Link Youtube: https://www.youtube.com/playlist?list=PL-eTPpRjlf10OnC3jwTjkwV2pCqb2lnrR

Computer Vision (Chapter 1-4)

1. Chapter 1

   Pada bagian feature matching diantara empat metode yaitu brute-froce search, brute-force with ORB, FLANN, dan LoFTR memiliki tantangan dan solusi masing-masing.

   Pada brute-force memiliki tantangan di proses pencocokan dengan membandingkan setiap fitur dari satu gambar dengna semua fitur di gambar ke 2 kurang cocok dengan gambar yang memiliki banyak fitur dan tidak efisien karena memakan banyak sumber daya komputasi, untuk solusinya menggunakan GPU atau parelisasi untuk mempercepat proses pencocokan fitur dan membatasi pasangan fitur yang dipertimbangan cocok hanya jika jaraknya lebih kecil dari nilai tertentu.

   Pada brute-force with ORB memiliki tantangan pada bagian kecocokan karena ORB menggunakan descriptor biner sehingga lebih sensitive terhadap noice atau perubahan pencahayaan, dan memiliki perfoma turun jika fiturnya lebih kompleks. Untuk solusinya yaitu menggunakan thresholdnya lebih adaptif dan mengkombinasi ORB dengan algoritma lain seperti LoFTR.

   Untuk FLANN memiliki tantangan pada parameter tuning yang rumit karena pencariannya pada pencocokan fiturnya membutuhkan konfigurasi manual atau adaptif untuk performa optimal, dan terkadang menghasilkan false positives. Solusinya menggunakan auto parameter tuning pada fungsi bawaan FLANN dan melakukan cross-checking agar meminimalisir false positive.

   Untuk LoFTR memiliki tantangan pada kebutuhan sumber daya komputasi yang tinggi dan cenderung overfitting. Solusinya menggunakan model pre-trained dan melakukan fine tuning.

2. Chapter 2

   Pada bagian CNN terdapat beberapa jenis model CNN seperti VGG19,Googlenet,mobilenet, Mobilenet with Timm(Transfer learning), dan Resnet memiliki tantangan dan solusi masing-masing.

   Pada VGG19 memiliki tantangan di ukuran model yang besar dan lambat karena memiliki 19 hidden layer sehingga lambat serta membutuhkan banyak memori dan daya komputasi, dan karena modelnya berukuran besar menjadi model VGG19 rentan terhadap overfitting. Solusinya yaitu menggunakan transfer learning seperti memakai model pre-trained dan fine-tune , menggunkana GPU yang lebih canggih atau Teknik parelisasi agar mempercepat training, dan melakukan kompresi model.

   Pada GoogLeNet memiliki tantangan pada arsitektur kompleks karena berbasis inception modul sehingga memerlukan desain khusus lebih rumit. Solusinya yaitu menggunakan versi model pre-trained Googlenet agar mempercepat training, dan menggunakan Automated

archicture search seperti Neural Architecture(NAS) agar dapat menemukan konfigurasi optimal secara otomatis.

Untuk MobileNet memiliki tantangan pada perfoma lebih rendah di data kompleks karena MobileNet dirancang untuk keterbatasan sumber daya seperi smartphone, dan tuning hyperparameter. Solusinya yaitu menggunakan transfer learning, fine-tuning depth/width multiplier, dan menggabungkan MobileNet dengan model lain agar dapat meningkatkan perfoma pada tugas-tugas pada fitur yang lebih kompelks,

Untuk MobileNet with Timm(Transfer Learning) memiliki tantangan pada kompleksitas integrasi, dan rtade-off antara kecepatan dan akurasi. Solusinya memilih varian MobileNet yang lebih kecil untuk device yang keterbatasan sumber daya, dan memanfaatkan fungsi-fungsi bawaan Timm untuk augmentasi, scheduler, dan pretrained weights untuk mempercepat pengembangan model.

Untuk ResNet memiliki tantangan pada latensi tinggi, Kebutuhan memori, degradasi performa untuk arsitektur yang sangat dalam karena bisa kesulitan dalam propagasi gradien. Solusinya yaitu menggunakan jenis model varian ResNet yang lebih kecil , menggunakan blok residual dengan bottleneck untuk mengurangi parameter, dan menggunakan gradient clipping untuk mengatasi vanishing gradients pada model dalam.

3. Chapter 3

Pada bagian Vision Transformer memiliki jenis atau metode Vision Transformer seperti Swin transformer, CvT, DiNAT, Vision transformer for object detection, DETR, transformer-based image segmentation memiliki tantangan dan solusi masing-masing.

Pada Swin Transfor memiliki tantangan pada kompleksitas komputasi karena Swin Transformer memanfaatkan shifted windows dan kesulitan training pada dataset kecil yang akan cenderung overfitting. Solusinya yaitu melakukan fine-tuning pada model pre-trained dan Hybrid modelling.

Pada CvT memiliki tantangan pada optimasi parameter dan Integrasi CNN dan Transformer karena desain CvT lebih rumit. Solusinya menggunakan automated hyperparameter tuning agar dapat menemukan kombinasi parameter yang optimal.

Pada DiNAT memiliki tantangan pada efisiensi memori karena arsitektur attention berbasis neighborhood membutuhkan pengelolaan memori yang baik pada resolusi tinggi. Solusinya menggunakan impelementasi Patch-wise attention agar mengurangi kebutuhan memori dan melakukan Task-specific tuning.

Untuk Vision Transformer for Object Detection memiliki tantangan pada Skabilitas di resolusi tinggi karena Vision Transformer tidak efisien jika input gambarnya yang beresolusi tinggi sehingga memerlukan banyak komputasi, dan label data yang memadai karena object detection memerlukan anotasi yang presisi. Solusinya menggunakan Teknik Data augmentation seperti cropping, dan flipping, dan menggunakan model Hierarchical ViT.

Untuk Transformer-Based Image Segmentation memiliki tantangan pada generalisasi model dan resolusi tinggi sehingga membutuhan sumber daya tinggi pada memproses gambar besar. Solusinya yaitu training pada gambar yang beresolusi rendah dan pendekatan multitask untuk melatih model pada berbagai dataset.

4. Chapter 4

Pada Multimodal Models memiliki jenis atau metode multimodal models seperti Multimodal Task and Models, CLIP,BLIP,OWL-VIT memiliki tantangan dan solusi masing-masing.

Pada Multimodal Tasks and Models memiliki tantangan pada integrasi data multimodal dan Efisiensi model karena menggabungkan informasi dari modalitas memerlukan Teknik yang kompleks. Solusinya menggunakan transfkrmer multi modal dan memakai tenik pengurangan parameter untuk meningkatkan efisiensi.

Pada CLIP memiliki tantangan pada kemampuan generalisasi, ketergantungan pada dataset besar, dan ketidakseimbangan modalitas. Solusinya melakukan fine-tuning, menggunakan augmentation data dan menggunakan dataset yang open-source.

Pada BLIP memiliki tantangan pada Efisiensi training, penggunaan data multimodal karena membutuhkan anotasi yang berkualitas tinggi. Solusinya menggunakan transfer learning dari model BLIP pre-trained, menggunakan semi-supervised atau auto dataset annotation.

Untuk OWL-ViT memiliki tantangan pada deteksi objek berbasis teks memerlukan pemetaan yang akurat antara deskripsi teks dan lokasi objek dalam gambar, dan ketergantungan pada data deskriptif. Solusinya menggunakan dataset yang beragam untuk meningkatkan generalisasi pemetaan teks ke objek.