Owl Detection

Using HOG and SVM

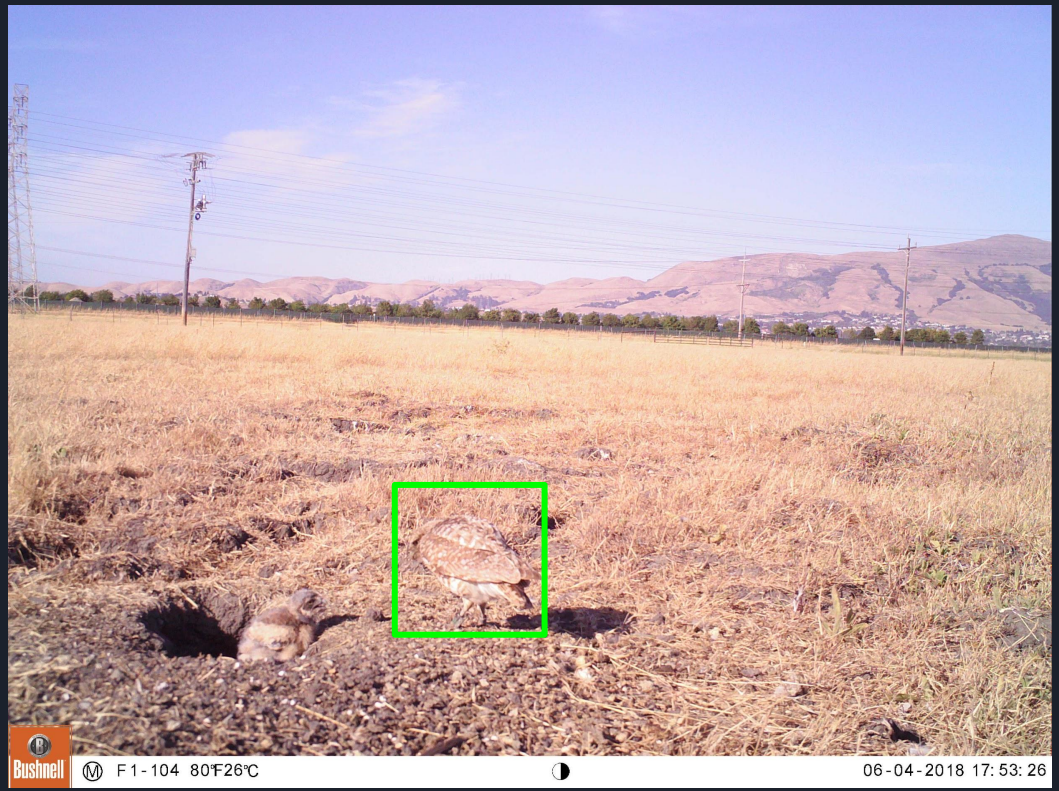
By,
Amogh Kamala Raj
Vineet Panchbhaiyye

Agenda

- Introduction
- Dataset
- Algorithm used and Implementation
- Results
- Conclusion

Can you find the Owl in this Picture?

Here it is!



Challenges

- Camouflage
- Occlusion
- Lighting Conditions

Data Set

- Training set : 1584 camera trap and flickr images containing owl and 15,000 negative sub sampled images which includes landscape and owl's habitat.
- Image size: 128 x 192 as compared to Dalal and Trigg's 64 x 128.
- Test set: 391 images.

