Feature Visualization

"How Neural Networks build up their understanding of Images"
source: https://distill.pub/2017/feature-visualization/

Sahil Arora
32953

HOCHSCHULE
RAVENSBURG-WEINGARTEN
UNIVERSITY
OF APPLIED SCIENCES

Overview

What we have ?

A pre-trained Convolutional Network (that classifies images).

In the work by Olah, Schubert and collaborators, GoogleLeNet[1] has been used which has been trained on ImageNet[2] dataset.

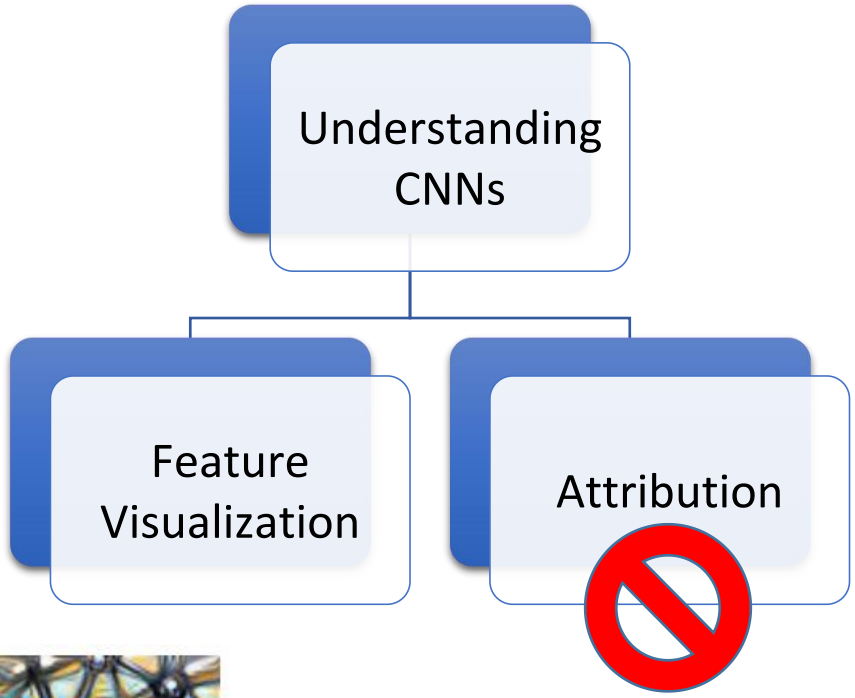
What is the Challenge ?

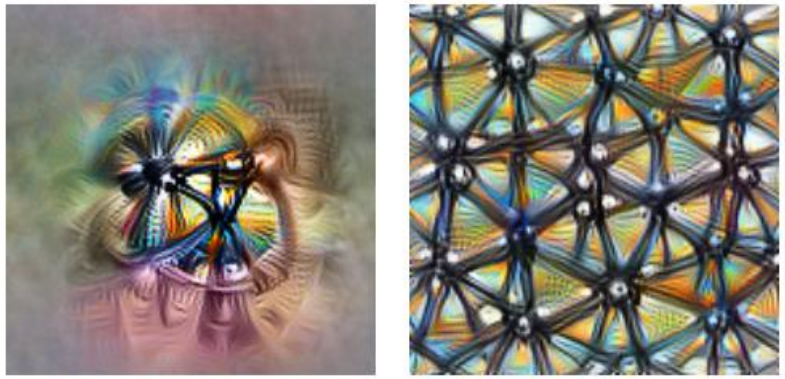
To figure out why CNN is making certain predictions.

Our Objective ?

Understand the theory on how to make a CNN interpretable to humans.

Areas of Research ?



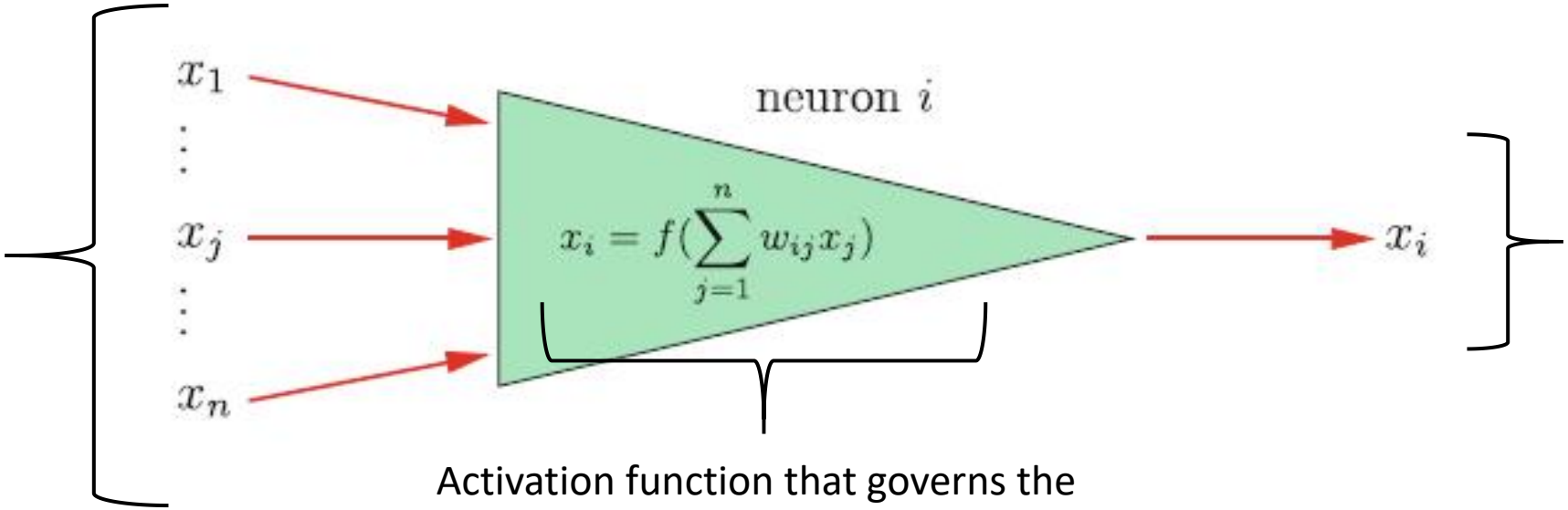


Deals with what a network (or parts of network) are looking for.

