

BAB V

PENUTUP

5.1. Kesimpulan

Penelitian ini secara fundamental bertujuan untuk merancang, mengimplementasikan, dan mengevaluasi sebuah sistem otomatis untuk estimasi abnormalitas pada spermatozoid manusia, dengan fokus spesifik pada analisis morfologi akrosom melalui pendekatan deep learning. Dengan memanfaatkan arsitektur Single Shot MultiBox Detector (SSD) yang diintegrasikan dengan backbone MobileNet, penelitian ini berhasil membangun sebuah model yang tidak hanya mampu mengidentifikasi spermatozoid dari citra mikroskopik tetapi juga mengklasifikasikannya secara real-time. Berdasarkan serangkaian eksperimen dan analisis mendalam yang telah diuraikan pada bab-bab sebelumnya, dapat ditarik beberapa kesimpulan esensial yang secara langsung menjawab rumusan masalah serta mengungkap wawasan teknis yang lebih mendalam.

5.1.1. Jawaban atas Rumusan Masalah

Bagian ini secara sistematis menyajikan jawaban definitif untuk setiap pertanyaan penelitian yang diajukan dalam Bab 1.2, dengan merujuk langsung pada bukti empiris dari metodologi dan hasil yang telah dipaparkan.

1. Jawaban Rumusan Masalah 1: Bagaimana menerapkan model pembelajaran deep learning dengan metode MobileNet dalam deteksi sperma berbasis video?

Implementasi model deep learning untuk deteksi sperma berbasis video berhasil direalisasikan melalui sebuah alur kerja (pipeline) yang terstruktur dan sistematis. Fondasi dari implementasi ini adalah adopsi arsitektur SSD-MobileNet V2 FPNLite, sebuah model yang dipilih secara strategis karena menawarkan keseimbangan optimal antara akurasi deteksi dan efisiensi komputasi, yang merupakan prasyarat krusial untuk aplikasi pemrosesan real-time.

Metodologi yang digunakan adalah transfer learning, di mana model yang telah melalui pra-pelatihan (pre-trained) pada dataset berskala besar dan umum, yaitu

COCO 2017, diadaptasi dan disesuaikan (fine-tuned) untuk tugas yang sangat spesifik: deteksi abnormalitas morfologi akrosom pada spermatozoid.

Proses implementasi ini dieksekusi dalam kerangka kerja TensorFlow Object Detection API dan mencakup beberapa tahapan kunci:

- Pengumpulan dan Persiapan Data: Dua dataset utama digunakan, yaitu VISEM (berbasis video) dan MHSMA (berbasis citra), yang menyediakan data visual untuk pelatihan dan validasi model.
- Anotasi Data: Proses pelabelan dilakukan secara manual menggunakan perangkat lunak khusus. LabelImg digunakan untuk data citra dari MHSMA, sementara OpenLabelling digunakan untuk menganotasi frame per frame dari dataset video VISEM. Proses ini menghasilkan ground truth berupa bounding box dan label kelas ('normal' atau 'abnormal') yang disimpan dalam format PASCAL VOC XML.
- Konversi ke Format TFRecord: Seluruh data citra dan anotasinya dikonversi ke dalam format TFRecord. Format biner ini dioptimalkan untuk operasi I/O (Input/Output) pada TensorFlow, sehingga secara signifikan mempercepat proses pemuatan data selama fase pelatihan.
- Konfigurasi dan Pelatihan: Seluruh parameter pelatihan, termasuk arsitektur model, learning rate, optimizer, dan jalur data, didefinisikan secara rinci dalam sebuah berkas konfigurasi (pipeline.config). Berkas ini menjadi cetak biru yang memandu proses fine-tuning model pra-pelatihan pada dataset sperma yang telah disiapkan.

