Type 1 : Code Reproduction

QUANTUM VISION TRANSFORMERS

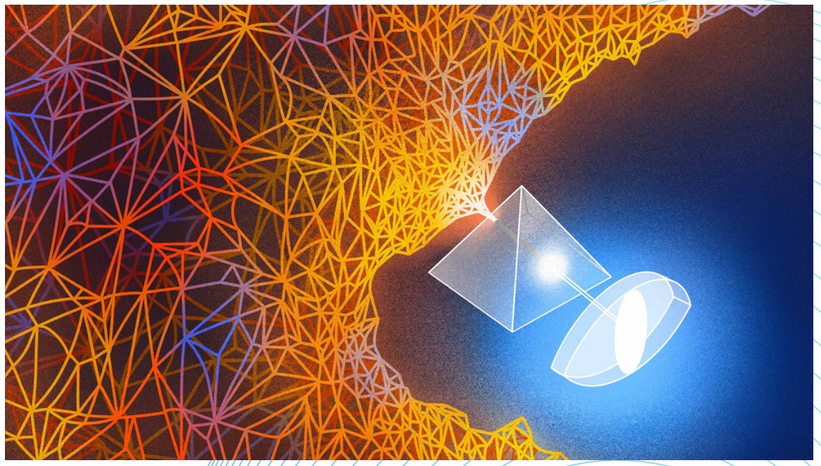
El Amine Cherrat, Iordanis Kerenidis, Natansh Mathur, Jonas Landman, Martin Strahm, and Yun Yvonna Li

AI Reverse Engineering Course

Arfeto Brian Estadimas (Student)
Professor Chaoning Zhang (Professor)

WHERE PUBLISHED

Published in Quantum Physics (quant-ph) (2022)



Is this a good paper :

Top paper with JCR Ranking <5%

Top conferenced paper in Quantum ML

Why select this paper :

- Transformer is the latest DL technology
- This is my research area (Quantum ML)
- Opportunity to maximize vision transformer with quantum technology

