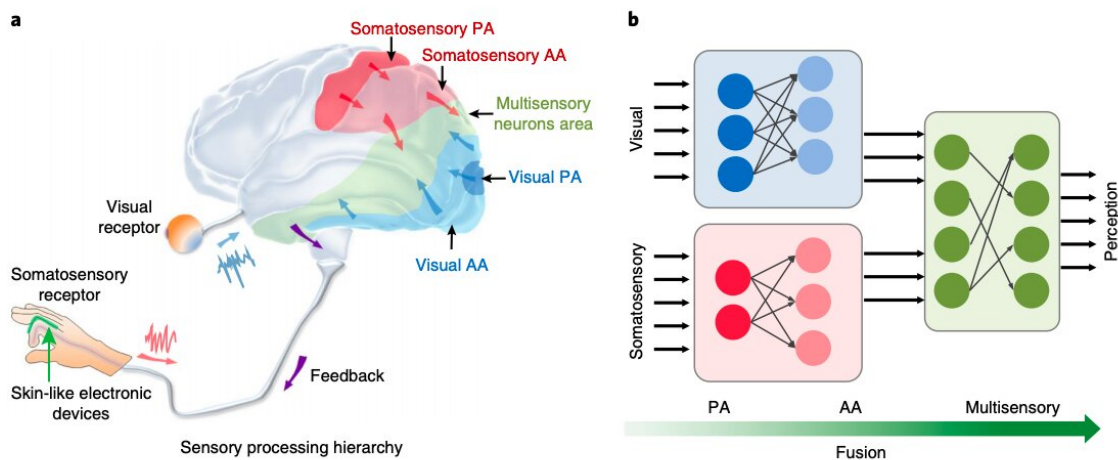# A brain-inspired architecture for human gesture recognition

July 14 2020, by Ingrid Fadelli

Bioinspired somatosensory–visual associated learning framework. Credit: Wang et al.

Researchers at Nanyang Technological University and University of Technology Sydney have recently developed a machine learning architecture that can recognize human gestures by analyzing images captured by stretchable strain sensors. The new architecture, presented in a paper published in *Nature Electronics*, is inspired by the functioning of the human brain.

"Our idea originates from how the human brain processes information,"

Xiaodong Chen, one of the researchers who carried out the study, told TechXplore. "In the human brain, high perceptual activities, such as thinking, planning and inspiration, do not only depend on specific sensory information, but are derived from a comprehensive integration of multi-sensory information from diverse sensors. This inspired us to combine visual information and somatosensory information to implement high-precision gesture recognition."

When humans are solving practical tasks, they typically integrate visual and somatosensory information gathered from their surrounding environment. These two types of information are complementary, as combined, they provide a better idea of all the elements involved in the problem that one is trying to solve.

