Automated Ship Detection using Satellite Photography

Juliana Antoninus, Data Scientist

Problem Statement

Using computer vision algorithms, develop a model to automatically identify the presence of ships in satellite images.

The Dataset

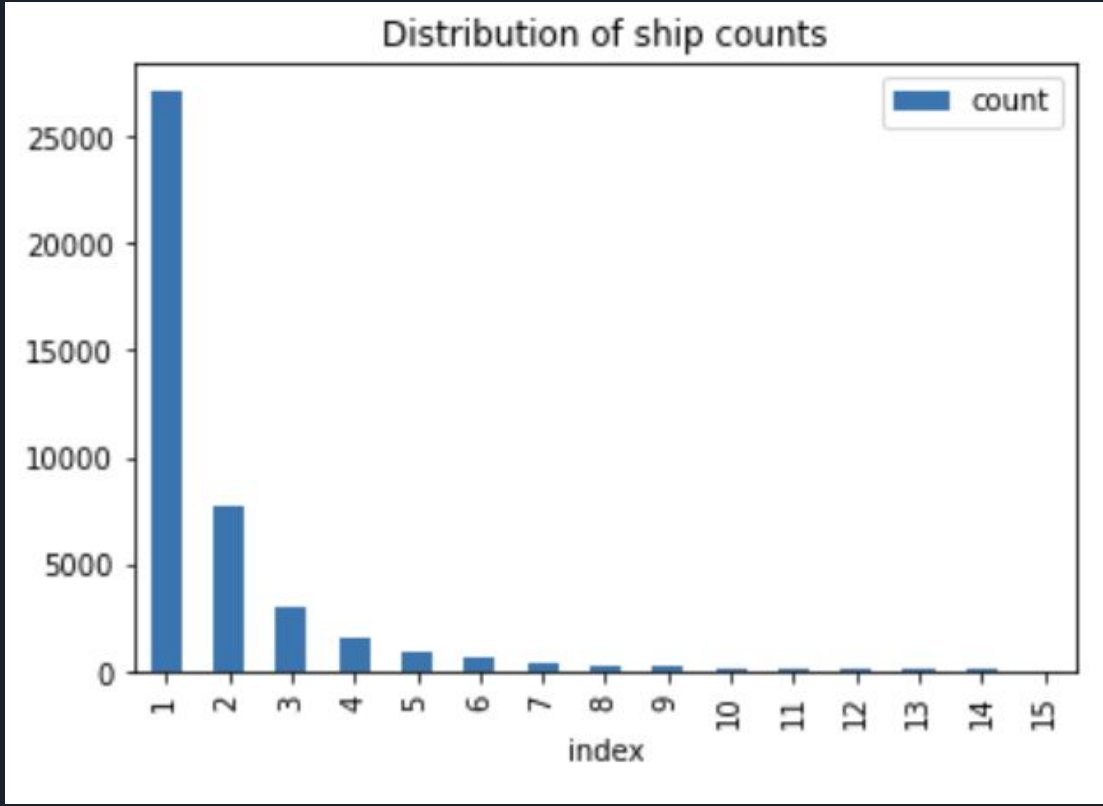
Airbus Ship Detection Challenge from Kaggle

- 29.25 GB of data
- 193k training images
- 15.6k test images
- Target result is a bounding box drawn around each ship in an image

