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 Generative Models and Advanced Applications

Chapter 13

Generative Adversarial Networks (GANs)

13.1 Generative Adversarial Networks

13.2 CycleGAN

13.3 Pix2Pix

Chapter 14

Style Transfer

14.1 Neural Style Transfer

14.2 Fast Neural Style Transfer

Chapter 15

Diffusion Models

15.1 Stable Diffusion

Chapter 16

Text-to-Image and Layout-to-Image Diffusion Models

16.1 Text-to-Image Diffusion Models

16.2 Layout-to-Image Diffusion Models

V Production Deployment

Chapter 17

Model Optimization

17.1 Model Compression

17.1.1 Quantization

17.1.2 Pruning

17.1.3 Knowledge Distillation

17.1.4 Neural Architecture Search

17.2 Model Acceleration

17.2.1 TensorRT

17.2.2 ONNX Runtime

17.2.3 OpenVINO

17.2.4 TensorFlow Lite

17.3 Model Deployment

17.3.1 TensorFlow Serving

17.3.2 TorchServe

17.3.3 NVIDIA Triton Inference Server

17.3.4 MLflow

17.3.5 Kubeflow

17.4 Model Monitoring

17.4.1 Prometheus

17.4.2 Grafana

VI Explainability and Interpretability

Chapter 18

Explainability in Computer Vision

18.1 Introduction to Explainability

18.1.1 What is Explainability?

18.1.2 Why is Explainability Important?

18.1.3 Types of Explainability

18.1.4 Challenges in Explainability

18.2 Methods for Explainability

18.2.1 Saliency Maps

18.2.2 Grad-CAM

18.2.3 Integrated Gradients

18.2.4 LIME

18.2.5 SHAP

18.3 Interpretable Models

18.3.1 Decision Trees

18.3.2 Rule-Based Models

18.3.3 Linear Models

18.3.4 Prototype-Based Models

18.4 Evaluating Explainability

18.4.1 Quantitative Evaluation