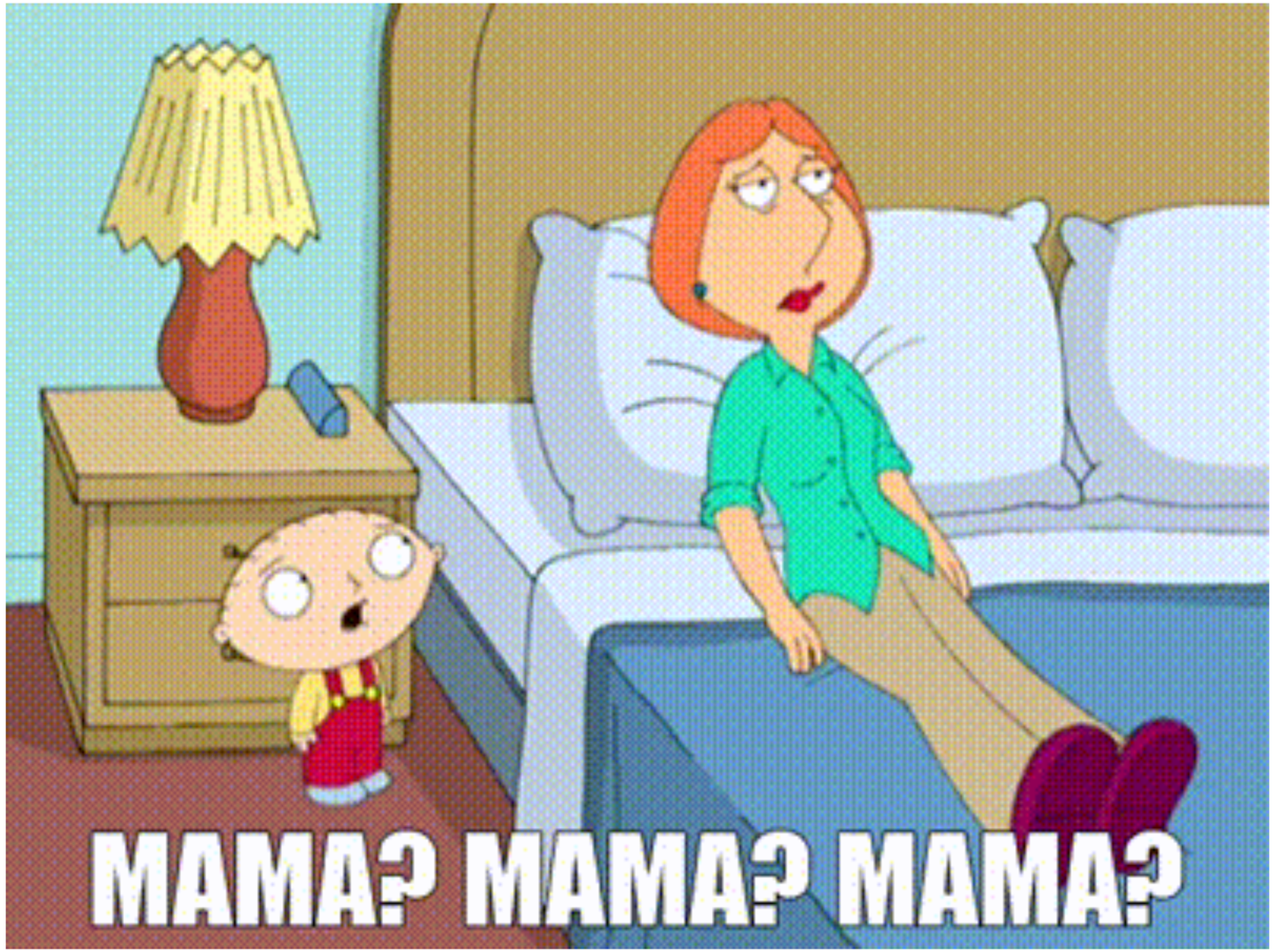
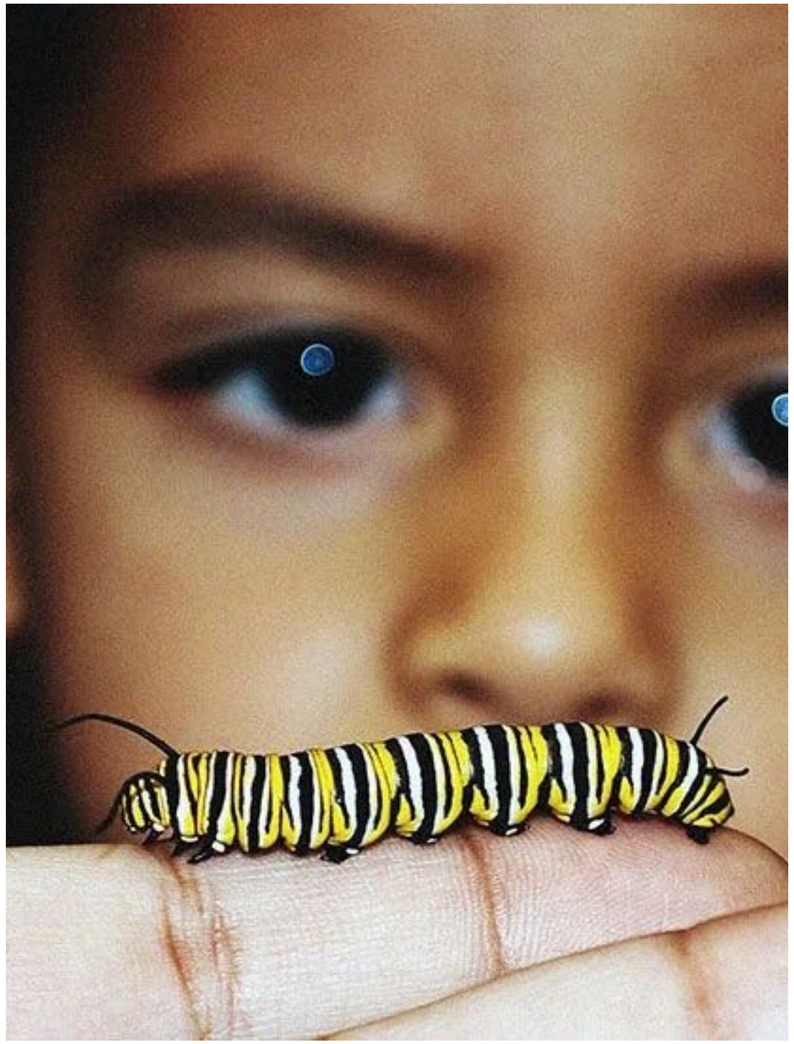
THE CATERPILLAR EFFECT

SHARIFAH, DSI-41-SG, 28 FEB 2024

https://stock.adobe.com/sg/images/butterfly-metamorphosis-development-stages-caterpillar-larva-pupa-adult-insect-hand-drawn-watercolor-illustration-isolated-on-white-background/360928306



https://memes.yarn.co/yarn-clip/f82142e6-4e4a-4ba6-8a40-778fd4d158fd/gif
https://encrypted-tbn2.gstatic.com/images?q=tbn:ANd9GcR5KERoP9qlFhDM67beYvj4uKboOqpAaR8FhWf2lh2CNpoGtxWV



