

COL 774 Assignment 4
Faraz Bukhari(2024AIZ8310), Hasan Mustafa(2024AIY7541)

BEST MODEL [LINK](#):

<https://drive.google.com/drive/folders/1MoUNyhA0LqJDJtDEH9pkhSbiBQ3RhNX9?usp=sharing>

Question 2: Data Preprocessing

1. Tokenization: used [bert-base-unencased](#)
2. Handling Variable length questions:
 - a. checked the maximum question length: [48](#)
 - b. Set max sequence length of encoder input: [64 \(next higher power of 2\)](#)
 - c. Handling variable length of questions: [padded smaller sequences](#)
3. Image Preprocessing:
 - a. Resized: to [224X224](#)
 - b. Normalization: [used mean and standard deviation of ImageNet-dataset](#)
4. Answer Preprocessing:
 - a. Tokenization: [custom made word tokenizer](#)
 - b. Vocab: [a dictionary of tokens](#)
 - c. Token-index-id: [position of token in vocabulary](#)

Question 3: Network Architecture

Implemented the complete architecture as described in the problem statement.

Question 4: Image Encoder

- Model: [Loaded RestNet101 from torchvision.models](#)
- Enhancements:
 - Removed last two layers (i.e. average pool and fully connected)
 - Projected the RestNet101 output from 2048 dimensions to 768 dimensions
 - Set the gradient parameters to True/False as required (in subsequent parts)

Question 5: Text Encoder

- Model: Transformer architecture implemented as per the problem statement

Question 6: Cross Attention

- Model:
 - Used text as Q
 - Used images as (K, V)

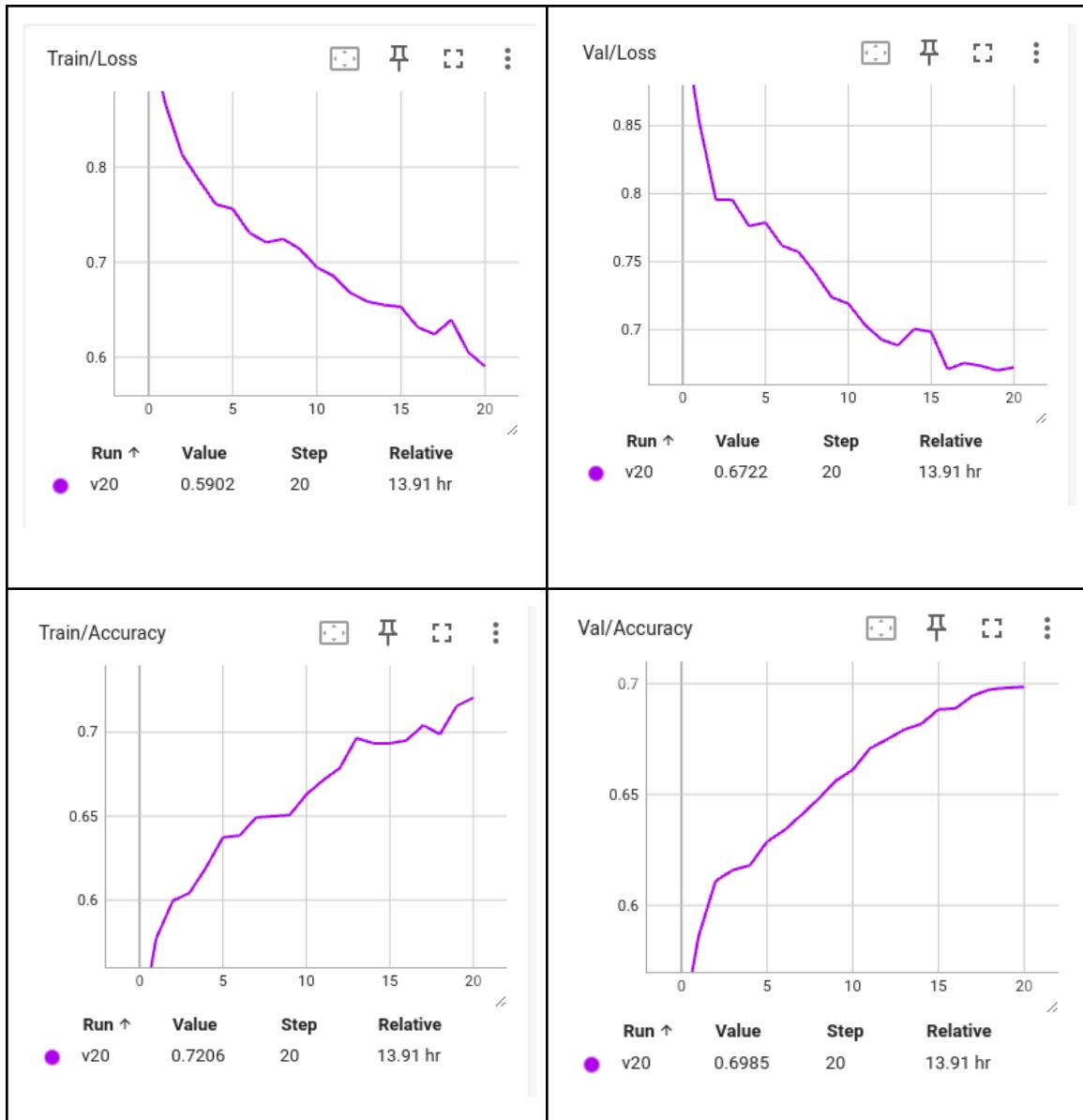
Question 7: Classifier

- Model:
 - Classifier implemented for 29 classes (corresponding to each answer token)

Question 8: Training and Evaluation

Required gradient updates for RestNet101 set to False.

