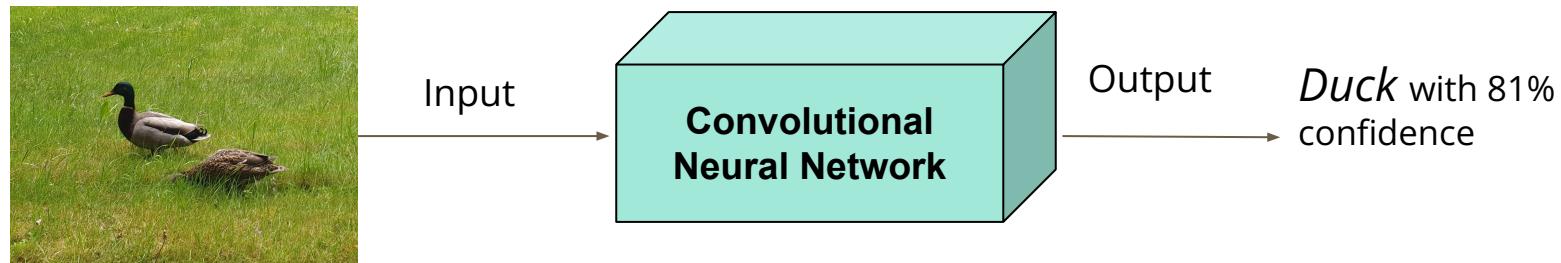


Analysis of visual receptive fields in deep neural networks

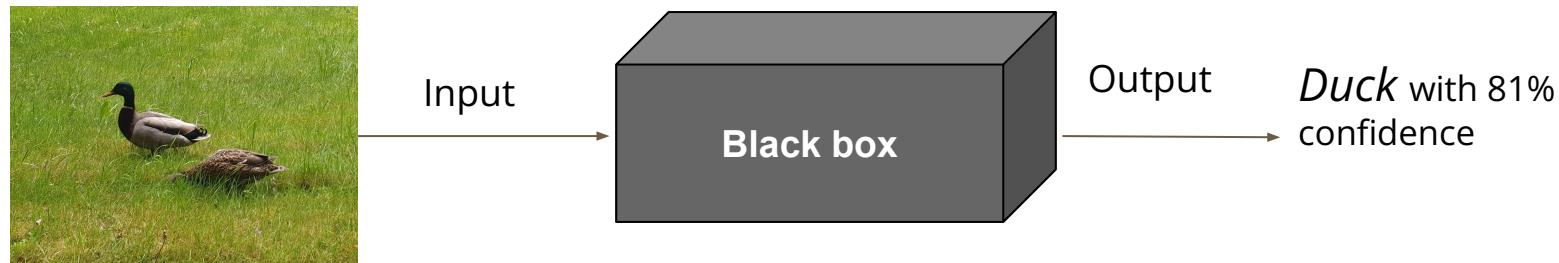
Khánh Nam NGUYỄN & David LAMBERT
M1 BIM 2021-2022