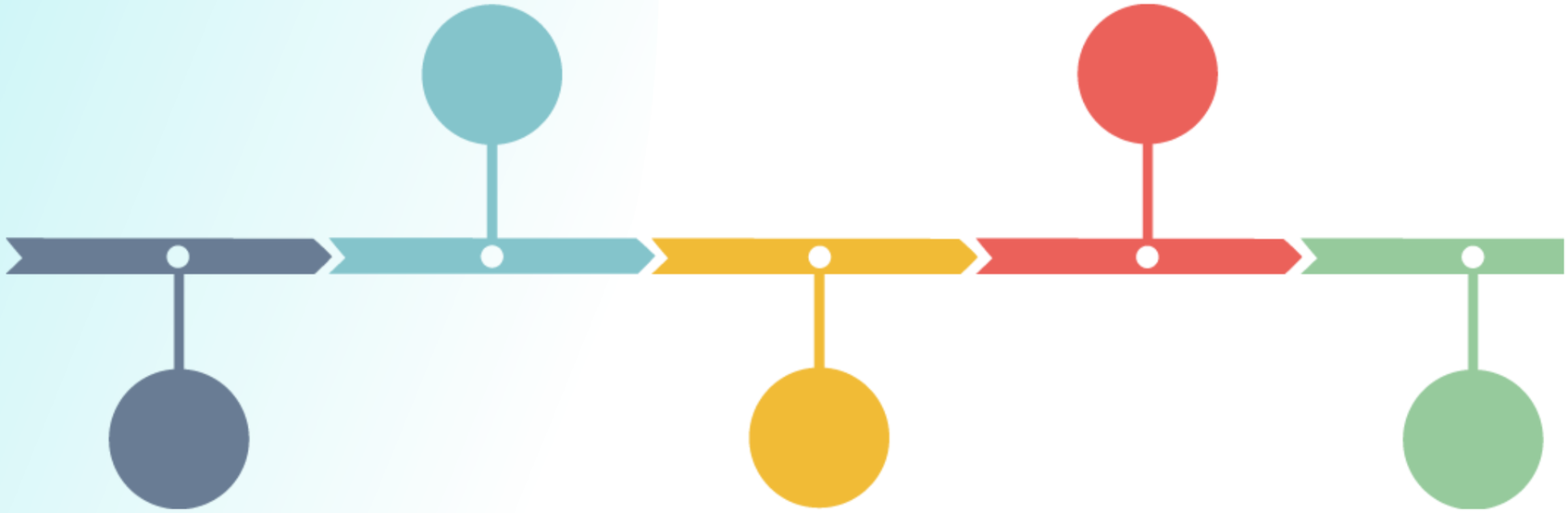
Vision Transformer

Bye-bye Convolutional

Reproduced Code for Experiment :

github.com/brianestadimas/quantum-VT

PRESENTATION CONTENTS



Explain Transformers
And Usage

Vision Transformers

Vision Transformers
Experiment
Result

Explain Quantum
Concept

Quantum Vision Transformers

TRANSFORMER IN NLP & AI

Some big companies currently are working competitively to build Deep Learning Transformer in their app...

OpenAI-GPT 3 (ChatGPT)

Well developed with Transformer

Google Translate

Machine translation system based on transformer architecture.

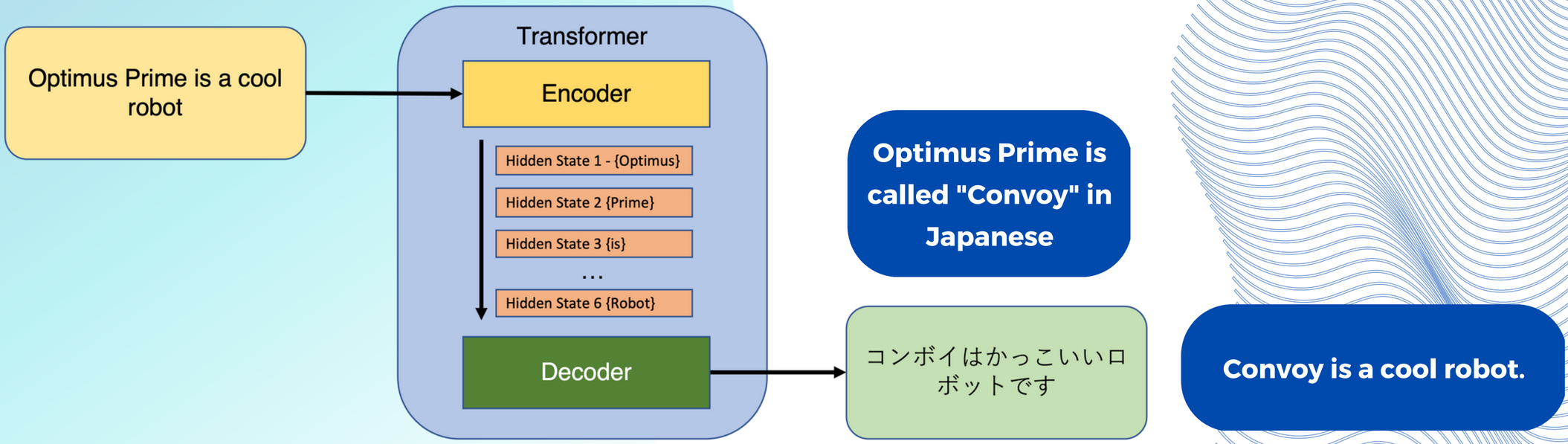
Youtube

Recommendation System

facebook research

NETFLIX

TRANSFORMER
CONCEPT



Simply translate from En to Jp will produce disinformation, because in Japanese the verbs are placed at the end of the utterance, whereas this never happens in English.