- a) Parameters:
 - i) Learning Rate: 1e-4 (with 1 epoch of warmup and rest decayed upon three consecutive iterations of non-decreasing validation loss);
 - ii) Batch Size=128;
 - iii) Epochs 20;
 - iv) Optimizer: Adam;
 - v) Loss: Cross Entropy
- b) Plots showing the train loss and Val Loss. The model has just started to overfit as is apparent by a bump and then a plateau in Val Loss.



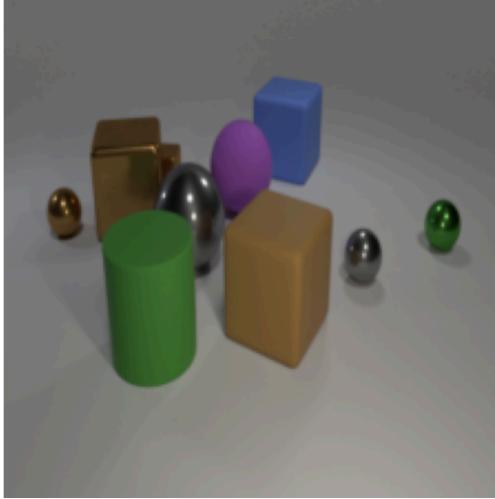
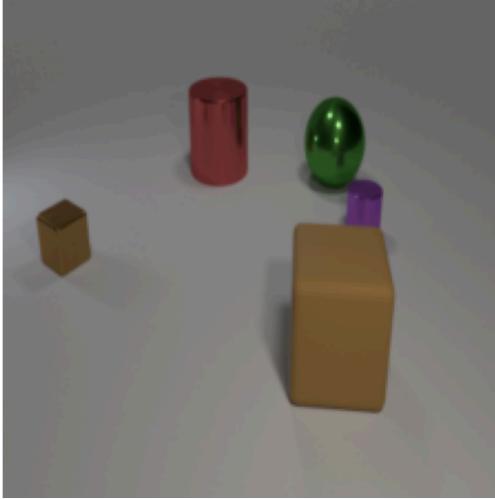
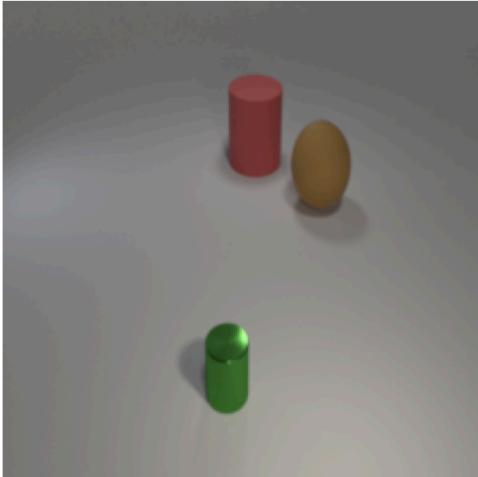
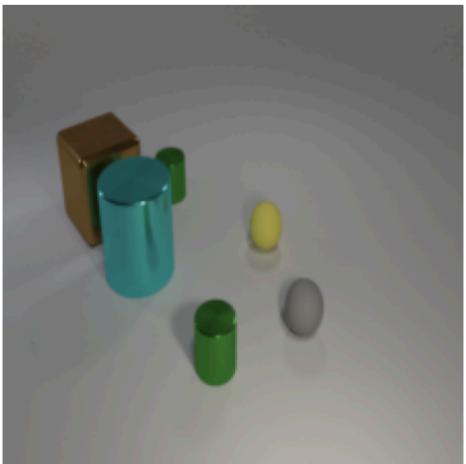
c) Evaluating on Validation A Set ---

```
{'loss': 0.6722125348790486, 'metrics': {'accuracy': 0.69848, 'precision': 0.5516858116268623, 'recall': 0.5672612497631663, 'f1': 0.544424897088961}}
```

