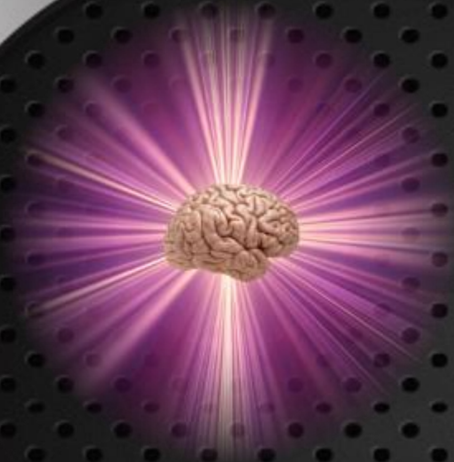
AGH University of Science and Technology
Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering
Department of Biocybernetics and Biomedical Engineering



Knowledge-based CI and DM in Biomedicine

Knowledge and Intelligence
Visualizing and Inspecting ConvNet Filters
Backpropagation and K-fold Cross Validation

Adrian Horzyk
horzyk@agh.edu.pl

Google: Adrian Horzyk

Information, Knowledge, Cognition, and Intelligence

What are knowledge and intelligence?

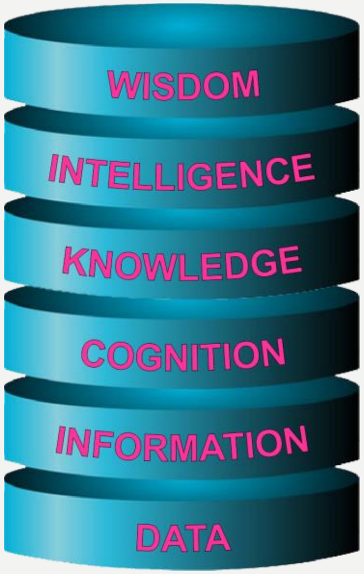
Facts and Information

Fact is a collection of related data that are arranged and ordered consistently.

Information is a collection of related data (facts) perceived by a receiver for whom these data have a certain meaning in the context of the already gained knowledge; and the state of the receiver is influenced by these data or the knowledge is updated, e.g. *the normal temperature of a human body is 36.6 ºC*.

Information creates new or modifies existing associations between known and new objects or data.

The **information receiver** must be able to associate data and pieces of information to understand the transmitted and received information.

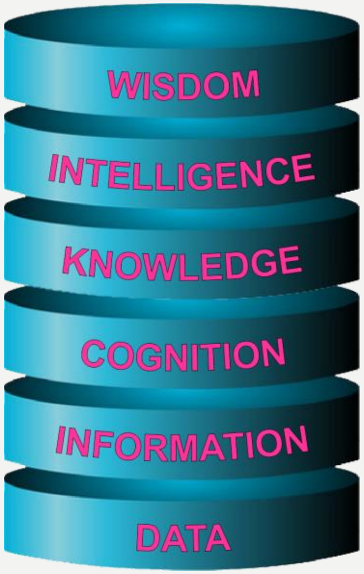


Cognition

Cognition is a mental action leading to the acquisition of knowledge from data and relationships and understanding them through thought, experience, and the senses.

It encompasses many aspects of intellectual functions and processes such as attention, the formation of knowledge, memory, judgment, evaluation, reasoning, problem solving, decision making, comprehension, processing, and using (production) of language.

Cognitive processes use existing knowledge and generate new knowledge for the processed data.

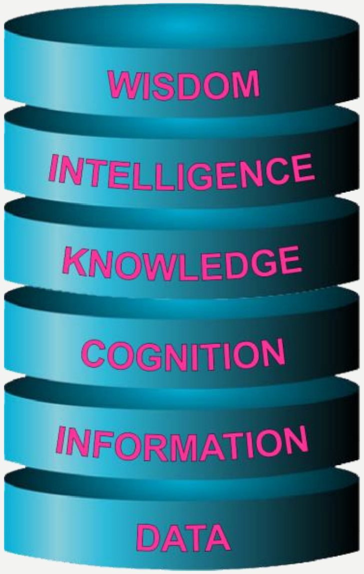


Knowledge

Knowledge is an abstract result of contextual, associative consolidation and representation of patterns, facts, and rules and their generalization, creating new methods, rules, and algorithms of data processing and inference.

In computer science, it can be perceived as a collection of contextually associated information and rules thanks to various relationships between various pieces of information collected over time.

Knowledge is closely related to intelligence because it allows for the inference and development of individual intelligence as well as the development of our individuality, being and nature.

