

# Contents

1	Introduction	13
2	Training Networks Using Multiple GPUs	15
2.1	How Many GPUs Do I Need?	15
2.2	Performance Gains Using Multiple GPUs	16
2.3	Summary	17
3	What Is ImageNet?	19
3.1	The ImageNet Dataset	19
3.1.1	ILSVRC	19
3.2	Obtaining ImageNet	21
3.2.1	Requesting Access to the ILSVRC Challenge	21
3.2.2	Downloading Images Programmatically	21
3.2.3	Using External Services	22
3.2.4	ImageNet Development Kit	22
3.2.5	ImageNet Copyright Concerns	23
3.3	Summary	25
4	Preparing the ImageNet Dataset	27
4.1	Understanding the ImageNet File Structure	27
4.1.1	ImageNet "test" Directory	28
4.1.2	ImageNet "train" Directory	29
4.1.3	ImageNet "val" Directory	30
4.1.4	ImageNet "ImageSets" Directory	31
4.1.5	ImageNet "DevKit" Directory	32

<b>4.2</b>	<b>Building the ImageNet Dataset</b>	<b>35</b>
4.2.1	Your First ImageNet Configuration File	35
4.2.2	Our ImageNet Helper Utility	40
4.2.3	Creating List and Mean Files	44
4.2.4	Building the Compact Record Files	48
<b>4.3</b>	<b>Summary</b>	<b>50</b>
<b>5</b>	<b>Training AlexNet on ImageNet</b>	<b>51</b>
<b>5.1</b>	<b>Implementing AlexNet</b>	<b>52</b>
<b>5.2</b>	<b>Training AlexNet</b>	<b>56</b>
5.2.1	What About Training Plots?	57
5.2.2	Implementing the Training Script	58
<b>5.3</b>	<b>Evaluating AlexNet</b>	<b>63</b>
<b>5.4</b>	<b>AlexNet Experiments</b>	<b>65</b>
5.4.1	AlexNet: Experiment #1	66
5.4.2	AlexNet: Experiment #2	68
5.4.3	AlexNet: Experiment #3	69
<b>5.5</b>	<b>Summary</b>	<b>72</b>
<b>6</b>	<b>Training VGGNet on ImageNet</b>	<b>73</b>
<b>6.1</b>	<b>Implementing VGGNet</b>	<b>74</b>
<b>6.2</b>	<b>Training VGGNet</b>	<b>79</b>
<b>6.3</b>	<b>Evaluating VGGNet</b>	<b>83</b>
<b>6.4</b>	<b>VGGNet Experiments</b>	<b>84</b>
<b>6.5</b>	<b>Summary</b>	<b>86</b>
<b>7</b>	<b>Training GoogLeNet on ImageNet</b>	<b>87</b>
<b>7.1</b>	<b>Understanding GoogLeNet</b>	<b>87</b>
7.1.1	The Inception Module	88
7.1.2	GoogLeNet Architecture	88
7.1.3	Implementing GoogLeNet	89
7.1.4	Training GoogLeNet	93
<b>7.2</b>	<b>Evaluating GoogLeNet</b>	<b>97</b>
<b>7.3</b>	<b>GoogLeNet Experiments</b>	<b>97</b>
7.3.1	GoogLeNet: Experiment #1	98
7.3.2	GoogLeNet: Experiment #2	99
7.3.3	GoogLeNet: Experiment #3	100
<b>7.4</b>	<b>Summary</b>	<b>101</b>
<b>8</b>	<b>Training ResNet on ImageNet</b>	<b>103</b>
<b>8.1</b>	<b>Understanding ResNet</b>	<b>103</b>
<b>8.2</b>	<b>Implementing ResNet</b>	<b>104</b>
<b>8.3</b>	<b>Training ResNet</b>	<b>110</b>
<b>8.4</b>	<b>Evaluating ResNet</b>	<b>114</b>