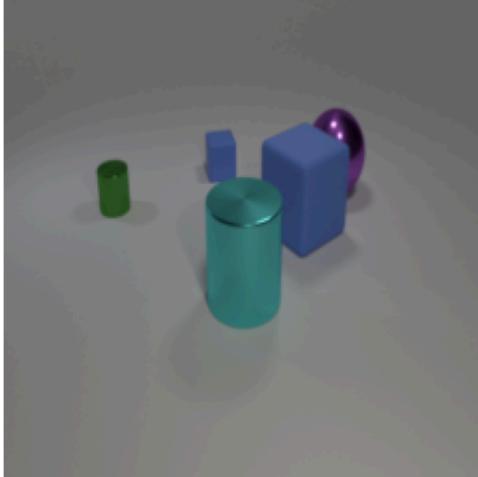
Evaluating on Test Set A —

```
{'loss': 0.6576629115168253, 'metrics': {'accuracy': 0.6974533333333334, 'precision': 0.5202734231491879, 'recall': 0.5276277589916654, 'f1': 0.5182716151688024}}
```

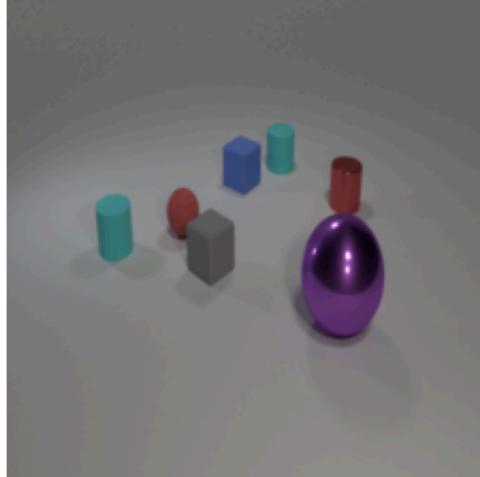
d) Visualizations:

d) Incorrect	e) Correct
<p>Q: how many balls are behind the big brown shiny block? A: 1 Pred: 2</p> 	<p>Q: what is the color of the large shiny cylinder? A: red Pred: red</p> 
<p>Q: what number of tiny objects are either green metallic objects or purple rubber spheres? A: 1 Pred: 0</p> 	<p>Q: there is a thing that is both on the right side of the brown thing and behind the small yellow rubber thing ; what material is it made of? A: metal Pred: metal</p> 

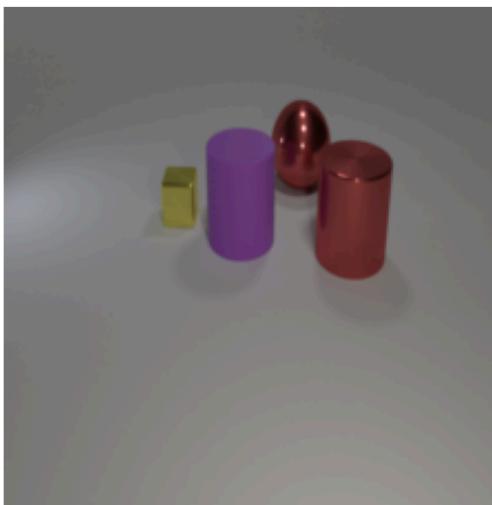
Q: what size is the other blue object that is the same shape as the tiny blue matte thing?
A: large
Pred: small



Q: how many other purple objects are the same shape as the large purple metallic thing?
A: 0
Pred: 0



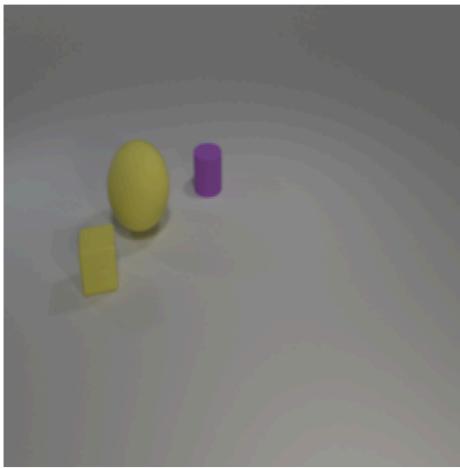
Q: do the yellow block and the large ball have the same material?
A: yes
Pred: no



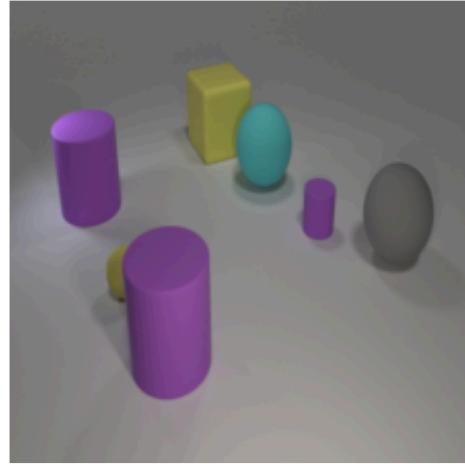
Q: what number of other things are there of the same shape as the big metal object?
A: 0
Pred: 0



Q: what is the color of the small object right of the small object to the left of the small purple rubber object that is to the right of the small yellow thing?
A: purple
Pred: yellow