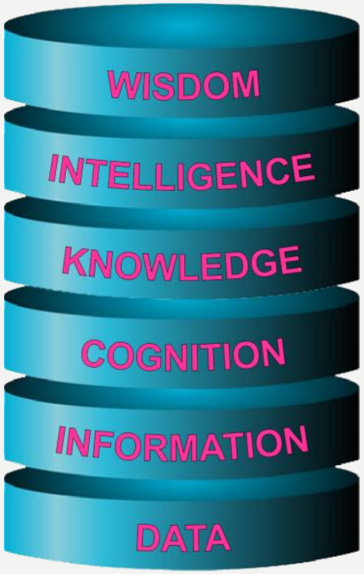


Intelligence

Intelligence is the mental ability to perceive information and use it to form knowledge to apply it to adapt to the environment, to solve a problem, or efficiently achieved goals.

Intelligence is the mental capability of reasoning, planning, solving problems, thinking abstractly, comprehending complex ideas, learn quickly, and use resources efficiently.

It encompasses processes such as learning, recognizing, classification, understanding, logic, planning, creativity, problem-solving, and self-awareness.



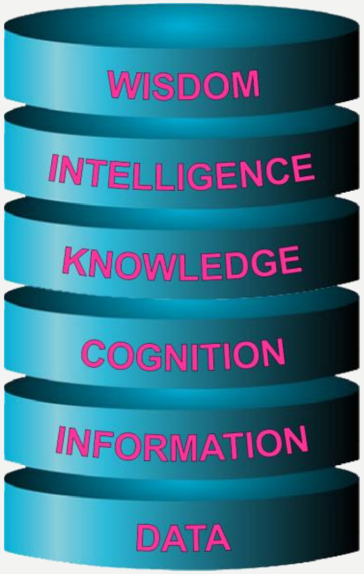
Wisdom

Wisdom is the ability to select the best, the most efficient or most profitable ways to reach the desired outcome or goals based on knowledge, needs, intelligence, and ethical priorities.

Wisdom allows for good judgment and a high quality of being.

Wisdom is usually a result of earlier attempts to reach a successful outcome on the basis of experience, knowledge, and intelligence.

Therefore, **wisdom** is treated as a manifestation of high intelligence and wide knowledge.



Artificial Intelligence

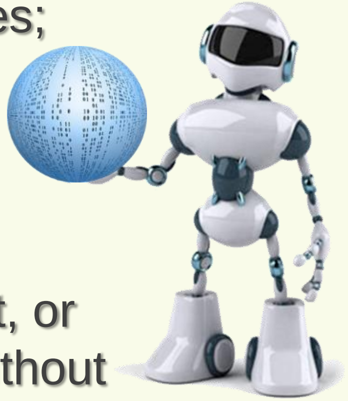
Artificial Intelligence should be able to:

- reproduce and imitate human intelligence;
- recognize and react to human needs and values;
- understand human psychology, personality, needs, and aspirations;
- adapt, learn, remember, and recall objects, facts, rules, and routines;
- recognize similar objects, facts, rules, and routines and generalize them;
- classify objects, associate them in different contexts and recall contextually;
- communicate with people logically and sensitively taking into account their needs and priorities;
- cooperate with people taking into account their weaknesses;
- replace people in frequent or arduous tasks;
- meet the needs of people as well as intelligent machines and define the needs to cooperate to satisfy them.

Computational Intelligence

Computational Intelligence:

- is a set of nature-inspired methodologies and approaches to address complex real-world problems to which mathematical modeling is useless or not efficient enough;
- usually refers to the ability of a computer to learn specific tasks from data or experimental observations, is focused on solving engineering tasks using adaptive mathematical models based on a human way of thinking or other biological processes;
- encompasses artificial neural networks, fuzzy logic, evolutionary computations, genetic algorithms, and various probabilistic methods;
- is used to recognize, classify, group (cluster), predict, or approximate efficiently in order to make decisions without human assistance or help people in the decision processes.



Deep Learning

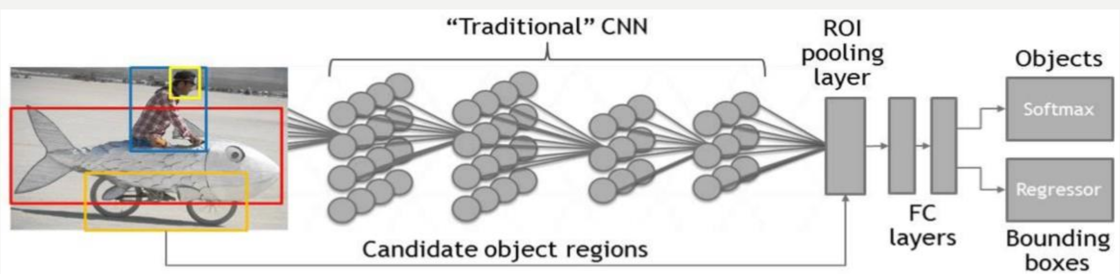
Deep learning develops representation of simple features to represent a hierarchy of more complex features which are finally used to represent objects, classes and solve problems.