In the transformer model, each word in the English sentence is first converted into a vector representation. (Encode)

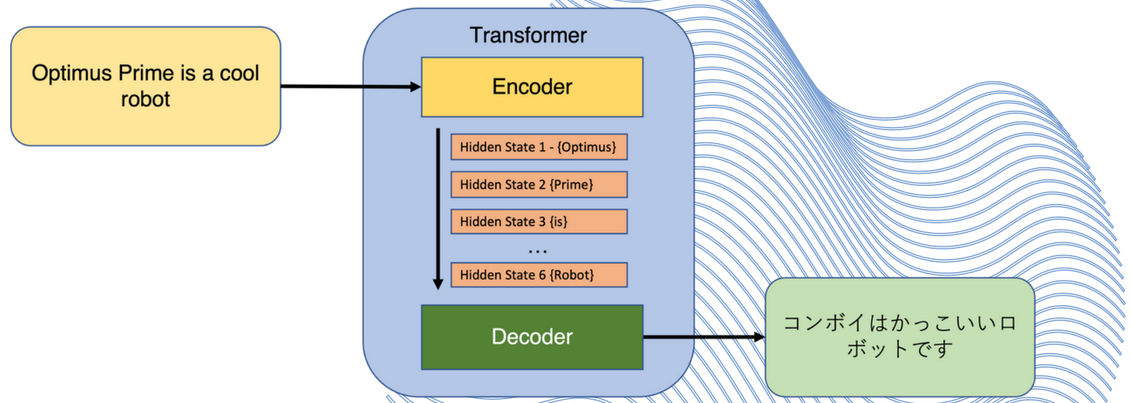
TRANSFORMER CONCEPT

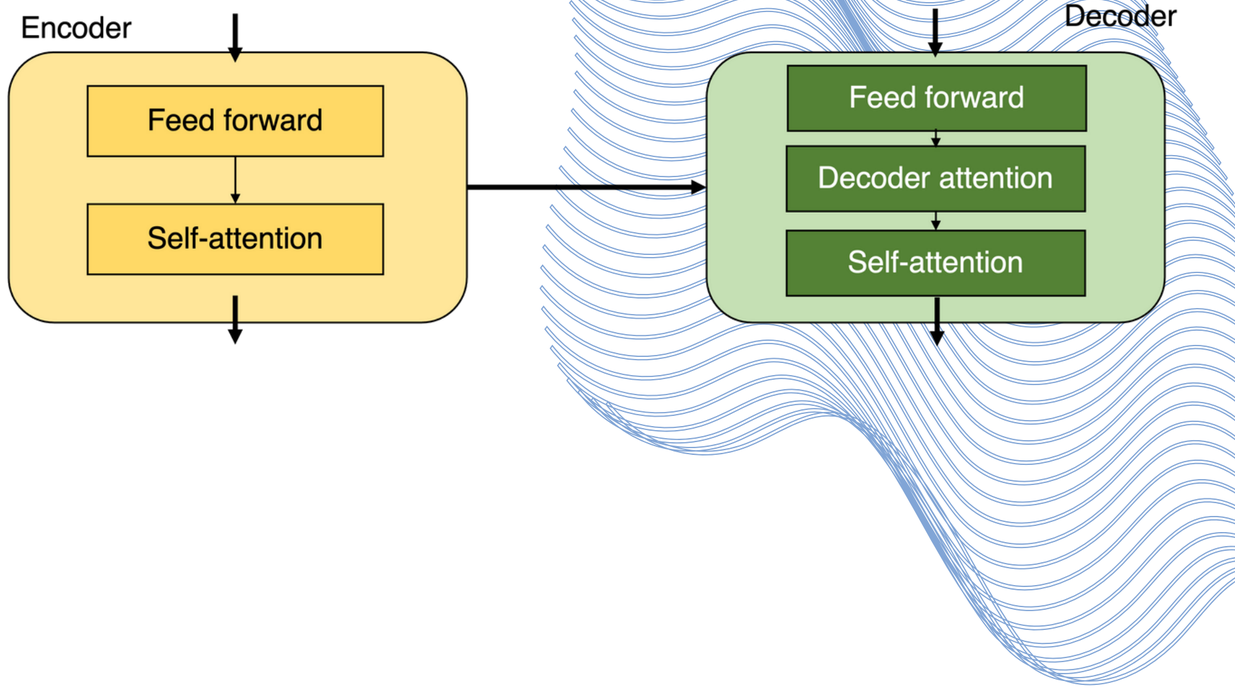
Self attention mechanism :

- Transformer attends to all the words in the input sequence simultaneously to compute a weighted representation of each word
- Weight based on relationship between sentences (That splited)
- Transformer to understand the context and meaning of each word in the sentence.