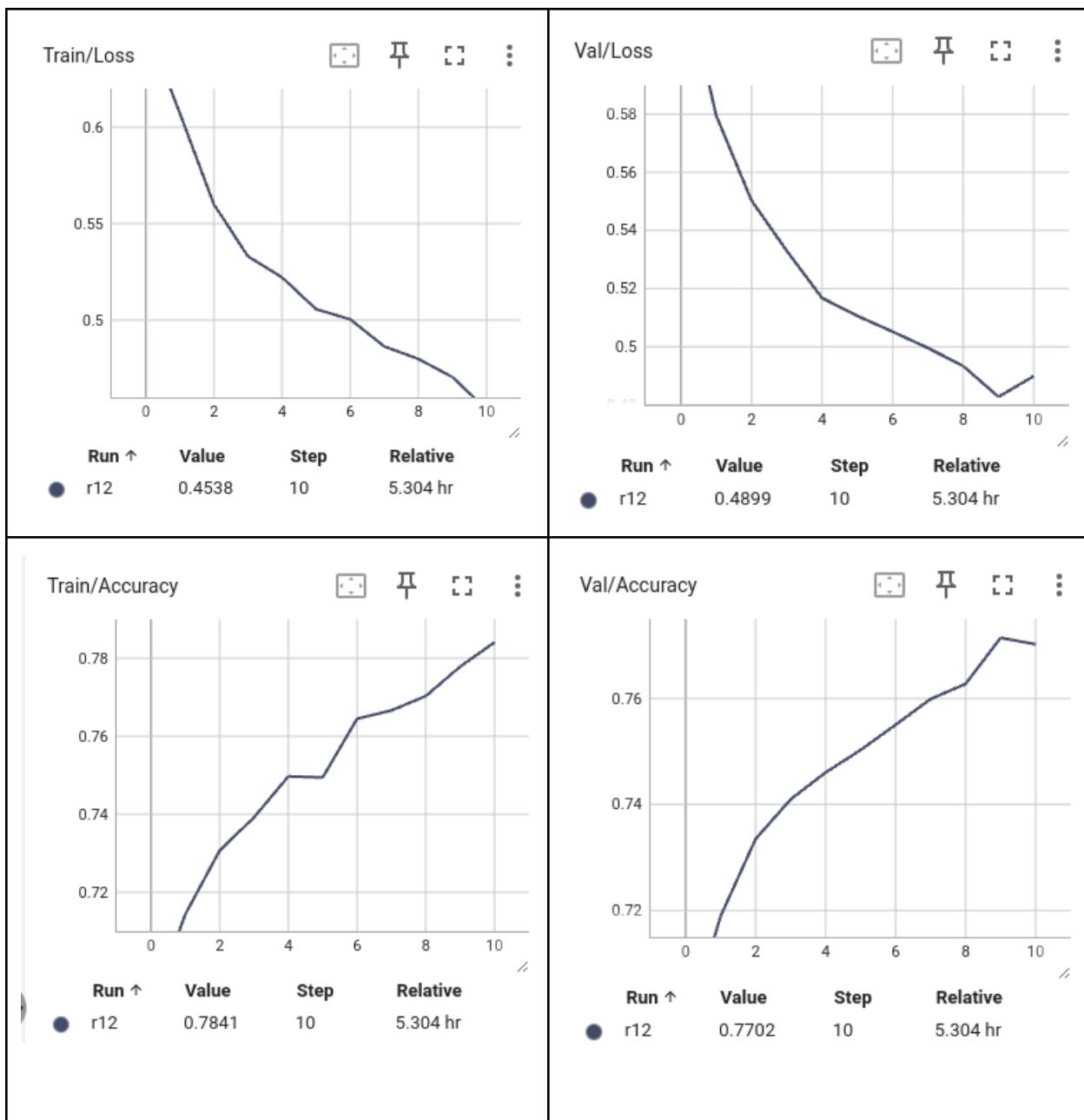
Q: there is an object that is both on the right side of the big cyan matte sphere and on the left side of the big gray ball ; what is it made of?
A: rubber
Pred: rubber



- e) Additionally shown 5 images which were misclassified (above). Based on these examples
- i) If objects are hidden behind other objects the model makes mistakes
 - ii) Model struggles with counting

Question 9: Fine Tuning Image Encoder

- a) Hyperparameters: same as in the part above the Starting checkpoint: above model was used as starting point for this model
- b) Plot is shown in figure, validation loss has started increasing and validation accuracy has started decreasing implying overfitting has started.



c) --- Evaluating on Validation Set A ---

```
{'loss': 0.4898894206746419, 'metrics': {'accuracy': 0.7702533333333333, 'precision': 0.6511770000090026, 'recall': 0.6213115016594808, 'f1': 0.6290820090126718}}
```