Today, we have many different deep learning systems which were already described and discussed during these lectures and implemented during laboratory and project classes.

Deep learning networks always have a hierarchical structure constructed from various layers, modules, or subnetworks. They are not always and not necessarily neural networks.

Deep learning systems are a part of:

• Learning systems • Representative learning systems • Hierarchical systems



(Fast) Region based Convolutional Networks (R-CNN)
Ross Girshick, Microsoft Research
https://github.com/rbgirshick/fast-rcnn

Visualization of Filters and Represented Patterns

Look inside the convolutional neural network?

Look Inside the Filters of CNN

How can we visualize what convnets have learned?

The answer to this question is especially important when we want to be sure that our network will generalize predict well based on really vivid features of the objects in the images, not on the basis of their surroundings!

Moreover, when our network does not generalize well, we can find out the reason for its looking into its internal representation of filters.

We can do three interesting visualizations:

1. Visualizing intermediate convnet outputs ("intermediate activations").
   This is useful to understand how successive convnet layers transform their input, and to get a first idea of the meaning of individual convnet filters.
2. Visualizing convnets filters.
   This is useful to understand precisely what visual pattern or concept each filter in a convnet is receptive to.
3. Visualizing heatmaps of class activation in an image.
   This is useful to understand which part of an image were identified as belonging to a given class, and thus allows to localize objects in images.

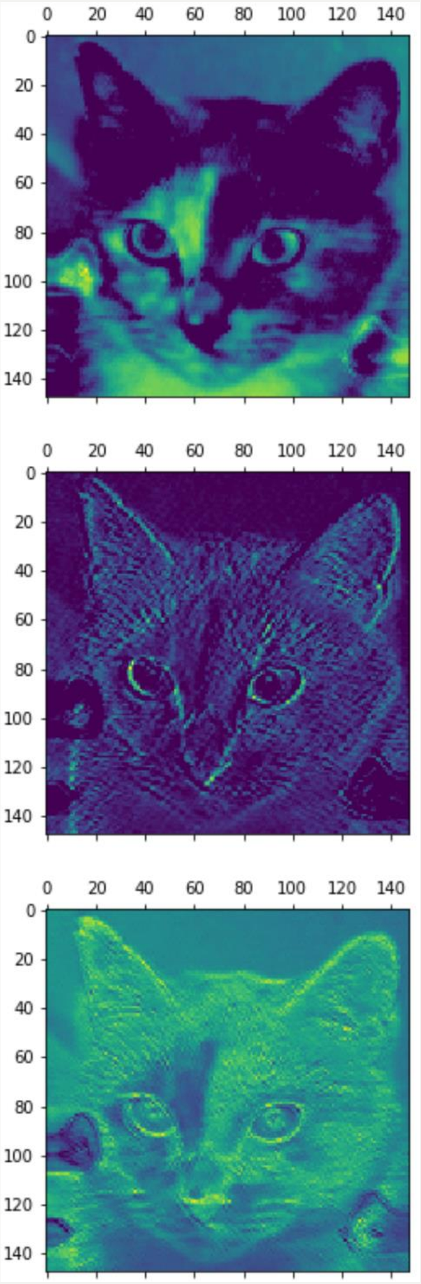
Visualizing intermediate activations

Visualizing intermediate activations consists in displaying the feature maps that are output by various convolution and pooling layers in a network for given a certain input.

The output of a layer is often called its activation, i.e., the output of the activation function.