Image source: https://distill.pub/2017/feature-visualization/

Background on NNs

(image source : W. Ertel, 'Introduction to Artificial Intelligence, 2nd Edition', Springer International Publishing AG 2017, pg 248)



Input values from all connected neurons

Output value of the concerned Neuron *i*

Activation function that governs the output of Neuron *i.*

f : Threshold func / Sigmoid func ...
w_{ij} : weight b/w 2 neurons 'i' and 'j'
x_j : Input from neuron j

1. Feature Visualization by Optimization

- Generally, neurons are differentiable wrt their inputs
- To achieve an activation goal of Neuron *i,* use derivatives (Gradient Ascent) to iteratively tweak the input towards that goal [3].

1. Feature Visualization by Optimization

- Generally, neurons are differentiable wrt their inputs
- To achieve an activation goal of Neuron *i,* use derivatives (Gradient Ascent) to iteratively tweak the input towards that goal [3].

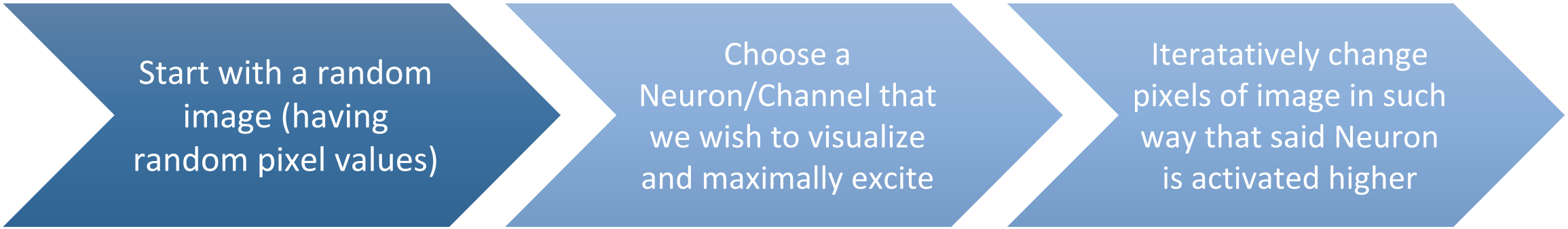
θ is network parameters (eg weight and bias)
x is input sample

- Problem reduces to :

$$x^* = \arg\max(h_{ij}(\theta, x))$$

1. Feature Visualization by Optimization

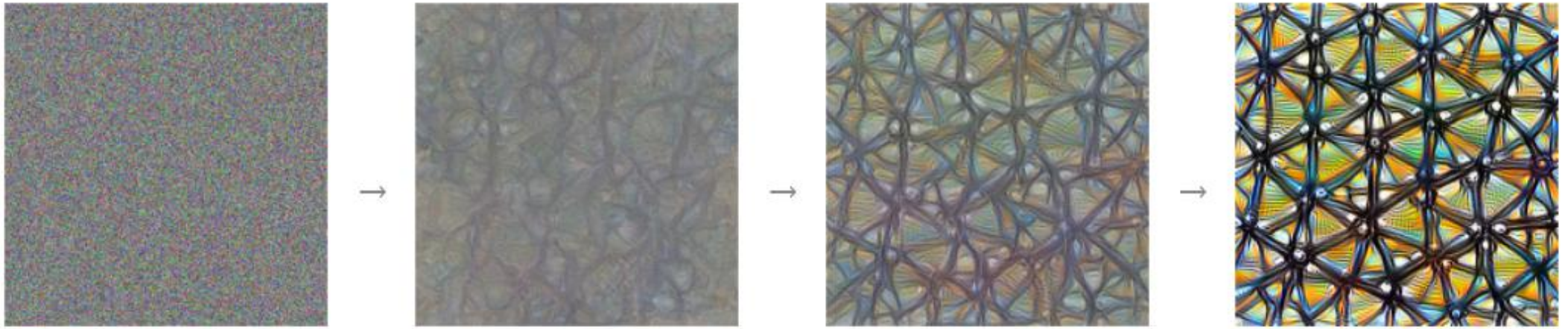
- Generally, neurons are differentiable wrt their inputs
- To achieve an activation goal of Neuron *i,* use derivatives (Gradient Ascent) to iteratively tweak the input towards that goal [3].



Start with a random image (having random pixel values)

Choose a Neuron/Channel that we wish to visualize and maximally excite

Iteratatively change pixels of image in such way that said Neuron is activated higher



This image contains the most prominent features that activates said Neuron *i.*

Image source: https://distill.pub/2017/feature-visualization/

1. Feature Visualization by Optimization

Optimization Objectives: What do we want examples of?

- Looking at individual Neurons / Channel
  - helps in understanding individual features

1. Feature Visualization by Optimization

Optimization Objectives: What do we want examples of?

- Looking at individual Neurons / Channel
  - helps in understanding individual features

- Looking at the whole Layer
  - using DeepDream, we can see what features are detected by the layer and can be used to further enhance the feature.

