



Computer Vision Projects with PyTorch

Design and Develop Production-Grade
Models

Akshay Kulkarni
Adarsha Shivananda
Nitin Ranjan Sharma

Apress®

Computer Vision Projects with PyTorch

**Design and Develop
Production-Grade Models**

**Akshay Kulkarni
Adarsha Shivananda
Nitin Ranjan Sharma**

Apress®

Computer Vision Projects with PyTorch: Design and Development-Grade Models

Akshay Kulkarni
Bangalore, Karnataka, India

Adarsha Shivananda
Hosanagara tq, Shimoga dt, Karnataka, India

Nitin Ranjan Sharma
Bangalore, India

ISBN-13 (pbk): 978-1-4842-8272-4
<https://doi.org/10.1007/978-1-4842-8273-1>

ISBN-13 (electronic): 978-1-4842-8273-1

Copyright © 2022 by Akshay Kulkarni, Adarsha Shivananda, and Nitin Ranjan Sharma

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Celestin Suresh John
Development Editor: Laura Berendson
Coordinating Editor: Shrikant Vishwakarma
Copy Editor: Kezia Endsley

Cover designed by eStudioCalamar

Cover image designed by Freepik (www.freepik.com)

Distributed to the book trade worldwide by Springer Science+Business Media New York, 1 New York Plaza, Suite 4600, New York, NY 10004-1562, USA. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail booktranslations@springernature.com; for reprint, paperback, or audio rights, please e-mail bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at <http://www.apress.com/bulk-sales>.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/978-1-4842-8272-4. For more detailed information, please visit <http://www.apress.com/source-code>.

Printed on acid-free paper

To our families.

Table of Contents

About the Authors	xi
About the Technical Reviewer	xiii
Introduction	xv
Chapter 1: The Building Blocks of Computer Vision	1
What Is Computer Vision	3
Applications.....	3
Channels.....	8
Convolutional Neural Networks	12
Understanding the CNN Architecture Type.....	23
Working with Deep Learning Model Techniques.....	33
Introduction to PyTorch.....	37
Summary.....	41
Chapter 2: Image Classification	43
Topics to Cover.....	44
Defining the Problem.....	44
Overview of the Approach.....	45
Creating an Image Classification Pipeline.....	46
First Basic Model.....	47
Data	47

TABLE OF CONTENTS

Data Exploration	51
Data Loader	53
Define the Model	55
The Training Process	63
The Second Variation of Model	69
The Third Variation of the Model.....	72
The Fourth Variation of the Model	80
Summary.....	83
Chapter 3: Building an Object Detection Model	85
Object Detection Using Boosted Cascade	86
R-CNN	90
The Region Proposal Network	92
Fast Region-Based Convolutional Neural Network.....	96
How the Region Proposal Network Works	98
The Anchor Generation Layer	101
The Region Proposal Layer	102
Mask R-CNN.....	103
Prerequisites	103
YOLO.....	104
YOLO V2/V3.....	106
Project Code Snippets.....	108
Step 1: Getting Annotated Data	110
Step 2: Fixing the Configuration File and Training	112
The Model File	113
Summary.....	128

Chapter 4: Building an Image Segmentation Model	129
Image Segmentation.....	130
Pretrained Support from PyTorch.....	133
Semantic Segmentation	133
Instance Segmentation.....	137
Fine-Tuning the Model	139
Summary.....	166
Chapter 5: Image-Based Search and Recommendation System	167
Problem Statement	168
Approach and Methodology	168
Implementation	170
The Dataset	171
Installing and Importing Libraries.....	172
Importing and Understanding the Data	173
Feature Engineering	178
Calculating Similarity and Ranking.....	186
Visualizing the Recommendations	189
Taking Image Input from Users and Recommending Similar Products	193
Summary.....	198
Chapter 6: Pose Estimation	199
Top-Down Approach.....	200
Bottom-Up Approach.....	200
OpenPose	201
Branch-1	201
Branch-2.....	201

TABLE OF CONTENTS

HRNet (High-Resolution Net).....	204
Higher HRNet	207
PoseNet.....	208
How Does PoseNet Work?	209
Pros and Cons of PoseNet	213
Applications of Pose Estimation	213
Test Cases Performed Retail Store Videos.....	214
Implementation	214
Step 1: Identify the List of Human Keypoints to Track	216
Step 2: Identify the Possible Connections Between the Keypoints.....	217
Step 3: Load the Pretrained Model from the PyTorch Library	218
Step 4: Input Image Preprocessing and Modeling	219
Step 5: Build Custom Functions to Plot the Output.....	220
Step 6: Plot the Output on the Input Image.....	223
Summary.....	227
Chapter 7: Image Anomaly Detection	229
Anomaly Detection.....	230
Approach 1: Using a Pretrained Classification Model	232
Step 1: Import the Required Libraries.....	233
Step 2: Create the Seed and Deterministic Functions	234
Step 3: Set the Hyperparameter	234
Step 4: Import the Dataset.....	235
Step 5: Image Preprocessing Stage	235
Step 6: Load the Pretrained Model	239
Step 7: Freeze the Model.....	241
Step 8: Train the Model.....	242
Step 9: Evaluate the Model.....	245

Approach 2: Using Autoencoder	255
Step 1: Prepare the Dataset Object	255
Step 2: Build the Autoencoder Network.....	256
Step 3: Train the Autoencoder Network	257
Step 4: Calculate the Reconstruction Loss Based on the Original Data	258
Step 5: Select the Most Anomalous Digit Based on the Error Metric Score	258
Output.....	258
Summary.....	259
Chapter 8: Image Super-Resolution.....	261
Up-Scaling Using the Nearest Neighbor Concept	263
Understanding Bilinear Up-Scaling	265
Variational Autoencoders	265
Generative Adversarial Networks.....	271
The Model Code	273
Model Development.....	274
Imports	275
Running the Application	291
Summary.....	294
Chapter 9: Video Analytics	297
Problem Statement	298
Approach.....	301
Implementation	302
Data	303
Uploading the Required Videos to Google Colab.....	304
Convert the Video to a Series of Images.....	304
Image Extraction.....	306

TABLE OF CONTENTS

Data Preparation.....306

Identify the Hotspots in a Retail Store308

Importing Images312

Getting Crowd Counts.....312

Security and Surveillance.....316

Identify the Demographics (Age and Gender).....321

Summary.....322

Chapter 10: Explainable AI for Computer Vision.....325

Grad-CAM.....326

Grad-CAM++.....328

NBDT330

 Step 1331

 Step 2332

 Steps 3 and 4332

Grad-CAM and Grad-CAM++ Implementation333

 Grad-CAM and Grad-CAM++ Implementation on a Single Image333

 NBDT Implementation on a Single Image.....337

Summary.....340

Index.....341