PROBLEM STATEMENT

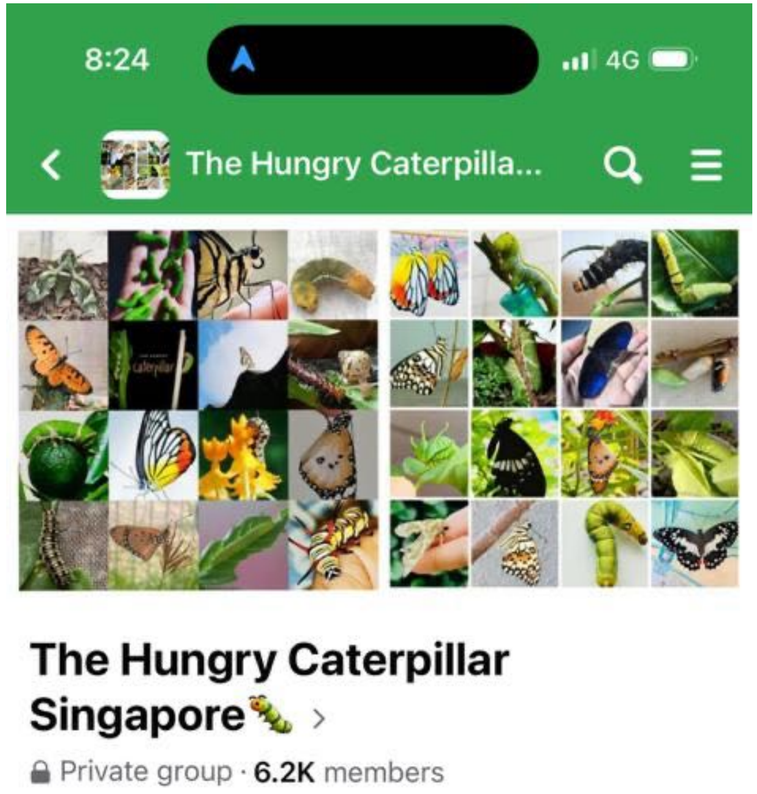
Your 6-year-old daughter found a caterpillar in the garden and wants to watch it grow into a butterfly. You want to support her, but you don’t know how to care for caterpillars.

You wish there was a way for your daughter to learn all about her caterpillar and its needs in a safe learning environment.

SHARIFAH, DSI-41-SG, 28 FEB 2024

https://www.medicalnewstoday.com/articles/caterpillar-rash-treatment

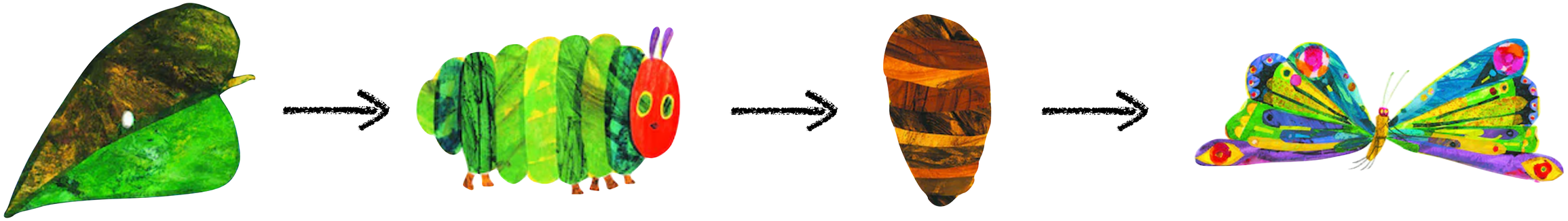
WHY THIS PROJECT MATTERS



- Caterpillars and butterflies are important to our biodiversity and ecology
- Highly-sensitive creatures
- Important to hobbyists, enthusiasts, educators

WORKFLOW

Future Work



Data Collection

Google Images
Data Augmentation

Data Cleaning

Image Preprocessing

Modelling

Convolution Neural Networks (CNN)
Transfer Learning

Application

Streamlit
OpenAI Assistant API

https://encrypted-tbn2.gstatic.com/images?q=tbn:ANd9GcQvBzlODt3GkK_7dpF93fs2LOMaa_73MX1yjitnetK6lHLwpyaX

Data

DATA COLLECTION & CLEANING

SHARIFAH, DSI-41-SG, 28 FEB 2024

https://2.bp.blogspot.com/-9ypAth7koY8/UjJ25yyKi_I/AAAAAAAABEQ/sJnvoXy6u8o/s1600/caterpillar_eborder1flipped.jpg

DATA AUGMENTATION

<u>Colour</u>

Gamma Contrast

Sigmoid Contrast

Linear Contrast

<u>Warping</u>

Cropping

Elastic

Polar

<u>Noise</u>

Jigsaw

Adding Noise

<u>Orientation</u>

Shear

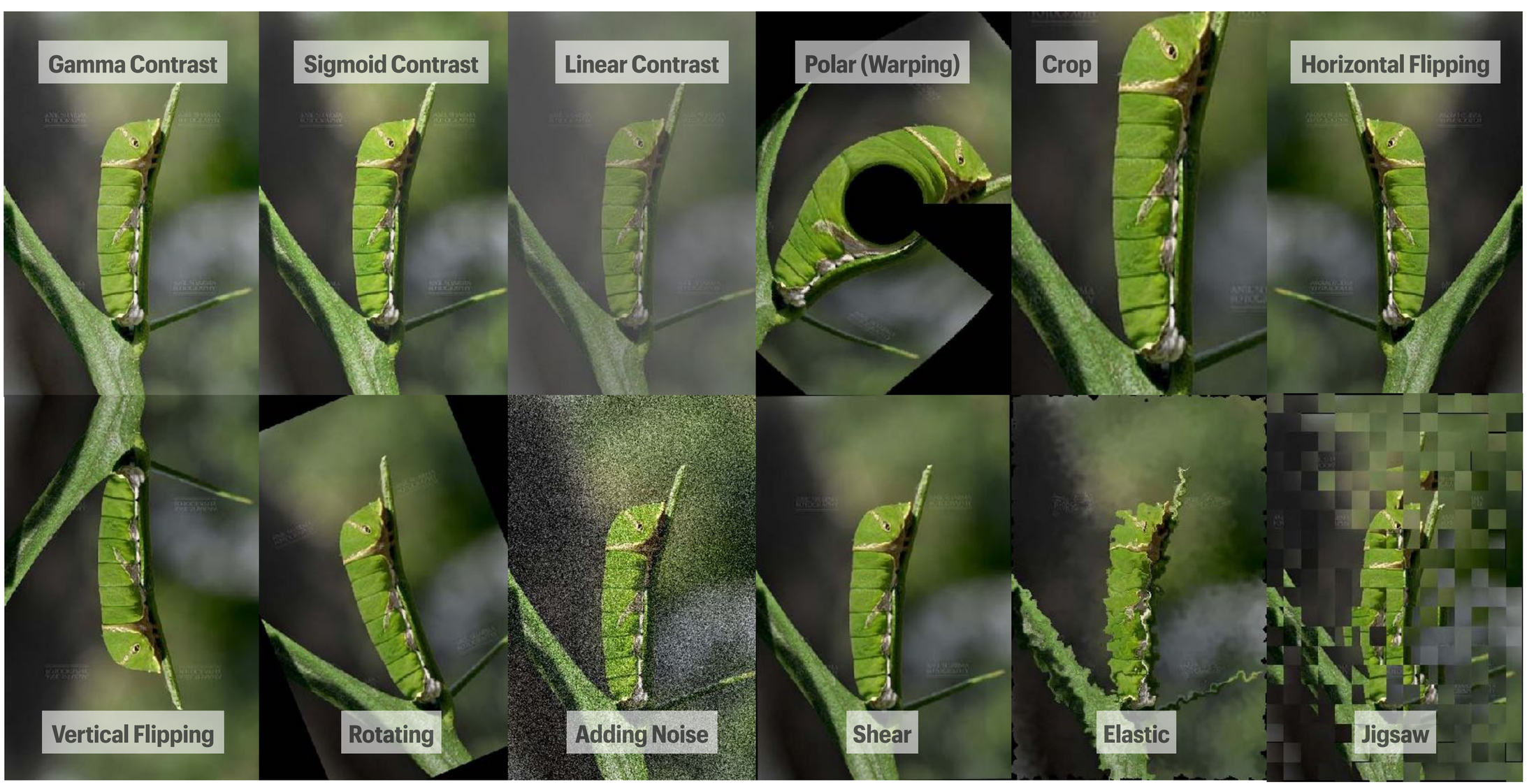
Rotation

Horizontal Flip

Vertical Flip

Applying different transformations to the original data image to increase the size and diversity of the dataset

