

Deep Learning CS577

IMAGE CLASSIFICATION WITH VISION TRANSFORMER

Mohammad Ausaf Ali Haqqani
(A20516413)

Bhoomika Panduranga
(A20503493)

Contents:

- Abstract
- Problem Statement
- Proposed Solution
- Implementation
- Results
- Conclusion
- Future work
- References



Abstract

TRANSFORMERS

NLP

COMPUTER VISION

CNN

Problem Statement

CNN Classifier << Pure Transformer

CNN Classifier << Pure Transformer

Proposed Solution

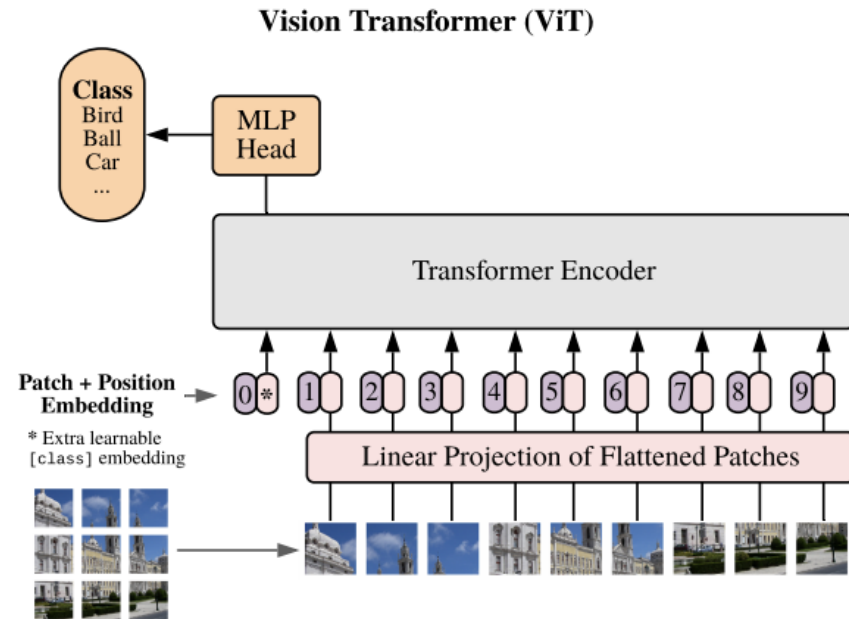
ViT

CIFAR100

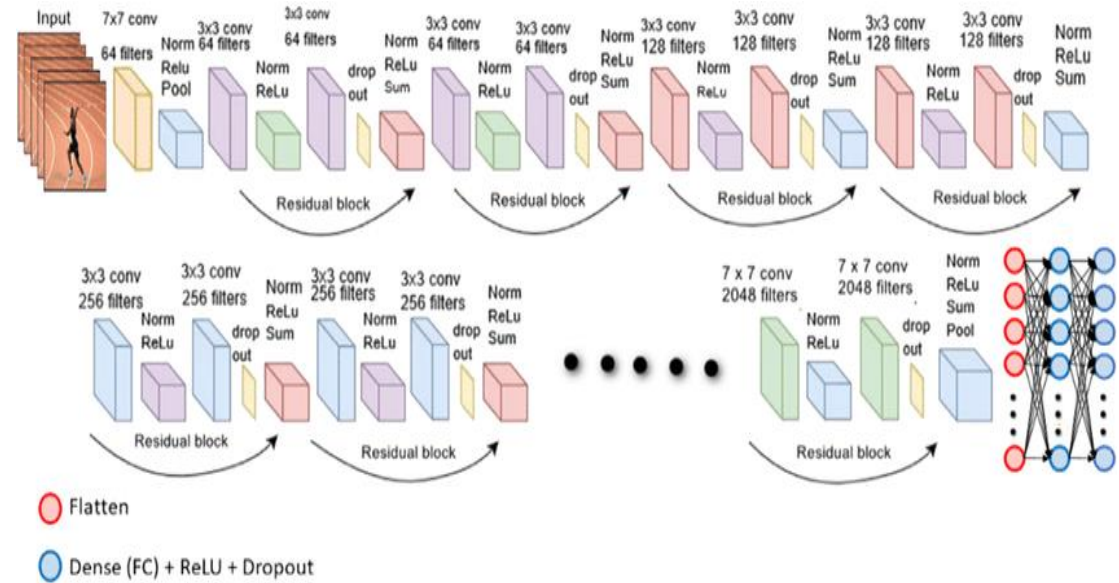
Tokens -> Patches

Supervised Learning

IMPLEMENTATION

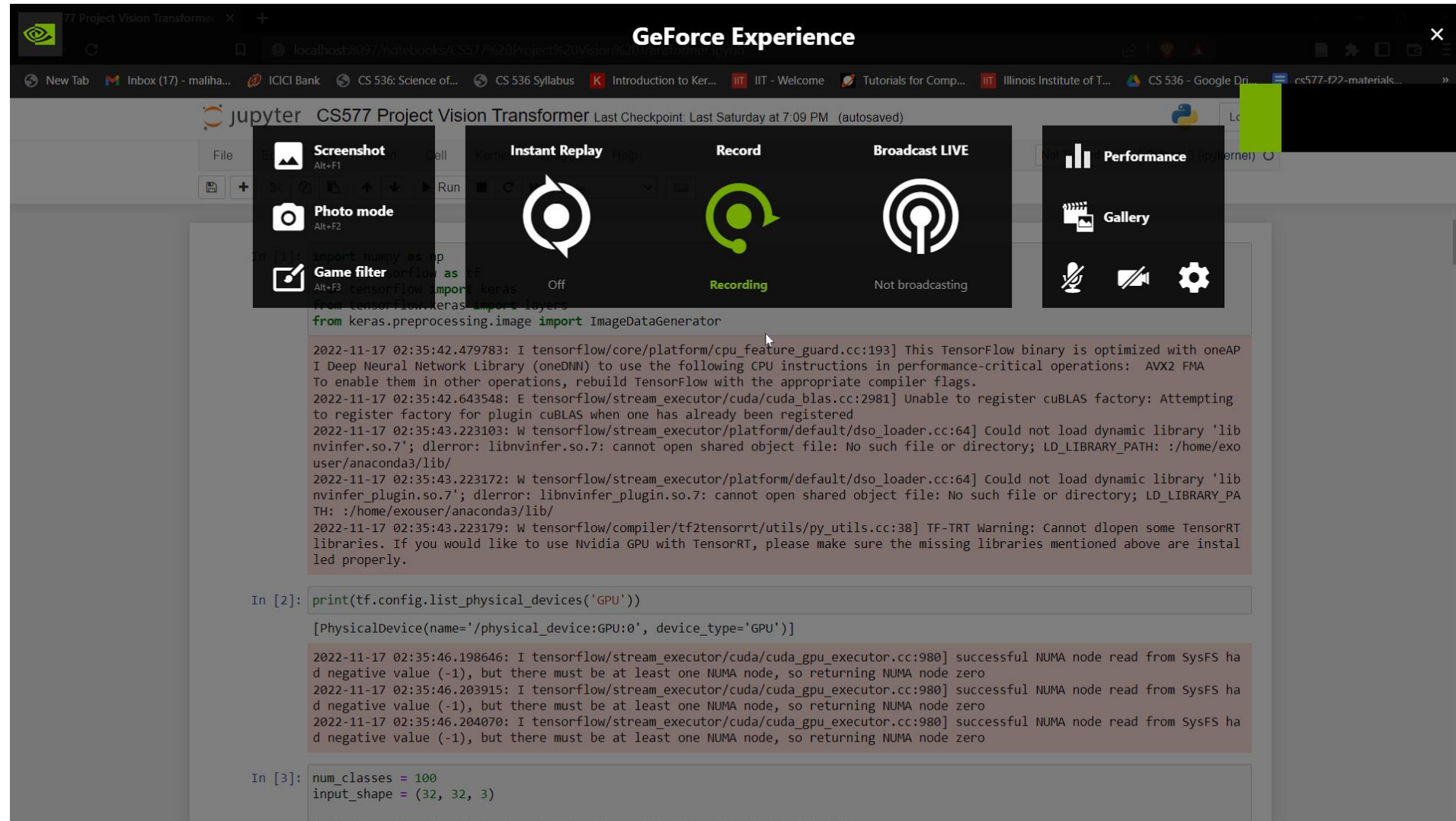


Transformer Architecture



ResNet101V2 Architecture

IMPLEMENTATION



The screenshot shows a Jupyter Notebook titled "CS577 Project Vision Transformer" running on a local host. The notebook is displaying code for importing TensorFlow and Keras, and a terminal window showing TensorFlow logs. The GeForce Experience interface is overlaid on the notebook, showing various controls like Screenshot, Photo mode, Game filter, Instant Replay, Record, Broadcast LIVE, Performance, and Gallery.

GeForce Experience Interface:

- Screenshot:** Alt+F1
- Photo mode:** Alt+F2
- Game filter:** Alt+F3
- Instant Replay:** Off
- Record:** Recording
- Broadcast LIVE:** Not broadcasting
- Performance:** Performance
- Gallery:** Gallery

Jupyter Notebook Code:

```
In [1]: import numpy as np
import tensorflow as tf