Advisor: Denis Sheynikhovich

Introduction

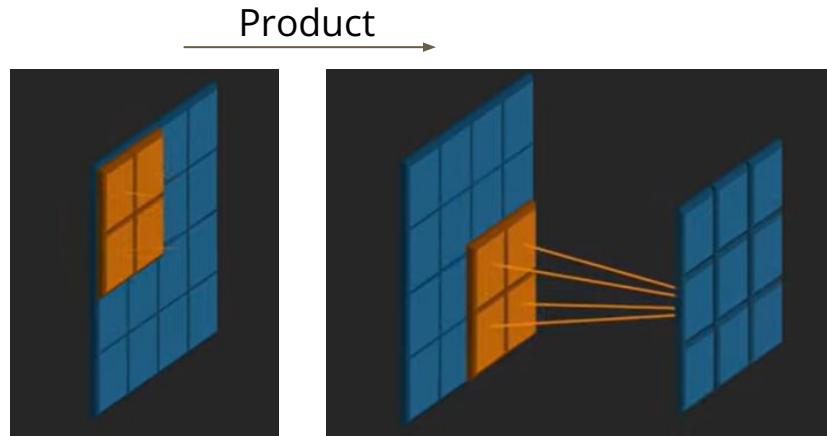


Introduction

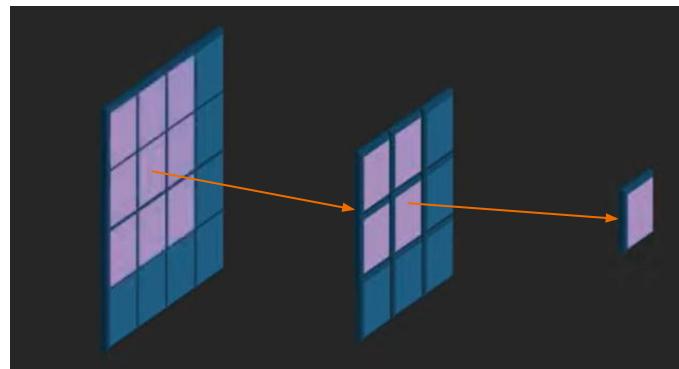


Methods: convolutions

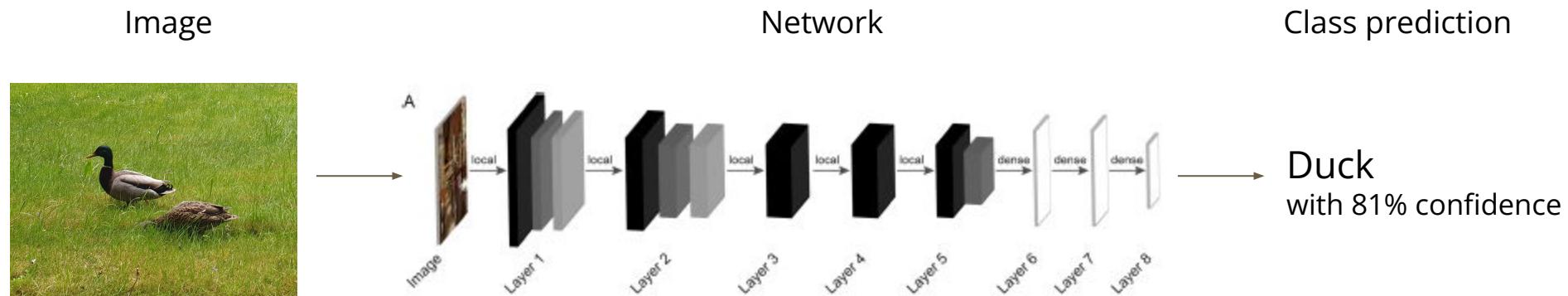
A convolution filter measures the local similarity of an input with a predefined pattern



