

# Visualizing the Image Features

(Notes while reading literature)

Aim : To increase the probability softmax gives to a class, it's easier to make the other outputs less likely rather than working towards making that class more likely.

Optimization can give us an example input that causes the desired behavior — but why bother with that? Couldn't we just look through the dataset for examples that cause the desired behavior?

It's because optimization approach enables us to know the features which actually excite a neuron. In dataset looking approach we may be misguided by things which merely just correlate to the main cause.

Have a look at the following cases :

**Dataset Examples** show us what neurons respond to in practice



**Optimization** isolates the causes of behavior from mere correlations. A neuron may not be detecting what you initially thought.



Baseball—or stripes?  
*mixed4a, Unit 6*

Animal faces—or snouts?  
*mixed4a, Unit 240*

Clouds—or fluffiness?  
*mixed4a, Unit 453*

Buildings—or sky?  
*mixed4a, Unit 492*

One generally adds a optimization term to the diversity objective to have an overall view and not make overfitting assumptions from one result.

To visualize features, we can't just go about optimizing the image. It generally leads to an image full of noise and nonsensical high frequency components. Mainly these high frequency components come from strided convolutions and pooling operations.

## Regularization

To overcome this high frequency noise in the optimized image we do regularization.