Data Collection



DATA AUGMENTATION

SPECIES	ORIGINAL NUMBER OF IMAGES	FINAL NUMBER OF IMAGES
Chocolate Pansy *(Junonia iphita)*	20	260
Lime Butterfly *(Papilio demoleus)*	20	260
Painted Jezebel *(Delias hyparete)*	20	260
Plain Tiger *(Danaus chrysippus)*	20	260









IMAGE PREPROCESSING

Accepted extensions:

jpeg, jpg, png, bmp

Images resized to 256 by 256

Image data is converted into train, test and validation

Classes are balanced

Data is scaled by dividing each vector by 255

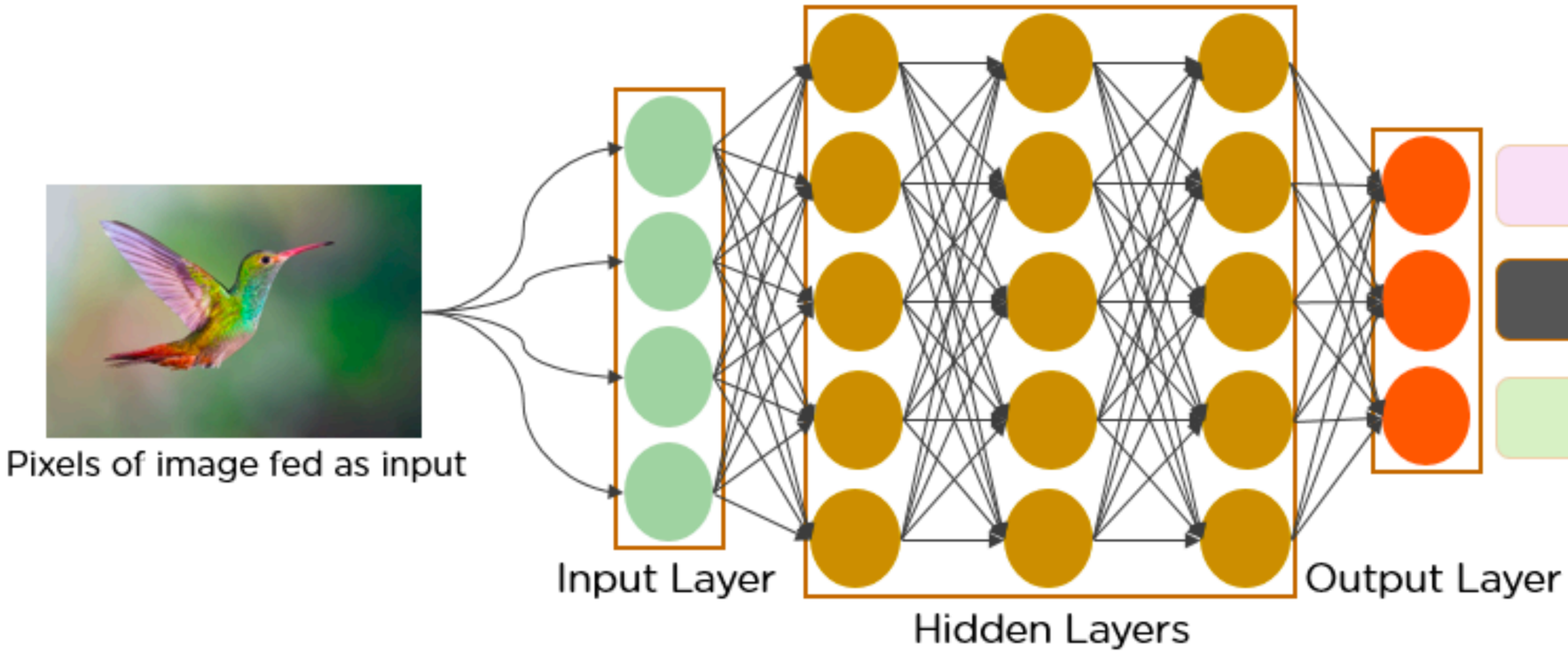
MODELLING

SHARIFAH, DSI-41-SG, 28 FEB 2024

https://2.bp.blogspot.com/-9ypAth7koY8/UjJ25yyKi_I/AAAAAAAABEQ/sJnvoXy6u8o/s1600/caterpillar_eborder1flipped.jpg

CONVOLUTION NEURAL NETWORKS (CNN)

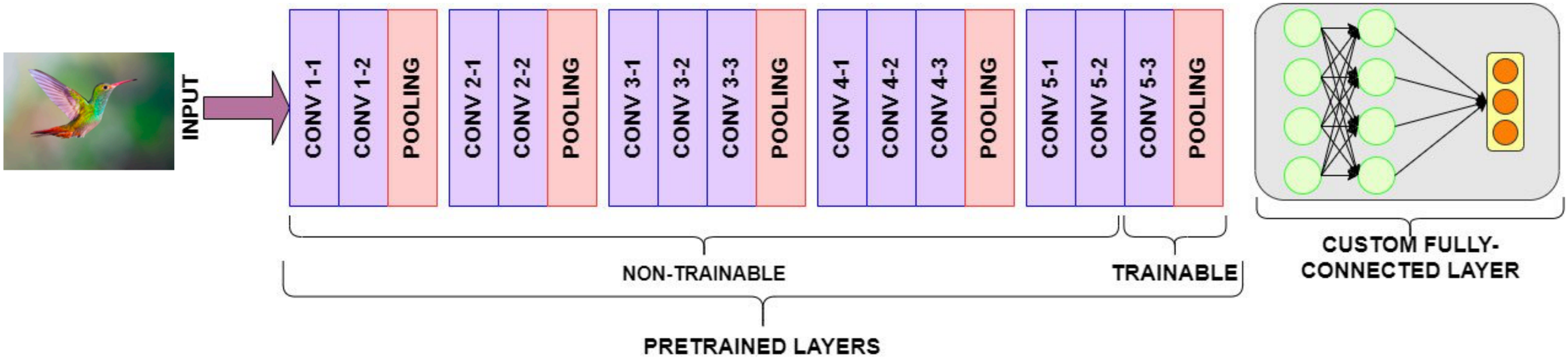
- **Extracts features and identify patterns**



https://editor.analyticsvidhya.com/uploads/25366Convolutional_Neural_Network_to_identify_the_image_of_a_bird.png

TRANSFER LEARNING (WITH VGG-16)

- Using a pre-trained model + customising the last few layers with new images



https://storage.googleapis.com/lds-media/images/transfer-learning-fine-tuning-approach.width-1200.jpg









MODELS WITHOUT REGULARISATION

Model	CNN	Transfer Learning using VGG16
Hidden Layers	6	19
Activation Functions	ReLU + Softmax (final layer)	
Optimiser	Adam	
Loss Function	Sparse Categorical CrossEntropy	

MODELS WITH REGULARISATION

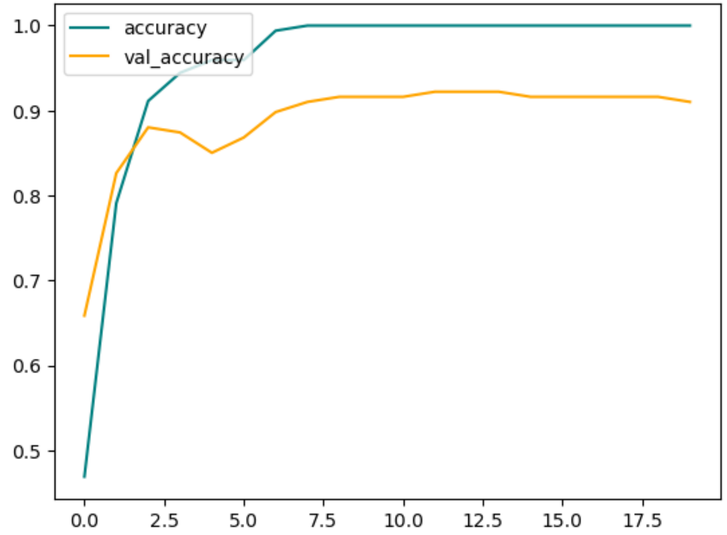
Model	CNN	Transfer Learning using VGG16
Final Epochs	8	7
Activation Functions	ReLU + Softmax (final layer)	
Optimiser	Adam	
Loss Function	Sparse Categorical CrossEntropy	
Regularisation	Early Stopping	