Output :

- Finally, the transformer produces a probability distribution over the Japanese vocabulary, allowing it to select the most likely translation of the English sentence.

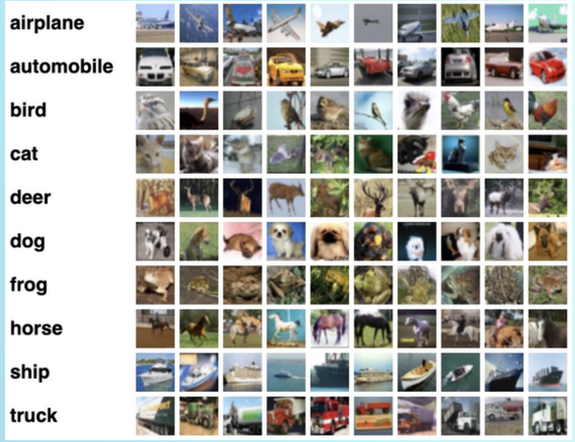




github.com/brianestadimas/quantum-VT

VISION TRANSFORMERS
USING TENSORFLOW-KERAS

With previous Transformers concept, we will be using it on image data. The Image will be encoded into smaller patches. Then it will be flattened and feed into Transformer Encoder.



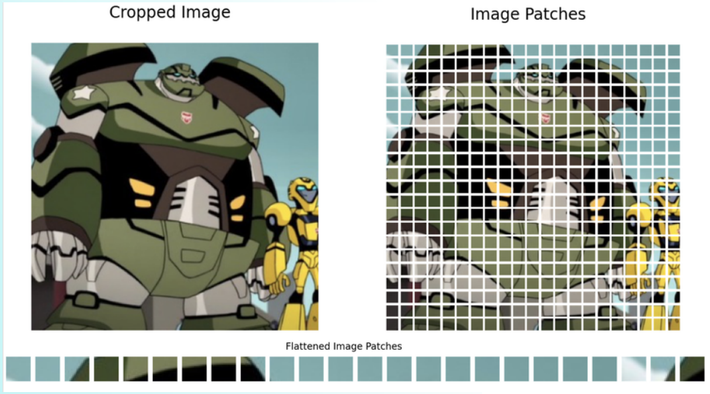

```
def data_augmentation(self):
    data_augmentation = keras.Sequential(
        [
            layers.Normalization(),
            layers.Resizing(self.image_size
            layers.RandomFlip("horizontal")
            layers.RandomRotation(factor=0.
            layers.RandomZoom(height_factor
        ],
        name="data_augmentation",
    )
    # Compute the mean and the variance of
    data_augmentation.layers[0].adapt(x_tra
```

1. Data Preprocessing

- We are using CIFAR-100 data in experiment, in original paper author's use MNIST dataset, which will be more challenging.
- Augment images. Image will be flipped, normalized, rotated and zoomed (arbitary)

VISION TRANSFORMERS
USING TENSORFLOW-KERAS

github.com/brianestadimas/quantum-VT




```python
## Implement the patch encoding layer
class PatchEncoder(layers.Layer):
    def __init__(self, num_patches, projection_
        super().__init__()
        self.num_patches = num_patches
        self.projection = layers.Dense(units=pr
        self.position_embedding = layers.Embedd
            input_dim=num_patches, output_dim=p
        )

    def call(self, patch):
        positions = tf.range(start=0, limit=sel
        encoded = self.projection(patch) + self
        return encoded
```

2. Patch and Encode the Images

- Patches is function that takes an input image and outputs a set of smaller image patches.
- Each images is extracted into 8x8 pixels, multiple images at the same time (batching)
- Then we will encode into function PatchEncoder which output is vectors.

