Vision Transformer for End-to-End particle reconstruction for the CMS experiment

Personal Details:

Name: Siddhesh Sharad Kashid

University: Indian Institute of Technology (BHU) Varanasi, India

Email: siddhesh.sharadk.cd.phy20@iitbhu.ac.in

Mobile No: (+91)9082384347

Country of Residence: India

Timezone: IST (GMT + 05:30)

Primary Language: English

I am a third year dual degree student pursuing B.Tech. + M.Tech. in Engineering Physics at Indian Institute of Technology (BHU) Varanasi. My semester will complete in late April leaving me enough time to get ready for my GSoC project. If I am selected, I shall be able to work around 40 hrs a week on the project, though am open to putting in more effort if the work requires.

Project Abstract:

The field of Computer Vision has for years been dominated by Convolutional Neural Networks (CNNs) which use filters and create features used by a multi-layer perceptron to perform the desired classification. But recently this field has been incredibly revolutionized by the architecture of Vision Transformers (ViT), which through the mechanism of self-attention has proven to obtain excellent results on many tasks. This project aims to use a Vision Transformer for end-to-end particle reconstruction for the CMS experiment.

Technical Details:

Below I’ve explained significant steps of the algorithm:

- ❖ Firstly, define a patch extractor function which inputs our 2 channel input (of shape (32,32,2)) and create patches of size 2x2x2.
- ❖ Created a patch encoder function, which encodes the sequence of the patch created by the patch extractor function. It outputs a linear sequence of embedded patches which further acts as an input to the transformer.
- ❖ The transformer encoder includes the following:
    - ➢ Multi-Head Self Attention Layer (MSA): This layer concatenates all the attention outputs linearly to the right dimensions. The many attention heads help train local and global dependencies in an image.

- ➢ Multi-Layer Perceptrons (MLP) Layer: This layer contains a two-layer Dense with Gaussian Error Linear Unit (GELU).
- ➢ Layer Norm (LN): This is added prior to each block as it does not include any new dependencies between the training images. This thereby helps improve the training time and overall performance.

❖ Trained the Vision Transformer model on GPU(Nvidia RTX 3060) for high speed training on a subset of the actual data (1000 images).

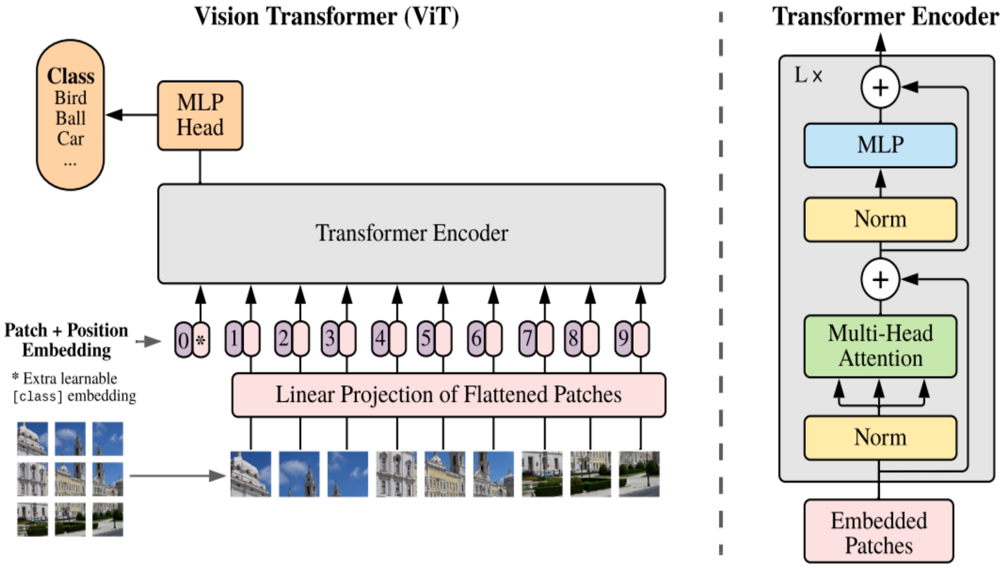

Image source: https://arxiv.org/pdf/2010.11929.pdf

Timeline:

- **Pre GSOC:** Implement the Vision Transformer model for high energy particle classification. I am even currently working on the algorithm as my personal project (Human Activity recognition using Vision Transformer) and this will have to be completed before April end, which means before the start of the GSOC project. This also easies my task in those 12 weeks and helps me concentrate on shipping professional code.