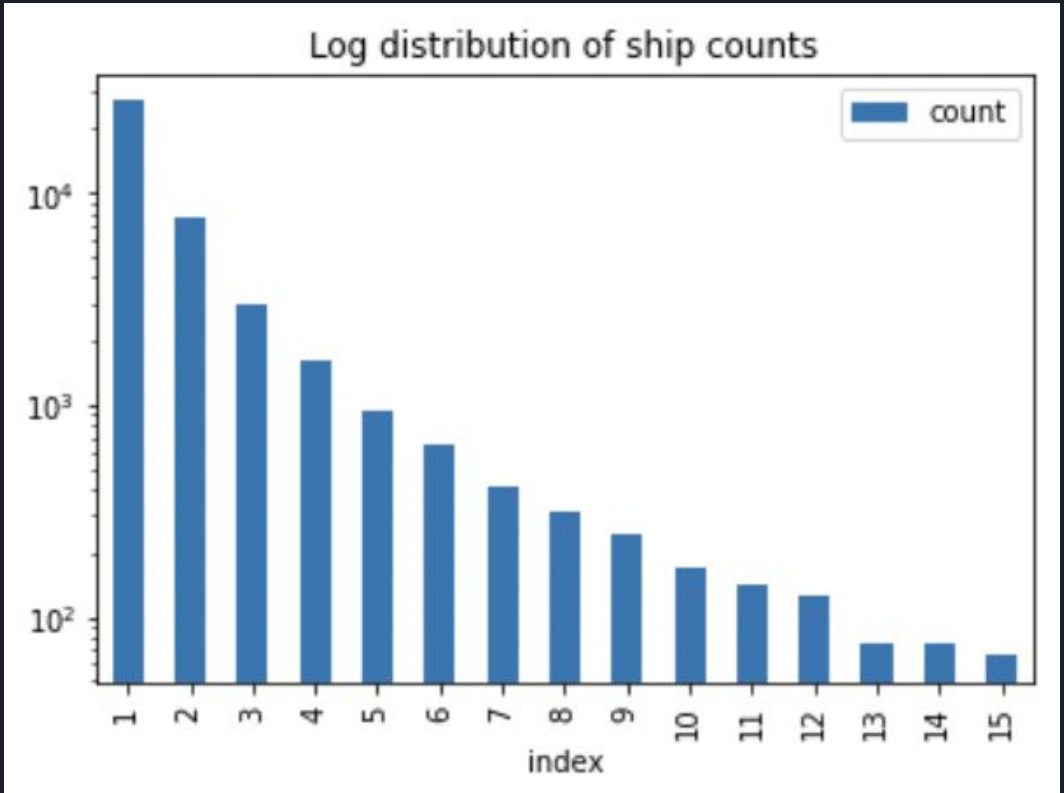
Exploratory Data Analysis

- Roughly 22% (42,556)of images have ships present
- 27104 images contain only a single ship
- The highest number of ships per image is 15, with 66 images containing 15 ships