CIFAR 100
32x32 px Images

→ **16 Patches** →

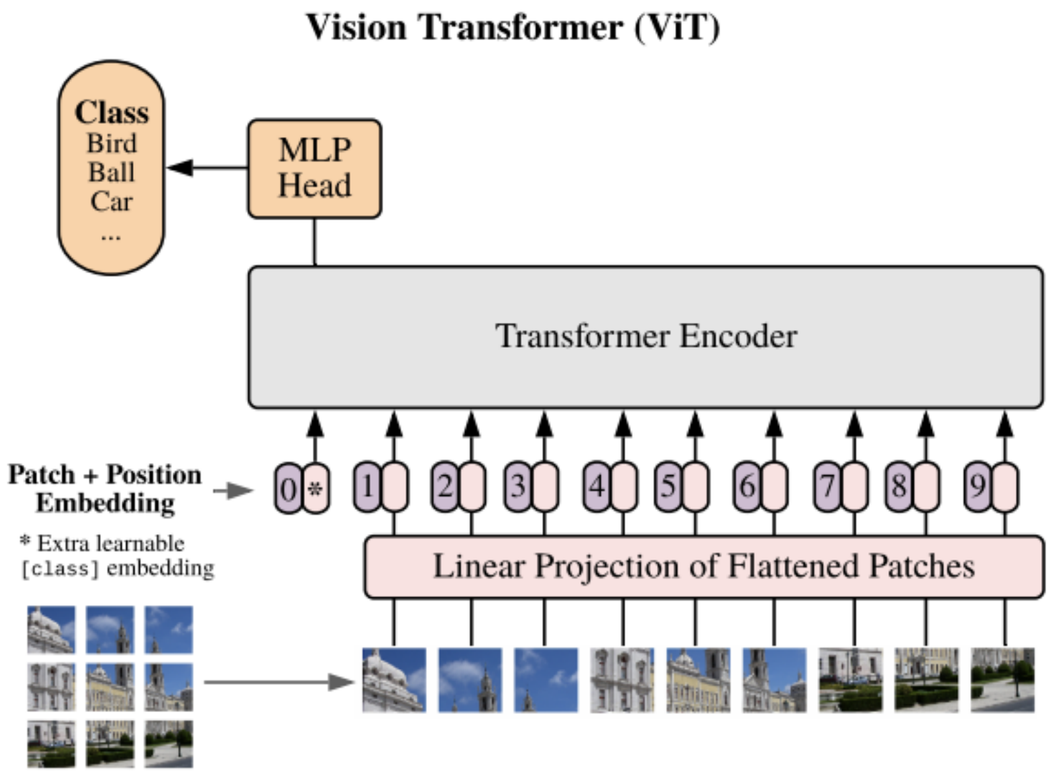
Patches
8x8

→

Encoded into Array
(Vector) - 64 Elements long

VISION TRANSFORMERS
USING TENSORFLOW-KERAS

github.com/brianestadimas/quantum-VT



3. Multi Head Attention

Multi-head attention layer takes the encoded patches image as both the query and key inputs, and computes an attention matrix that weights the values of each image patch based on how relevant they are to another images patches.

```
    x1 = layers.LayerNormalization(epsilon=1e-6)(encoded_patches)
    # Create a multi-head attention layer.
    attention_output = layers.MultiHeadAttention(
        num_heads=self.num_heads, key_dim=self.projection_dim, dropout=0.1
    )(x1, x1)
```

4. Construct MLP Layers

- Uses the GELU (Gaussian Error Linear Unit) activation function in between.
- FC layers are designed to apply linear transformations to the input data, and the GELU activation function applies a non-linear transformation to the output of the first FC layer.

VISION TRANSFORMERS
EXPERIMENT RESULTS

github.com/brianestadimas/quantum-VT

Model Name	Dataset	Avg Accuracy	Source
VIT	CIFAR-100	81.82 %	Experiment
ResNet-110	CIFAR-100	77.27 % (Highest)	Wide Residual Networks Sergey Zagoruyko and Nikos Komodakis
ResNet-50	CIFAR-100	75.92 %	Deep Residual Learning for Image Recognition. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun

```
================] - ETA: 0s - loss: 0.6031 - accuracy: 0.176/176 [==============
==============] - 343s 2s/step - loss: 0.6031 - accuracy: 0.8182 - top-5-accuracy
: 0.9750 - val_loss: 1.9577 - val_accuracy: 0.5586 - val_top-5-accuracy: 0.8230
Epoch 200/200
```

