GSoC 2023 Project Proposal

Organization: ML4SCI

Vision Transformers for End-to-End Particle Reconstruction for the CMS Experiment

Mentors

Emanuele Usai
Ruchi Chudasama
Shravan Chaudhari
Sergei Gleyzer
Eric Reinhardt
Samuel Campbell

Author

Ch Pavan

Contents

Introduction and Student Information

Name: Ch Pavan Harshit
Email: chpavan2003@gmail.com, ugs205131_cse.pavan@cbit.org.in
Github ID: Github
LinkedIn: LinkedIn
Phone No: +919346542987
College: Chaitanya Bharathi Institute of Technology Hyderabad, India
Degree: Bachelor of Engineering
Branch: Computer Science Engineering
Year: 3rd year(6th semester)
CGPA: 8.5
Expected Graduation: May 2024
Time Zone: GMT +5:30 (India)
Resume: Resume

Project Description

To develop Transformer based architectures for the classification of high-energy particles.

Originally meant for NLP tasks ViTs divide an input image into small patches and treat each patch as a token, like how words are tokens in NLP. ViTs then use self-attention mechanisms to capture the relationships between different patches and learn high-level features from the image. ViTs can achieve state-of-the-art performance on image recognition tasks with less computational cost and more accuracy than conventional convolutional neural networks (CNNs).

Benefits to Community

ViTs can handle large-scale and high-resolution images while also having less computational cost. They do not need downsampling or pooling layers which lose information. ViTs can help us understand complex patterns among the images which could help in identifying features or anomalies. ViTs can be easily pretrained on large datasets and fine-tuned on specific domain tasks which could improve their generalization and robustness.

There have been many innovations in the field of ViTs and using these to create models with high accuracies. ViTs offer a novel way of analyzing data from particle detectors and experiments, challenging the traditional methods that rely on hand-crafted features and domain knowledge. ViTs have the potential to uncover new patterns and insights from complex and high-dimensional data, opening new horizons for scientific discovery.

E2E evaluation summary

Github repo

Task 1

- The task was to train a deep learning network on a dataset of electron and photon interactions with a calorimeter that measured energy and time in two channels. The desired performance metric was an AUC score of 0.80 or higher. I experimented with various network architectures and hyperparameters to optimize the model, but the best AUC score I obtained was 0.7934.

- Some of the factors that influenced the model performance were the amount of data available, the number of convolution layers and filters in the CNN model, and the presence or absence of maxpool layers. However, none of these variations could overcome the plateau of 72% accuracy that the model reached after some epochs.
- A possible explanation for this limitation is that the CNN model was not able to extract meaningful features from the data. Therefore, a different network architecture might be more suitable for this task, such as a graph convolution network (GCN).
- A GCN might have an advantage over a CNN because it can represent data as nodes and edges instead of 2D grids, which might capture more information about the electron and photon interactions.

Task 2

- The second task involved training a deep learning network on a dataset of quarks and gluons that interacted with a calorimeter and produced three-channel images.
- I used a CNN model with several convolutional layers, maxpooling and dropout layers. I have also tried using resnet architecture but it was not that effective. The toughest part of the task which I faced was to load the entire data because it was overfitting. To overcome this I have loaded the dataset in chunks.
- I achieved an ROC-AUC score of 72.01, this could be improved by different architectures such as efficientnet. Even graph neural networks might lead to better accuracy.

Task 3

- The third task was to train a vision transformer model that could achieve comparable performance to the first convolutional neural network model.
- A vision transformer (ViT) is a novel approach for image classification that uses a Transformer-like architecture over image patches. It splits an image into fixed-size patches and embeds them linearly. Then it adds position embeddings and passes the resulting sequence of vectors to a standard Transformer encoder. For classification, it adds an extra learnable "classification token" to the sequence.
- I trained a Swin transformer model and achieved an AUC-ROC score of 75.77

Observations

- When I applied the standard ViT model, I obtained poor results with only 50% accuracy. I tried to modify the model by replacing the dense layer with a convolution layer, but it did not improve the results significantly.
- Therefore, I came across other variants of ViT-based models such as Swin transformer, Detr transformer and CSwin transformer. I trained a Swin transformer model and achieved an AUC-ROC score of 0.75, which was much better than the standard ViT model.
- The Swin model uses a shifted window mechanism that enhances the efficiency and effectiveness of ViT.
- Though this was less than the CNN model, a vision transformer is more effective on pre-trained weights meaning that it can leverage existing knowledge from large-scale datasets and adapt to new domains with less data and computation.