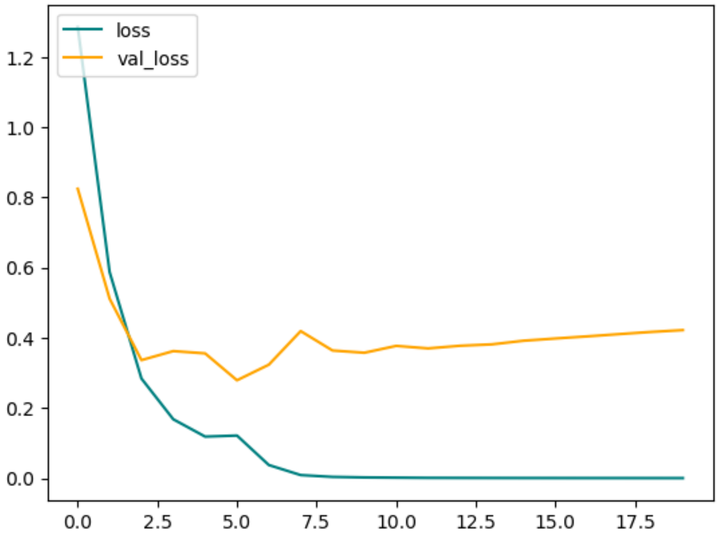
CNN

<u>Legend</u>
Teal >> Training data
Orange >> Validation data

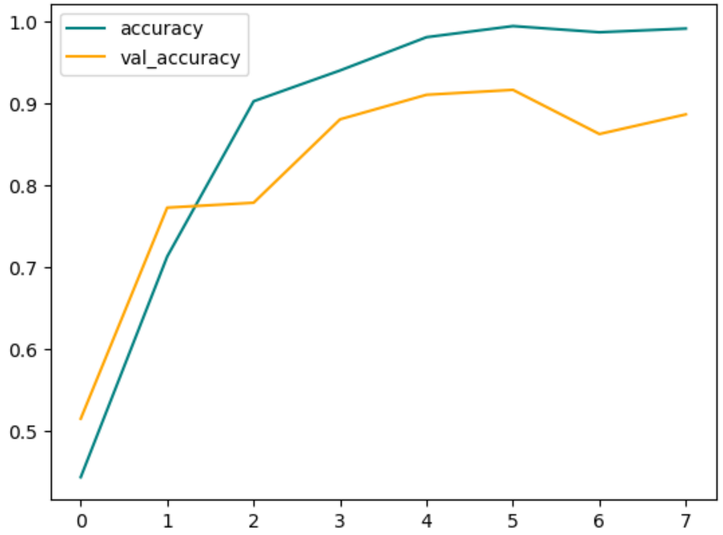
<u>Preliminary insights</u>

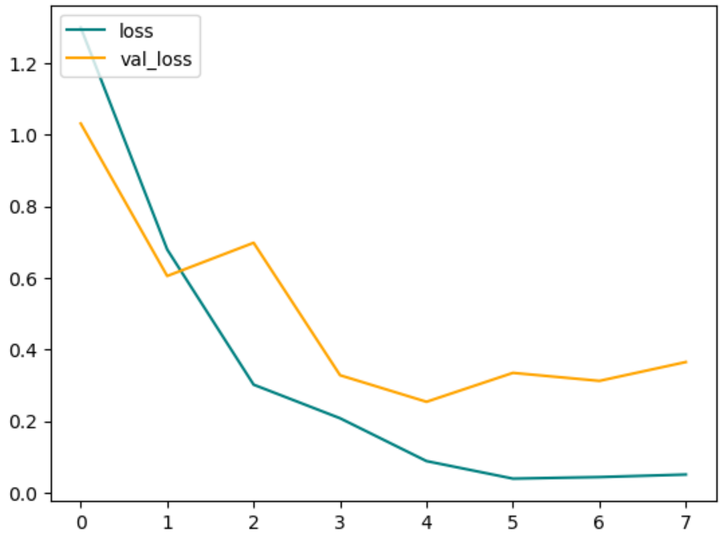
Some evidence of overfitting despite early stopping