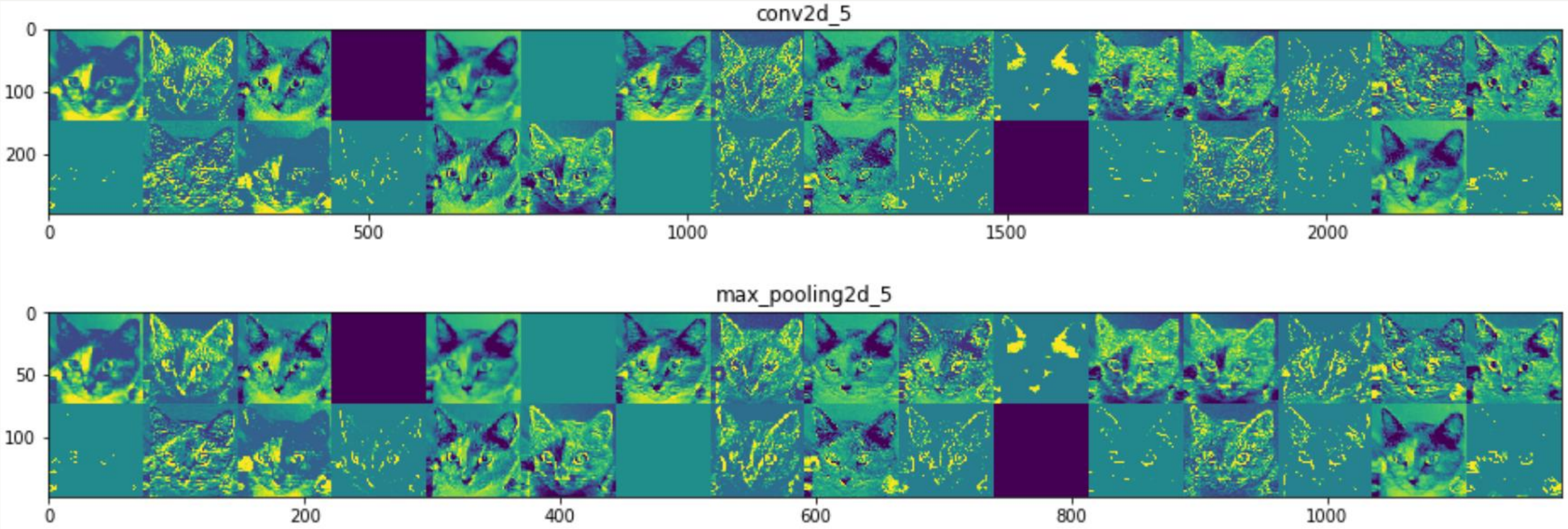
This gives a view into how an input is decomposed unto the different filters learned by the network. These feature maps we want to visualize have 3 dimensions: width, height, and depth (channels).

Each channel encodes relatively independent features, so the proper way to visualize these feature maps is by independently plotting the contents of every channel, as a 2D image.



Visualizing intermediate activations

We can plot intermediate activations of all channels and all layers:

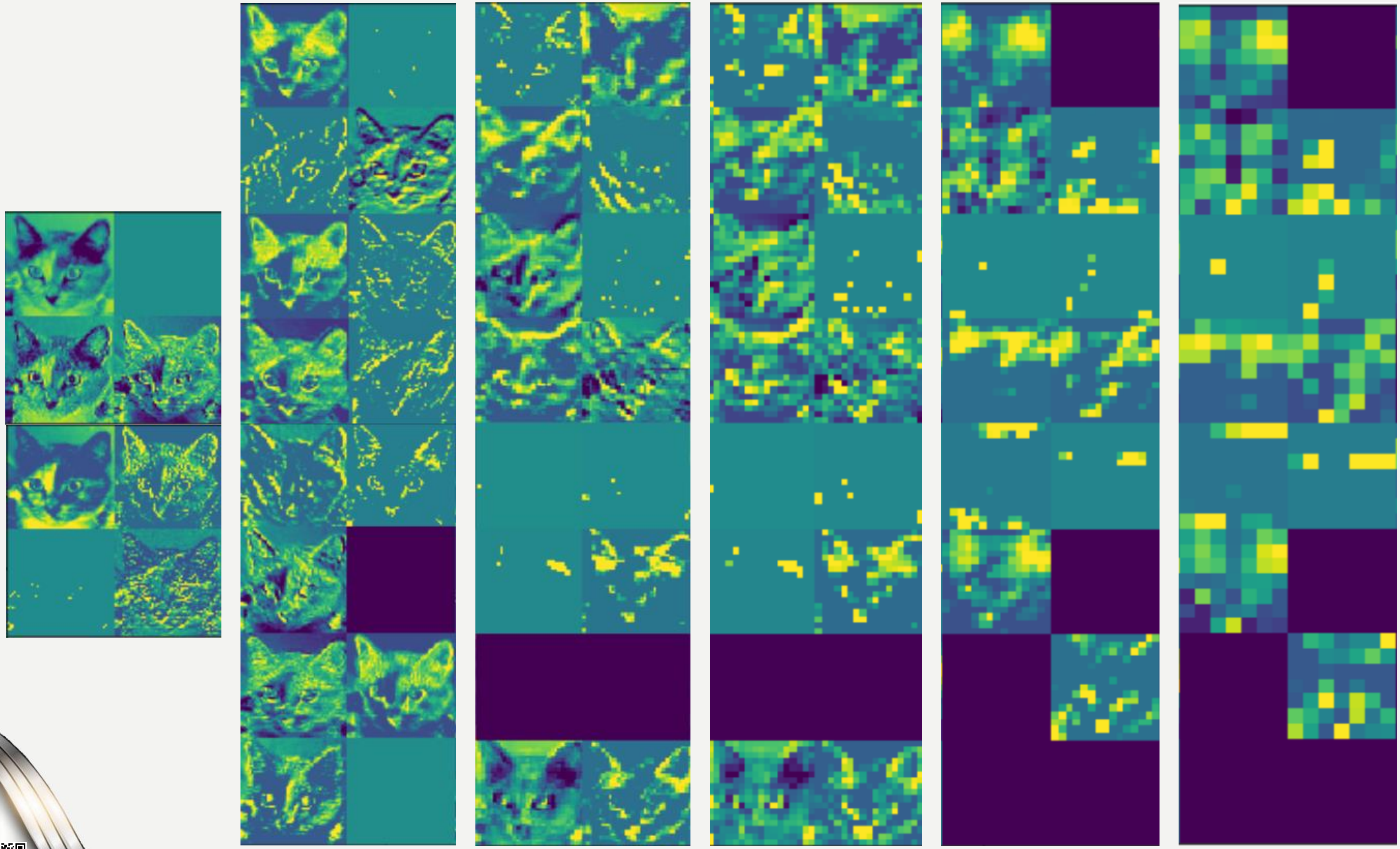


Black squares show the null intermediate values for some filters.

It means that these filters have not reacted to the presented image with activation, so no interesting feature was present in the input image from these filters point of view.

Visualizing intermediate activations

Subsequent layers presents still more abstract features of the input image:



Remarks

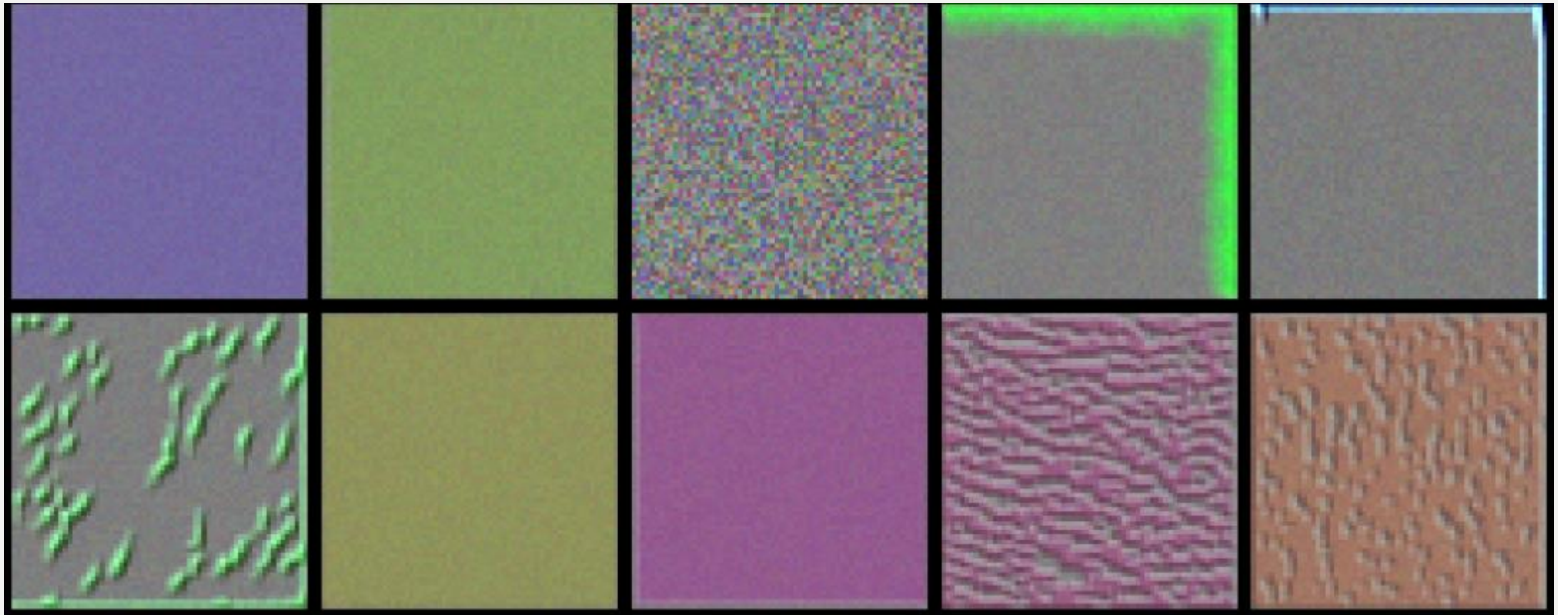
- The first layer acts as a collection of various edge detectors.
  At that stage, the activations are still retaining almost all of the information present in the initial picture.
- As we go higher-up, the activations become increasingly abstract and less visually interpretable.
  They start encoding higher-level concepts such as "cat ear" or "cat eye".
  Higher-up presentations carry increasingly less information about the visual contents of the image, and increasingly more information related to the class of the image.
- The sparsity of the activations is increasing with the depth of the layer:
  in the first layer, almost all filters are activated by the input image,
  but in the following layers more and more filters are usually blank (black rectangles).
  This means that the pattern encoded by the filter isn't found in the input image.
- We can notice that sometimes some filters are very similar to each other;
  as a result, they code similar features. This might limit the discrimination properties of the network, so maybe the layers should be unfrozen to train better representation and discrimination of the dataset features.
- Sometimes the bright pixels are presented outside the body of the cat.
  It means that the classification of the cats is processed based on the environment features where the cats were usually photographed!
  Sometimes it might be an advantage, but sometimes it might be a problem, so bear in mind to pay attention to such bright features outside the classified objects.