- **Observation 1**

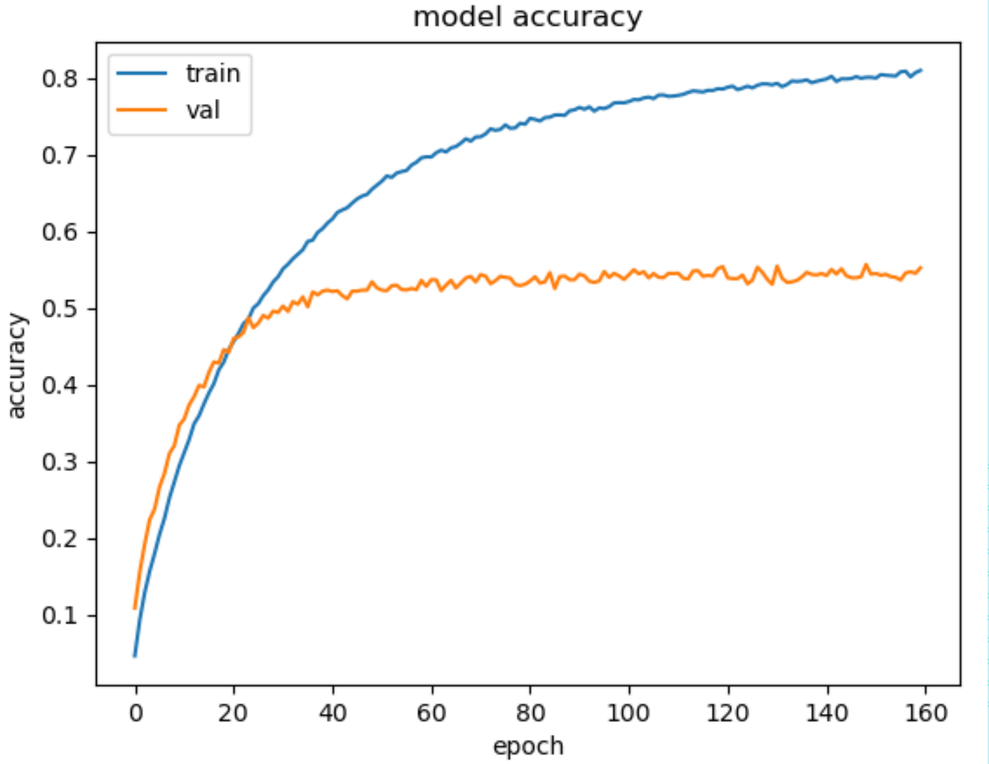
With only 200 epochs, experiment accuracy can break the ResNet highest record (CIFAR-100). While as ResNet is highly well developed model.

- **Observation 2**

ViT has a more modular architecture than ResNet, easier to modify the code for different tasks

github.com/brianestadimas/quantum-VT

VISION TRANSFORMERS
EXPERIMENT RESULTS



- **Observation 3**

In performance, ResNet speed and RAMs are better than ViT for smaller datasets, and vice versa.

MNIST -> ResNet performs better overall

CIFAR-100 -> ViT performs better

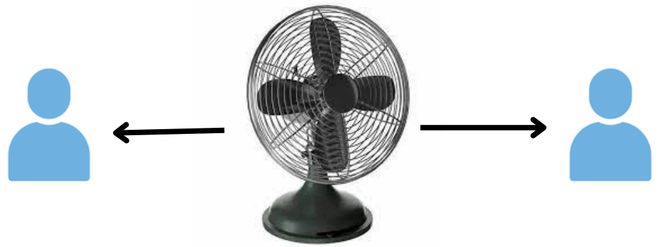
- **Limitation**

This experiment is time limited. So by changing hyperparameter and more experiments, can get alot better results.

```
================] - ETA: 0s - loss: 0.6031 - accuracy: 0.176/176 [==============
==============] - 343s 2s/step - loss: 0.6031 - accuracy: 0.8182 - top-5-accuracy
: 0.9750 - val_loss: 1.9577 - val_accuracy: 0.5586 - val_top-5-accuracy: 0.8230
Epoch 200/200
```