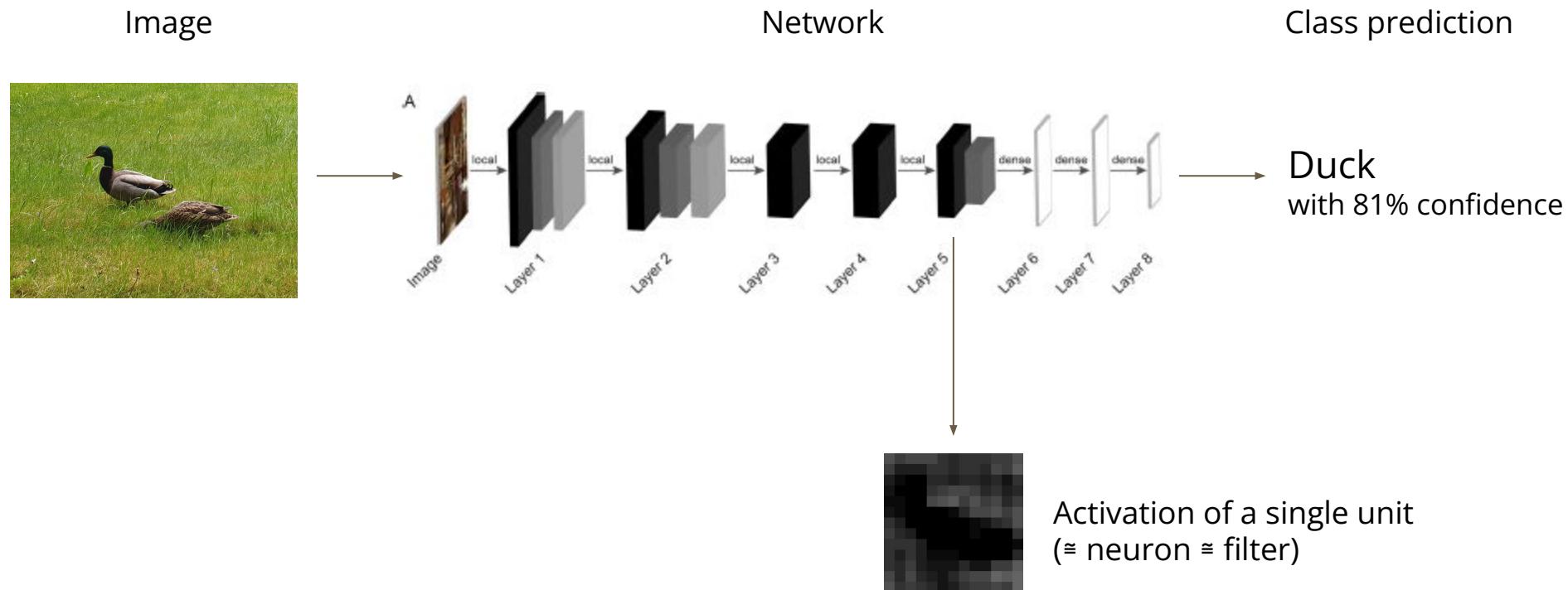
Conservation of information in smaller and smaller outputs in successive layers



Methods: deep convolutional neural networks (deep CNNs)

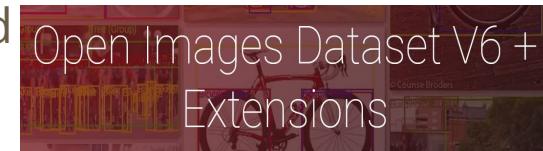


Methods: deep convolutional neural networks (deep CNNs)



Method: Collecting data

- COCO Dataset - 2017 Unlabeled images [123K/19GB]:
 - Common objects in Context
- Open Images Dataset V6 [382K/80GB]: Crowdsourced Extension
 - Household objects, plants, animals, food and people ...
- A subset of the Places dataset containing 4000 images



Methods: Top K images

- Top K images is a set of K images that have the strongest activation in each unit.
- Methods to calculate activation map:
 - max value as criteria
 - mean value as criteria
 - largest activation map

Methods: Top K images

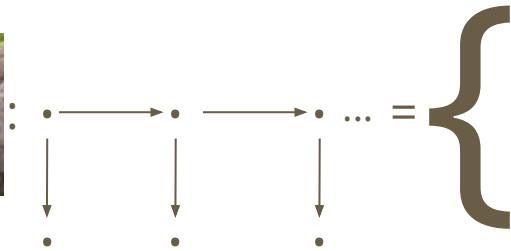
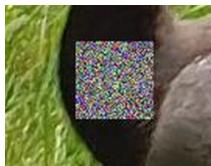
- Top K images is a set of K images that have the strongest activation in each unit.
- Methods to calculate activation map:
 - max value as criteria
 - mean value as criteria
 - largest activation map