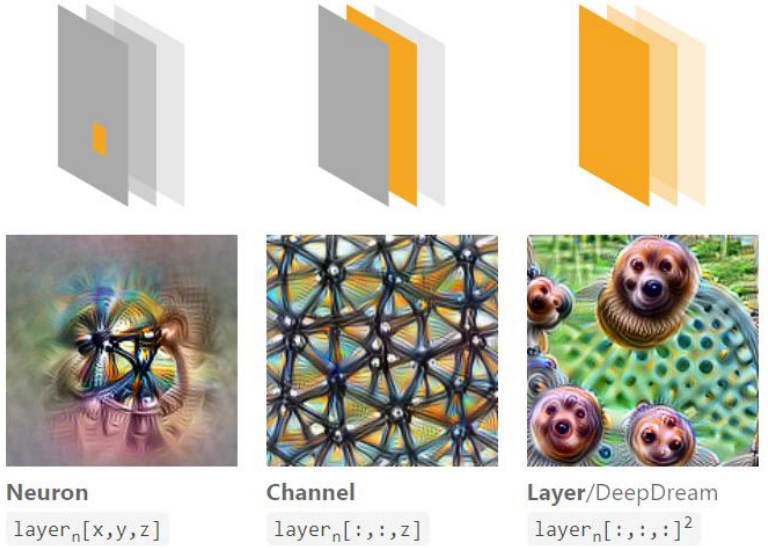
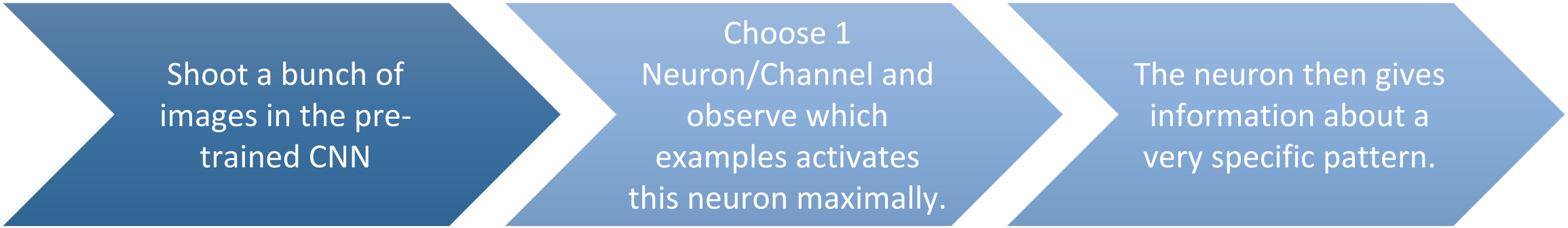


Image source: https://distill.pub/2017/feature-visualization/

1a. Feature Visualization by Dataset

Find 9 exemplary Images

- Let us first look at Feature Visualization with dataset examples



1a. Feature Visualization by Dataset

Find 9 exemplary Images

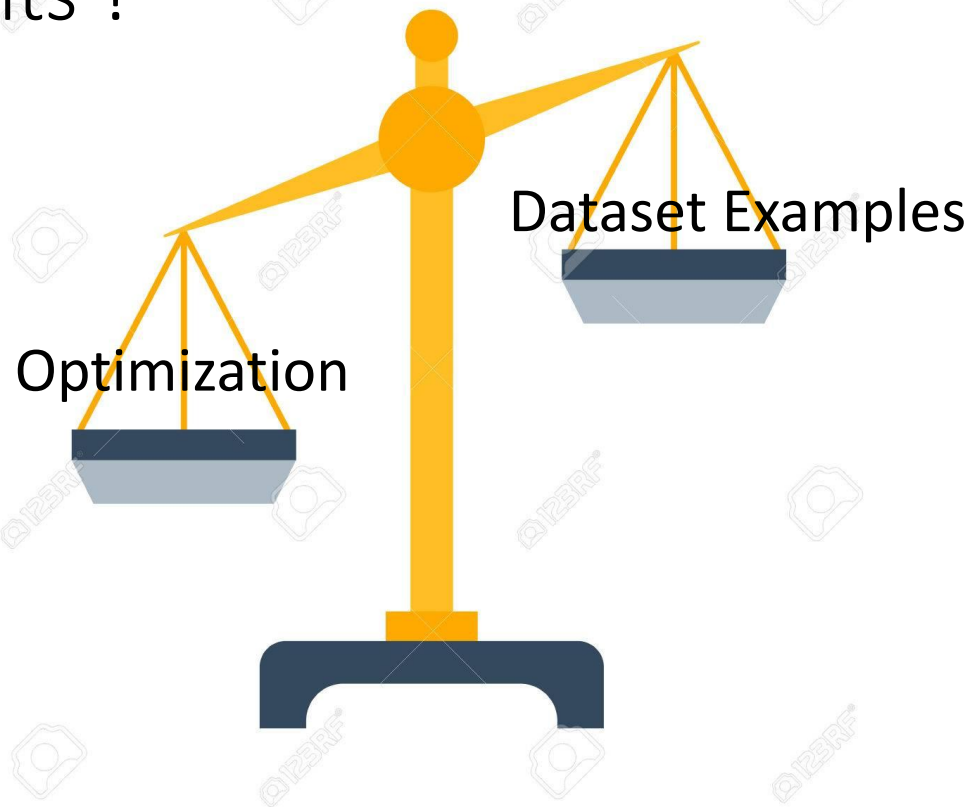
- Let us first look at Feature Visualization with dataset examples

Visualizing a neuron

Corresponding dataset examples

1. Feature Visualization by Optimization

Which one gives better results ?



1. Feature Visualization by Optimization

Which one gives better results ?

- It separates the actual feature causing the behaviours from ones that merely co-relate with the cause.