Without early stopping





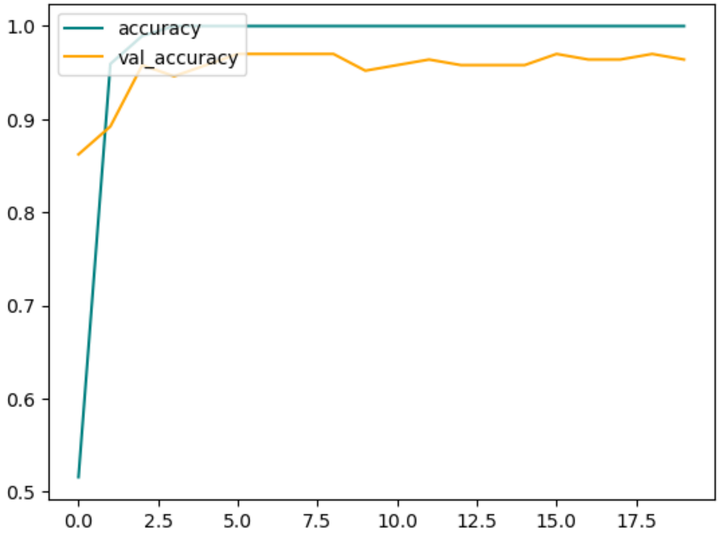
With early stopping

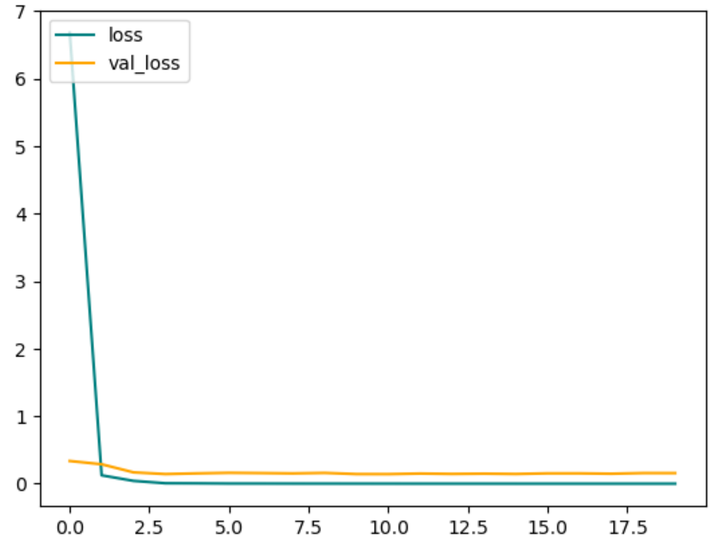
VGG-16

Legend
Teal >> Training data
Orange >> Validation data

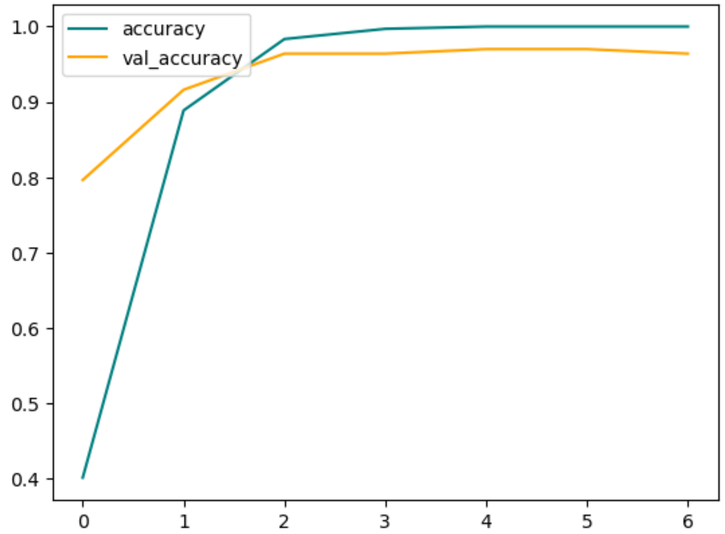
Preliminary insights

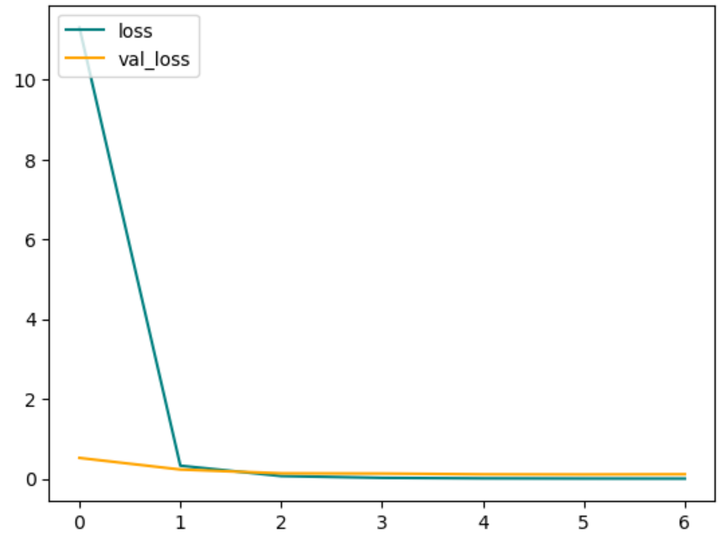
Seems to be more well-fitted compared to the CNN model





Without early stopping





With early stopping

EVALUATION OF THE TRAIN AND TEST DATA

- Data input from the test set (from train test split)
- Comparison of the difference in values from the training to the testing data
    - presence of overfitting, if any
    - good / bad performance

Metrics

High precision score: Whenever it predicts a positive outcome, it is more likely to be correct.

High recall score: The higher the recall score, the less likely that a false negative was detected.

High F1-score: F1-score strikes a balance between precision and recall

EVALUATING THE MODELS

Class / Metrics	Model with Regularisation	Chocolate Pansy	Lime	Painted Jezebel	Plain Tiger
Precision	CNN	-0.12	-0.04	**-0.01**	-0.09
	VGG-16-TL	**-0.1**	**0.02**	-0.02	**-0.04**
Recall	CNN	-0.07	-0.07	-0.13	**0**
	VGG-16-TL	**-0.01**	**-0.06**	**-0.06**	-0.03
F1-Score	CNN	-0.09	-0.05	-0.07	-0.07
	VGG-16-TL	**-0.05**	**-0.02**	**0.04**	**0.04**
Accuracy	CNN	-0.07			
	VGG-16-TL	**-0.03**			

INSIDE THE HIDDEN LAYERS

SHARIFAH, DSI-41-SG, 28 FEB 2024

https://2.bp.blogspot.com/-9ypAth7koY8/UjJ25yyKi_I/AAAAAAAABEQ/sJnvoXy6u8o/s1600/caterpillar_eborder1flipped.jpg

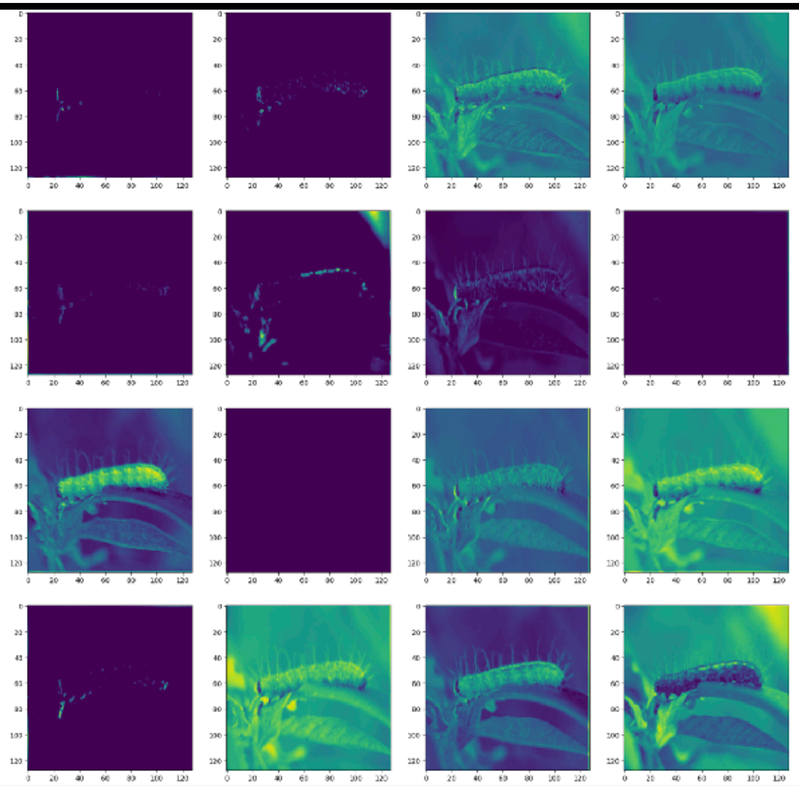
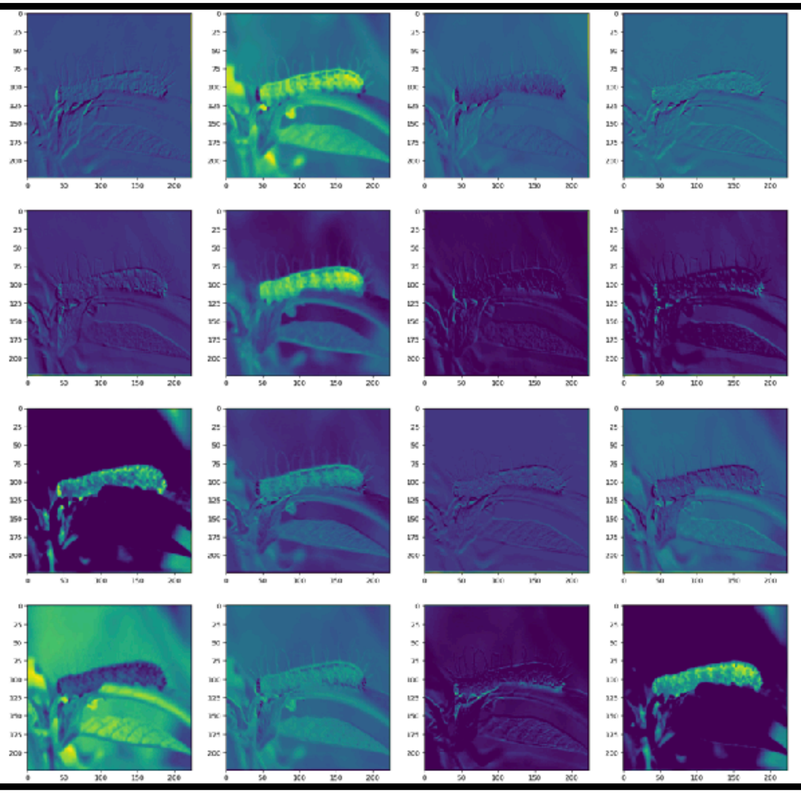