Methods: Discrepancy maps and receptive fields

- Creation of a set of occluded (locally disturbed) images



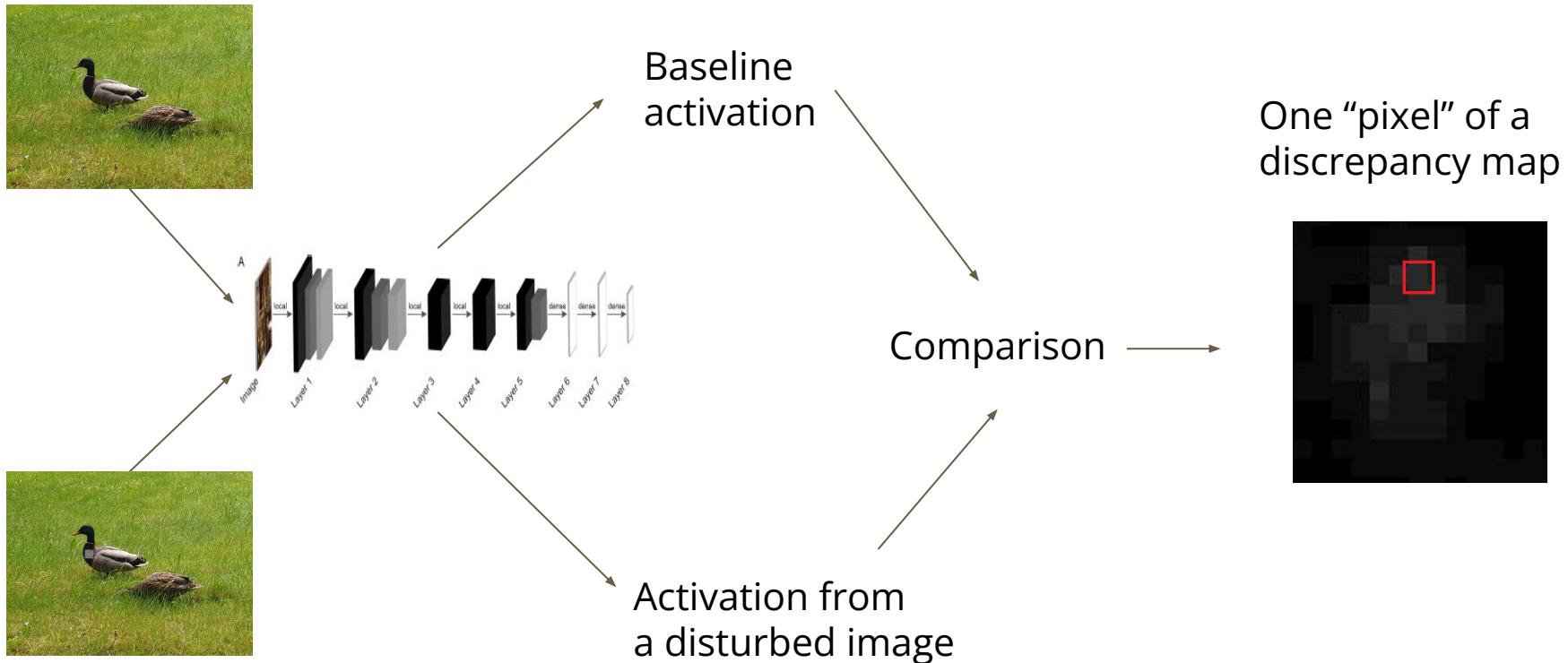
+



- Discrepancy map : comparing activations of original image and occluded image



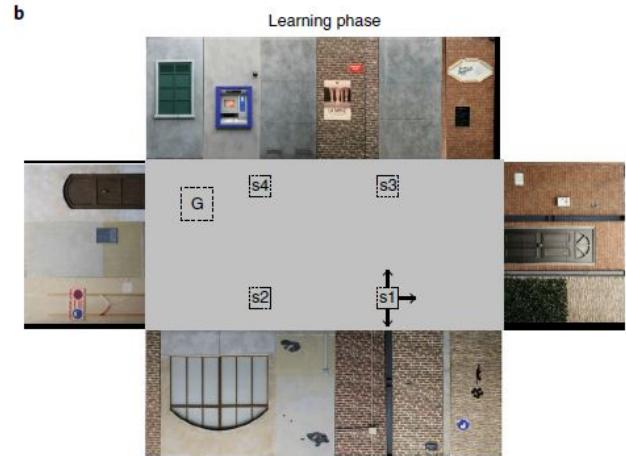
Methods: Discrepancy maps and receptive fields



Methods: the Avatar experiment

A visual navigation experiment, in a controlled setting

Reconstructed images from the point of view of the subject