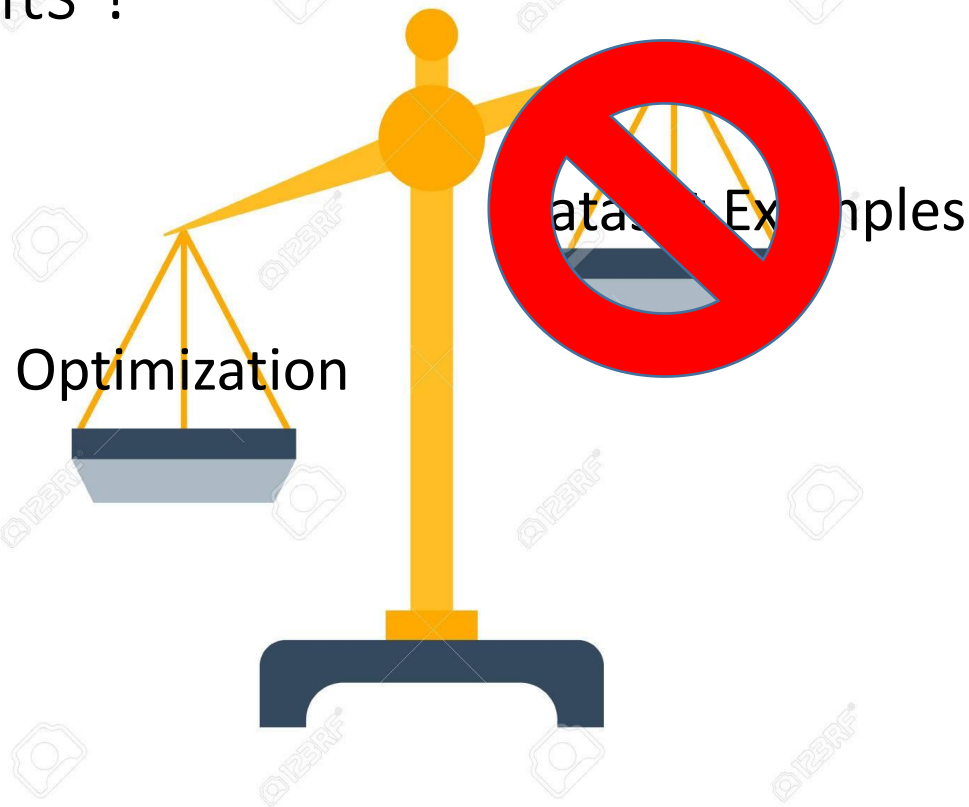
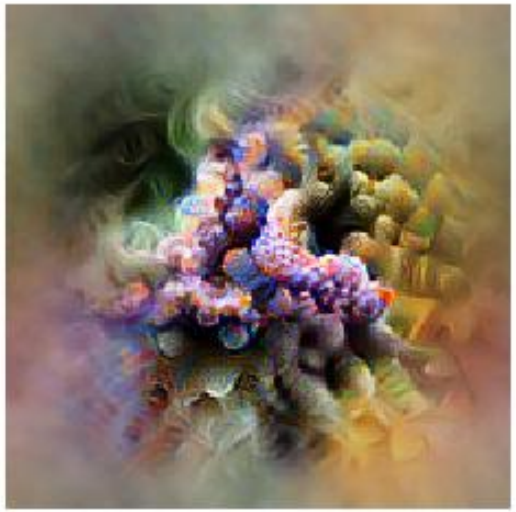


Image source: https://distill.pub/2017/feature-visualization/

Dataset Examples

Optimization



What does the concerned neuron see ?

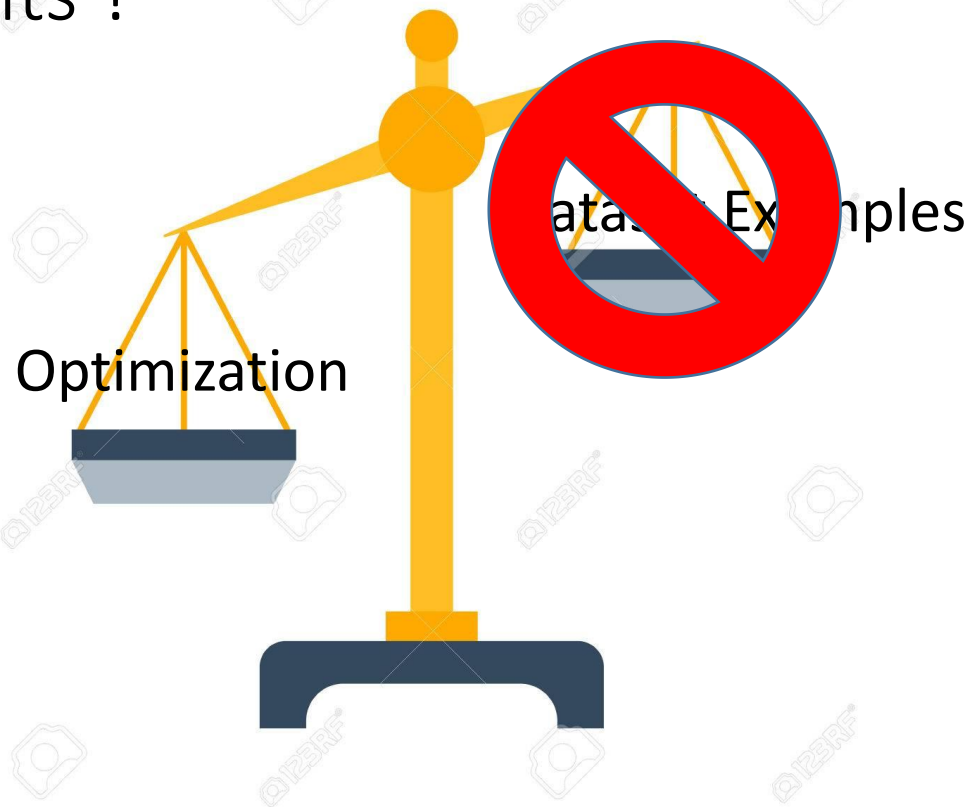
Animal faces - or *Snouts?*

Clouds - *or Fluffiness ?*

1. Feature Visualization by Optimization

Which one gives better results ?

- It separates the actual feature causing the behaviours from ones that merely co-relate with the cause.
- It offers flexibility: we can easily configure a particular image in order to make an additional neuron activated.



What is Feature Visualization ?

- Limit focus to one single Neuron/Channel/Layer
- Get to know what this single neuron detects

Sub-topics ?

- Feature Visualization by Optimization
- **Diversity**
- The Enemy of Feature Visualization

2. Diversity

Question boils down to: “Are we seeing the whole picture ?”

- It is entirely possible for a genuine example to still mislead by showing us only one facet of what that feature represents.
- Optimization generally just gives us one extremely positive example, and a negative one too.

2. Diversity

Question boils down to: “Are we seeing the whole picture ?”

- It is entirely possible for a genuine example to still mislead by showing us only one facet of what that feature represents.
- Optimization generally just gives us one extremely positive example, and a negative one too.

Is there method by which optimization gives us diversity ?

BTW, datasets have a big advantage here.

- We can look at whole spectrum of activations and see what activates the neuron to what extent.

2. Diversity

How to achieve diversity by Optimization ?

- Approach by Nguyen, Yosinski & Other[4]:
  - Search through dataset for diverse examples
  - Use them as starting points for optimization process

  This initiates optimization in different facets of features.

