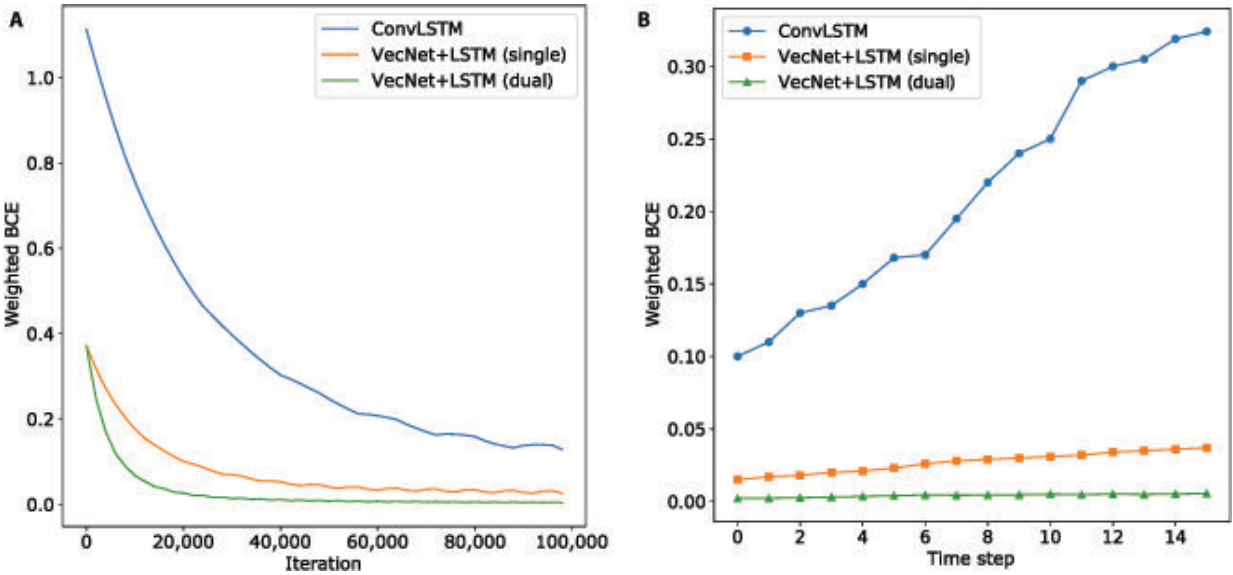# When it comes to neural networks learning motion, it's all relative

March 29 2023

Quantitative comparison of existing method ConvLSTM and new method VecNet+LSTM. (A) Change of prediction accuracy with respect to training iterations. Lower weighted BCE means higher prediction accuracy. (B) Change of prediction accuracy over time steps. Credit: Hehe Fan et al.

Seeking to explore the capabilities of neural networks for recognizing and predicting motion, a group of researchers led by Hehe Fan developed and tested a deep learning approach based on relative change in position encoded as a series of vectors, finding that their method worked better than existing frameworks for modeling motion. The

group's key innovation was to encode motion separately from position.

The group's research was published in *Intelligent Computing*.

The new method, VecNet+LSTM, scored higher than six other artificial neural network frameworks within the field of video research when tested on recognition of motion. Some of the other frameworks were merely weaker, while others were totally unsuitable for modeling motion.

When measured against the common ConvLSTM method for motion prediction, the new method was more accurate, required less time to train and did not lose accuracy as quickly when making additional predictions.

The paper concludes that "modeling relative position change is necessary for motion recognition and makes motion prediction easier."

This research suggests future directions for machine learning for video analysis, since motion recognition, together with object recognition, is the basis for recognizing actions. In other words, even if a neural network can recognize a door, if it cannot learn the motion "open," then it cannot learn the action of opening a door. The method also holds promise for video prediction, though it deals with the motion of individual points rather than of whole systems.

A good model for motion is necessary for artificial intelligence approaches that try to build up a holistic picture of the world by integrating different forms of knowledge. In other words, if a neural network cannot learn motion, then it cannot learn the characteristic action of an object, such as a door opening.

The researchers consider motion as a sequence of arrows or "vectors,"

each one of a certain length, pointing in a certain direction. Each [vector](#) in their experiment can be thought of as a pair of image frames showing the "before" and "after" positions of a small white dot moving on a black surface during one unit of time. The vectors can also be thought of as a pair of two numbers representing movement in two dimensions, a horizontal movement and a [vertical movement](#).

The researchers' neural network, VecNet, first had to learn from a series of examples how the "before" and "after" frames given to it change the position of the white dot. There are separate VecNet components that learn the starting position, horizontal movement, vertical movement and final position of the dot.

Since one vector is not enough for motion recognition, another component was introduced for adding together the vectors over time. This "long short-term memory" component can remember multiple individual movements and thus guess what the next movement step or steps will be, so it can be used for motion prediction as well as motion recognition. The combined system for recognizing and/or predicting motion is thus called VecNet+LSTM.

The advantage of using vectors is that they represent [motion](#) and speed in the most abstract, dictionary sense: they show the amount of change in the position of an object in a period of time, separately from any set of coordinates in the spatial environment. Thus, for example, if the white dot moves in a circle in the top left corner of the black surface, the network can recognize this situation as somewhat the same as the one in which the white dot moves in a circle in the bottom right corner of the black surface.

More information: Hehe Fan et al, How Deep Neural Networks Understand Motion? Toward Interpretable Motion Modeling by Leveraging the Relative Change in Position, *Intelligent Computing*

(2023).

Provided by Intelligent Computing