Visualizing Convnet Filters

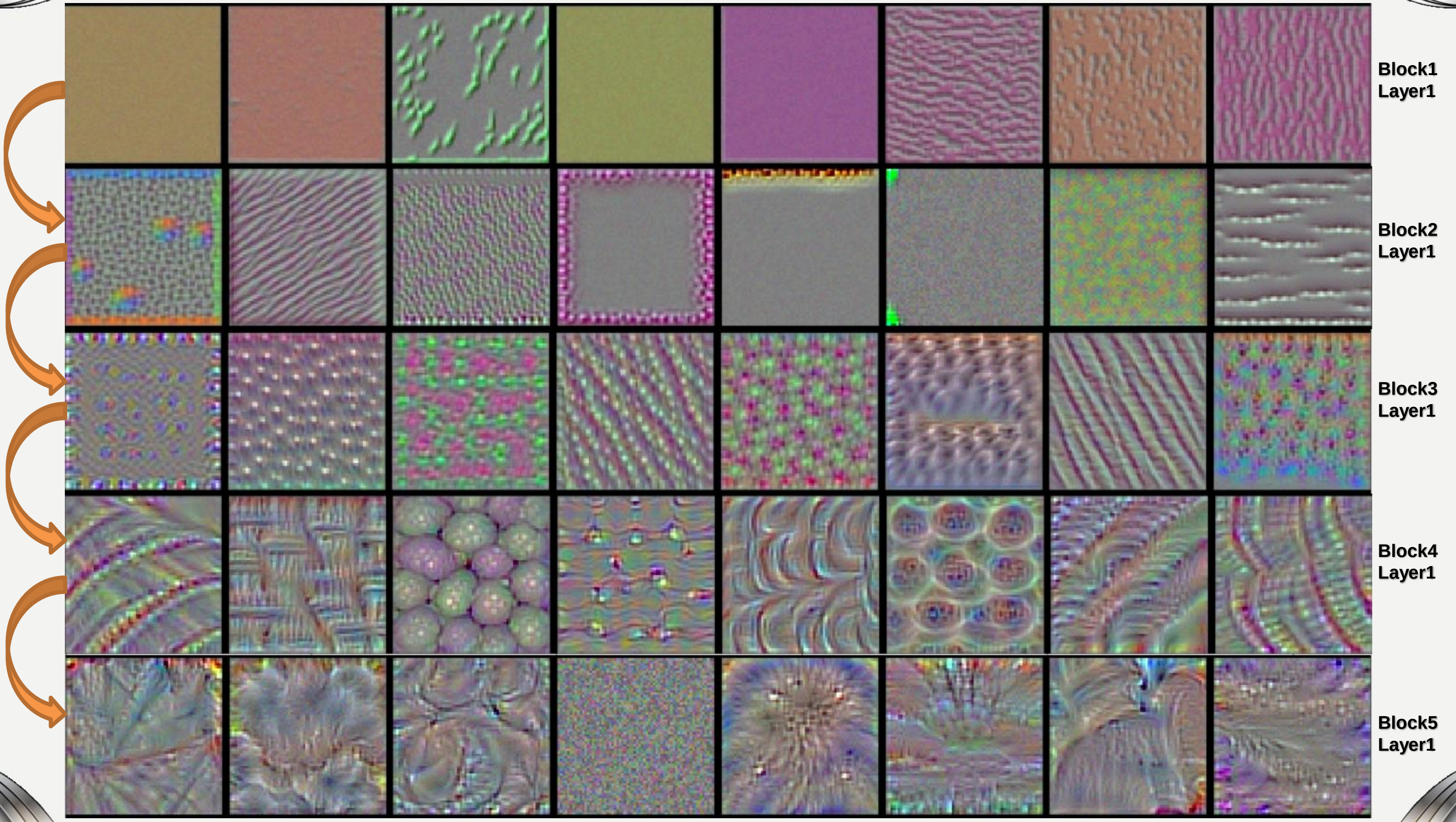
Another essential thing about how to inspect **the filters learned by a convnet** is to **display the visual pattern** that each filter is meant to respond to.

