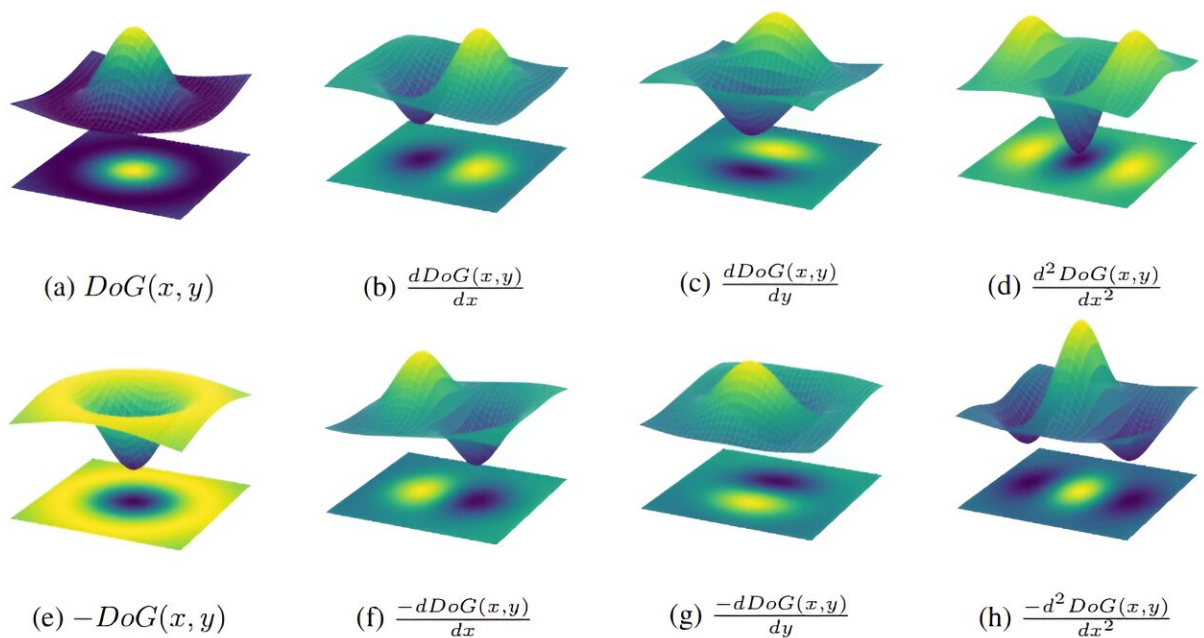


# Researchers investigating how AI categorizes images find similarities to visual systems in nature

June 5 2024



Complete diagram of the On and Off DoG filters and their derivatives. Credit: <https://openreview.net/pdf?id=4VgBjsOC8k>

Researchers at TU Wien have investigated how an artificial intelligence categorizes images. The results show astonishing similarities to visual systems in nature.

How do you teach a machine to recognize objects in images? Huge progress has been made in this area in recent years. With the help of neural networks, for example, images of animals can be assigned to the respective species with a very high success rate. This is achieved by training a neural network with the help of many sample images. The network is adapted step by step until it ultimately yields the correct answers as precisely as possible.

Usually, however, it remains a mystery as to which structures are formed in the process and which mechanisms develop in the neural network that ultimately lead to the goal.

A team from TU Wien, led by Prof. Radu Grosu, and a team from MIT, led by Prof. Daniela Rus, have now investigated precisely this question and arrived at some astonishing results: Structures are formed in the artificial neural network that bear a striking resemblance to structures that occur in the nervous system of animals or humans.

The team presented their [research](#) at the International Conference on Learning Representations ([ICLR 2024](#)) in Vienna in May.



Peyman M. Kiasari, Zahra Babaiee and Radu Grosu (left to right). Credit: Vienna University of Technology

## Several layers of neurons

"We work with so-called convolutional neural networks. These are artificial neural networks that are often used to process [image data](#)," says Zahra Babaiee from the Institute of Computer Engineering at TU Wien. She is the first author of the paper and did part of her work together with Daniela Rus at MIT and part of her work with Peyman M. Kiasari and Radu Grosu at TU Wien.

The design of these networks was inspired by the biological neural networks in our eyes and brain. There, visual impressions are processed

by several layers of neurons. Certain neurons become active, for example, when they are activated by light signals in the eye, and transmit signals to neurons in the layer behind them.

In artificial neural networks, this principle is digitally imitated on a computer: the desired input—for example a digital image—is transferred pixel by pixel to the first layer of artificial neural networks. The activity of the neurons in this first layer simply depends on whether they are presented with a lighter or darker pixel.

These activity values of the neurons in the first layer are then used to determine the activity of the neurons in the next layer: Each of the neurons in the subsequent layer combines the signals from the first layer according to a very specific individual pattern (according to a very specific formula), and this yields a value that is then used to determine the activity of the neuron in the next layer.

## **Astonishing similarity to biological neural networks**

"In convolutional neural networks, not all neurons in one layer play a role for every neuron in the next layer," explains Zahra Babaiee. "Even in the brain, not every neuron in a layer is connected to all neurons in the previous layer without exception, but only to the neighboring neurons in a very specific area."

In [convolutional neural networks](#), so-called "[filters](#)" are therefore used to decide which neurons have an influence on a particular subsequent neuron and which do not. These filters are not predetermined, but are shaped automatically during the training of the [neural network](#).

"While the network is being trained with many thousands of images, these filters and other parameters are constantly being adjusted. The algorithm tries out which weighting of the neurons from the previous

layer leads to the best result until the images are assigned to the correct category with the highest possible reliability," says Zahra Babaiee. "The algorithm does this automatically, we have no direct influence on it."

However, at the end of the training, it is possible to analyze which filters have developed in this way. And this reveals interesting patterns: the filters do not take on completely random forms, but fall into several simple categories.

"Sometimes the filters develop in such a way that one neuron is particularly strongly influenced by the neuron directly in front of it and hardly at all by others," says Zahra Babaiee.

Other filters look cross-shaped, or they show two opposite areas—one whose neurons have a strongly positive influence on the activity of the neuron in the next layer, and another whose neurons have a strongly negative influence.

"The amazing thing is that exactly these patterns have already been observed in biological nervous systems, for example in monkeys or cats," says Zahra Babaiee. In humans, the processing of visual data is likely to work in the same way.

It is probably no coincidence that [biological evolution](#) has produced the same filter functions that arise in an automated machine learning process. "If you know that precisely these structures are formed again and again during visual learning, then you can already take this into account in the training process and develop machine learning algorithms that reach the desired result much faster than before," says Zahra Babaiee.

**More information:** Unveiling the Unseen: Identifiable Clusters in Trained Depthwise Convolutional Kernels.

[openreview.net/pdf?id=4VgBjsOC8k](https://openreview.net/pdf?id=4VgBjsOC8k)

Provided by Vienna University of Technology

Citation: Researchers investigating how AI categorizes images find similarities to visual systems in nature (2024, June 5) retrieved 16 August 2024 from <https://techxplore.com/news/2024-06-ai-categorizes-images-similarities-visual.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.