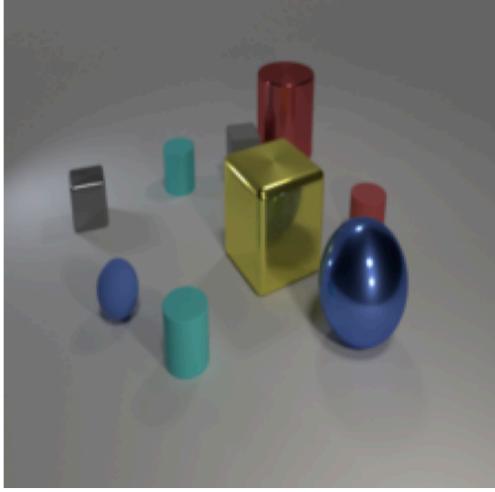
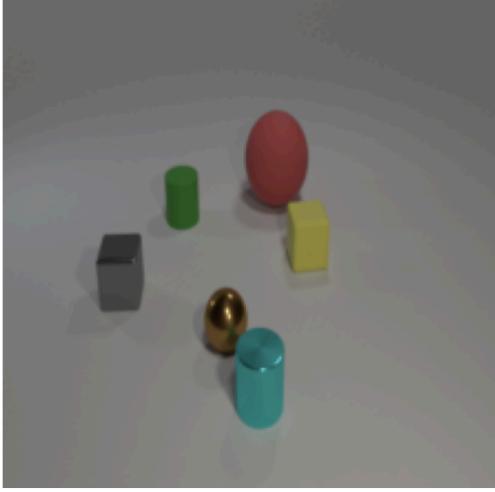
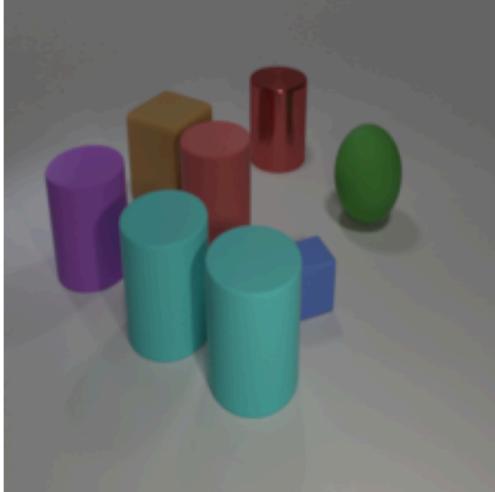
--- Evaluating on Test Set A ---

```
{'loss': 0.4953877480888367, 'metrics': {'accuracy': 0.7706533333333333, 'precision': 0.6799755246816224, 'recall': 0.6258364429135816, 'f1': 0.6350491537395316}}
```

d) Shown 5 images with questions, correct answers and predicted answers.

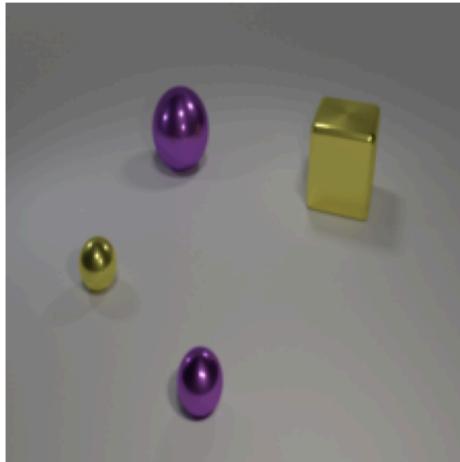
e) Additionally shown 5 images which were misclassified. Based on these examples

- i) If objects are hidden behind other objects the model makes mistakes
- ii) Model struggles with counting

e) Incorrect Predictions	d) Correct Prediction
<p>Q: how many rubber things are tiny gray cubes or small blue things? A: 2 Pred: 1</p> 	<p>Q: is the large matte object the same color as the tiny matte cube? A: no Pred: no</p> 
<p>Q: what number of things are cylinders or tiny brown cubes? A: 5 Pred: 3</p> 	<p>Q: there is a large ball to the right of the shiny sphere behind the cylinder behind the brown ball ; what is its material? A: metal Pred: metal</p> 

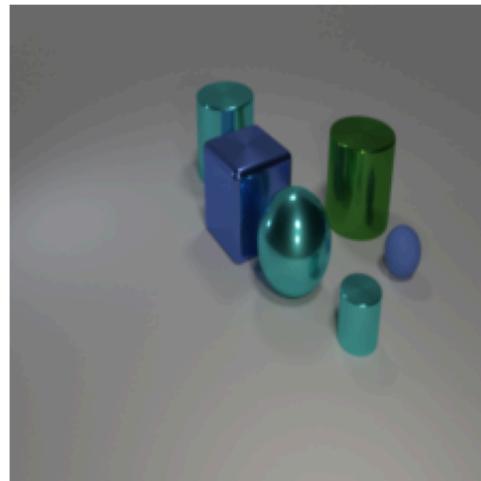
Q: are there an equal number of purple metallic spheres that are in front of the small yellow thing and metallic balls to the left of the big metal cube?