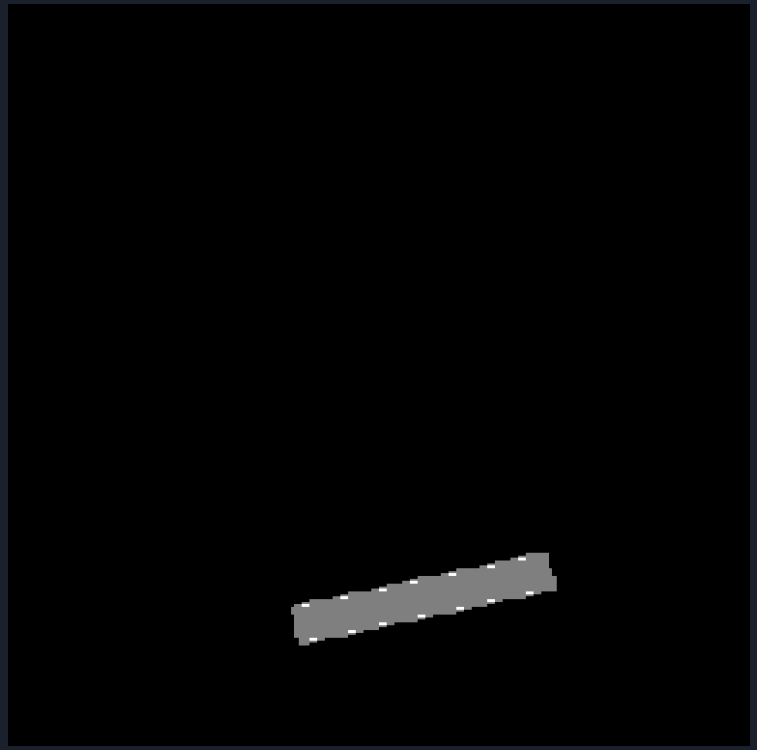
Distribution of ships per image



Log distribution of ships per image



Target value



Run-Length Encoding

264661 17 265429 33 266197 33 266965 33 267733 33 268501 33 269269 33 270037 33
270805 33 271573 33 272341 33 273109 33 273877 33 274645 33 275413 33 276181 33
276949 33 277716 34 278484 34 279252 33 280020 33 280788 33 281556 33 282324 33
283092 33 283860 33 284628 33 285396 33 286164 33 286932 33 287700 33 288468 33
289236 33 290004 33 290772 33 291540 33 292308 33 293076 33 293844 33 294612 33
295380 33 296148 33 296916 33 297684 33 298452 33 299220 33 299988 33 300756 33
301524 33 302292 33 303060 33 303827 34 304595 34 305363 33 306131 33 306899 33
307667 33 308435 33 309203 33 309971 33 310739 33 311507 33 312275 33 313043 33
313811 33 314579 33 315347 33 316115 33 316883 33 317651 33 318419 33 319187 33
319955 33 320723 33 321491 33 322259 33 323027 33 323795 33 324563 33 325331 33
326099 33 326867 33 327635 33 328403 33 329171 33 329938 34 330706 34 331474 33
332242 33 333010 33 333778 33 334546 33 335314 33 336082 33 336850 33 337618 33
338386 33 339154 33 339922 33 340690 33 341458 33 342226 33 343003 24 343787 8