2. Diversity

Simple
Optimization

Dataset eg. that
caused high
activations

Optimization with diversity which reveals four different facets

Image source: https://distill.pub/2017/feature-visualization/

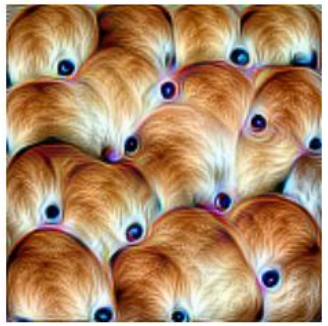
2. Diversity

Benefits of Diverse Feature Visualization

- It allows us to more closely pinpoint what activates a neuron.

By looking at dataset examples, we can make and check predictions about what inputs will activate the neuron.

2. Diversity

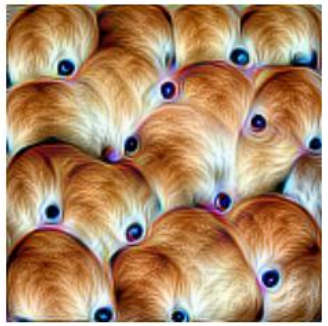


Optimization shows two eyes and downward curved faces.

Maybe neuron activates on seeing top of dog heads.

Image source: https://distill.pub/2017/feature-visualization/

2. Diversity



Optimization shows two eyes and downward curved faces.

Maybe neuron activates on seeing top of dog heads.

Optimization with diversity : Includes examples without eyes and upward curved edges

Maybe the neuron activates mostly on fur and texture.

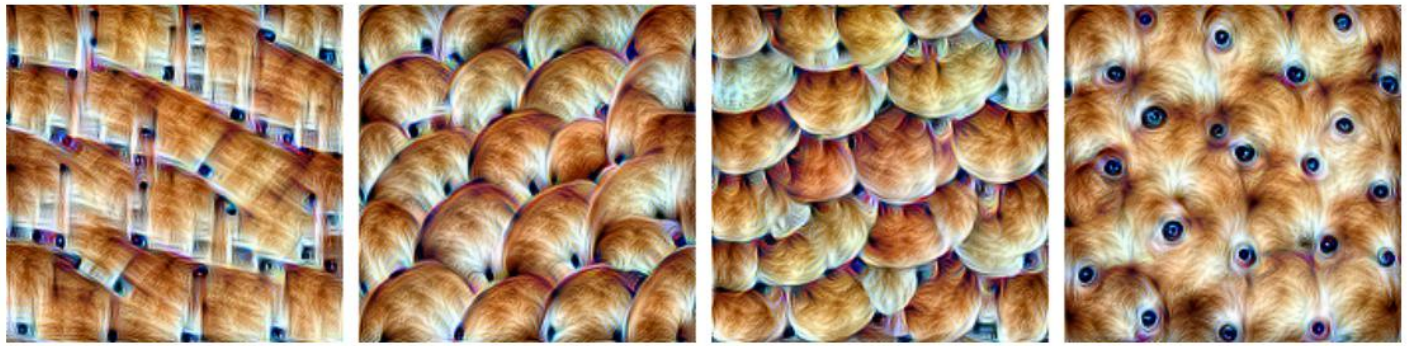
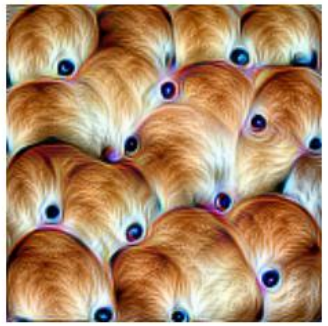


Image source: https://distill.pub/2017/feature-visualization/

2. Diversity

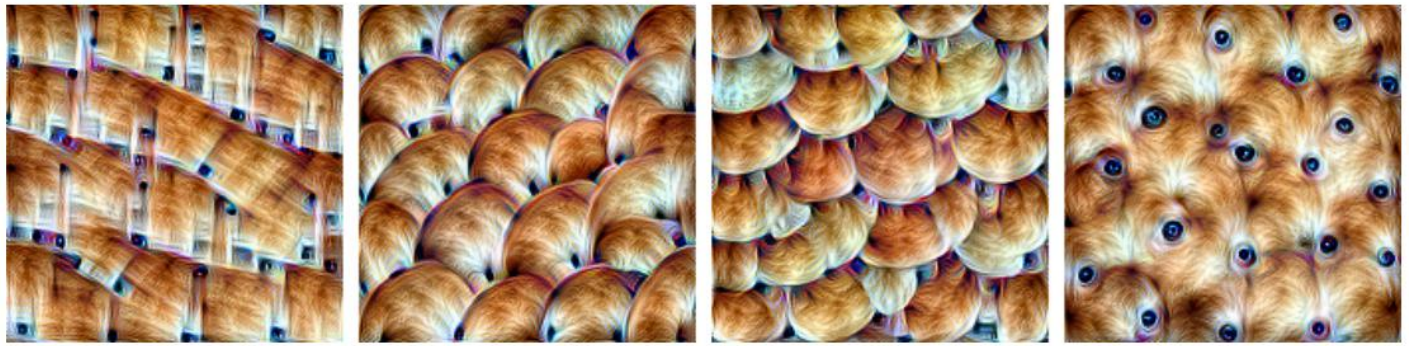
Spoon with a similar color also activates the neuron.



Optimization shows two eyes and downward curved faces.

Optimization with diversity : Includes examples without eyes and upward curved edges

Maybe the neuron activates mostly on fur and texture.



Dataset Examples

Image source: https://distill.pub/2017/feature-visualization/

2. Diversity

Shortcomings?

- Optimization with diversity may make examples different in an un-natural way.
- This can result in strange mixture of ideas.

2. Diversity

Shortcomings?

- Optimization with diversity may make examples different in an un-natural way.
- This can result in strange mixture of ideas

It suggest that neurons are not right semantic units for understanding Neural Networks.

What is Feature Visualization ?

- Limit focus to one single Neuron/Channel/Layer
- Get to know what this single neuron detects

Sub-topics ?