from tensorflow.keras import layers
from keras.preprocessing.image import ImageDataGenerator
```

Terminal Output:

```
2022-11-17 02:35:42.479783: I tensorflow/core/platform/cpu_feature_guard.cc:193] This TensorFlow binary is optimized with oneAPI Deep Neural Network Library (oneDNN) to use the following CPU instructions in performance-critical operations: AVX2 FMA
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
2022-11-17 02:35:42.643548: E tensorflow/stream_executor/cuda/cuda_blas.cc:2981] Unable to register cuBLAS factory: Attempting to register factory for plugin cuBLAS when one has already been registered
2022-11-17 02:35:43.223103: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libnvinfer.so.7'; dlderror: libnvinfer.so.7: cannot open shared object file: No such file or directory; LD_LIBRARY_PATH: /home/exouser/anaconda3/lib/
2022-11-17 02:35:43.223172: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libnvinfer_plugin.so.7'; dlderror: libnvinfer_plugin.so.7: cannot open shared object file: No such file or directory; LD_LIBRARY_PATH: /home/exouser/anaconda3/lib/
2022-11-17 02:35:43.223179: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Cannot dlopen some TensorRT libraries. If you would like to use Nvidia GPU with TensorRT, please make sure the missing libraries mentioned above are installed properly.
```

Jupyter Notebook Code:

```
In [2]: print(tf.config.list_physical_devices('GPU'))
```