This can be done with gradient ascent in input space:
applying gradient descent to the value of the input image of
a convnet so as to maximize the response of a specific filter,
starting from a blank input image.

The resulting input image would be one that the chosen filter is maximally responsive to:



Visualizing Convnet Filters of VGG-16 trained with ImageNet



These filter visualizations tell us a lot about how convnet layers see the world: each layer in a convnet simply learns a collection of filters such that their inputs can be expressed as a combination of the filters.

Visualizing Heatmaps of Class Activation

We will introduce a visualization technique that is useful for understanding which parts of a given image led a convnet to its final classification decision.

This is helpful for "debugging" the decision process of a convnet, in particular in case of classification mistakes.

This techniques is called Class Activation Map" (CAM) visualization, and consists in producing heatmaps of "class activation" over input images.

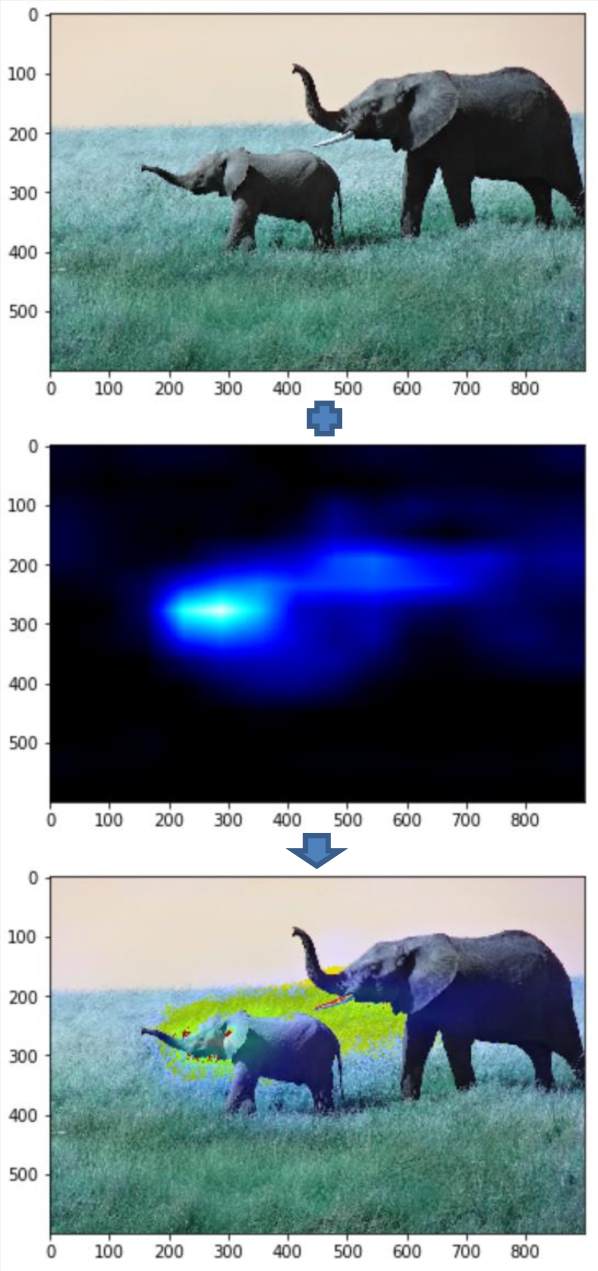
A "class activation" heatmap is a 2D grid of scores associated with a specific output class, computed for every location in any input image, indicating how important each location is with respect to the class considered.

Visualizing Heatmaps of Class Activation

This visualisation technique answers two important questions:

- Why did the network think this image contained an African elephant?
- Where is the African elephant located in the image?

In particular, it is interesting to note that the ears of the elephant cub are strongly activated: this is probably how the network can recognize the difference between African and Indian elephants.



BIBLIOGRAPY

1. A. Horzyk, J. A. Starzyk, J. Graham, Integration of Semantic and Episodic Memories, IEEE Transactions on Neural Networks and Learning Systems, 2017, DOI: 10.1109/TNNLS.2017.2728203.
2. A. Horzyk and J.A. Starzyk, Fast Neural Network Adaptation with Associative Pulsing Neurons, IEEE Xplore, In: 2017 IEEE Symposium Series on Computational Intelligence, 2017.
3. Basawaraj, Janusz A. Starzyk and A. Horzyk, Lumped Mini-Column Associative Knowledge Graphs, IEEE Xplore, In: 2017 IEEE Symposium Series on Computational Intelligence, 2017.
4. A. Horzyk, Deep Associative Semantic Neural Graphs for Knowledge Representation and Fast Data Exploration, Proc. of KEOD 2017, SCITEPRESS Digital Library, 2017.