- **Community Bonding:**

  Getting acquainted with the mentors of Ml4Sci and the procedure that needs to be followed to submit code and get it reviewed. Discussing with the team on what exactly needs to be the problem statement(minute details, like the kind of features to be used from the dataset we have, like modification in the original Vision Transformer architecture, what should be algorithmically decided based on input data features, etc).

- **Week 1-2:**

  Understand the relevant parts of the CMS experiment and the data collected from it, and try to figure out what features can be extracted from the data and construct the model accordingly. For example the data has 2 channels (hit energy and time), we can extract different features such as spatial features from the hit energy channel and temporal features from the time channel. Code the input pipelines for the large input data.

- **Week 3:**

  Explore the existing state-of-the-art model for high energy particle classification. Begin building the Vision Transformer.

- **Week 4-6:**

  Complete the vanilla Vision Transformer Model, train and test it on the features extracted/data. Code the transformer using tensorflow framework and research the scope of developing custom loss functions and custom layers to be included in the transformer. Fine tune the model using existing optimization frameworks such as Keras-tuner, HyperOpt, etc for optimal performance on specific use cases.

- **Week 7-8:**

  Code the transformer for all different kinds of features and make use of ensemble learning for better aggregated probabilities of each class. My personal choice would be to construct a different transformer model for each individual channel and then try out different aggregating techniques on the features extracted from each of the transformers to give probability of the individual class.

- **Week 9–10:**

Take feedback from the mentors and iterate on the designs and improvise on use cases. Ensure code quality by adding more test cases and working on creating output pipelines for deploying the vision transformer to use.

- **Week 11:**

  Spare week in case of some work getting delayed, in case of any emergency or otherwise.

Implementation:

The proposed model will be implemented using Tensorflow and trained on the HPC clusters provided by Ml4Sci or on the local machine with CUDA support. The trained model will be evaluated on real CMS data and compared against existing methods.

Personal Inspiration for the project:

The Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) generates massive amounts of particle collision data. The ability to accurately reconstruct these particles is essential for understanding the physics behind these collisions. In recent years, the Transformer architecture has shown remarkable success in the natural language processing field. As a Physics student at Indian Institute of Technology Varanasi, with a keen interest in Machine Learning, I Am very excited to work on the idea of implementing the latest developments in the field of

Deep learning i.e the Vision Transformers for the CMS experiment. I've been exploring various fields of Physics trying to find out the scope of application of machine learning techniques in it. This project can be a great opportunity to work in the domain of high energy particle physics and explore the scope of usage of latest deep learning methodologies in it.

Conclusion:

This project proposes the use of a Vision Transformer for end-to-end particle reconstruction for the CMS experiment. The proposed approach has the potential to significantly improve the accuracy and efficiency of particle reconstruction in high-energy physics. By leveraging the power of self-attention mechanisms, the Vision Transformer can handle the complex spatial and temporal relationships between particles, leading to more accurate reconstruction.

Results:

The CNN model achieves an ROC-AUC score of 0.81 on the training dataset when trained on 270000 samples. Whereas the Vision Transformer model achieves a score of 0.947 on training data when trained on just 1000 samples. The solutions to the tasks has been submitted via given email id.

References:

1. https://arxiv.org/pdf/2010.11929.pdf

2. https://www.tensorflow.org/api_docs/python/tf/image/extract_patches