When developing their technique for human gesture recognition, therefore, Chen and his colleagues ensured that it could integrate different types of sensory information gathered by multiple sensors. Ultimately, their goal was to build an architecture that could recognize human gestures with remarkably high accuracy.
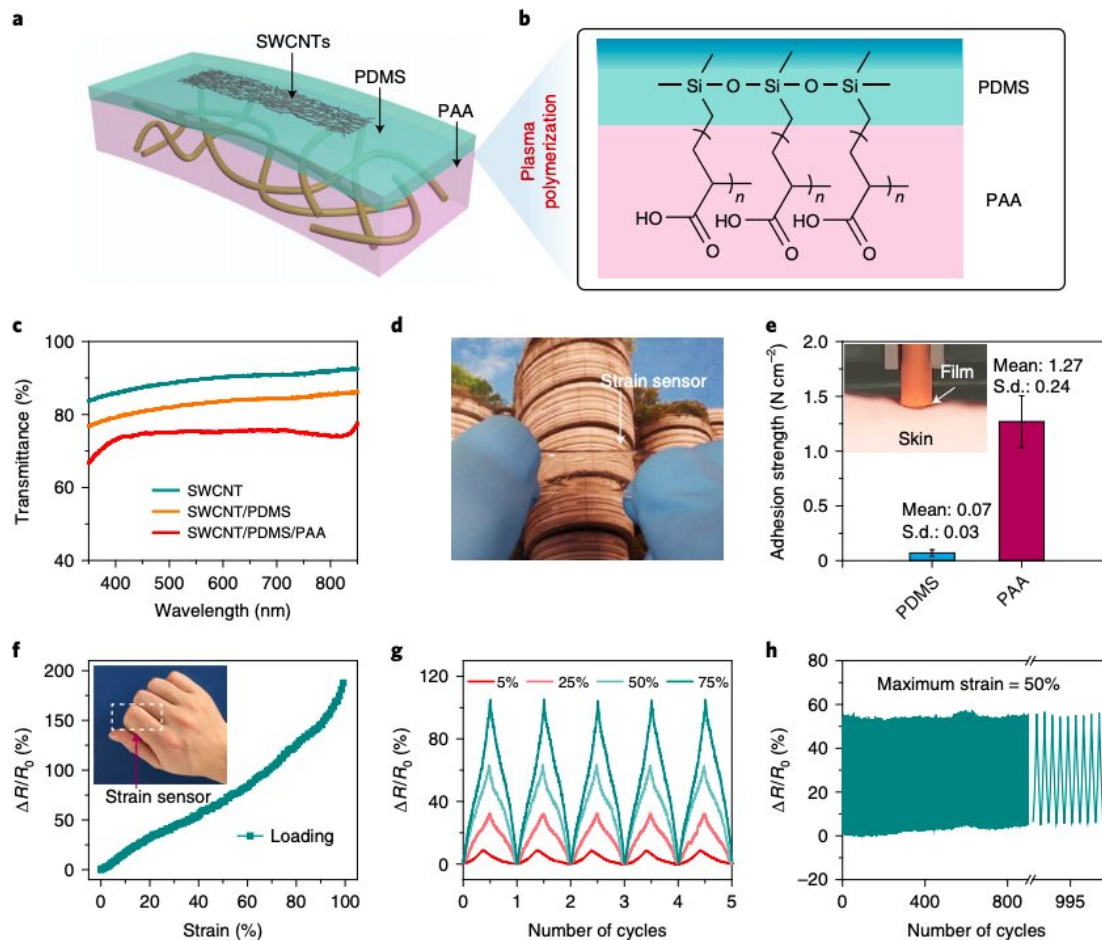
Conformable, transparent and adhesive stretchable strain sensor. Credit: Wang et al.

"To achieve our aim, we improved sensor data quality, by designing and fabricating stretchable and conformable sensors that could gather more accurate somatosensory data of hand gestures compared to current wearable sensors," Chen said. "In addition, we developed a bioinspired somatosensory-visual (BSV) learning architecture that can rationally fuse visual information and somatosensory information, resembling the somatosensory-visual fusion hierarchy in the brain."

The BSV learning architecture developed by Chen and his colleagues replicates how the human brain fuses somatosensory and visual information in several ways. First, its multilayer and hierarchical structure mimics that of the brain, with [artificial neural networks](link) instead of biological ones.

In addition, some of the sectional networks within the architecture process the same modal sensory data processed by neural networks in the brain. For instance, a sectional convolutional neural network (CNN) specifically performs convolution operations, artificially replicating the function of the local receptive field within biological nervous systems and thus mimicking the initial [visual information](link) processing that takes place in parts of the [human brain](link) responsible for vision.