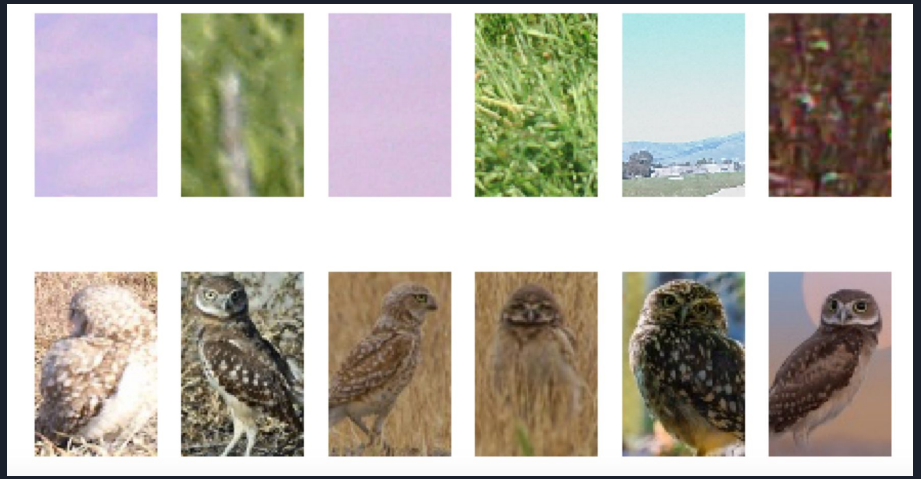
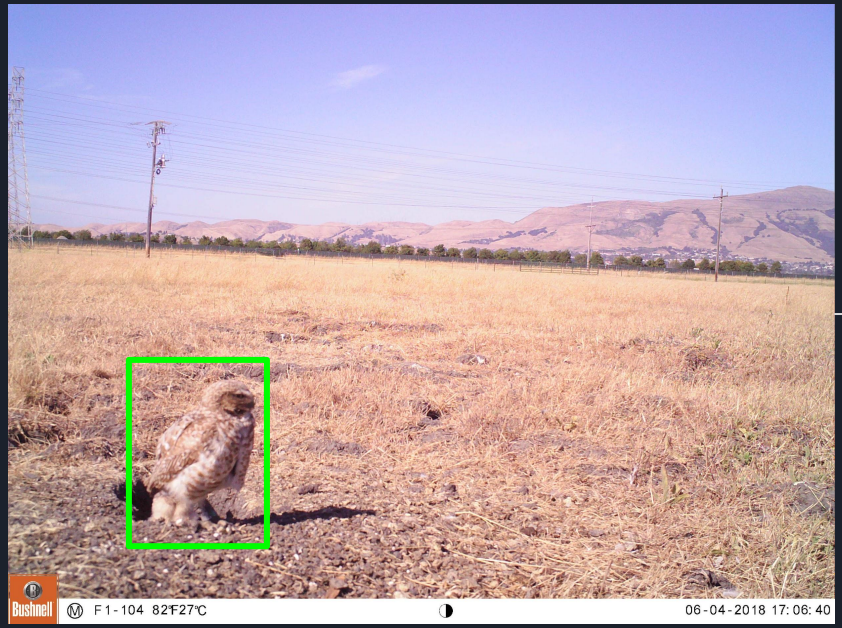


Image preprocessing



Crop

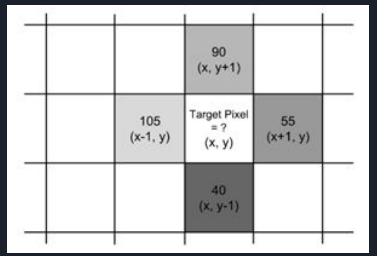
Resize
128x192

Histograms of Oriented Gradients

- Originally designed for Human detection.
- Compute horizontal and vertical gradients of 128 x 192 image.
- Compute the gradient orientation and magnitude .
- Divide the image into 16x16 block with 50% overlap.
  - 23 x 15 = 345 block in total.
- Each block should consist of 2x2 cells with size 8x8.
- Quantize the gradient orientation into 9 bins.
  - Vote is gradient magnitude.
  - Interpolate votes between neighbouring bin center.
- Concatenate histograms.

Computing Gradient vector

- Sobel Filter masks in X and Y direction



$$\nabla f(x,y) = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} f(x+1,y) - f(x-1,y) \\ f(x,y+1) - f(x,y-1) \end{bmatrix}$$