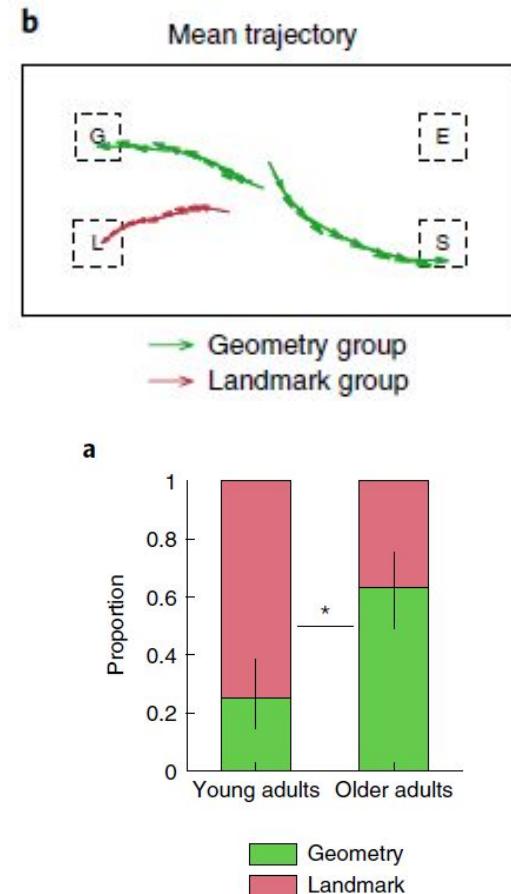
Methods: the Avatar experiment

Two groups emerge based on the predominant category of visual cues they use:

- geometric cues (corners, borders, ...)
- landmark cues (colored objects, panels with text, ...)

Link between predominant visual cue category (orientation behavior) and age

→ Modeling this behavior with an artificial neural network ?



Result: Top K images

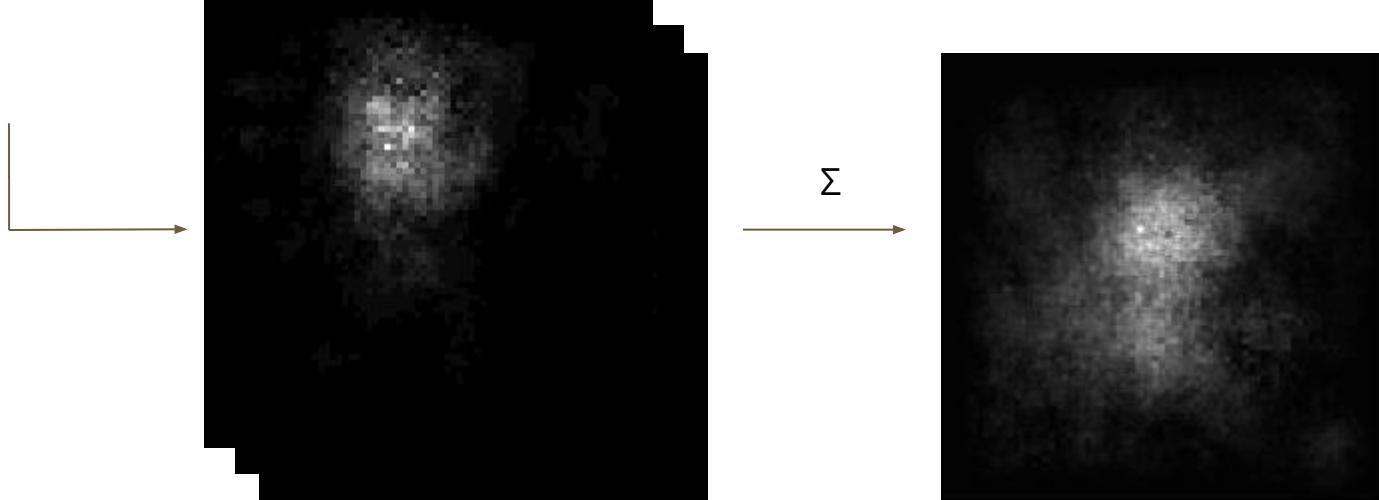
Unit 9 of layer 11:



Unit 0 of layer 11:



Results: receptive fields in AlexNet



Results: AvatarNet

- Tuning the model
 - Redefine the fully connected layer dimension for 2 classes instead of the default one
- Validating the model
 - Using confusion matrix: accuracy = 93,4%

	Geometry	Landmark
Geometry	24821	1637
Landmark	1884	25714

Results: discrepancy maps from the Avatar experiment

Unit 30 of layer 11:
("geometry" labeled
data)

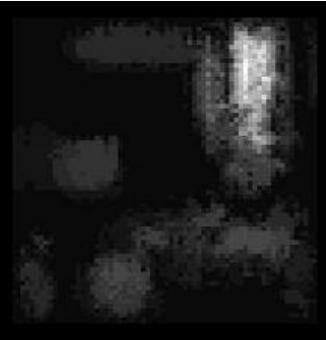


Results: discrepancy maps from the Avatar experiment

Unit 4 of layer 11:
("landmark" labeled
data)



Discussion: feature localization v. identification

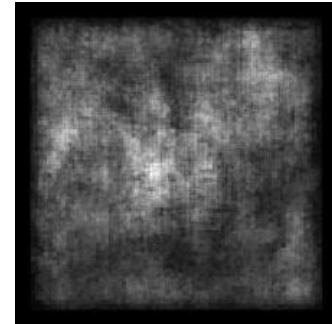


{}

→

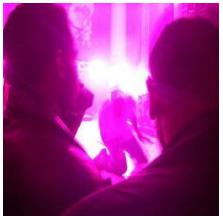
Σ

→

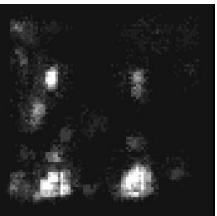
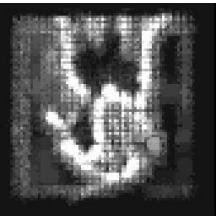


Discussion: feature localization v. identification

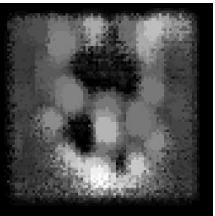
Dataset
images



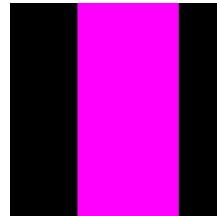
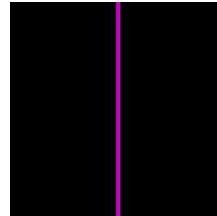
Layer 4,
unit 6



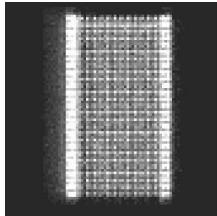
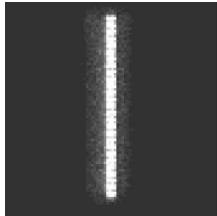
Layer 4,
unit 122