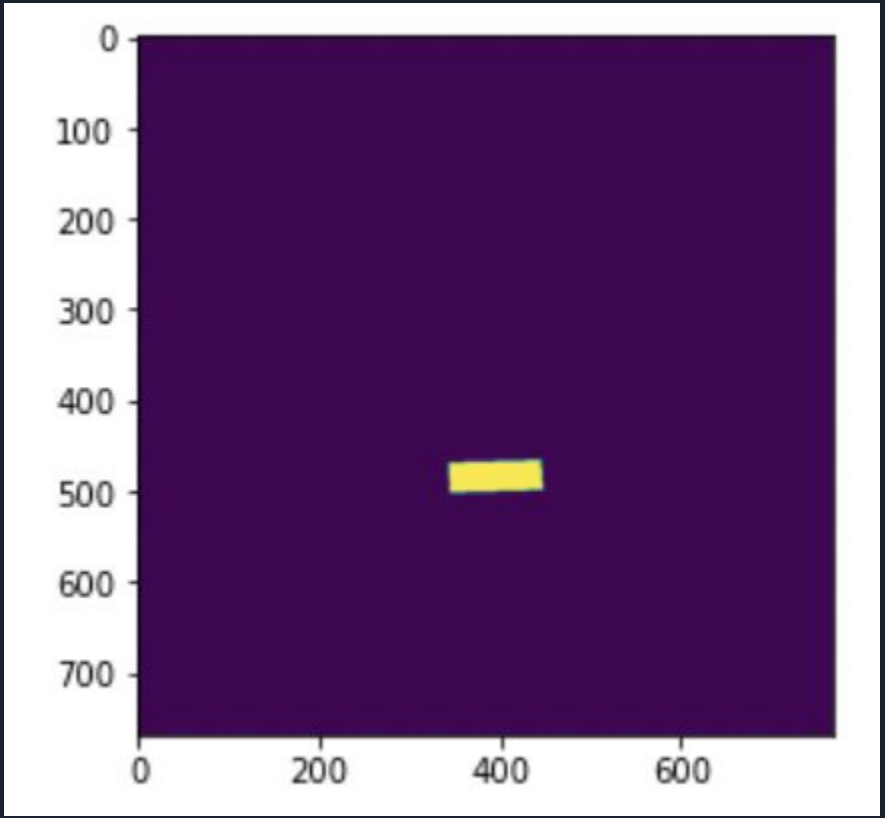
Run-Length Encoding

264661 17

264661 % 768 = 469

264661 // 768 = 344

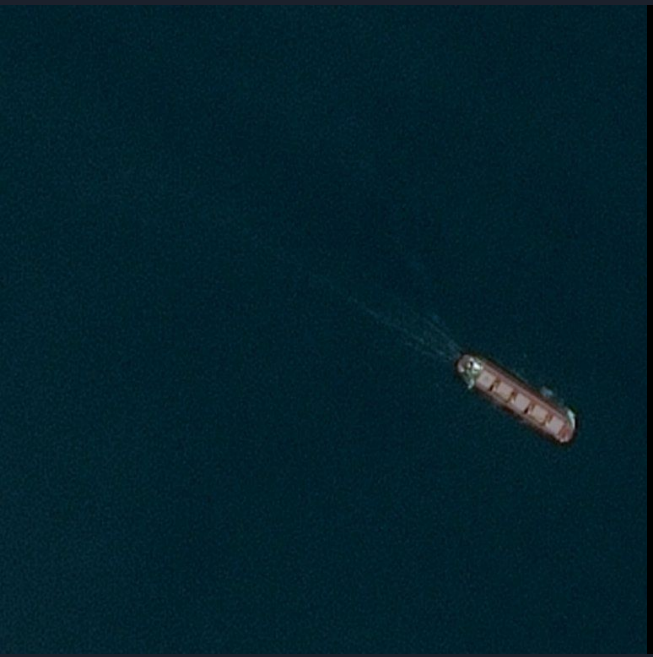
Start coordinate: (469, 344)

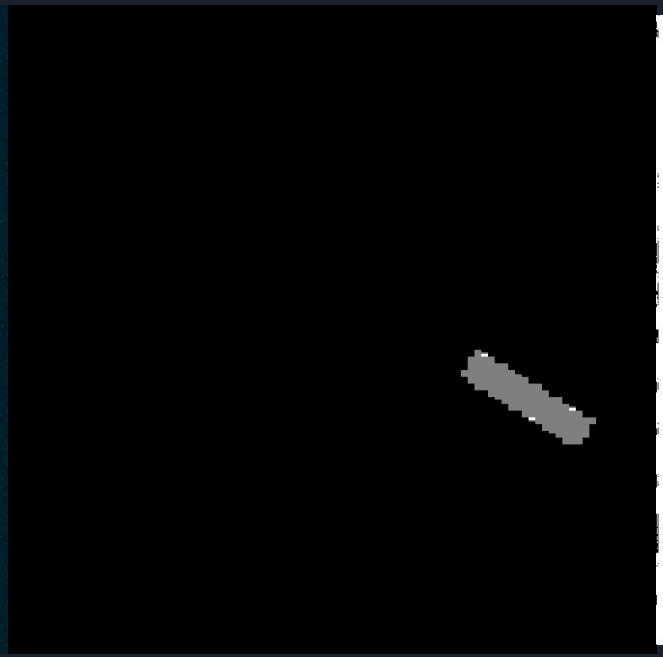


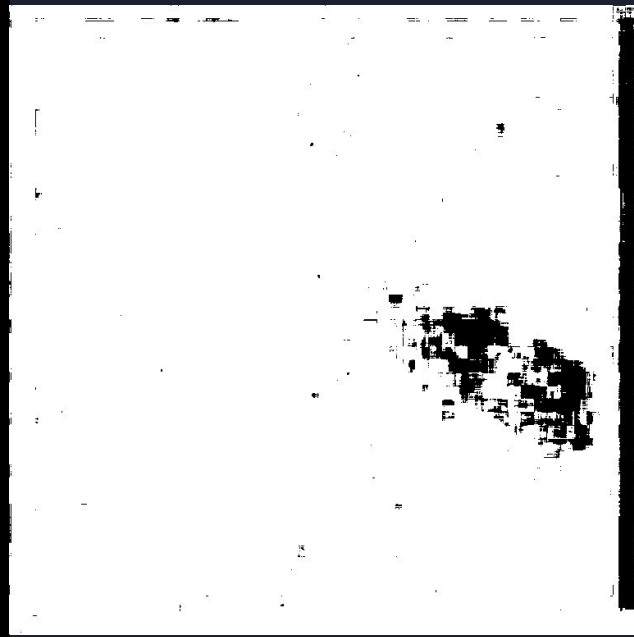
Ship detection network

- Trained a CNN to identify presence of ships
- 2000 training images, 1000 validation images, with a 50:50 split
- Took a significant amount of time and memory due to large size of images
- Performed better than baseline, with a 62% accuracy rate among the validation images, although it plateaued around this point
- 76% accuracy rate with the training images