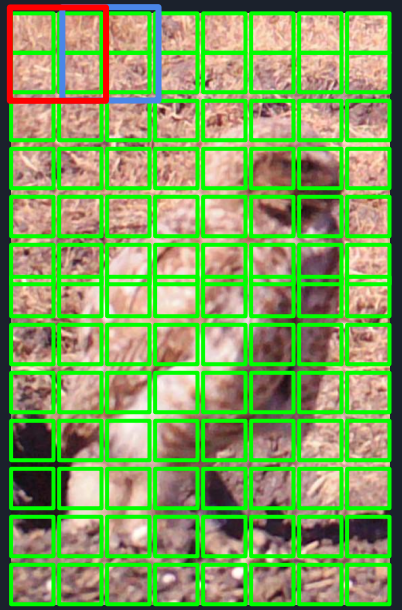
- Gradient magnitude and Orientation

$$\theta = \arctan\left(g_y / g_x\right)$$

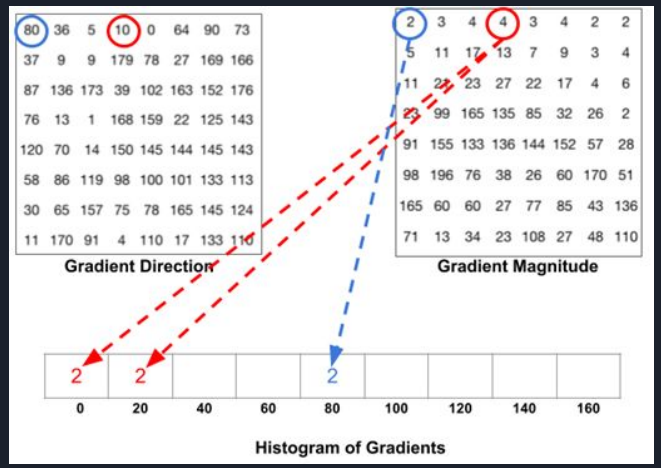
$$g = \sqrt{g_x^2 + g_y^2}.$$

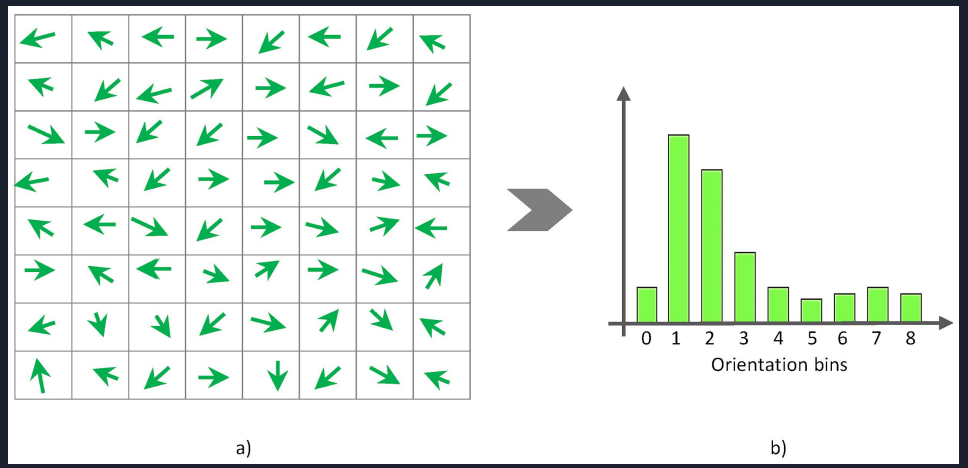
Blocks and Cell

- 16x16 blocks of 50% overlap
  - 23 x 15 = 345 blocks
- Each block should consist of 2x2 cells with size 8x8



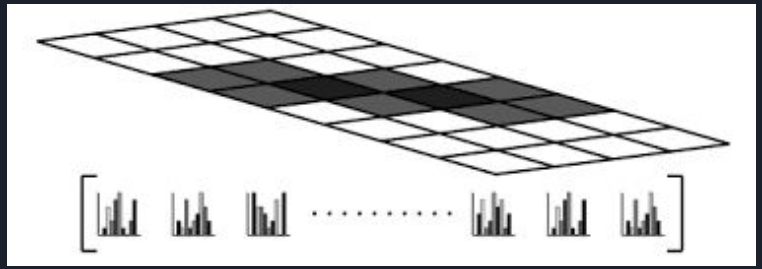
Quantizing gradient orientation in to histogram





Final Feature Vector

- Concatenate the histograms
  - Make it into a 1D matrix of length 12,420



- Feed the feature vector to machine learning algorithm like SVM for Training and testing purposes.