Related Work

- My experience in Machine Learning spans several projects that demonstrate my proficiency in TensorFlow and Pytorch. I have implemented YOLO, ResNet and EfficientNet architectures for various tasks such as sleep detection and object recognition.
- I have participated in several hackathons that challenged me to apply my skills and creativity. One of the most memorable ones was a national-level AI hackathon by Dr Reddy, where our team reached the finals. In that project, I used TensorFlow to build an RNN model for an NLP model.
- I also have made a disease prediction website which employs decision trees.
- My college coursework includes NLP, machine learning and Image Processing and Computer vision which has further improved my understanding of deep learning.

These projects have also taught me how to learn new technologies quickly, communicate effectively with team members and manage time efficiently.

Biographical Information

- I am a third-year undergraduate student pursuing a Bachelor of Engineering degree in computer science.
- My academic curriculum covers various topics such as image processing and computer vision, Machine Learning, Natural language Processing as well as web development, data structures, object-oriented programming and database management systems.
- I have also gained practical experience through two internships in different domains. In my first internship at Cloudbox99, I developed a python web scraping script to collect data from potential clients and an outgoing IVR bot using twilio api, node.js and sentiment analysis.
- In my second internship at Sparkcognition, I applied an efficientNet model to train on a dataset for image classification.
- I have experience in python and c++ as well both tensorflow and pytorch frameworks.

Why ML4SCI?

Machine learning is a fascinating area that keeps evolving with new developments. I share this passion and curiosity for this field, as well as for particle physics. Transformers are amazing models that achieve remarkable results, and I would love to learn more about their inner workings through this project. This opportunity would give me valuable insights into the mechanisms and applications of vision transformers. I believe that ML has a lot of potential to enhance scientific discovery and innovation in various domains. I believe open source can facilitate innovation, learning and collaboration among researchers around the world. My view is that ML4Sci is a remarkable organization that applies machine learning to scientific problems and promotes open-source culture. I aim to support the work of others by sharing these models.

Proposed Deliverables

- Implementing vision transformers for end-to-end particle reconstruction for CMS experiment. Different architectures can be implemented based on the project requirement.

May 4 - 28

- Discussing the project with mentors, read documentation, and get up to speed to begin working on their projects.
- Look into relevant information useful for the project and brush up on the relevant physics-based concepts. I will learn more into transformer architectures and their implementations.

May 29 - June 12

- Improving upon the Swin transformer which used in task 3 uses the shifting window mechanism.

June 12 - June 26

- Implementing the PVT transformer which is a type of vision transformer that utilizes a pyramid structure to make it an effective backbone for dense prediction tasks.
- Specifically it allows for more fine-grained inputs (4 x 4 pixels per patch) to be used, while simultaneously shrinking the sequence length of the Transformer as it deepens - reducing the computational cost.

June 26 - July 10

- Applying Transformer variants such as DeIT and perform adequate testing to improve the ROC-AUC score. DeiT is a data-efficient image transformer that is used for image classification tasks. It has several advantages over other image transformers.

July 14 - July 28

- Implementing Cswin transformer, it is a novel variant of the vision transformer that uses cross-shaped windows to capture local and global dependencies in an image.
- I believe that this model can achieve a higher AUC-ROC score than other models because it can better exploit the spatial structure of the calorimeter images. Since most of the energy deposition occurs in the center of the image, the cross-shaped windows can focus on this region and ignore irrelevant background information. This way, the model can learn more discriminative features and improve its performance.

July 28 - August 21

- Testing and bug removal.
- Thorough documentation.

Post GSoC

After the proposed timeline, I would love to start implementing any additional features and contribute to ML4SCI even after GSoC and given an opportunity, would love to participate in scientific research with the mentors.

Other Information

I will be taking my end semester exams in June and this may affect my availability for some days. I appreciate your understanding and support during this period. I assure you that I will work diligently and efficiently to meet the deadlines and deliver high-quality work. I will utilize my skills and abilities to the best of my potential to fulfill the assigned tasks.