

Nama: Alfikri
NIM: 1103223015
Kelas: TK-46-06

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?

Jawaban:

Fenomena vanishing gradient pada lapisan awal arsitektur CNN yang memiliki banyak lapisan konvolusi bisa terjadi ketika gradien yang dihitung selama backpropagation menjadi sangat kecil saat sampai di lapisan pertama, menyebabkan pembaruan parameter yang sangat kecil dan pelatihan menjadi lambat atau terhenti. Ini sering terjadi pada fungsi aktivasi seperti sigmoid atau tanh yang cenderung mengurangi nilai gradien pada lapisan-lapisan awal. Untuk mengatasi hal ini, kita bisa menggunakan fungsi aktivasi yang lebih stabil seperti ReLU yang tidak mengalami masalah vanishing gradient yang signifikan. Penambahan Batch Normalization setelah lapisan konvolusi ke-Y justru bisa memperburuk generalisasi karena bisa memperkenalkan ketergantungan yang lebih besar pada distribusi mini-batch, yang mungkin berkontribusi pada model yang terlalu terikat pada data training dan kurang generalizable. Alternatifnya, kita bisa menggunakan teknik seperti Dropout atau memperkenalkan model dengan arsitektur residual (seperti ResNet) untuk mengatasi masalah ini.

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?

Jawaban:

Stagnasi loss training pada nilai tinggi setelah banyak epoch bisa disebabkan oleh tiga hal utama: learning rate yang terlalu tinggi yang membuat model melompati minimum global, inisialisasi berat yang buruk sehingga model memulai pelatihan pada titik yang kurang optimal, atau kompleksitas model yang tidak sesuai dengan data. Cyclic Learning Rate (CLR) dapat membantu model keluar dari local minima dengan secara berkala meningkatkan dan menurunkan learning rate, sehingga memberikan peluang bagi model untuk melompat keluar dari minimum lokal. Selain itu, momentum dalam optimizer SGD membantu mempercepat konvergensi dengan mengurangi fluktuasi dan memberikan dorongan tambahan dalam arah yang tepat berdasarkan momentum dari pembaruan sebelumnya.

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!

Jawaban:

Dying ReLU adalah fenomena yang terjadi ketika unit-unit dalam jaringan CNN mengalami "mati" dan outputnya menjadi nol terus menerus. Ini biasanya disebabkan oleh inisialisasi berat yang buruk atau learning rate yang terlalu besar, yang dapat menyebabkan neuron tidak aktif selama pelatihan. Ketika hal ini terjadi, gradien tidak dapat mengalir melalui neuron yang mati selama backpropagation, menghambat pembelajaran. ReLU yang mati akan mengganggu aliran gradien dan menghalangi pembaruan parameter model, sehingga menghambat pelatihan dan memengaruhi performa model.

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!

Jawaban:

Ketika menggunakan fungsi loss berbobot kelas (class-weighted loss) dan kelas Spesies X stagnan pada nilai AUC-ROC 0.55, sementara kelas lainnya mencapai lebih dari 0.85, bisa jadi ada ketidakseimbangan yang lebih mendalam dalam data atau masalah terkait model. Fungsi loss berbobot kelas dirancang untuk mengatasi masalah ketidakseimbangan kelas, tetapi bisa gagal jika distribusi data terlalu terdistorsi atau jika model tidak cukup kuat untuk menangani data tersebut. Faktor-faktor potensial yang menyebabkan masalah ini antara lain ketidakseimbangan jumlah data untuk setiap kelas, arsitektur model yang tidak cukup kompleks untuk menangani variasi dalam data, atau data yang memiliki noise tinggi yang membuat model kesulitan mempelajari pola yang relevan.

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

Jawaban:

Overfitting terjadi ketika model terlalu kompleks untuk dataset yang diberikan, sehingga meskipun akurasi training sangat tinggi, akurasi validasi turun drastis. Hal ini menunjukkan bahwa model tidak mampu menggeneralisasi dengan baik ke data yang tidak terlihat dan terjebak pada fitur yang hanya relevan untuk data

training. Menambah kapasitas model dengan menambah lapisan atau neuron tidak selalu meningkatkan generalisasi. Tiga kesalahan desain arsitektur yang dapat memicu overfitting adalah: 1) Tidak cukupnya data untuk mendukung kompleksitas model, 2) Tidak adanya regularisasi (seperti Dropout), dan 3) Terlalu banyak parameter yang menyebabkan model mempelajari noise daripada pola yang berguna.