Finally, the architecture devised by the researchers fuses features using a newly developed sparse neural network. This network replicates how multisensory neurons in the brain represent early and energy-efficient interactions between visual and somatosensory information.
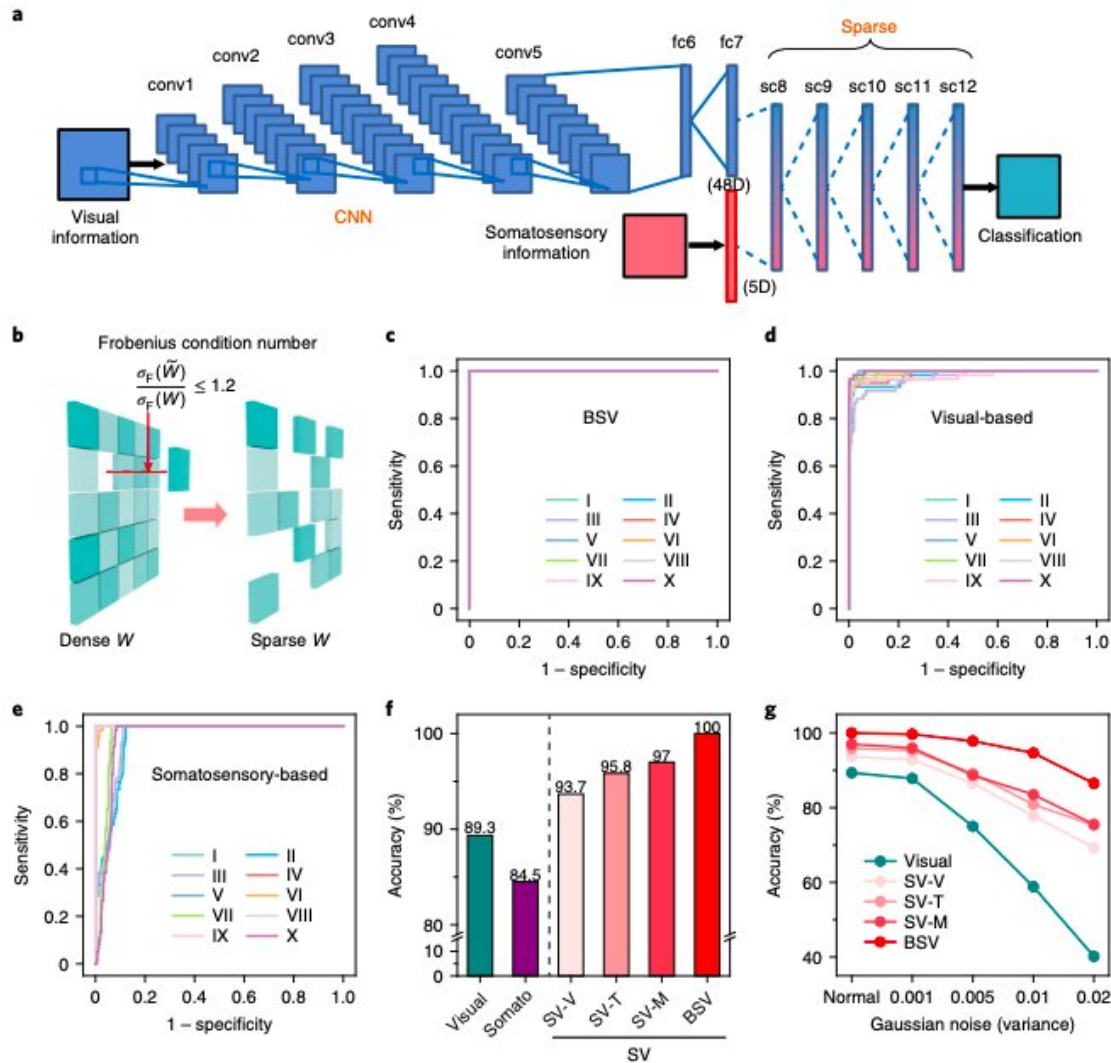
BSV associated learning for classification. Credit: Wang et al.

"The technique we developed has three unique characteristics," Chen explained. "Firstly, it can process early interactions of visual and somatosensory information. Secondly, the convolution operations carried out by the CNN resemble the function of the local receptive field in biological nervous systems, which can automatically learn hierarchical deep spatial features and extract shift-invariant features from original

images. Finally, we introduced a new pruning strategy that depends on Frobenius condition number to achieve the energy-efficient sparse neural network."

In a series of initial evaluations, the BSV learning architecture devised by Chen and his colleagues outperformed unimodal recognition approaches (i.e., approaches that only process visual or somatosensory data, rather than considering both). Remarkably, it was also able to recognize human gestures more accurately than three multimodal recognition techniques developed in the past, namely weighted-average fusion (SV-V), weighted-attention fusion (SV-T) and weighted-multiplication fusion (SV-M) architectures.

"Our bioinspired learning architecture can achieve the best recognition accuracy, compared to unimodal recognition approaches (visual-based, somatosensory-based) and common multimodal recognition approaches (SV-V, SV-T, and SV-M)," Chen said. "It also maintains a high recognition accuracy (the accuracy shows a slight decrease for BSV, which is much better than others) in non-ideal conditions, where images are noisy and under- or over-exposed."

The brain-inspired architecture developed by this team of researchers could ultimately be introduced in a number of real-world settings. For instance, it could be used to develop healthcare robots that can read a patient's body language or it could help to create more advanced virtual reality (VR), augmented reality (AR) and entertainment technology.

"Its unique biomimetic characteristics make our architecture superior to most existing approaches, which was verified by our experimental results," Chen said. "Our next step will be to build a VR and AR system based on the bioinspired fusion of the visual data and sensor data."

More information: Ming Wang et al. Gesture recognition using a

bioinspired learning architecture that integrates visual data with somatosensory data from stretchable sensors, *Nature Electronics* (2020). DOI: 10.1038/s41928-020-0422-z