A: no
Pred: yes



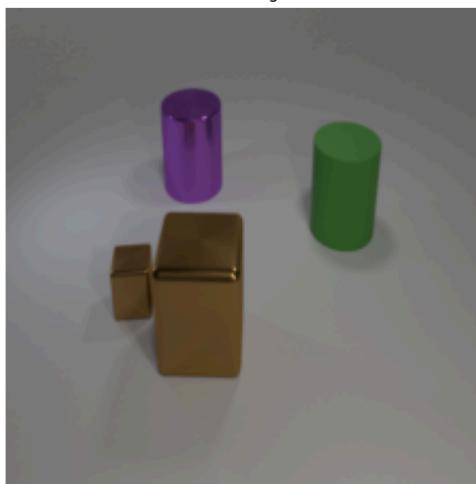
Q: there is a metallic object that is to the right of the large blue shiny block and behind the shiny block ; what is its size?

A: large
Pred: large



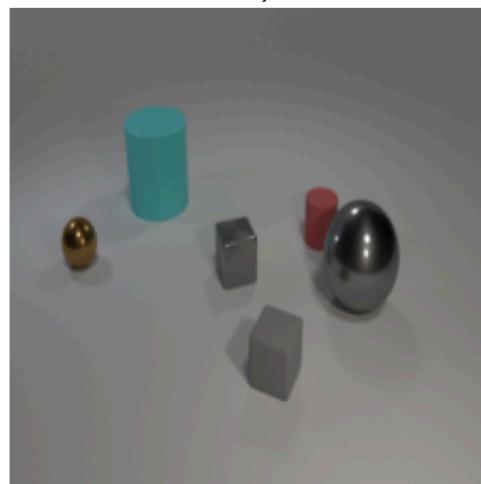
Q: how big is the metal thing that is both behind the big brown object and in front of the rubber object?

A: small
Pred: large

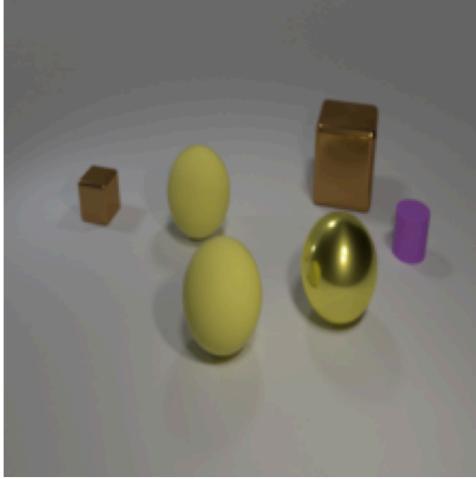


Q: there is a large rubber object that is the same shape as the small red rubber object ; what is its color?

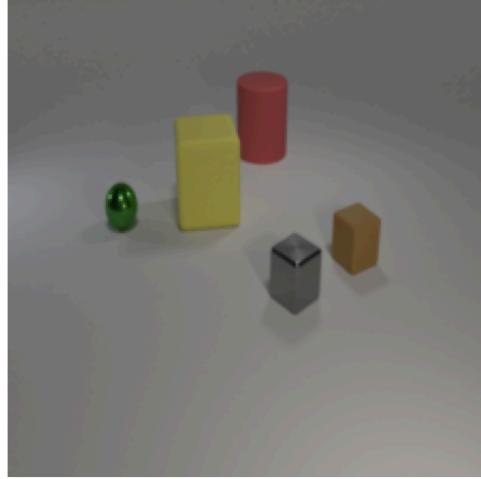
A: cyan
Pred: cyan



Q: what number of big matte things are the same shape as the yellow shiny object?
A: 2
Pred: 1



Q: are there the same number of big matte cubes and tiny gray matte balls?
A: no
Pred: no



Question 10: Further Enhancements (BERT + Focal Loss)

Best Model used: Model from Question 9 (unfrozen ResNet101)