VISUALISING THE LAYERS [1]

CNN	VGG-16-TL-CNN
	

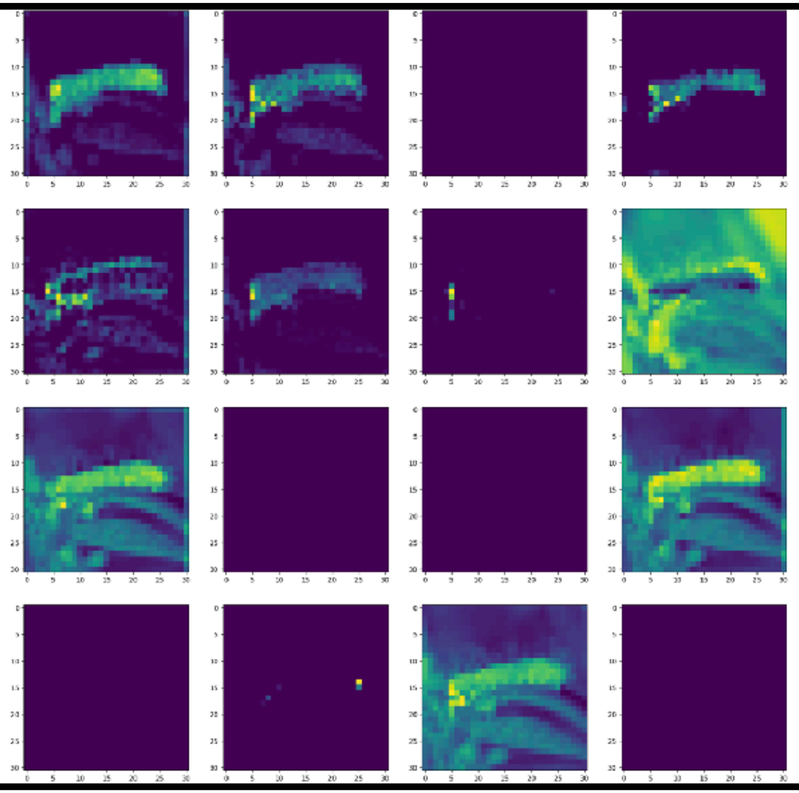
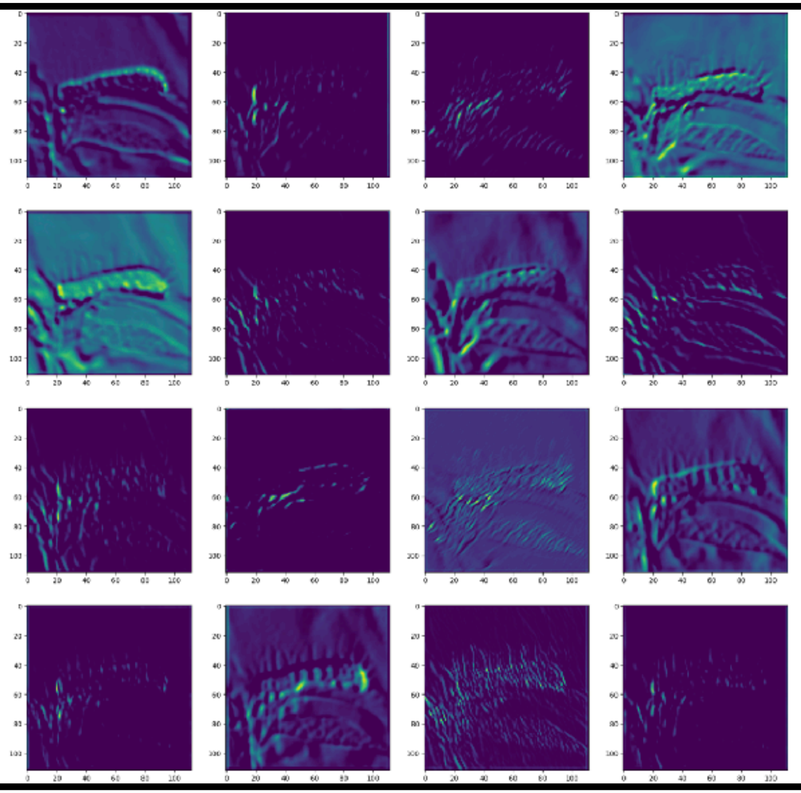








VISUALISING THE LAYERS [5]

CNN	VGG-16-TL-CNN
	

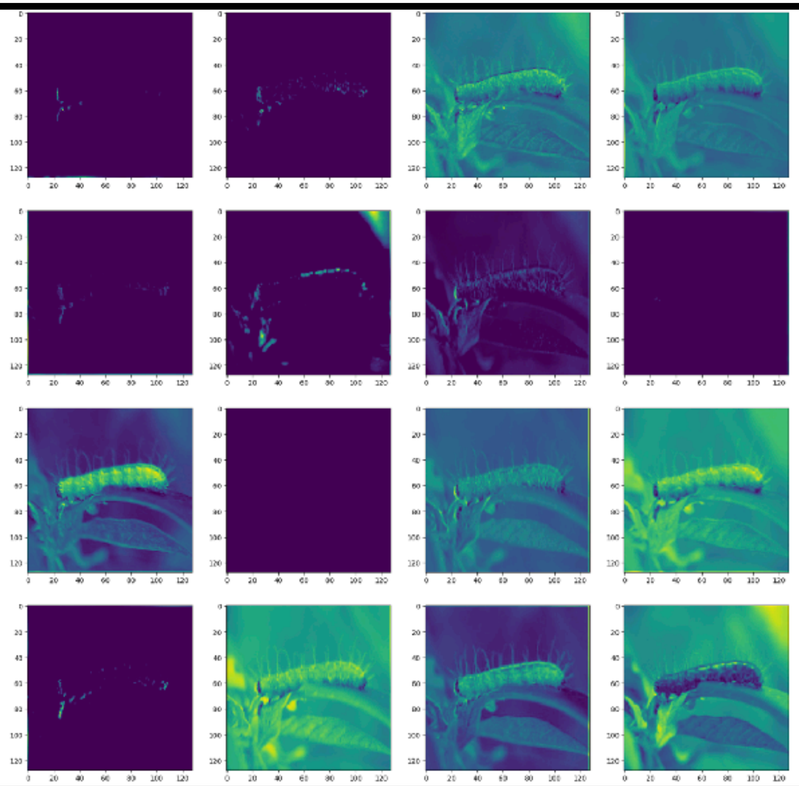
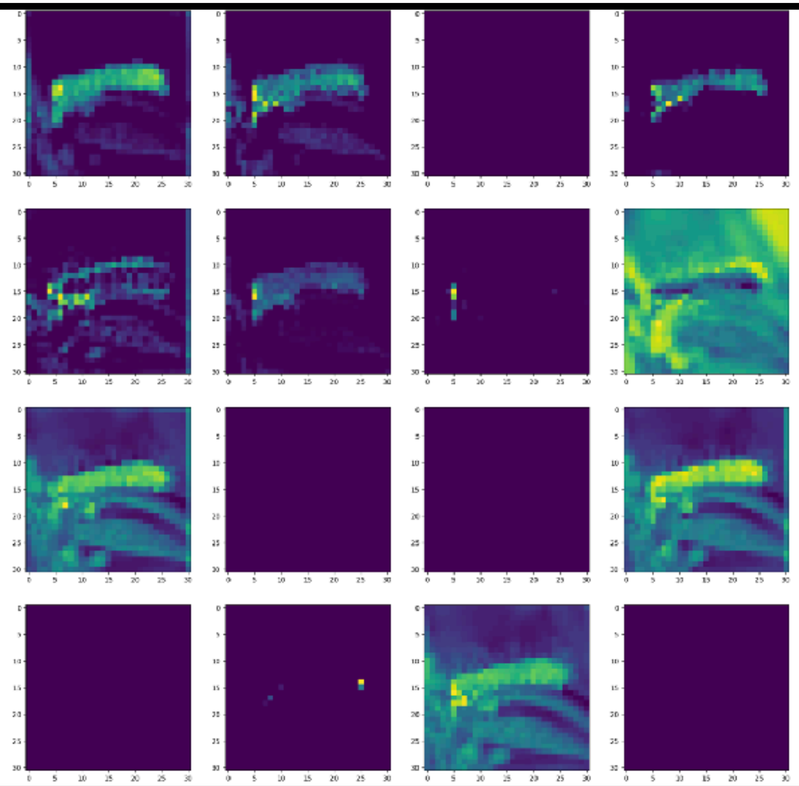