Detecting Owl in large images

- Initial scan
  - Sliding window of size 128 x 192 with 50% overlap in the next stride.
- Subsequent scans
  - The size of the sliding window will be doubled i.e. 256 x 384
- The scanning is repeated until the size of the window becomes half of the image height or width

AP matrix for linear and polynomial SVM classifier

Linear:

4194 (TN)	60 (FN)
88 (FP)	304 (TP)

Poly with degree 3:

4234 (TN)	20 (FN)
84 (FP)	308 (TP)

SVM results: Linear vs polynomial function

	Classifier	Precision	Recall	F1 score	Support Vectors
No owl (Negative)	Linear	0.98	0.99	0.98	4502
Owl (Positive)	Linear	0.84	0.8	0.82	394

	Classifier	Precision	Recall	F1 score	Support Vectors
No owl (Negative)	Poly with degree 3	0.98	1	0.99	4503
Owl (Positive)	Poly with degree 3	**0.95**	0.8	0.87	395

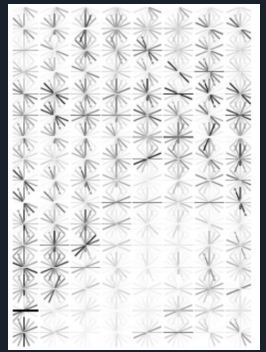
Sample results

Conclusion

- Difficult task because of too many edges.
- HOG not good representation for birds/animals with camouflage.
- Tried 64x96 images size 128x192 performs better.
- More data!!
- CNN performs better

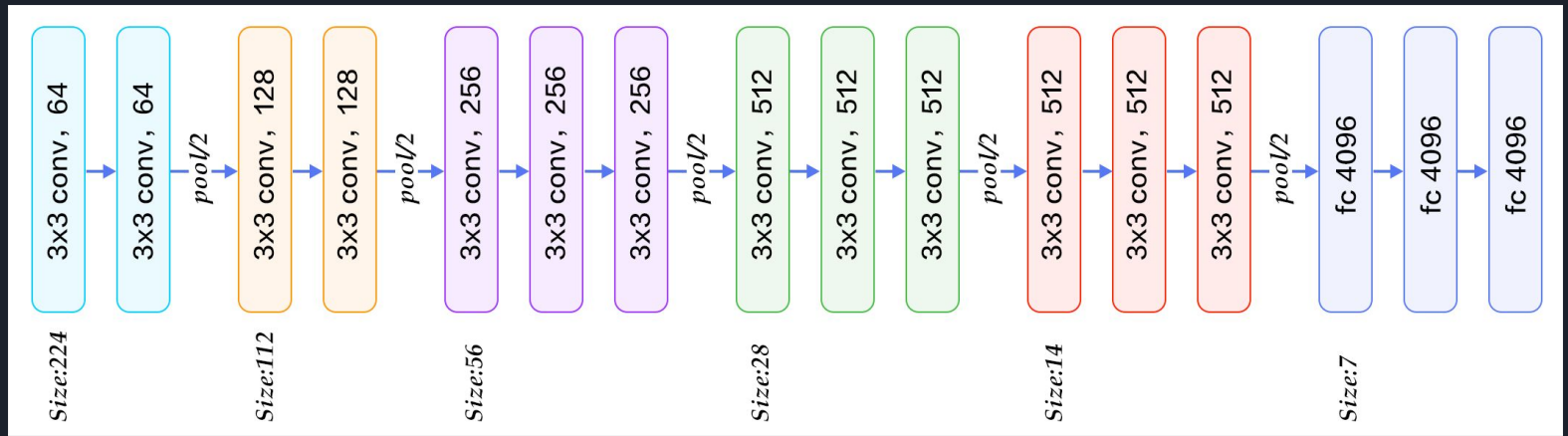




Retraining VGG16

- Has 16 layers, input image size= 224x224
- Trained the last 2 convolution layers + 1 fully connected layer of size 128
- Training set positive class - 1584 negative class 1442
- Coded implemented using Keras library
- Test accuracy 0.9727% [True positives + True negatives]
- Sliding window implemented for 224x224
- Trained for 20 epochs, using Adam optimizer.



VGG16 results

References

1. Histogram of Oriented Gradients for Human Detection, Dalal and Triggs, CVPR 2005, https://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf
2. Very Deep Convolutional Networks for Large-Scale Image Recognition, Karen Simonyan, Andrew Zisserman, ICLR 2015, https://arxiv.org/pdf/1409.1556.pdf