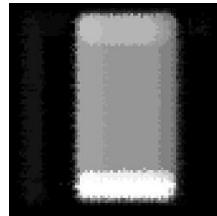
Manufactured
images



Layer 4,
unit 6



Layer 4,
unit 122



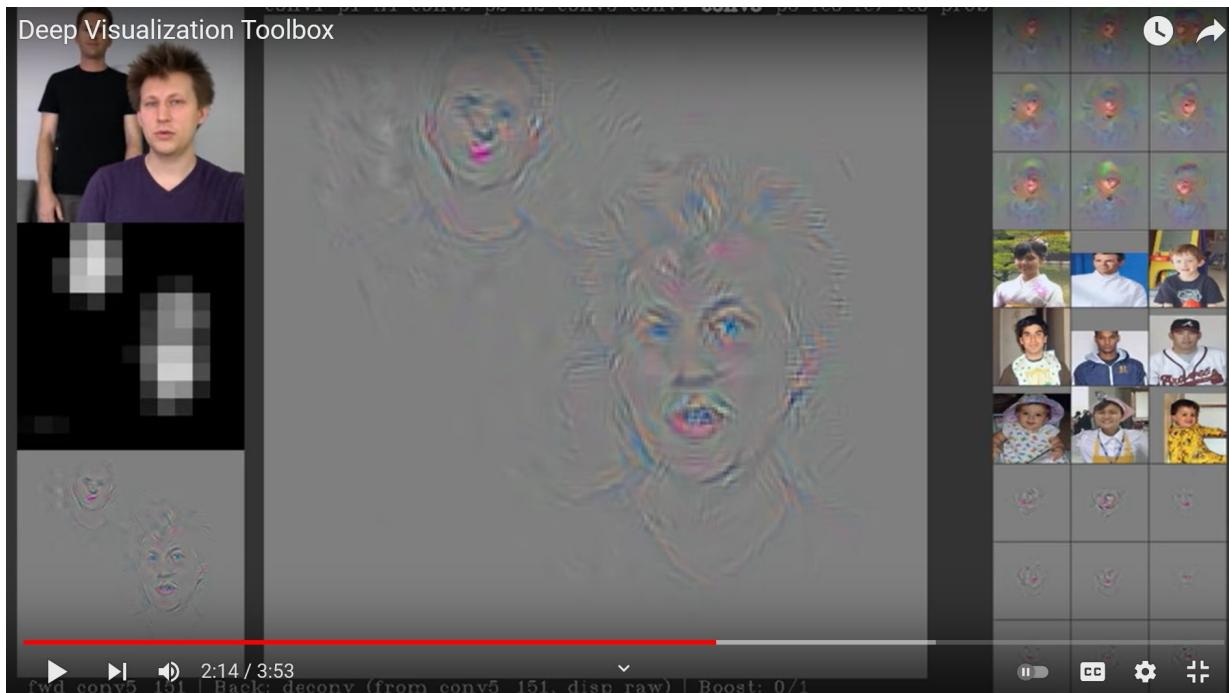
Discussion: possible alternatives

Deepvis toolbox:

- automatic generation of high-activation images
- deconvnets

Other networks :

- Transformers arch. ?
- SqueezeNet ?



Discussion: beyond “weird tricks”

One weird trick for parallelizing convolutional neural networks

Alex Krizhevsky
Google Inc.
akrizhevsky@google.com

April 29, 2014

“Despite the impending widespread deployment of foundation models, *we currently lack a clear understanding of how they work*, when they fail, and what they are even capable of due to their emergent properties.”

Bommasani et al., 2021. *On the Opportunities and Risks of Foundation Models*.
DOI: 10.48550/ARXIV.2108.07258

Discussion: explainability, responsibility, accountability



Replying to [REDACTED] ...

Bruh it's a picture of a fictional Shrimptaur, what ethics are involved here

[REDACTED] May 24, 2022 · Twitter Web App

1 Retweet 61 Likes

Спасибо!