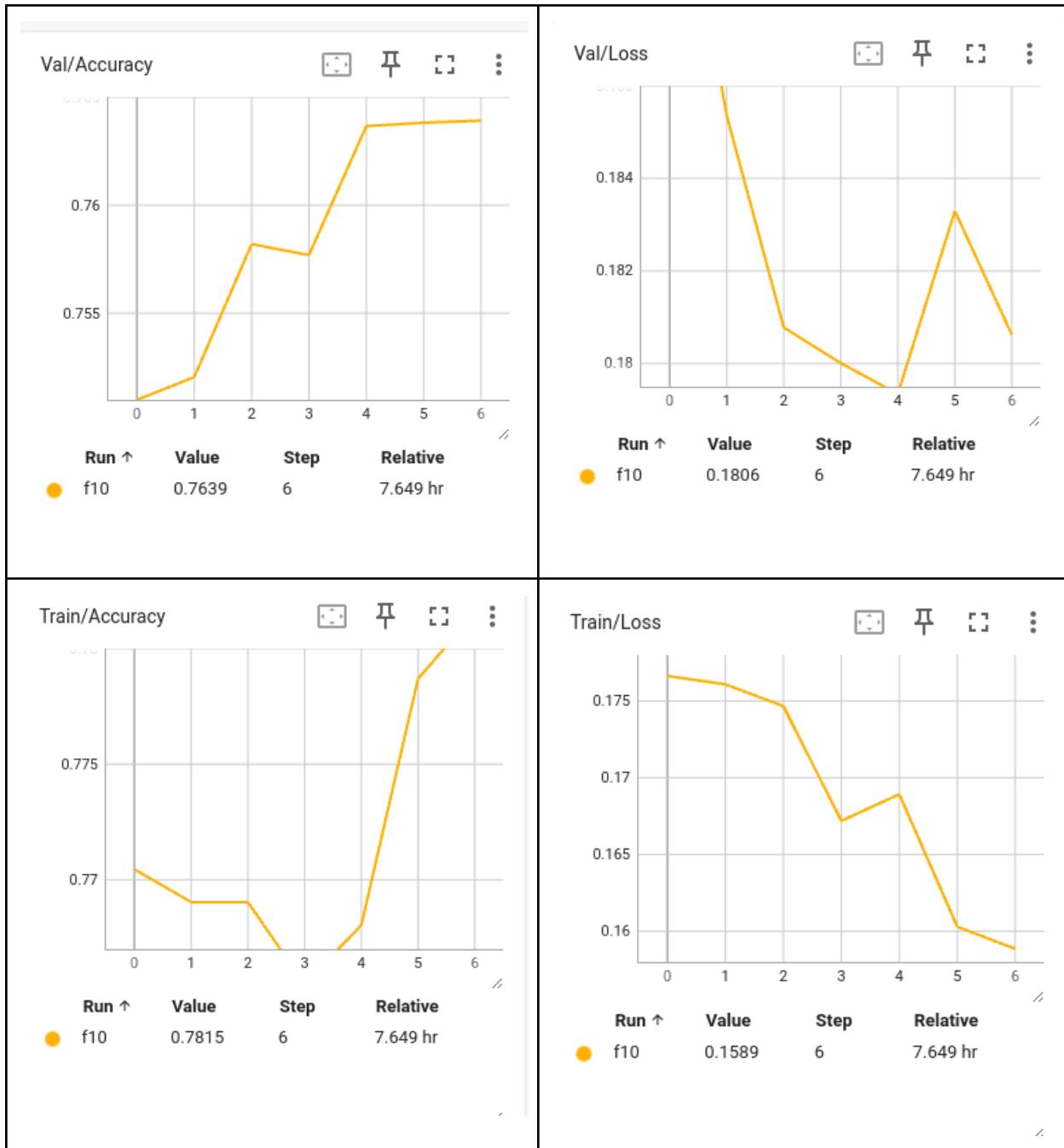
1. Focal loss

a. **Hyperparameters:**

- i. Same as above,
- ii. alphas in focal loss were set to 1, gamma was taken to be 2.

b. **Plots are as shown below.**

Training and validation loss are both fluctuating making inference about overfitting/underfitting difficult. Model achieved slight improvement and may improve further on continued training.



c. **Evaluating on Test Set A**

```
{'loss': 0.5153809331385295, 'metrics': {'accuracy': 0.7663466666666666, 'precision': 0.6491471931532515, 'recall': 0.670012014661121, 'f1': 0.6487186083798541}}
```

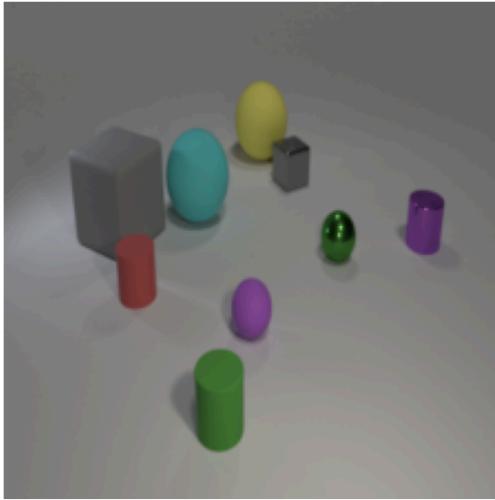
d. **Visualization:** Images, Question, Answer, Predictions

Incorrect	Correct
-----------	---------

Q: how many shiny things are either purple things or tiny gray balls?

A: 1

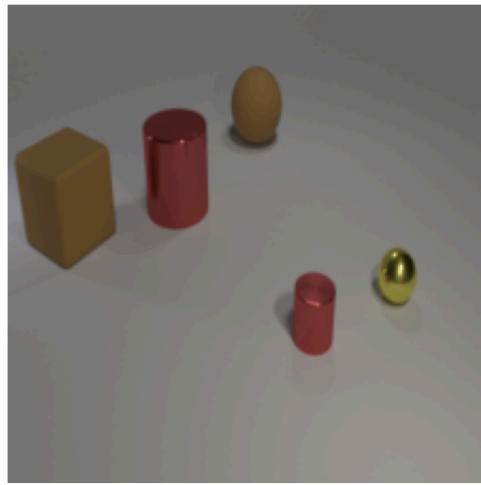
Pred: 2



Q: there is a large brown object that is the same material as the large brown ball ; what shape is it?

