Deep Learning: Modern Artificial Vision

Jaume Ivars Grimalt

March 2025

Contents

Acknowledgements

A special thanks to all those who contributed to this work...

Preface

This is the preface of the document...

I Foundations of Modern Computer Vision

Chapter 1

What is Computer Vision?

1.1 Baisic Concepts

Computer vision is a field of artificial intelligence that focuses on enabling machines to interpret and understand visual information from the world, similar to how humans do. It involves the development of algorithms and models that can analyze images and videos, extract meaningful features, and make decisions based on visual data. Computer vision has applications in various domains, including autonomous vehicles, medical imaging, surveillance, robotics, and augmented reality.

1.2 The Role of Deep Learning in Computer Vision

Chapter 2

Convolutional Neural Networks (CNNs)

2.1 The Basics of Convolution

2.2 Pooling Layers

2.3 Wide Convolutions and Point-wise Convolutions

2.4 Activation Functions

Chapter 3

Optimization in Deep Learning

3.1 Gradient Descent and Backpropagation

3.2 Optimizers

3.3 Challenges in Optimization

Chapter 4

Building Blocks of Modern CNNs

4.1 Batch Normalization

4.2 UpSampling and Transposed Convolutions

4.3 Gradient Flow Through Layers

II Core Architectures and Building Blocks

Chapter 5

Backbone Networks

5.1 CNN-based Backbones

5.1.1 ResNet

5.1.2 ResNeXt

5.1.3 EfficientNet

5.1.4 MobileNet

5.2 Introduction to Transformers in Vision

5.2.1 Vision Transformer

5.3 Advanced Transformer Backbones

5.3.1 Swin Transformer

Chapter 6

Neck Networks

6.1 FPNs

6.2 PANs

6.3 BiFPN

Chapter 7

Head Networks

7.1 TOOD

Chapter 8

Attention Mechanisms

8.1 Attention with CNNs

8.1.1 Deformable Convolutions

8.2 Attention with ViTs

III Core Tasks in Computer Vision

Chapter 9

Image Classification

Chapter 10

Object Detection

10.1 Key Models

10.1.1 YOLO

10.1.2 R-CNN

10.1.3 Faster R-CNN

10.1.4 DETR

10.1.5 RT-DETR

10.1.6 RetinaNet

10.2 The State of the Art Models

10.2.1 YOLOv12

10.2.2 CO-DETR

Chapter 11

Semantic Segmentation

11.1 Key Models

11.1.1 U-Net

11.1.2 U2-Net

11.1.3 SegFormer

11.2 The State of the Art Models

Chapter 12

Density Map Estimation

12.1 Key Models

12.1.1 CSRNet

12.1.2 Cascaded CSRNet

12.2 The State of the Art Models

IV Production Deployment

Chapter 13

Model Optimization

13.1 Model Compression

13.1.1 Quantization

13.1.2 Pruning

13.1.3 Knowledge Distillation

13.1.4 Neural Architecture Search

13.2 Model Acceleration

13.2.1 TensorRT

13.2.2 ONNX Runtime

13.2.3 OpenVINO

13.2.4 TensorFlow Lite

13.3 Model Deployment

13.3.1 TensorFlow Serving

13.3.2 TorchServe

13.3.3 NVIDIA Triton Inference Server

13.3.4 MLflow

13.3.5 Kubeflow

13.4 Model Monitoring

13.4.1 Prometheus

13.4.2 Grafana

V Explainability and Interpretability

Chapter 14

Explainability in Computer Vision

14.1 Introduction to Explainability

14.1.1 What is Explainability?

14.1.2 Why is Explainability Important?

14.1.3 Types of Explainability

14.1.4 Challenges in Explainability

14.2 Methods for Explainability

14.2.1 Saliency Maps

14.2.2 Grad-CAM

14.2.3 Integrated Gradients

14.2.4 LIME

14.2.5 SHAP

14.3 Interpretable Models

14.3.1 Decision Trees

14.3.2 Rule-Based Models

14.3.3 Linear Models

14.3.4 Prototype-Based Models

14.4 Evaluating Explainability

14.4.1 Quantitative Evaluation