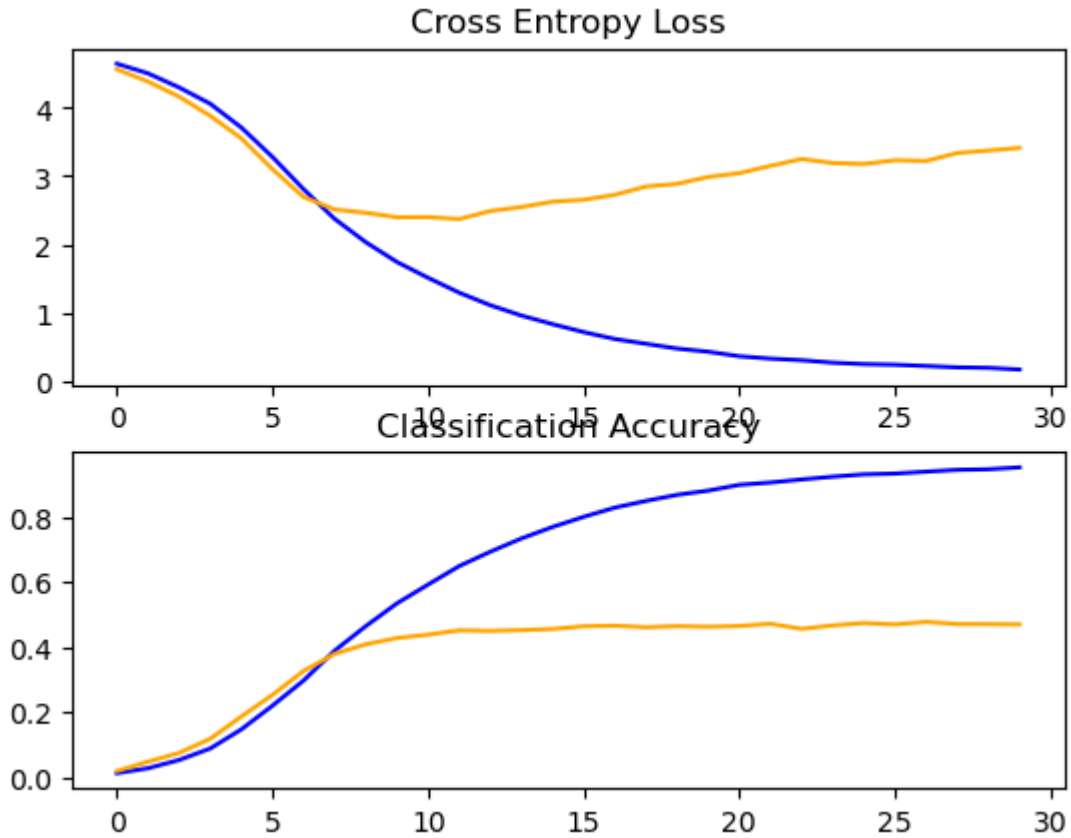
Terminal Output:

```
[PhysicalDevice(name='/physical_device:GPU:0', device_type='GPU')]
2022-11-17 02:35:46.198646: I tensorflow/stream_executor/cuda/cuda_gpu_executor.cc:980] successful NUMA node read from SysFS had negative value (-1), but there must be at least one NUMA node, so returning NUMA node zero
2022-11-17 02:35:46.203915: I tensorflow/stream_executor/cuda/cuda_gpu_executor.cc:980] successful NUMA node