Compsci 682 - Neural Networks

Object discovery and localization using an image colorization model

- Abhay Mittal and Ayush Sharma

Problem Statement

What: Analyze deep hidden layers of a colorization network and build an object detector from their activations.
Why: To build detectors in scenarios where less annotated training data is available.

Approach

Stage 1: Identification of hidden units

- Run the colorizer on Caltech 101 dataset.
- For every image:
  - Extract the activations from 'fc7' layer (1x1 convolutional layer).
  - Store the maximum activation for each hidden unit (convolutional filter).
  - Extract the receptive field for the maximum activation position determined above.
- Once all the images are done, sort the activations in descending order and take the top 1000 activations for each hidden unit in the fc7 layer. Also, keep track of the images which generated those activations.
- Select k best neurons for the target object category.

Stage 2: Localization of the object in a test image

- For the test image, store the maximum activations for the k best neurons selected earlier denoted by **b**).
- Initialize :
  - best object position ≔ top left corner of image
  - best difference ≔ 0
- Now run a sliding occlusion window over the test image. For each sliding window position:
  - Measure the maximum activations for the k best neurons selected, denoted by **t**.
  - Measure the l^2 norm of the difference vector ($||\mathbf{b}-\mathbf{t}||_2$)
  - If (l^2 norm > best difference)
    - update best difference and the object location to be the current position.

Network

- A pretrained VGG 16 based colorizer network by Larsson et. al. [1]
- The network is trained over Image Net dataset.

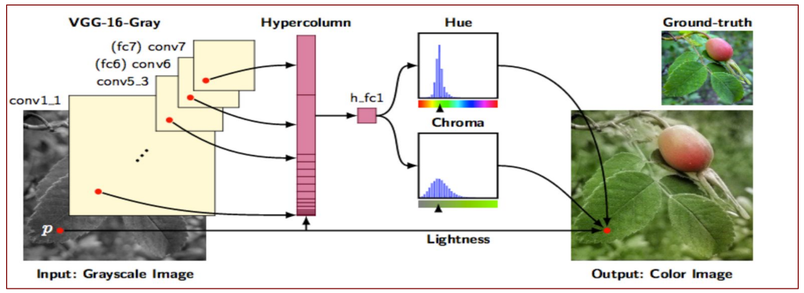
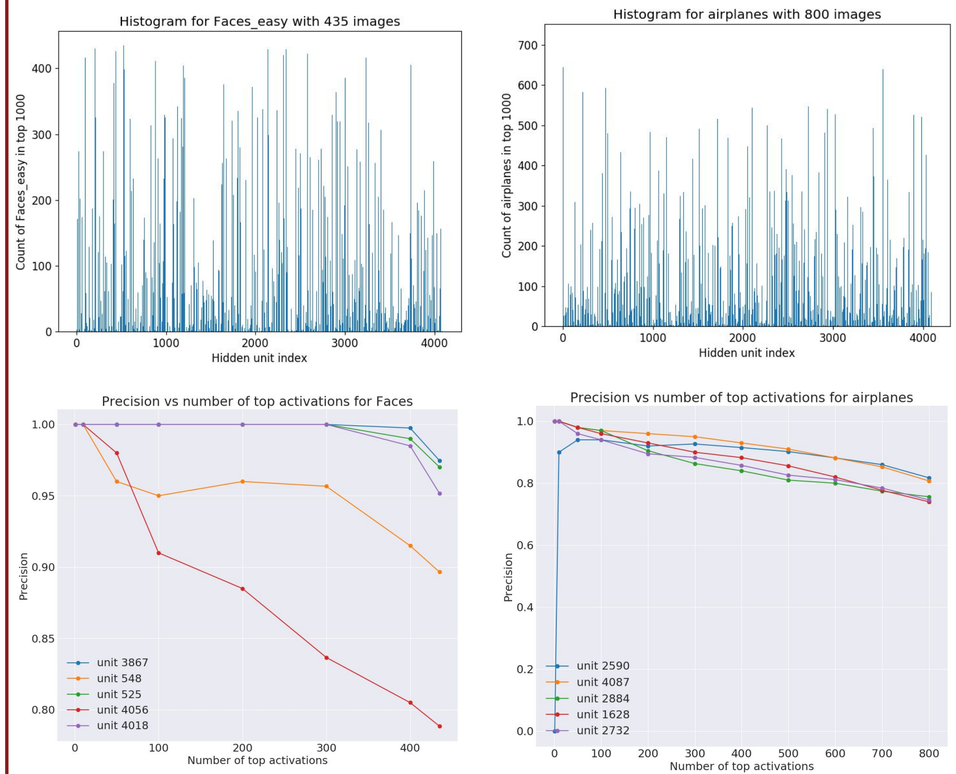


Image Credit: [1]

Hidden Unit Identification



Top row shows the frequency of faces and airplanes respectively in the top 1000 activations of neurons in fc7. Bottom row displays the precision for faces and airplanes respectively.l

Object Localization

Conclusion

- Some hidden units learn to detect objects.
- An occluding window method can be used to detect.
- Limitations of occlusion window: object must be in image, determination of correct window size and position is slow.

Future Work

- Bigger dataset and more targets.
- An efficient way of determine occlusion window size and stride.
- Other approaches for localization.
- Train a detector on labels generated by our method.

References

1. Larsson, G., Maire, M., & Shakhnarovich, G. (2016, October). Learning representations for automatic colorization. In *European Conference on Computer Vision* (pp. 577-593). Springer International Publishing.
2. Larsson, G., Maire, M., & Shakhnarovich, G. (2017). Colorization as a Proxy Task for Visual Understanding. *arXiv preprint arXiv:1703.04044*.