Dengan demikian, dapat disimpulkan bahwa penerapan model deep learning telah berhasil dilaksanakan melalui sebuah pipeline yang komprehensif, mulai dari persiapan data hingga pelatihan model. Pemanfaatan transfer learning pada arsitektur SSD-MobileNet V2 FPNLite terbukti menjadi pendekatan yang efektif untuk mengadaptasi model visi komputer generalis ke dalam domain analisis mikroskopik yang sangat terspesialisasi, sehingga secara tuntas menjawab rumusan masalah pertama.

2. Jawaban Rumusan Masalah 2: Bagaimana menerapkan real-time detection menggunakan metode MobileNet SSD dalam deteksi sperma?

Penerapan deteksi objek menggunakan model MobileNet SSD diwujudkan melalui sebuah pipeline inferensi yang dikembangkan dengan bahasa pemrograman

Python dan memanfaatkan pustaka OpenCV sebagai komponen utamanya.¹ Alur kerja ini dirancang untuk memproses data video, baik dari berkas yang sudah ada maupun dari aliran kamera secara real-time, dengan menjalankan siklus inferensi pada setiap frame.

Proses inti inferensi untuk kedua mode implementasi adalah sebagai berikut:

- **Pemuatan Model:** Sistem menginisialisasi dan memuat checkpoint model TensorFlow terlatih (berkas .ckpt) yang berisi semua weight yang telah dipelajari untuk tugas deteksi abnormalitas sperma.
- **Proses Inferensi per Frame:** Setiap frame yang diakuisisi dari sumber video diperlakukan sebagai citra tunggal dan diteruskan ke model. Model melakukan proses feedforward untuk menghasilkan prediksi, yang mencakup koordinat bounding box dan label kelas ('normal' atau 'abnormal') beserta skor kepercayaannya.³
- **Visualisasi Hasil:** Hasil prediksi divisualisasikan pada frame asli dengan menggambar bounding box dan menambahkan label kelas menggunakan fungsi dari OpenCV.

Dengan demikian, penerapan deteksi real-time dan berbasis video berhasil diimplementasikan melalui dua alur kerja spesifik yang dibangun di atas fondasi yang sama: pemuatan model TensorFlow dan pemrosesan frame-by-frame menggunakan OpenCV. Berdasarkan alur kerja inti tersebut, implementasi dibagi menjadi dua mode operasional yang berbeda sesuai dengan kode sumber yang dikembangkan:

- **Inferensi pada Berkas Video:**

Mode ini dirancang untuk memproses berkas video yang telah direkam sebelumnya. Sistem akan membaca video frame demi frame, menjalankan proses inferensi pada setiap frame, dan kemudian menyusun kembali frame-frame yang telah dianotasi menjadi sebuah berkas video baru sebagai output. Mode ini sangat berguna untuk analisis post-hoc dan pengujian kuantitatif model pada dataset video standar.

- **Inferensi Real-Time dari Kamera:**

Mode ini dirancang untuk aplikasi interaktif dan demonstrasi langsung. Sistem secara kontinu menangkap frame dari kamera web yang terhubung. Setiap frame yang ditangkap langsung diproses melalui pipeline inferensi, dan hasilnya (video dengan bounding box dan label) ditampilkan secara langsung di layar. Ini membuat deteksi

yang terjadi secara real-time, mendemonstrasikan kemampuan model untuk beroperasi dalam lingkungan langsung.

Dengan demikian, penerapan deteksi real-time dan berbasis video berhasil diimplementasikan melalui dua alur kerja spesifik yang dibangun di atas fondasi yang sama: pemuatan model TensorFlow dan pemrosesan frame-by-frame menggunakan OpenCV. Perbedaan fungsional antara inferensi pada video dan deteksi real-time dari kamera secara tuntas menjawab bagaimana metode ini dapat diterapkan dalam skenario yang berbeda.

3. Jawaban Rumusan Masalah 3: Seberapa besar tingkat presisi, recall, dan F1-Score model menggunakan metode MobileNet SSD dalam deteksi abnormalitas sperma berdasarkan akrosom?