<b>8.5</b>	<b>ResNet Experiments</b>	<b>114</b>
8.5.1	ResNet: Experiment #1	114
8.5.2	ResNet: Experiment #2	114
8.5.3	ResNet: Experiment #3	115
<b>8.6</b>	<b>Summary</b>	<b>118</b>
<b>9</b>	<b>Training SqueezeNet on ImageNet</b>	<b>119</b>
<b>9.1</b>	<b>Understanding SqueezeNet</b>	<b>119</b>
9.1.1	The Fire Module	119
9.1.2	SqueezeNet Architecture	121
9.1.3	Implementing SqueezeNet	122
<b>9.2</b>	<b>Training SqueezeNet</b>	<b>126</b>
<b>9.3</b>	<b>Evaluating SqueezeNet</b>	<b>130</b>
<b>9.4</b>	<b>SqueezeNet Experiments</b>	<b>130</b>
9.4.1	SqueezeNet: Experiment #1	130
9.4.2	SqueezeNet: Experiment #2	132
9.4.3	SqueezeNet: Experiment #3	133
9.4.4	SqueezeNet: Experiment #4	134
<b>9.5</b>	<b>Summary</b>	<b>136</b>
<b>10</b>	<b>Case Study: Emotion Recognition</b>	<b>139</b>
<b>10.1</b>	<b>The Kaggle Facial Expression Recognition Challenge</b>	<b>139</b>
10.1.1	The FER13 Dataset	139
10.1.2	Building the FER13 Dataset	140
<b>10.2</b>	<b>Implementing a VGG-like Network</b>	<b>145</b>
<b>10.3</b>	<b>Training Our Facial Expression Recognizer</b>	<b>148</b>
10.3.1	EmotionVGGNet: Experiment #1	151
10.3.2	EmotionVGGNet: Experiment #2	151
10.3.3	EmotionVGGNet: Experiment #3	152
10.3.4	EmotionVGGNet: Experiment #4	153
<b>10.4</b>	<b>Evaluating our Facial Expression Recognizer</b>	<b>155</b>
<b>10.5</b>	<b>Emotion Detection in Real-time</b>	<b>157</b>
<b>10.6</b>	<b>Summary</b>	<b>161</b>
<b>11</b>	<b>Case Study: Correcting Image Orientation</b>	<b>163</b>
<b>11.1</b>	<b>The Indoor CVPR Dataset</b>	<b>163</b>
11.1.1	Building the Dataset	164
<b>11.2</b>	<b>Extracting Features</b>	<b>168</b>
<b>11.3</b>	<b>Training an Orientation Correction Classifier</b>	<b>171</b>
<b>11.4</b>	<b>Correcting Orientation</b>	<b>173</b>
<b>11.5</b>	<b>Summary</b>	<b>175</b>
<b>12</b>	<b>Case Study: Vehicle Identification</b>	<b>177</b>
<b>12.1</b>	<b>The Stanford Cars Dataset</b>	<b>177</b>
12.1.1	Building the Stanford Cars Dataset	178