Process Name	User	% CPU	ID	Memory	Disk read tota
python	briancqi	80.20	179111	6.5 GB	738.8 MB

QUANTUM COMPUTING CONCEPT



FAN ANALOGY

Classic :

- It is not possible to use a fan for 2 people in different position, either we change person position, or simultaneously change fan direction
- Not possible to have fan power 0.75 (between medium and high) directly

Quantum :

- It is possible to have 2 different fan directions at the same time!
- It is possible to set fan power between medium and high (setting amplitudes)

WHY QUANTUM?

IBM

QUANTUM COMPUTER

Has been proven better than classical by IBM and Google. This will be used in the Future IBM development.

QUANTUM MECHANICS

Potentially lead to exponentially faster than classical! This is still in early stage of development

MILESTONE

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's when an unknown printer took a galley of type and scrambled it to make a type specimen book.

2020

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate vel ipsum ac fringilla. Nunc cursus, arcu nec pretium aliquet.

2021

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate vel ipsum ac fringilla. Nunc cursus, arcu nec pretium aliquet.

2022

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate vel ipsum ac fringilla. Nunc cursus, arcu nec pretium aliquet.

STAY INSPIRED. NEVER STOP CREATING.

hello@reallygreatsite.com

THE PROCESS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate vel ipsum ac fringilla. Nunc cursus, arcu nec pretium aliquet.

1

Process 01

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate vel ipsum ac fringilla.

2

Process 02

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate vel ipsum ac fringilla.