Analisis kinerja model menunjukkan bahwa performa model sangat bergantung pada durasi pelatihan, di mana titik optimal tercapai sebelum model mengalami overfitting yang parah. Model yang dilatih selama 5.000 langkah (steps) pada dataset VISEM teridentifikasi sebagai konfigurasi yang paling efektif dan seimbang. Berdasarkan evaluasi kualitatif manual yang ketat terhadap video uji standar yang belum pernah dilihat oleh model, metrik kinerja untuk model optimal ini adalah sebagai berikut:

- Presisi (Precision): 65.52%
- Recall: 57.58%
- F1-Score: 61.29%

Hasil ini secara empiris mengonfirmasi bahwa melanjutkan pelatihan melampaui 5.000 langkah justru bersifat kontraproduktif. Terjadi penurunan drastis pada nilai Recall, yang mengindikasikan bahwa model menjadi terlalu "ragu-ragu" dan gagal mendeteksi banyak objek yang seharusnya dikenali. Penurunan ini adalah manifestasi dari overfitting, di mana model kehilangan kemampuannya untuk menggeneralisasi pada data baru.

Dengan demikian, jawaban definitif untuk rumusan masalah ketiga adalah bahwa model SSD-MobileNet SSD mencapai tingkat kinerja praktis yang terukur dengan F1-Score sebesar 61.29%, presisi 65.52%, dan recall 57.58%. Kinerja ini dicapai pada kondisi pelatihan optimal (5.000 langkah), yang menyoroti pentingnya

menemukan keseimbangan yang tepat dalam proses pelatihan untuk memaksimalkan kapabilitas generalisasi model.

5.1.2. Implikasi Atas Penelitian

Berdasarkan serangkaian pengujian yang telah dilakukan, dapat ditarik beberapa implikasi penting.

1. Pertama, **kualitas dan konsistensi anotasi data lebih berpengaruh daripada kuantitas data semata**. Penggunaan dataset gabungan dengan gaya anotasi yang berbeda terbukti merusak proses belajar model.
2. Kedua, untuk dataset yang relatif homogen, model cenderung cepat mengalami **overfitting**. Observasi kualitatif dan kuantitatif menunjukkan bahwa performa deteksi terbaik (diukur dengan F1-Score) dicapai pada tahap training yang relatif awal (5.000 langkah). Melanjutkan training lebih jauh justru menurunkan kemampuan generalisasi model.
3. Ketiga, terdapat **diskrepansi yang jelas antara evaluasi metrik mAP dan performa praktis**. Sebuah temuan penting dalam penelitian ini adalah adanya perbedaan signifikan antara hasil evaluasi kuantitatif (metrik mAP dari TensorBoard yang sangat rendah) dan evaluasi kualitatif (observasi visual yang baik). Temuan ini menyoroti bahwa metrik evaluasi otomatis seperti mAP, meskipun merupakan standar industri, dapat memberikan gambaran yang menyesatkan dan tidak lengkap tanpa divalidasi oleh observasi manusia, terutama dalam kasus-kasus yang menantang.

Diskrepansi ini terjadi karena sifat ketat dari metrik evaluasi mAP. Pada kasus deteksi objek yang kecil, padat, dan sering tumpang tindih seperti sperma, sedikit saja pergeseran pada *bounding box* dapat menyebabkan nilai *Intersection over Union* (IoU) turun di bawah ambang batas 0.5. Akibatnya, meskipun secara visual deteksi terlihat benar oleh mata manusia, skrip evaluasi akan menilainya sebagai *False Positive*, yang secara drastis menurunkan skor mAP. Lebih jauh lagi, seperti yang ditunjukkan pada skenario dataset gabungan, mAP yang tinggi juga bisa menipu jika model hanya berkinerja baik pada sebagian kecil data yang mudah dikenali, sementara mengabaikan mayoritas data yang lebih sulit. Maka dari itu, Metrik mAP yang rendah tidak selalu

berarti model gagal total, melainkan bisa disebabkan oleh ketidakakuratan minor pada penempatan *bounding box* yang dihukum berat oleh ambang batas IoU.

