
Table of Contents

Foreword.....	xv
---------------	----

Preface.....	xvii
--------------	------

Part I. Introduction to Generative Deep Learning

1. Generative Modeling.....	3
What Is Generative Modeling?	4
Generative Versus Discriminative Modeling	5
The Rise of Generative Modeling	6
Generative Modeling and AI	8
Our First Generative Model	9
Hello World!	9
The Generative Modeling Framework	10
Representation Learning	12
Core Probability Theory	15
Generative Model Taxonomy	18
The Generative Deep Learning Codebase	20
Cloning the Repository	20
Using Docker	21
Running on a GPU	21
Summary	21
2. Deep Learning.....	23
Data for Deep Learning	24
Deep Neural Networks	25

What Is a Neural Network?	25
Learning High-Level Features	26
TensorFlow and Keras	27
Multilayer Perceptron (MLP)	28
Preparing the Data	28
Building the Model	30
Compiling the Model	35
Training the Model	37
Evaluating the Model	38
Convolutional Neural Network (CNN)	40
Convolutional Layers	41
Batch Normalization	46
Dropout	49
Building the CNN	51
Training and Evaluating the CNN	53
Summary	54

Part II. Methods

3. Variational Autoencoders.....	59
Introduction	60
Autoencoders	61
The Fashion-MNIST Dataset	62
The Autoencoder Architecture	63
The Encoder	64
The Decoder	65
Joining the Encoder to the Decoder	67
Reconstructing Images	69
Visualizing the Latent Space	70
Generating New Images	71
Variational Autoencoders	74
The Encoder	75
The Loss Function	80
Training the Variational Autoencoder	82
Analysis of the Variational Autoencoder	84
Exploring the Latent Space	85
The CelebA Dataset	85
Training the Variational Autoencoder	87
Analysis of the Variational Autoencoder	89
Generating New Faces	90

Latent Space Arithmetic	91
Morphing Between Faces	92
Summary	93
4. Generative Adversarial Networks	95
Introduction	96
Deep Convolutional GAN (DCGAN)	97
The Bricks Dataset	98
The Discriminator	99
The Generator	101
Training the DCGAN	104
Analysis of the DCGAN	109
GAN Training: Tips and Tricks	110
Wasserstein GAN with Gradient Penalty (WGAN-GP)	113
Wasserstein Loss	114
The Lipschitz Constraint	115
Enforcing the Lipschitz Constraint	116
The Gradient Penalty Loss	117
Training the WGAN-GP	119
Analysis of the WGAN-GP	121
Conditional GAN (CGAN)	122
CGAN Architecture	123
Training the CGAN	124
Analysis of the CGAN	126
Summary	127
5. Autoregressive Models	129
Introduction	130
Long Short-Term Memory Network (LSTM)	131
The Recipes Dataset	132
Working with Text Data	133
Tokenization	134
Creating the Training Set	137
The LSTM Architecture	138
The Embedding Layer	138
The LSTM Layer	140
The LSTM Cell	142
Training the LSTM	144
Analysis of the LSTM	146
Recurrent Neural Network (RNN) Extensions	149
Stacked Recurrent Networks	149

Gated Recurrent Units	151
Bidirectional Cells	153
PixelCNN	153
Masked Convolutional Layers	154
Residual Blocks	156
Training the PixelCNN	158
Analysis of the PixelCNN	159
Mixture Distributions	162
Summary	164
6. Normalizing Flow Models.....	167
Introduction	168
Normalizing Flows	169
Change of Variables	170
The Jacobian Determinant	172
The Change of Variables Equation	173
RealNVP	174
The Two Moons Dataset	174
Coupling Layers	175
Training the RealNVP Model	181
Analysis of the RealNVP Model	184
Other Normalizing Flow Models	186
GLOW	186
FFJORD	187
Summary	188
7. Energy-Based Models.....	189
Introduction	189
Energy-Based Models	191
The MNIST Dataset	192
The Energy Function	193
Sampling Using Langevin Dynamics	194
Training with Contrastive Divergence	197
Analysis of the Energy-Based Model	201
Other Energy-Based Models	202
Summary	203
8. Diffusion Models.....	205
Introduction	206
Denoising Diffusion Models (DDM)	208
The Flowers Dataset	208

