

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan, penelitian ini menghasilkan beberapa kesimpulan sebagai berikut.

1. *Mixed-Precision QAT* meningkatkan efisiensi komputasi pada aplikasi konversi SDR ke HDR *real-time* karena mengurangi beban presisi model yang dijalankan melalui TensorRT pada GPU NVIDIA. Penurunan presisi pada sebagian layer membuat kebutuhan memori, bandwidth, dan operasi inferensi menjadi lebih ringan dibanding FP32. Resolusi proses juga berpengaruh langsung terhadap efisiensi karena resolusi yang lebih rendah menghasilkan ukuran *tensor* dan jumlah operasi per *frame* yang lebih kecil. Hal ini terlihat dari peningkatan FPS pada 720p dibanding 1080p pada seluruh konfigurasi model.
2. *Mixed-Precision QAT* memberikan *trade-off* paling seimbang antara efisiensi dan akurasi visual. Konfigurasi ini mampu mencapai *playback real-time* dengan 56,54 FPS pada 720p dan 31,67 FPS pada 1080p, serta tetap menjaga kualitas hasil konversi dalam batas toleransi yang ditetapkan. Dibanding INT8 Full, *Mixed-Precision QAT* sedikit lebih lambat, tetapi lebih aman karena layer yang sensitif terhadap tone, warna, *highlight*, dan *output* akhir tetap dipertahankan pada presisi yang lebih tinggi.

5.2 Saran

Berdasarkan keterbatasan dan hasil penelitian, beberapa saran untuk pengembangan selanjutnya adalah sebagai berikut.

1. Aplikasi konversi SDR ke HDR *real-time* yang dikembangkan pada penelitian ini dapat digunakan kembali sebagai platform pengujian untuk model atau konfigurasi optimisasi lain. Pengembangan selanjutnya dapat dilakukan dengan mengganti model, *checkpoint*, atau TensorRT engine yang digunakan, tanpa perlu membangun ulang seluruh *pipeline playback*, compare, *benchmark*, dan logging dari awal.
2. Evaluasi akurasi dapat dikembangkan dengan menambahkan fitur temporal testing pada aplikasi. Pada penelitian ini, *benchmark* akurasi masih berfokus pada evaluasi *frame-level* menggunakan PSNR, SSIM, DeltaEITP, dan HDR-VDP3, sehingga belum secara khusus mengukur flicker, perubahan tone antar-*frame*, atau konsistensi visual selama *playback*. Fitur temporal testing dapat menambahkan perhitungan perbedaan antar-*frame*, visualisasi stabilitas temporal, serta logging metrik temporal untuk membandingkan konfigurasi model secara lebih lengkap.

3. Penelitian selanjutnya dapat mengevaluasi pendekatan *Mixed-Precision QAT* pada berbagai arsitektur GPU NVIDIA, versi CUDA, dan TensorRT yang berbeda untuk mengkaji konsistensi peningkatan efisiensi inferensi pada lingkungan *deployment* yang lebih beragam.
4. Pengujian dapat diperluas pada resolusi yang lebih tinggi, seperti 4K, serta pada varian HDRTVNet++ dengan Hallucination Generator. Pengujian ini dapat menunjukkan apakah *Mixed-Precision QAT* tetap memberikan *trade-off* yang baik ketika beban komputasi dan kebutuhan memori meningkat.
5. Recipe *QAT* dapat dikembangkan lebih lanjut dengan variasi loss, strategi pemilihan data, dan skema proteksi layer yang lebih adaptif. Pengembangan ini penting untuk menjaga akurasi warna, luminansi, dan *highlight* pada konten dengan karakter visual yang lebih ekstrem.
6. Penelitian selanjutnya dapat menambahkan uji persepsi subjektif pengguna untuk melengkapi metrik objektif. Hal ini penting karena kualitas HDR tidak hanya dipengaruhi oleh kedekatan numerik terhadap ground truth, tetapi juga oleh persepsi visual pengguna pada display HDR yang sebenarnya.