BIBLIOGRAPY

6. A. Horzyk, Neurons Can Sort Data Efficiently, Proc. of ICAISC 2017, Springer-Verlag, LNAI, 2017, pp. 64-74, ICAISC BEST PAPER AWARD 2017 sponsored by Springer.
7. A. Horzyk, J. A. Starzyk and Basawaraj, Emergent creativity in declarative memories, IEEE Xplore, In: 2016 IEEE Symposium Series on Computational Intelligence, Greece, Athens: Institute of Electrical and Electronics Engineers, Curran Associates, Inc. 57 Morehouse Lane Red Hook, NY 12571 USA, 2016, ISBN 978-1-5090-4239-5, pp. 1-8, DOI: 10.1109/SSCI.2016.7850029.
8. A. Horzyk, Human-Like Knowledge Engineering, Generalization and Creativity in Artificial Neural Associative Systems, Springer-Verlag, AISC 11156, ISSN 2194-5357, ISBN 978-3-319-19089-1, ISBN 978-3-319-19090-7 (eBook), DOI 10.1007/978-3-319-19090-7, Springer, Switzerland, 2016, 39-51.

BIBLIOGRAPY

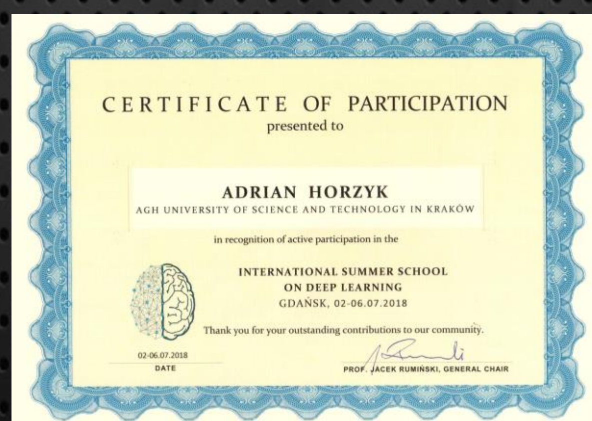
9. A. Horzyk, Innovative Types and Abilities of Neural Networks Based on Associative Mechanisms and a New Associative Model of Neurons - Invited talk at ICAISC 2015, Springer-Verlag, LNAI 9119, 2015, pp. 26-38, DOI 10.1007/978-3-319-19324-3_3.
10. Horzyk, A., How Does Generalization and Creativity Come into Being in Neural Associative Systems and How Does It Form Human-Like Knowledge?, Neurocomputing, 2014.
11. Horzyk, A., Human-Like Knowledge Engineering, Generalization and Creativity in Artificial Neural Associative Systems, Springer, AISC 11156, 2014.
12. A. Horzyk and J.A. Starzyk, Associative Fine-Tuning of Biologically Inspired Active Neuro-Associative Knowledge Graphs, In: 2018 IEEE Symposium Series on Computational Intelligence (SSCI 2018), IEEE Xplore, pp. 2068-2075, 2018.

BIBLIOGRAPY

13. Francois Chollet, “Deep learning with Python”, Manning Publications Co., 2018.
14. Ian Goodfellow, Yoshua Bengio, Aaron Courville, “Deep Learning”, MIT Press, 2016, ISBN 978-1-59327-741-3.
15. Home page for this course: http://home.agh.edu.pl/~horzyk/lectures/ahdydci.php
16. Nikola K. Kasabov, Time-Space, Spiking Neural Networks and Brain-Inspired Artificial Intelligence, In Springer Series on Bio- and Neurosystems, Vol 7., Springer, 2019.
17. Holk Cruse, Neural Networks as Cybernetic Systems, 2nd and revised edition
18. R. Rojas, Neural Networks, Springer-Verlag, Berlin, 1996.
19. Convolutional Neural Network (Stanford)
20. Visualizing and Understanding Convolutional Networks, Zeiler, Fergus, ECCV 2014.

BIBLIOGRAPY

20. Home page for this course:
http://home.agh.edu.pl/~horzyk/lectures/ahdydkbcidmb.php



CERTIFICATE OF PARTICIPATION

presented to

ADRIAN HORZYK

AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN KRAKÓW

in recognition of active participation in the

INTERNATIONAL SUMMER SCHOOL
ON DEEP LEARNING
GDAŃSK, 02-06.07.2018

Thank you for your outstanding contributions to our community.

02-06.07.2018
DATE

PROF. JACEK RUMIŃSKI, GENERAL CHAIR