read from SysFS had negative value (-1), but there must be at least one NUMA node, so returning NUMA node zero
2022-11-17 02:35:46.204070: I tensorflow/stream_executor/cuda/cuda_gpu_executor.cc:980] successful NUMA node read from SysFS had negative value (-1), but there must be at least one NUMA node, so returning NUMA node zero
```

Jupyter Notebook Code:

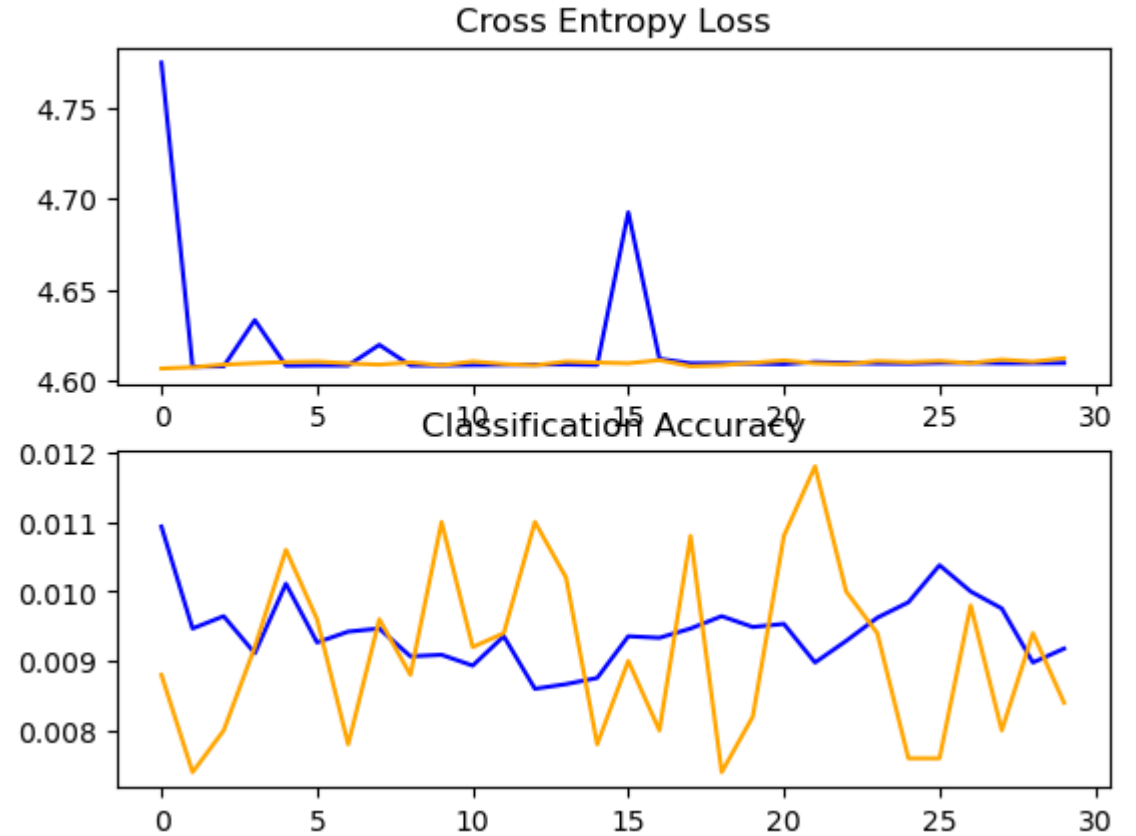
```
In [3]: num_classes = 100
input_shape = (32, 32, 3)
```