VISUALISING THE LAYERS (CNN)

Model.layers[1]	Model.layers[5]
	

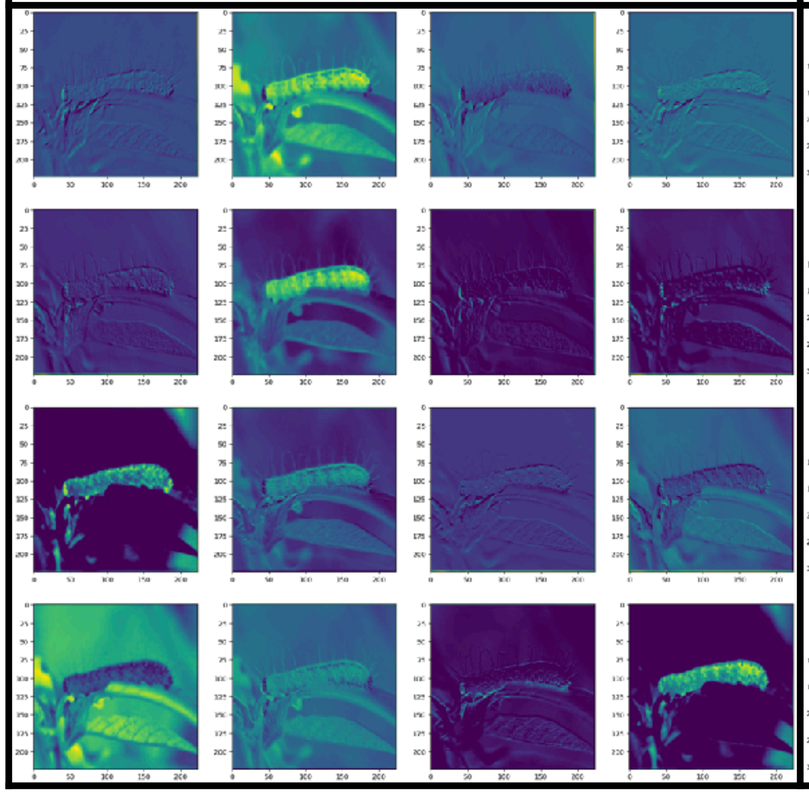
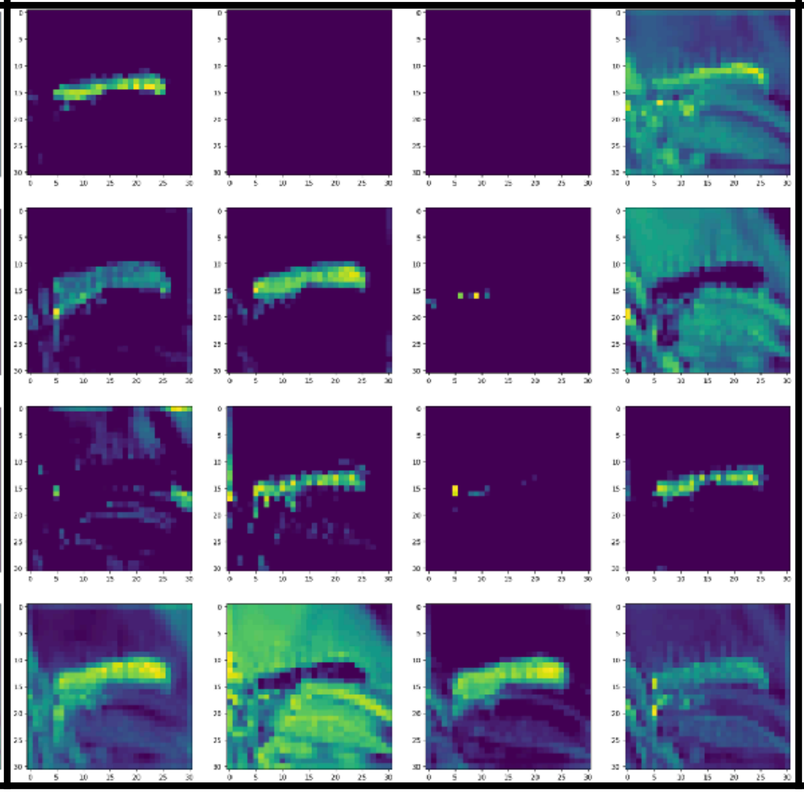
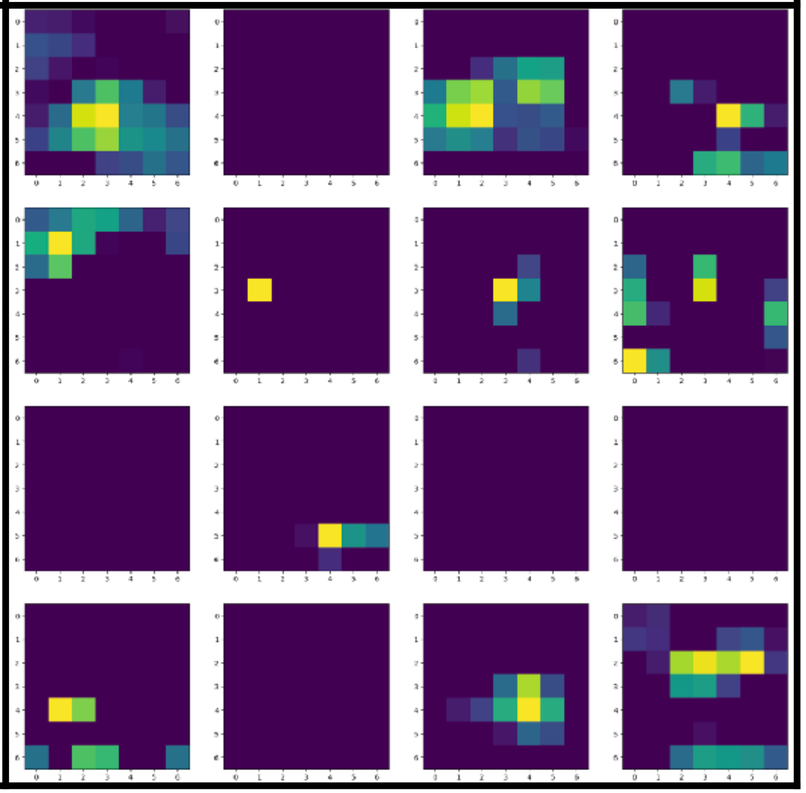








VISUALISING THE LAYERS (VGG-16-TL-CNN)

Model.layers[1]	Model.layers[5]	Model.layers[18]
		

