

1. Arsitektur CNN dengan X lapisan konvolusi menghasilkan akurasi training 98% tetapi akurasi validasi 62%. Jelaskan fenomena vanishing gradient yang mungkin terjadi pada lapisan awal, dan bagaimana cara memitigasinya! Mengapa penambahan Batch Normalization setelah lapisan konvolusi ke-Y justru memperburuk generalisasi, serta strategi alternatif untuk menstabilkan pembelajaran?

= Vanishing gradient terjadi ketika gradien yang dihitung selama backpropagation menjadi sangat kecil di lapisan awal CNN, menyebabkan pembaruan bobot yang tidak signifikan dan menghambat pembelajaran fitur dasar. Hal ini sering terjadi pada arsitektur dalam dengan aktivasi seperti sigmoid atau tanh, yang memiliki rentang gradien sempit. Untuk memitigasi, penggunaan fungsi aktivasi ReLU/Leaky ReLU, inisialisasi bobot (He/Xavier), atau teknik residual connection (seperti di ResNet) dapat menjaga gradien tetap stabil. Namun, penambahan Batch Normalization (BN) setelah lapisan konvolusi ke-Y bisa memperburuk generalisasi jika BN memperkenalkan noise berlebihan atau mengurangi kemampuan model beradaptasi dengan variasi data uji. Alternatifnya, penggunaan Layer Normalization, dropout selektif, atau mengurangi kompleksitas model lebih efektif menstabilkan pembelajaran tanpa over-regularisasi.

2. Ketika melatih CNN dari nol, loss training stagnan di nilai tinggi setelah XXX(3 digit epoch) epoch. Identifikasi tiga penyebab potensial terkait laju pembelajaran (learning rate), inisialisasi berat, atau kompleksitas model! Mengapa penggunaan Cyclic Learning Rate dapat membantu model keluar dari local minima, dan bagaimana momentum pada optimizer SGD memengaruhi konvergensi?

= Stagnasi loss training disebabkan oleh (1) learning rate terlalu tinggi/rendah (gradien "melompat" atau tidak konvergen), (2) inisialisasi bobot tidak optimal (misalnya, semua bobot awal nol menghambat propagasi sinyal), atau (3) kompleksitas model tidak sesuai (terlalu sederhana untuk menangkap pola data). Cyclic Learning Rate (CLR) membantu dengan mengubah learning rate secara periodik, memungkinkan model "melarikan diri" dari local minima dengan meningkatkan gradien sementara. Momentum pada SGD mempercepat konvergensi dengan mengakumulasi arah gradien konsisten sekaligus mengurangi osilasi, sehingga membantu navigasi permukaan loss yang kompleks.

3. Pada klasifikasi spesies ikan menggunakan CNN, penggunaan fungsi aktivasi ReLU tidak menunjukkan peningkatan akurasi setelah 50 epoch, meskipun learning rate telah dioptimasi. Jelaskan fenomena dying ReLU yang mungkin terjadi dan bagaimana hal ini mengganggu aliran gradien selama backpropagation!

= Dying ReLU terjadi ketika neuron hanya menghasilkan output nol untuk semua input, menyebabkan gradiennya nol selama backpropagation. Akibatnya, neuron tersebut tidak lagi memperbarui bobot dan menjadi tidak aktif permanen, mengurangi kapasitas model. Fenomena ini mengganggu aliran gradien karena lapisan berikutnya kehilangan informasi dari neuron "mati", sehingga pembelajaran terhambat. Solusinya termasuk menggunakan Leaky ReLU/Parametric ReLU (memperkenalkan slope kecil untuk input negatif) atau monitoring distribusi aktivasi untuk menyesuaikan learning rate.

4. Pada pelatihan CNN untuk klasifikasi XX spesies ikan, grafik AUC-ROC menunjukkan satu kelas (Spesies X) stagnan di 0.55 sementara kelas lain mencapai  $>0.85$  setelah YYY epoch. Analisis mengapa class-weighted loss function gagal meningkatkan kinerja Spesies X, dan identifikasi tiga faktor penyebab potensial terkait karakteristik data dan arsitektur model!

= Class-weighted loss mungkin gagal meningkatkan kinerja Spesies X karena (1) ketidakseimbangan ekstrem (jumlah sampel Spesies X sangat kecil bahkan setelah penimbangan), (2) fitur Spesies X ambigu atau tumpang tindih dengan kelas lain (misalnya, pola warna atau bentuk mirip), atau (3) noise pada label/data (misalnya, anotasi salah). Selain itu, arsitektur model mungkin tidak cukup dalam mengekstrak ciri unik Spesies X, atau pooling agresif menghilangkan detail kritis. Pendekatan alternatif seperti data augmentation spesifik kelas, feature engineering, atau ensembling model bisa lebih efektif.

5. Pada arsitektur CNN untuk klasifikasi ikan, peningkatan kompleksitas model justru menyebabkan penurunan akurasi validasi dari 85% ke 65%, meskipun akurasi training mencapai 98%. Jelaskan fenomena overfitting yang terjadi, dan mengapa penambahan kapasitas model tidak selalu meningkatkan generalisasi! Identifikasi 3 kesalahan desain arsitektur yang memicu degradasi performa.

= Overfitting terjadi ketika model terlalu kompleks sehingga menghafal noise/data training alih-alih mempelajari pola umum. Penambahan kapasitas model (lapisan/filter berlebih) tanpa regularisasi meningkatkan risiko ini karena parameter lebih banyak menangkap variasi nonsensial. Tiga kesalahan desain umum: (1) Kurangnya mekanisme regularisasi (dropout, L2 regularization), (2) penggunaan pooling berlebihan yang merusak informasi spasial relevan, dan (3) kedalaman tidak seimbang tanpa shortcut connection (seperti skip connections di ResNet) yang memperparah vanishing gradient. Solusinya melibatkan optimasi trade-off kapasitas model, augmentasi data, dan integrasi teknik regularisasi adaptif.