**The latest research in Optical Neural Network**

**Introduction**

With the development of machine learning and Deep learning, artificial intelligence has become increasingly accessible to the public though platform like ChatGPT, copilot and so on. However, training those large language model requires significant amount of electricity, the problem of energy consumption is gaining attention from major companies. In recent years, researchers try to improve the computational efficiency to reduce the use of energy. At the same time, some researchers found that optical devices also can be used to construct neural network, which called Optical Neural Network. By replacing the original electron-based calculation with a photon-based calculation, the properties of photons, such as energy savings and the fact that photons propagate much faster than electrons, can be utilized to save energy usage and at the same time make the calculation of the model faster.

**The latest research**

**Hybrid Optical-Electronic Convolutional Neural Networks**

In 2018, Julie Chang’s group at first proposed a hybrid neural network. The team focused on reducing the workload of the electronic processor in the inference process by introducing optics into the first convolutional layer of the CNN. Assuming that light is spatially incoherent and assuming monochromatic illumination, the team proposes the 4f system shown in the figure to implement the first convolutional layer. The Fourier planes within the system allow for modulation in terms of amplitude and phase, like a bandpass filter in a digitizing system, and in this respect change the point spread function of the system. This optical 2D convolutional layer can be viewed as a convolutional layer, and the resulting flipped PSF serves as the convolutional kernel.

图示

描述已自动生成

Fig.1 The 4f System that Replaces the First Convolutional Layer of CNN

**an optical neural network based on optical dot products**

In 2022, Tianyu Wang and his groups have proposed an optical neural network based on optical dot products and proved the efficiency of ONN. They used a smartphone display as an incoherent light source to encode the input vectors and displayed the bitmap image through the OLED by running a python program written in advance on the computer side. (Fig.2) The pixel light from the OLED was resized using a zoom lens system so that the resulting pixels matched the pixel size of the prepared SLM, and the light from the OLED display was intensity modulated using a phase reflective liquid crystal spatial light modulator (SLM) coupled with a polarizing beamsplitter and a Lightwave plate, thus realizing scalar multiplication by non-negative vector elements. The intensity modulated pixel values are then mapped to 8-bit resolution transmittance using a pre-created intensity lookup table. This is then done by integrating the photon flux that effectively hits the detector within a developed time window. The reading of the results of a single dot product is realized by exploiting the defects of the MPCC.

图示

描述已自动生成

Fig.2 Schematic of Dot Product of Optical Vectors

After determining the time-point accuracy of the vector multiplication operation through the experimental setup described above, the experimental setup was utilized in the core part of the ONN. It was realized that optical computation was used in the portion of the linking layer but was still passed through an electronically grounded digital processor between layers. The team ultimately used handwritten digit classification from the MNIST dataset as a validation experiment (Fig.3), which yields different accuracy detection results at different photon budgets. When scalar multiplication detects 3.1 photons, the accuracy can reach 99%.

图示

描述已自动生成

Fig.3 Schematic Diagram of Handwritten Digit Classification by ONN

**Target Users and Applications (GPT)**

Matrix multiplication is a core operation in neural networks, and recent research has focused on accelerating this process using optical systems. Optical matrix multiplication can be achieved by modulating light through optical devices such as phase shifters and beam splitters. Researchers at various institutions are working on integrating optical matrix multiplication into existing machine learning frameworks like TensorFlow. This integration could dramatically speed up training and inference in AI models. Researchers have also demonstrated the potential for performing FFTs optically, which are essential for image and signal processing tasks in AI models.

As research in optical neural networks advances, a wide range of applications is emerging across various industries. ONNs can process images at speeds far exceeding conventional processors, making them ideal for applications in autonomous vehicles, robotics, and medical diagnostics. Real-time image recognition systems based on ONNs are now being tested for deployment in critical applications. Combining quantum computing principles with optical neural networks is a hot topic in current research. Quantum-inspired ONNs have the potential to solve highly complex problems more efficiently than classical AI approaches. The ability to process massive datasets quickly makes ONNs suitable for high-performance computing, where they can be used to accelerate scientific simulations, financial modeling, and big data analytics.

**Challenges (GPT)**

Despite the rapid advancements in optical neural networks, several challenges remain that must be addressed before they can be widely adopted. Some of the primary challenges include.

Optical systems are more susceptible to noise and signal interference, which can impact the accuracy of computations. Researchers are working on developing more robust systems that can mitigate these issues. Fabricating photonic circuits and components for ONNs is a complex and costly process. Improving the scalability of production is essential for making ONNs commercially viable. While there have been efforts to integrate ONNs with existing machine learning frameworks, more work is needed to develop software tools and platforms that support the design, training, and deployment of ONNs.

About GPT Part:

In my assessment, the text generated by GPT tends to be more fluent and coherent compared to my own writing. However, it does not always fully align with my specific requests, particularly when tasked with identifying new research. The generated content often appears to be derived from general information or news sources rather than from academic research or cutting-edge studies.