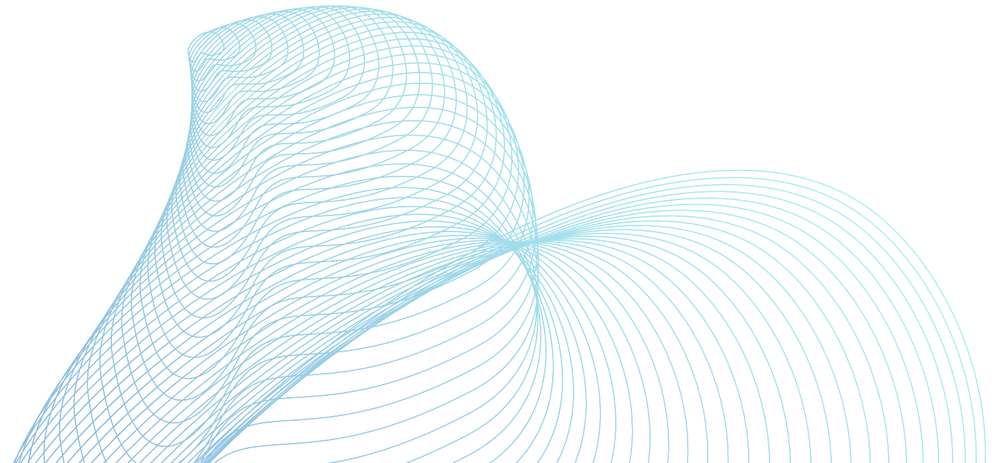
3

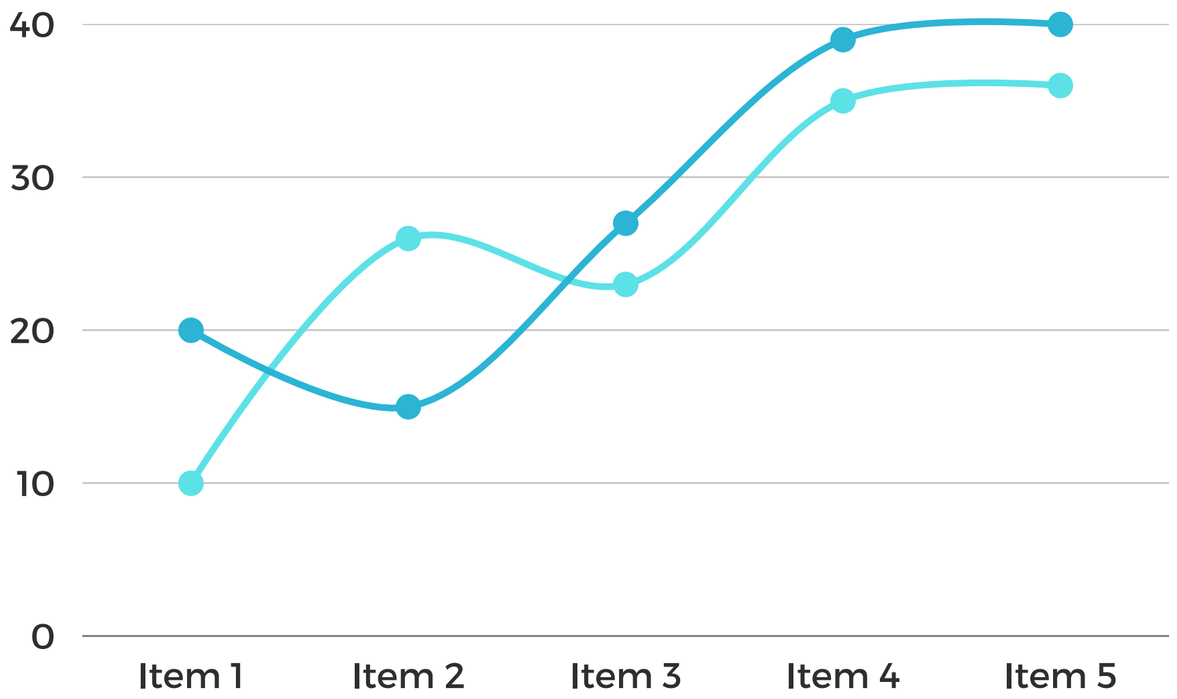
Process 03

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate vel ipsum ac fringilla.

MARKET ANALYSIS

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's when an unknown printer took a galley of type and scrambled it to make a type specimen book.





OUR GOALS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate vel ipsum ac fringilla.

Goals 01

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate vel ipsum ac fringilla.

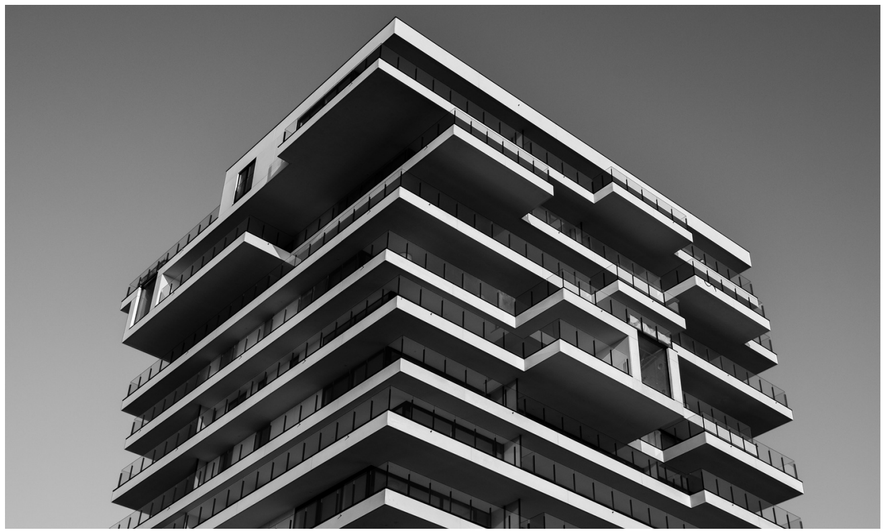
Goals 02

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate vel ipsum ac fringilla.

OUR PROJECT

Client : Borcelle Company

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate vel ipsum ac fringilla. Nunc cursus, arcu nec pretium aliquet, dui metus fermentum sapien, id cursus libero orci ac augue.



LET'S CONNECT WITH US!

+123-456-7890

hello@reallygreatsite.com

www.reallygreatsite.com