Generating the Bounding Boxes



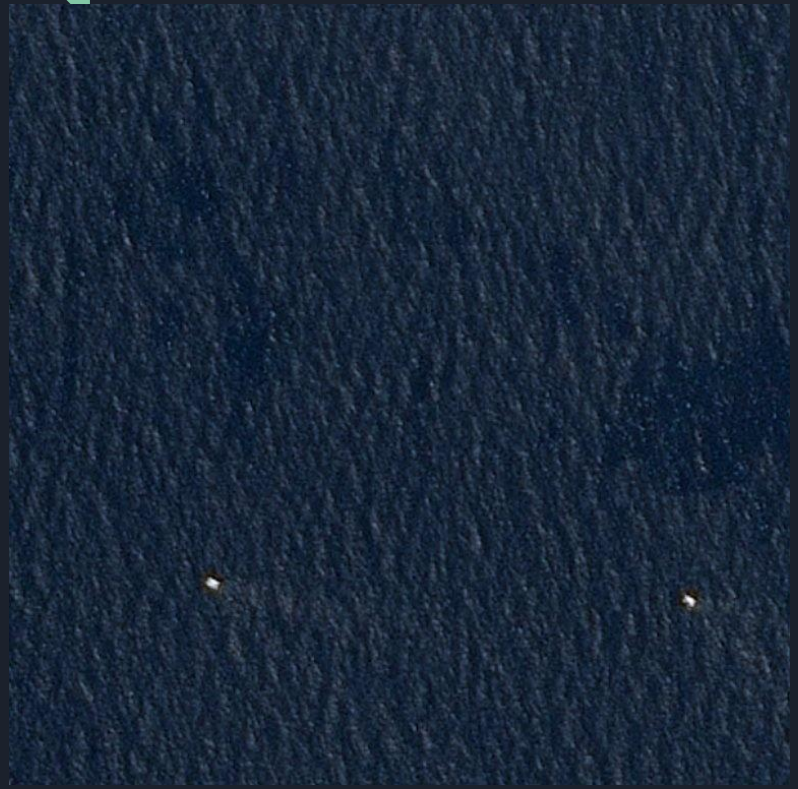


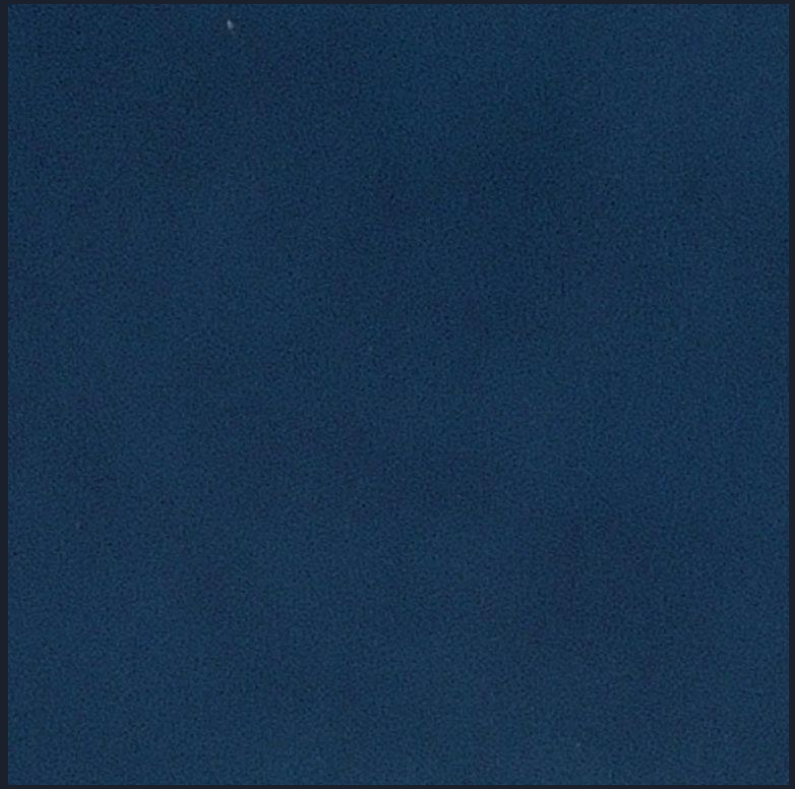


Challenges

Size Variation

Inconsistent Contrast





Cloud cover

Non-ocean geography

Recommendations

- Use a larger training set on a more powerful computer
- Train a network to segment images based on size of ship
- Explore auto-contrast options

Questions?