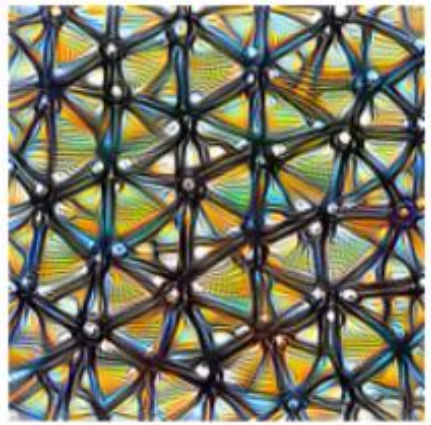
- Feature Visualization by Optimization
- Diversity
- **The Enemy of Feature Visualization**

3. The Enemy of Feature Visualization

Unfortunately...

Process of Optimization has a few tricks:

What we expected in previous topic ?



Expected result of
Iterative Visualization

Image source:
https://distill.pub/2017/
feature-visualization/

3. The Enemy of Feature Visualization

Unfortunately...

What we actually get as result ?

- image full of noise and high-frequency patterns that the network responds strongly to

Actual result of Iterative Visualization

Image source: https://distill.pub/2017/feature-visualization/

3. The Enemy of Feature Visualization

How to deal with Noises?

CURRENT RESEACH TOPICS...

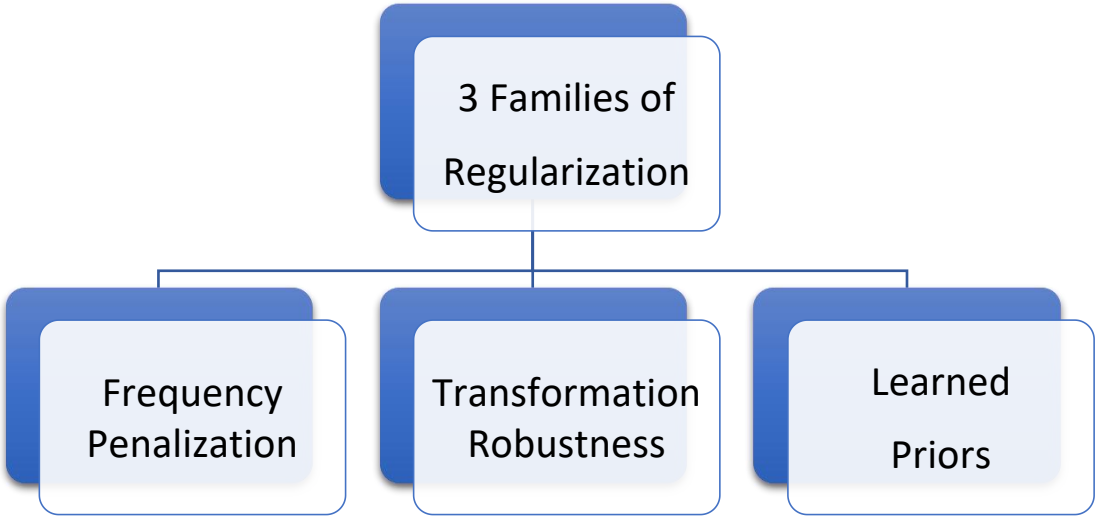
To impose a more natural structure using regularizer or constraints.

3. The Enemy of Feature Visualization

How to deal with Noises?

CURRENT RESEACH TOPICS...

To impose a more natural structure using regularizer or constraints.



3. The Enemy of Feature Visualization

a. Frequency Penalization

- One approach is to penalize variance between neighbouring pixels.
- Other approach is to penalize high-frequency noises by blurring the image in each optimization step.[5]
- We could loose legitimate features with blurring but process can be improved by using Bilateral Filter.

3. The Enemy of Feature Visualization

a. Frequency Penalization

Without Blurring

With Blurring

Image source:
https://distill.pub/2017/feature-visualization/

3. The Enemy of Feature Visualization

b. Transformation Robustness

This approach find examples that still activate the neuron even when we slightly transform the inputs.

- Scale
- Rotate
- ....

3. The Enemy of Feature Visualization

b. Transformation Robustness

This approach find examples that still activate the neuron even when we slightly transform the inputs.

- Scale
- Rotate
- ....

Simple Optimization

Rotated 45°
Scaled 1.1x

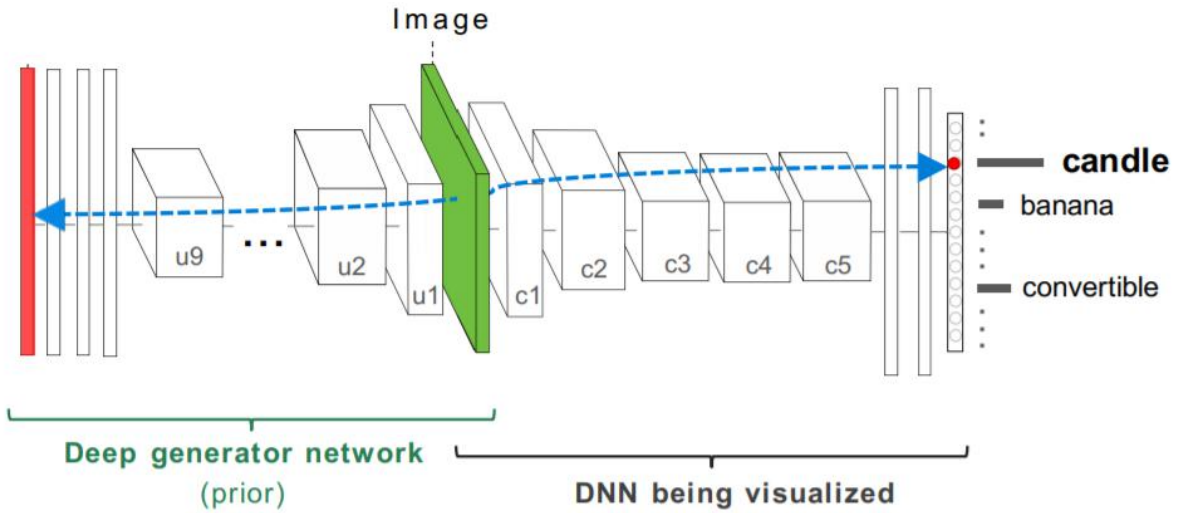
Image source:
https://distill.pub/2017/feature-visualization/

3. The Enemy of Feature Visualization

c. Learned Priors

This approach is to learn a generator that maps points in latent space to examples of your data, then optimize within that space.