The Forward Diffusion Process	209
The Reparameterization Trick	210
Diffusion Schedules	211
The Reverse Diffusion Process	214
The U-Net Denoising Model	217
Training the Diffusion Model	224
Sampling from the Denoising Diffusion Model	225
Analysis of the Diffusion Model	228
Summary	231

Part III. Applications

9. Transformers.....	235
Introduction	236
GPT	236
The Wine Reviews Dataset	237
Attention	238
Queries, Keys, and Values	239
Multihead Attention	241
Causal Masking	242
The Transformer Block	245
Positional Encoding	248
Training GPT	250
Analysis of GPT	252
Other Transformers	255
T5	256
GPT-3 and GPT-4	259
ChatGPT	260
Summary	264
10. Advanced GANs.....	267
Introduction	268
ProGAN	269
Progressive Training	269
Outputs	276
StyleGAN	277
The Mapping Network	278
The Synthesis Network	279
Outputs from StyleGAN	280
StyleGAN2	281

Weight Modulation and Demodulation	282
Path Length Regularization	283
No Progressive Growing	284
Outputs from StyleGAN2	286
Other Important GANs	286
Self-Attention GAN (SAGAN)	286
BigGAN	288
VQ-GAN	289
ViT VQ-GAN	292
Summary	294
11. Music Generation.....	297
Introduction	298
Transformers for Music Generation	299
The Bach Cello Suite Dataset	300
Parsing MIDI Files	300
Tokenization	303
Creating the Training Set	304
Sine Position Encoding	305
Multiple Inputs and Outputs	307
Analysis of the Music-Generating Transformer	309
Tokenization of Polyphonic Music	313
MuseGAN	317
The Bach Chorale Dataset	317
The MuseGAN Generator	320
The MuseGAN Critic	326
Analysis of the MuseGAN	327
Summary	329
12. World Models.....	331
Introduction	331
Reinforcement Learning	332
The CarRacing Environment	334
World Model Overview	336
Architecture	336
Training	338
Collecting Random Rollout Data	339
Training the VAE	340
The VAE Architecture	341
Exploring the VAE	343
Collecting Data to Train the MDN-RNN	346

Training the MDN-RNN	346
The MDN-RNN Architecture	347
Sampling from the MDN-RNN	348
Training the Controller	348
The Controller Architecture	349
CMA-ES	349
Parallelizing CMA-ES	351
In-Dream Training	353
Summary	356
13. Multimodal Models.....	359
Introduction	360
DALL·E 2	361
Architecture	362
The Text Encoder	362
CLIP	362
The Prior	367
The Decoder	369
Examples from DALL·E 2	373
Imagen	377
Architecture	377
DrawBench	378
Examples from Imagen	379
Stable Diffusion	380
Architecture	380
Examples from Stable Diffusion	381
Flamingo	381
Architecture	382
The Vision Encoder	382
The Perceiver Resampler	383
The Language Model	385
Examples from Flamingo	388
Summary	389
14. Conclusion.....	391
Timeline of Generative AI	392
2014–2017: The VAE and GAN Era	394
2018–2019: The Transformer Era	394
2020–2022: The Big Model Era	395
The Current State of Generative AI	396
Large Language Models	396

Text-to-Code Models	400
Text-to-Image Models	402
Other Applications	405
The Future of Generative AI	407
Generative AI in Everyday Life	407
Generative AI in the Workplace	409
Generative AI in Education	410
Generative AI Ethics and Challenges	411
Final Thoughts	413
Index.....	417