Results



Resnet101v2- 30 epochs

313/313 [=====] - 3s
10ms/step - loss: 3.3324 - accuracy: 0.4793 the
accuracy is: 47.929999232292175



Transformer – 30 epochs

313/313 [=====] - 5s
17ms/step - loss: 4.6079 - accuracy: 0.0100 -
top-5-accuracy: 0.0500 Test accuracy: 1.0% Test
top 5 accuracy: 5.0%

Conclusion

Large Dataset

Scalable

Economical

TPU

Future work



Pretrain on large datasets



ViT – Segmentation and detection



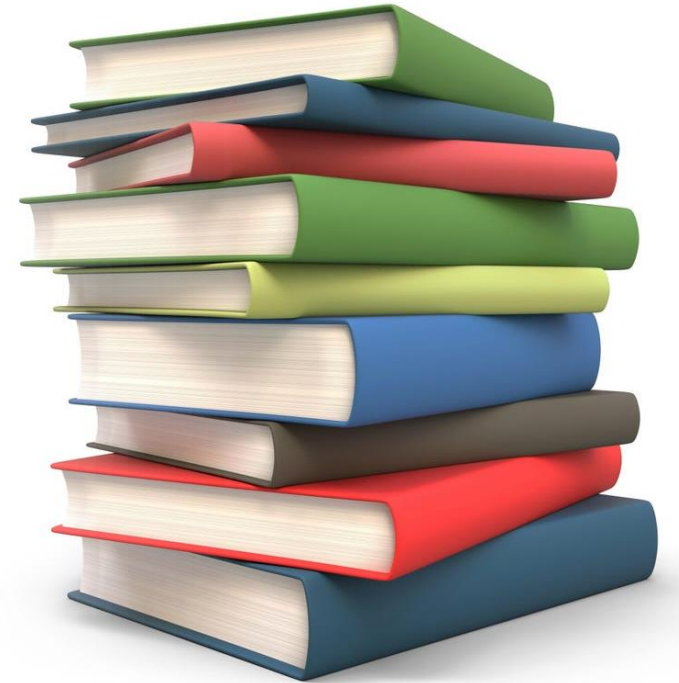
Explore self supervised pre- training methods



Further scaling

References

- [1]. https://keras.io/examples/vision/image_classification_with_vision_transformer/
- [2]. <https://arxiv.org/abs/2010.11929>
- [3]. Bello, B. Zoph, Q. Le, A. Vaswani, and J. Shlens. Attention augmented convolutional networks. In ICCV, 2019.
- [4]. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In NAACL, 2019.



Thank you