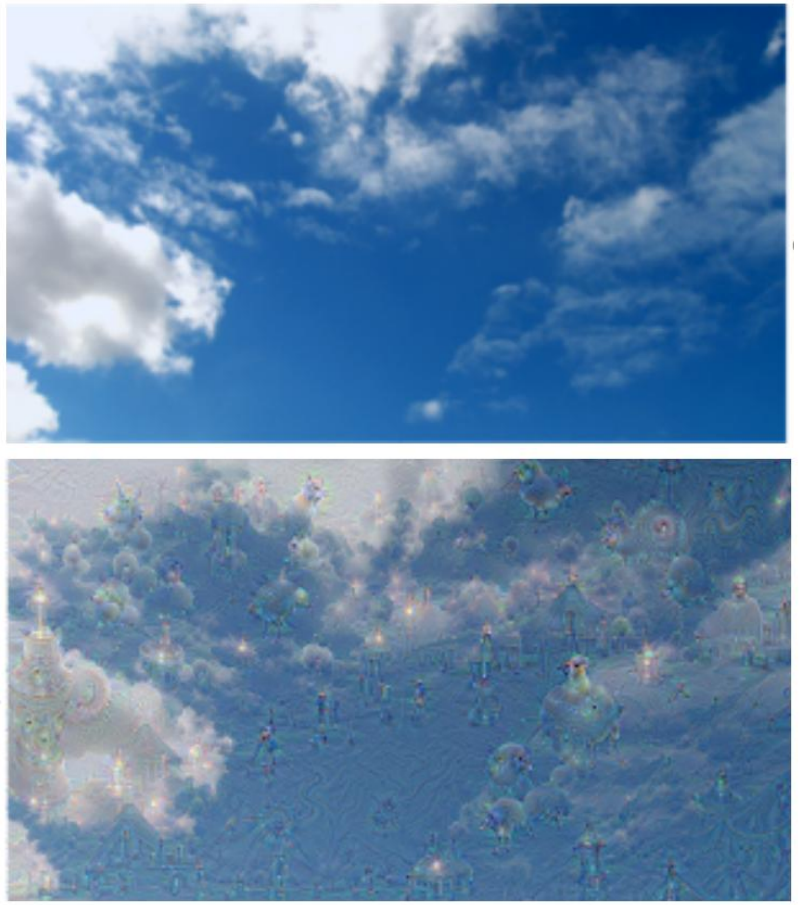
Approach by Nguyen and Yosinski comprises of a GAN with a 'to-be-visualized' Neural Network [6].

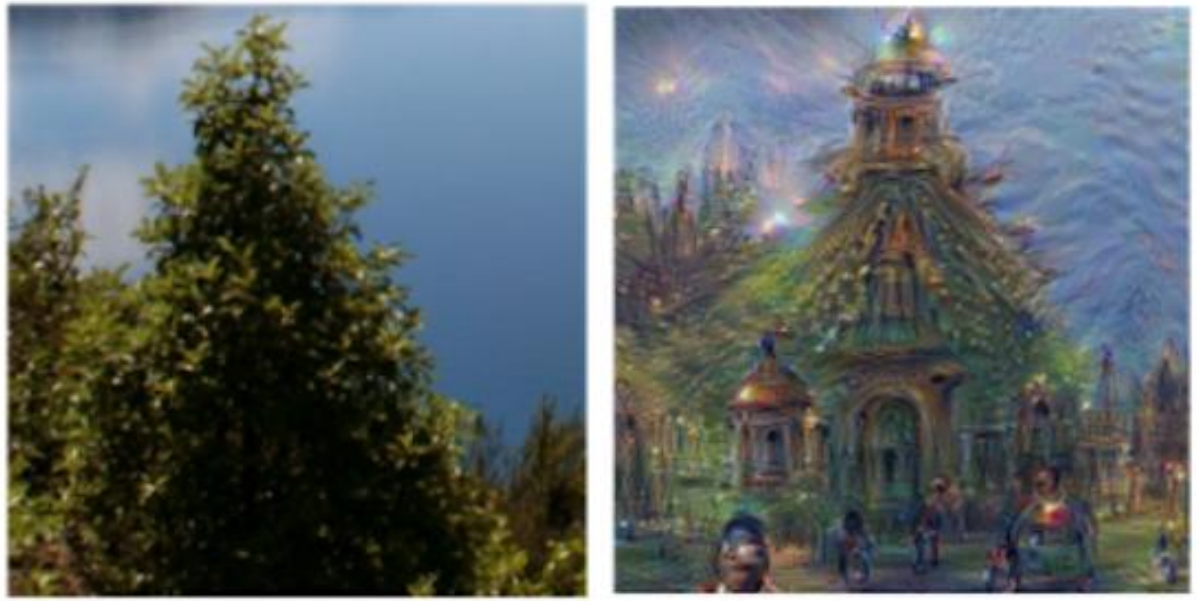


(image source : Nguyen et al. , 'Synthesizing the preferred inputs for neurons in neural networks via deep generator networks', Advances in Neural Information Processing Systems 29 (NIPS 2016)