Bisa ditarik kesimpulan, strategi optimal untuk kasus ini bukanlah sekadar memaksimalkan jumlah langkah training, melainkan menemukan "sweet spot" di mana model telah cukup belajar tanpa mulai menghafal. Berdasarkan hasil evaluasi, **checkpoint yang dihasilkan pada sekitar 5.000 langkah merupakan model yang paling efektif** untuk aplikasi deteksi sperma secara praktis. Untuk pengembangan di masa depan, fokus sebaiknya tidak pada penambahan langkah training, melainkan pada pengayaan **augmentasi data** untuk membuat proses training lebih "menantang" dan memaksa model untuk belajar fitur yang lebih robust, sehingga dapat menunda terjadinya overfitting.

5.2 Saran

Berdasarkan temuan dan kesimpulan yang diperoleh, terdapat beberapa saran yang dapat dijadikan acuan untuk penelitian selanjutnya:

1. **Standarisasi Anotasi Data:** Penelitian selanjutnya disarankan untuk lebih memperhatikan konsistensi gaya anotasi antar dataset, terutama saat melakukan penggabungan data dari berbagai sumber. Ketidaksesuaian dalam format atau kriteria anotasi terbukti dapat mengganggu proses pembelajaran model dan menurunkan kinerjanya secara signifikan.
2. **Eksplorasi Teknik Augmentasi Lanjut:** Untuk mengatasi overfitting pada dataset yang homogen, disarankan agar peneliti mengeksplorasi teknik augmentasi data yang lebih variatif dan kompleks. Augmentasi seperti *cutout*, *random erasing*, *color jitter*, atau *mixup* dapat digunakan untuk menciptakan skenario pelatihan yang lebih menantang, sehingga model belajar fitur yang lebih general dan robust.
3. **Evaluasi Multi-Metrik dan Validasi Visual:** Mengingat keterbatasan metrik mAP dalam mendeteksi objek kecil dan tumpang tindih, penelitian selanjutnya sebaiknya menggunakan pendekatan evaluasi yang lebih komprehensif. Kombinasi antara metrik kuantitatif (seperti F1-score, precision, recall) dan validasi kualitatif melalui observasi visual sangat dianjurkan untuk memperoleh penilaian yang lebih realistis terhadap performa model.
4. **Analisis Dinamis Checkpoint Training:** Alih-alih mengandalkan pelatihan

hingga langkah maksimum, pendekatan yang lebih adaptif dalam menentukan waktu terbaik untuk menyimpan model (*early stopping* berbasis evaluasi periodik) patut dipertimbangkan. Penelitian lebih lanjut dapat mengevaluasi bagaimana *sweet spot* pelatihan dapat diidentifikasi secara otomatis berdasarkan tren metrik evaluasi.

5. Peningkatan Post-Processing dan Toleransi IoU: Mengingat sensitivitas metrik mAP terhadap sedikit kesalahan posisi *bounding box*, pengembangan sistem post-processing seperti *non-maximum suppression* yang lebih toleran terhadap pergeseran kecil dapat menjadi solusi. Alternatif lainnya adalah melakukan eksperimen dengan ambang batas IoU yang lebih fleksibel atau dinamis, disesuaikan dengan karakteristik objek yang diamati.
6. Ekspansi Dataset dan Diversifikasi Kasus: Untuk meningkatkan generalisasi model, disarankan untuk memperluas cakupan dataset, termasuk variasi morfologi sel sperma dari berbagai individu, kondisi medis, atau teknik pengambilan gambar. Diversifikasi ini dapat membantu model belajar mengenali lebih banyak variasi bentuk dan latar belakang yang kompleks.

Halaman ini sengaja dikosongkan