Artificial intelligence: Towards a better understanding of the underlying mechanisms

November 27 2019

AI will serve to develop a network control system that not only detects and reacts to problems but can also predict and avoid them. Credit: CC0 Public Domain

The automatic identification of complex features in images has already become a reality thanks to artificial neural networks. Some examples of software exploiting this technique are Facebook's automatic tagging system, Google's image search engine and the animal and plant

recognition system used by iNaturalist. We know that these networks are inspired by the human brain, but their working mechanism is still mysterious.

New research, conducted by SISSA in association with the Technical University of Munich and published for the 33rd Annual NeurIPS Conference, proposes a new approach for studying deep neural networks and sheds new light on the image elaboration processes that these networks are able to carry out.

Similar to what happens in the visual system, neural networks used for automatic image recognition analyse the content progressively, through a chain of processing stages. However, to date, it is not completely clear which mechanisms allow deep networks to reach their extraordinary levels of accuracy.

"We have developed an innovative method to systematically measure the level of complexity of the information encoded in the various layers of a deep network—the so-called intrinsic dimension of image representations," say Davide Zoccolan and Alessandro Laio, respectively neuroscientist and physicist at SISSA. "Thanks to a multidisciplinary work that has involved the collaboration of experts in physics, neurosciences and machine learning, we have succeeded in exploiting a tool originally developed in another area to study the functioning of deep neural networks."

SISSA scientists, in association with Jakob Macke of the Technical University of Munich, have examined how the information acquired from neural networks used for image classification is processed: "We have found that image representations undergo a progressive transformation. In the early processing stages, image information is faithfully and exhaustively represented, giving rise to rich and complex representations. In the final processing stages, the information is

radically simplified, producing image representations that are supported by a few dozen parameters," explain the two scientists. "Surprisingly we found that the classification accuracy of a neural network tightly depends on its ability to simplify: the more it simplifies the information, the more accurate it is."

This is an especially important result for SISSA that has recently launched a new research program in data science, with the goal of studying and developing innovative algorithms for the processing of complex and large data sets.

More information: Alessio Ansuini et al. Intrinsic dimension of data representations in deep neural networks: papers.nips.cc/paper/8843-intr … -neural-networks.pdf

Provided by International School of Advanced Studies (SISSA)