Untuk memastikan keandalan model, dilakukan tahap **evaluasi performa** menggunakan metrik kuantitatif yang terdiri atas *mean Average Precision (mAP)* dan *Intersection over Union (IoU)* untuk menilai akurasi deteksi kendaraan, serta *Root Mean Square Error (RMSE)* dan *Mean Absolute Error (MAE)* untuk menilai ketepatan prediksi kemacetan. Evaluasi ini bertujuan untuk mengukur efektivitas sistem *hybrid* dibandingkan model konvensional seperti *Mask R-CNN* tunggal atau *LSTM* tunggal.

Secara konseptual, penelitian ini diharapkan dapat menjawab dua tantangan utama dalam sistem manajemen lalu lintas cerdas (*intelligent traffic management system*): (1) kemampuan mendeteksi objek kendaraan dengan presisi tinggi pada kondisi lalu lintas yang kompleks, dan (2) kemampuan memprediksi perubahan kepadatan lalu lintas secara akurat dan real-time. Model *Hybrid Mask R-CNN–ResNet* dan *LSTM* yang dioptimasi ini menjadi solusi komprehensif yang menggabungkan aspek visual, *spasial*, dan *temporal* dalam satu kesatuan sistem prediktif.

Kerangka konseptual ini menegaskan bahwa integrasi *spasial–temporal* dilakukan melalui transformasi fitur hasil deteksi kendaraan menjadi representasi sekuensial yang dioptimasi untuk memprediksi kemacetan di persimpangan jalan