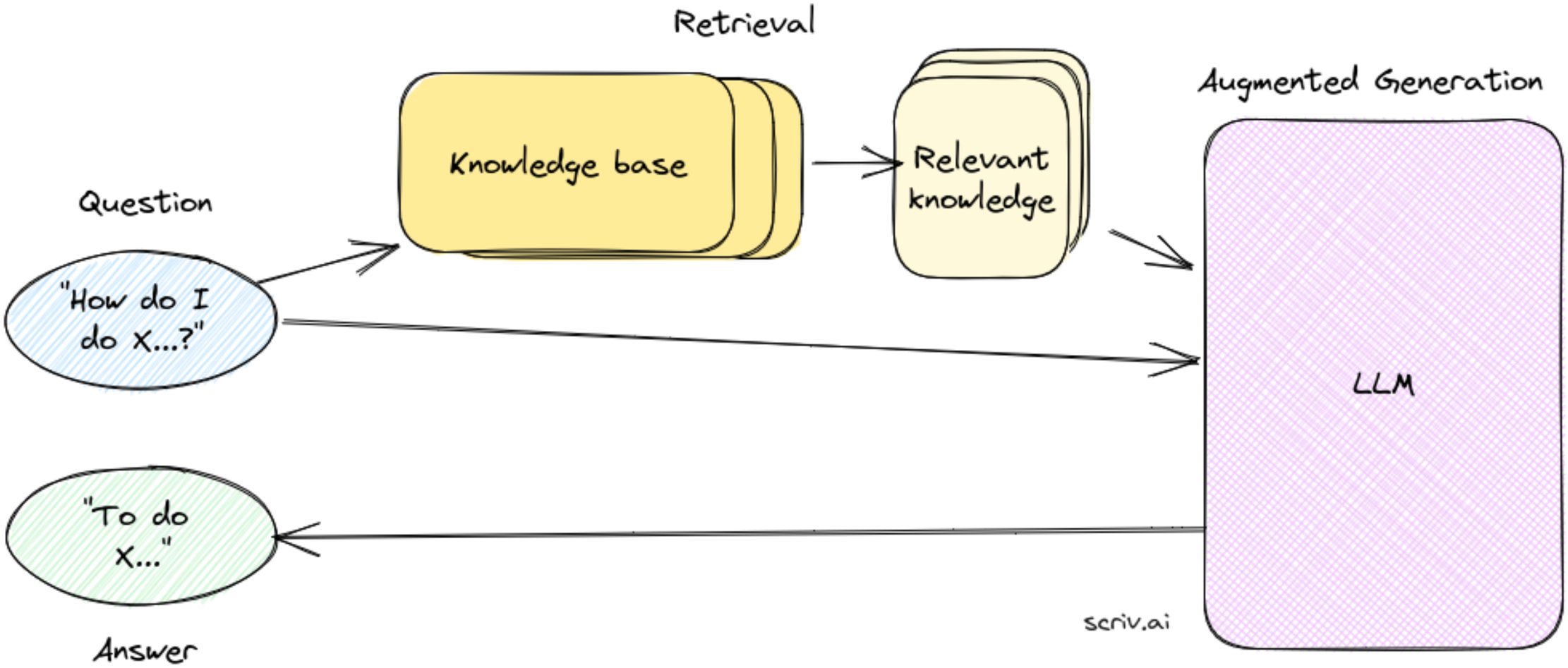
Application

STREAMLIT DEMO

SHARIFAH, DSI-41-SG, 28 FEB 2024

https://2.bp.blogspot.com/-9ypAth7koY8/UjJ25yyKi_I/AAAAAAAABEQ/sJnvoXy6u8o/s1600/caterpillar_eborder1flipped.jpg

RAG USING OPEN AI ASSISTANTS API



https://communitykeeper-media.s3.amazonaws.com/media/images/rag-overview.original.png

Application

You are 7 years old.

Ask a question about the Plain Tiger Caterpillar!

Here, you can ask questions about your caterpillar 🐛

✓ We're done! Here's the information you wanted! Ask another question if you wish 🌱

What do they eat?

The Plain Tiger caterpillar eats leaves of the Crown Flower plant. It enjoys munching on these leaves to grow and survive.

What would you like to know about Plain Tiger Caterpillars?

Application

You are 7 years old.

Ask a question about the Plain Tiger Caterpillar!

Here, you can ask questions about your caterpillar 🐛

✓ We're done! Here's the information you wanted! Ask another question if you wish 🌱

What do they grow into?

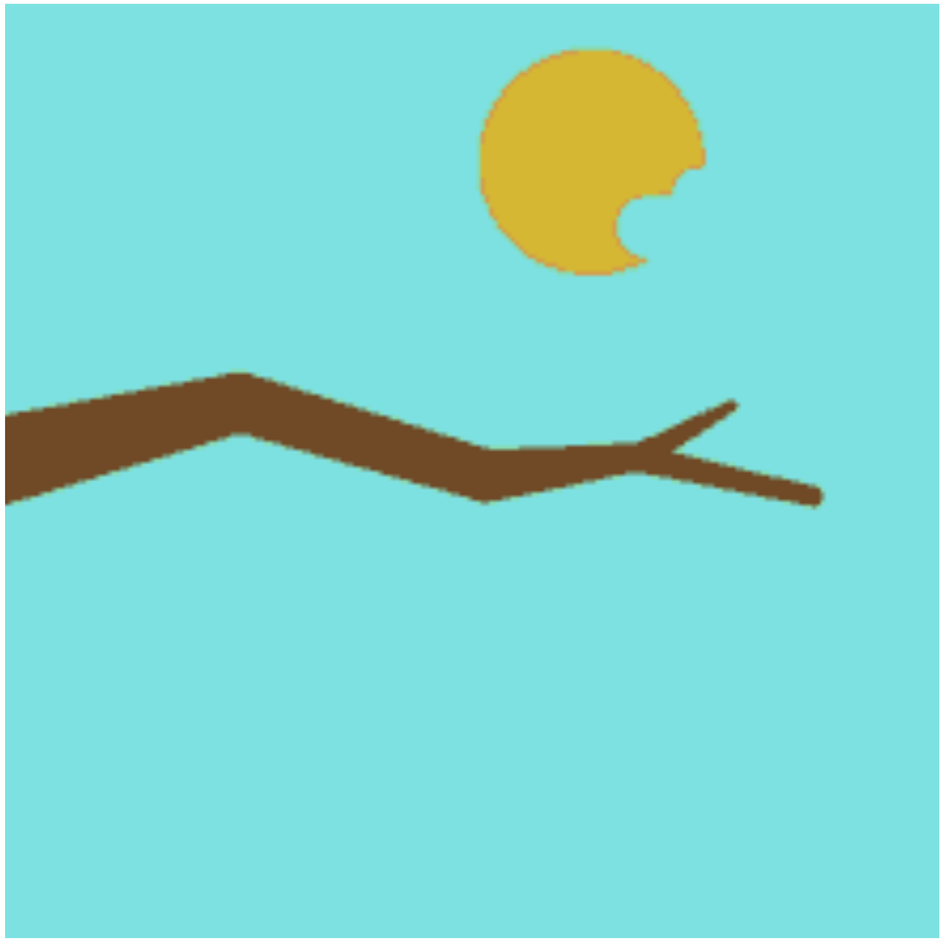
The Plain Tiger Caterpillar grows into a butterfly with bright orange wings that have black veins and borders. The butterfly also has white spots on the edges of the wings and black lines on its body .

What would you like to know about Plain Tiger Caterpillars?

FUTURE WORK

- To be able to identify the species at any stage of its life cycle
- To expand to more species of butterflies and moths for classification
- Chatbot to include giving images as responses
- Chatbot to include recommended websites that the user can visit to find out more about their caterpillar
- Future applications of this model
    - Can expand to different types of animals like birds, snakes or other insects

CONCLUDING THOUGHTS



- More exploration into the use of RAG using OpenAI's Assistant API (currently still in BETA)
- Is transfer learning always the best way? Does it distract or does it enhance?
- How could this be applied to other systems (e.g. recommender systems)?

THANK YOU

REFERENCES

- PPT is inspired by and uses images from Eric Carle's "The Very Hungry Caterpillar"
- Images downloaded and saved from Google Images
- Information on Caterpillars from Nepal Desk, Butterfly Circle
- All other references are found on the slides or in notebooks