<b>12.2</b>	<b>Fine-tuning VGG on the Stanford Cars Dataset</b>	<b>185</b>
12.2.1	VGG Fine-tuning: Experiment #1	190
12.2.2	VGG Fine-tuning: Experiment #2	191
12.2.3	VGG Fine-tuning: Experiment #3	192
<b>12.3</b>	<b>Evaluating our Vehicle Classifier</b>	<b>193</b>
<b>12.4</b>	<b>Visualizing Vehicle Classification Results</b>	<b>195</b>
<b>12.5</b>	<b>Summary</b>	<b>199</b>
<b>13</b>	<b>Case Study: Age and Gender Prediction</b>	<b>201</b>
<b>13.1</b>	<b>The Ethics of Gender Identification in Machine Learning</b>	<b>201</b>
<b>13.2</b>	<b>The Adience Dataset</b>	<b>202</b>
13.2.1	Building the Adience Dataset	203
<b>13.3</b>	<b>Implementing Our Network Architecture</b>	<b>217</b>
<b>13.4</b>	<b>Measuring “One-off” Accuracy</b>	<b>219</b>
<b>13.5</b>	<b>Training Our Age and Gender Predictor</b>	<b>222</b>
<b>13.6</b>	<b>Evaluating Age and Gender Prediction</b>	<b>225</b>
<b>13.7</b>	<b>Age and Gender Prediction Results</b>	<b>228</b>
13.7.1	Age Results	228
13.7.2	Gender Results	229
<b>13.8</b>	<b>Visualizing Results</b>	<b>231</b>
13.8.1	Visualizing Results from Inside Adience	232
13.8.2	Understanding Face Alignment	236
13.8.3	Applying Age and Gender Prediction to Your Own Images	238
<b>13.9</b>	<b>Summary</b>	<b>242</b>
<b>14</b>	<b>Faster R-CNNs</b>	<b>245</b>
<b>14.1</b>	<b>Object Detection and Deep Learning</b>	<b>245</b>
14.1.1	Measuring Object Detector Performance	246
<b>14.2</b>	<b>The (Faster) R-CNN Architecture</b>	<b>249</b>
14.2.1	A Brief History of R-CNN	250
14.2.2	The Base Network	253
14.2.3	Anchors	254
14.2.4	Region Proposal Network (RPN)	256
14.2.5	Region of Interest (ROI) Pooling	257
14.2.6	Region-based Convolutional Neural Network	258
14.2.7	The Complete Training Pipeline	259
<b>14.3</b>	<b>Summary</b>	<b>259</b>
<b>15</b>	<b>Training a Faster R-CNN From Scratch</b>	<b>261</b>
<b>15.1</b>	<b>The LISA Traffic Signs Dataset</b>	<b>261</b>
<b>15.2</b>	<b>Installing the TensorFlow Object Detection API</b>	<b>262</b>
<b>15.3</b>	<b>Training Your Faster R-CNN</b>	<b>263</b>
15.3.1	Project Directory Structure	263
15.3.2	Configuration	265
15.3.3	A TensorFlow Annotation Class	267

15.3.4	Building the LISA + TensorFlow Dataset .....	269
15.3.5	A Critical Pre-Training Step .....	274
15.3.6	Configuring the Faster R-CNN .....	275
15.3.7	Training the Faster R-CNN .....	280
15.3.8	Suggestions When Working with the TFOD API .....	282
15.3.9	Exporting the Frozen Model Graph .....	286
15.3.10	Faster R-CNN on Images and Videos .....	286
<b>15.4</b>	<b>Summary</b> .....	<b>290</b>
<b>16</b>	<b>Single Shot Detectors (SSDs)</b> .....	<b>293</b>
<b>16.1</b>	<b>Understanding Single Shot Detectors (SSDs)</b> .....	<b>293</b>
16.1.1	Motivation .....	293
16.1.2	Architecture .....	294
16.1.3	MultiBox, Priors, and Fixed Priors .....	295
16.1.4	Training Methods .....	296
<b>16.2</b>	<b>Summary</b> .....	<b>297</b>
<b>17</b>	<b>Training a SSD From Scratch</b> .....	<b>299</b>
<b>17.1</b>	<b>The Vehicle Dataset</b> .....	<b>299</b>
<b>17.2</b>	<b>Training Your SSD</b> .....	<b>300</b>
17.2.1	Directory Structure and Configuration .....	300
17.2.2	Building the Vehicle Dataset .....	302
17.2.3	Training the SSD .....	307
17.2.4	SSD Results .....	310
17.2.5	Potential Problems and Limitations .....	311
<b>17.3</b>	<b>Summary</b> .....	<b>312</b>
<b>18</b>	<b>Conclusions</b> .....	<b>313</b>
<b>18.1</b>	<b>Where to Now?</b> .....	<b>314</b>