4. Interesting Projects

a. Deep Dream [7] by Google



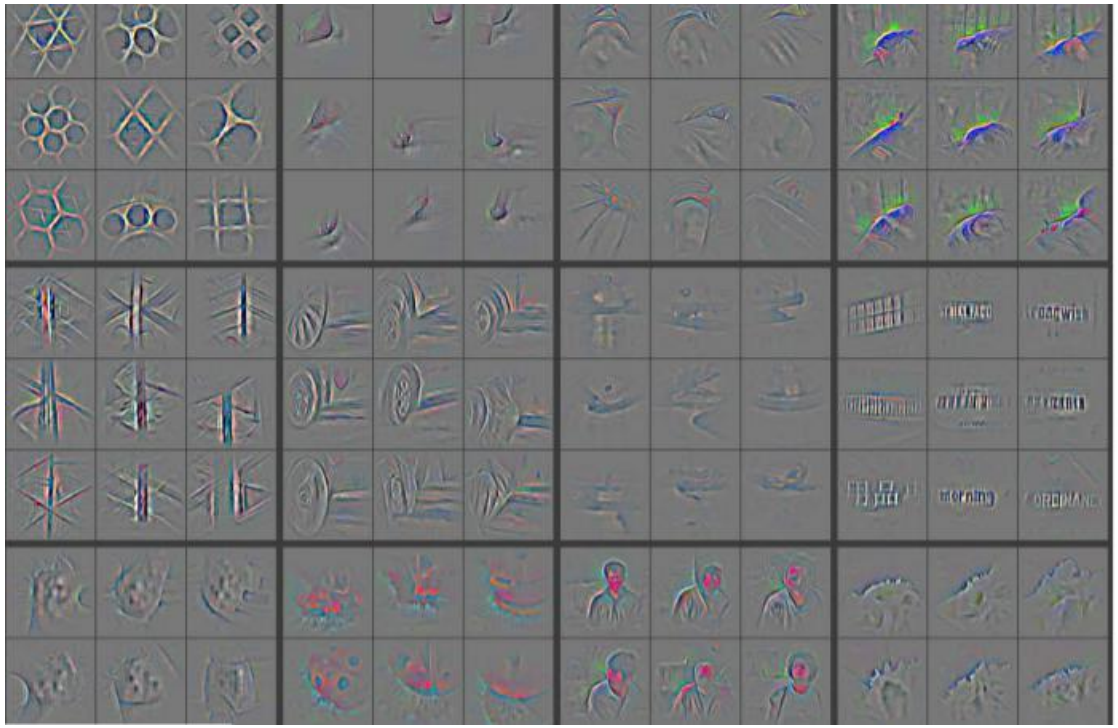


(image source : https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

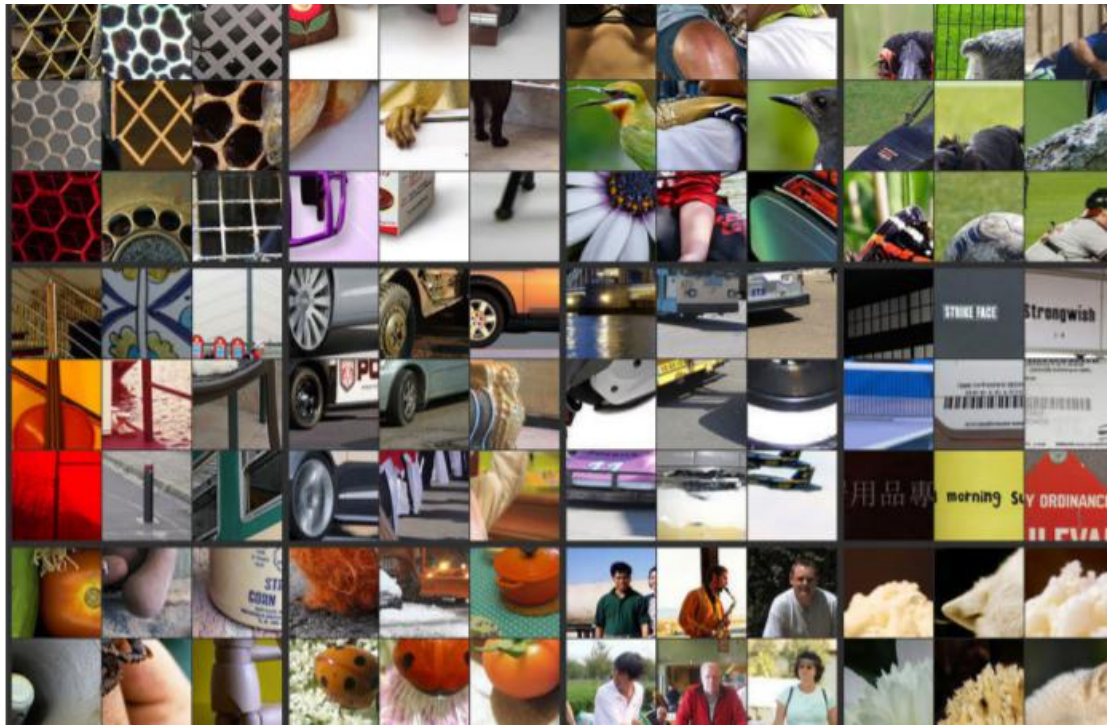
4. Interesting Projects

a. Deep Visualization Toolbox [8] by J. Yosinski

Github :
https://github.com/yosinski/deep-visualization-toolbox



Layer 3



Dataset Images

(image source : Zeiler and Fergus, 'Visualizing and Understanding Convolutional Networks'. Lecture Notes in Computer Science, vol 8689. Springer, Cham

Conclusion

Today's News

- We are able to see what our CNN is 'seeing'.
- We can configure the CNN or the example images to highlight features important for image classification.
- By itself, feature visualization hasn't given a complete satisfactory understanding.

Conclusion

Today's News

- We are able to see what our CNN is 'seeing'.
- We can configure the CNN or the example images to highlight features important for image classification.
- By itself, feature visualization hasn't given a complete satisfactory understanding.

- Some issues that stand out:
  - Understanding neuron interactions.
  - Finding which units are most meaningful for understanding neural net activations.
  - Giving a holistic view of all facets of a feature.

Questions ?

References

[1] Szegedy et al., 'Going deeper with convolutions', IEEE conference on computer vision and pattern recognition, 2015

[2] Deng et al., 'Imagenet: A large-scale hierarchical image database', Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on

[3] Erhan et al., 'Visualizing higher-layer features of a deep network', 2009. University of Montreal, Vol 1341

[4] Nguyen, Yosinski et al., 'Multifaceted feature visualization: Uncovering the different types of features learned by each neuron in deep neural networks', 2016. arXiv preprint arXiv:1602.03616

RW U

References

[5] Nguyen, Yosinski et al., 'Deep neural networks are easily fooled: High confidence predictions for unrecognizable images', 2015, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition

[6] Nguyen, Yosinski et al., 'Synthesizing the preferred inputs for neurons in neural networks via deep generator networks', Advances in Neural Information Processing Systems 29 (NIPS 2016)

[7] Google AI Blog (2020, June 16, 20:46), https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

[8] Github (2020, June 16, 20:47), https://github.com/yosinski/deep-visualization-toolbox