A: cube

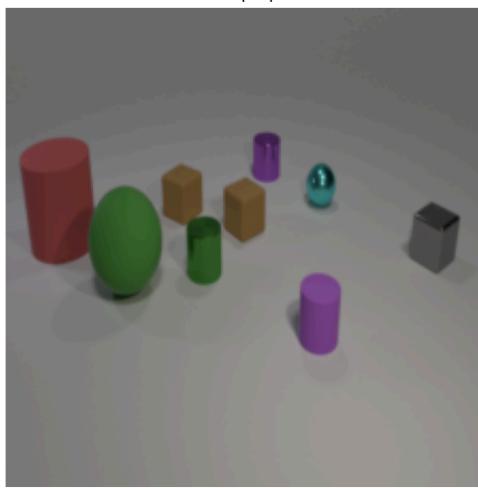
Pred: cube



Q: there is a small object that is in front of the small gray shiny object and behind the green matte thing ; what color is it?

A: green

Pred: purple



Q: is the material of the big object that is behind the small red shiny cylinder the same as the ball in front of the small shiny cylinder?

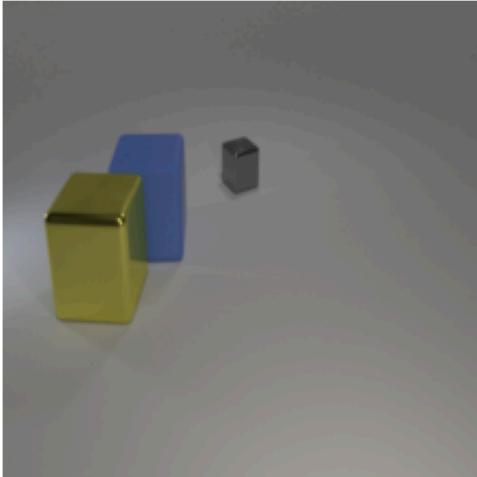
A: no

Pred: no



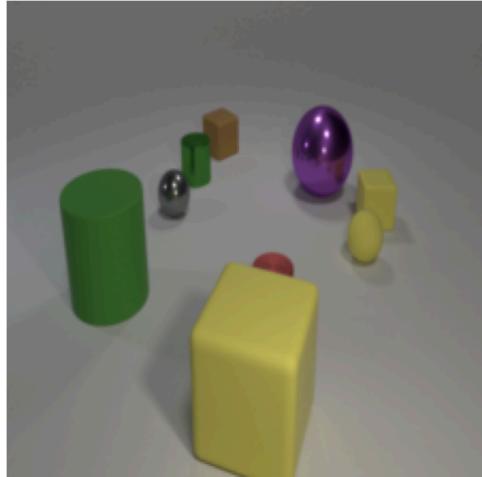
Q: what is the material of the big block behind the metal object that is in front of the block that is right of the large blue object?

A: rubber
Pred: metal



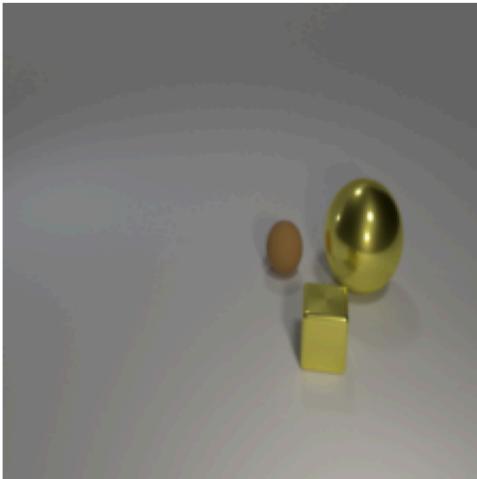
Q: do the metallic thing that is in front of the tiny gray shiny ball and the tiny brown rubber thing have the same shape?

A: no
Pred: no



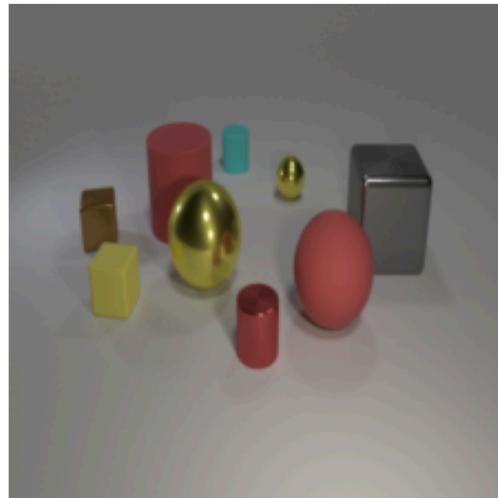
Q: what is the color of the small thing that is the same shape as the big yellow metal object?

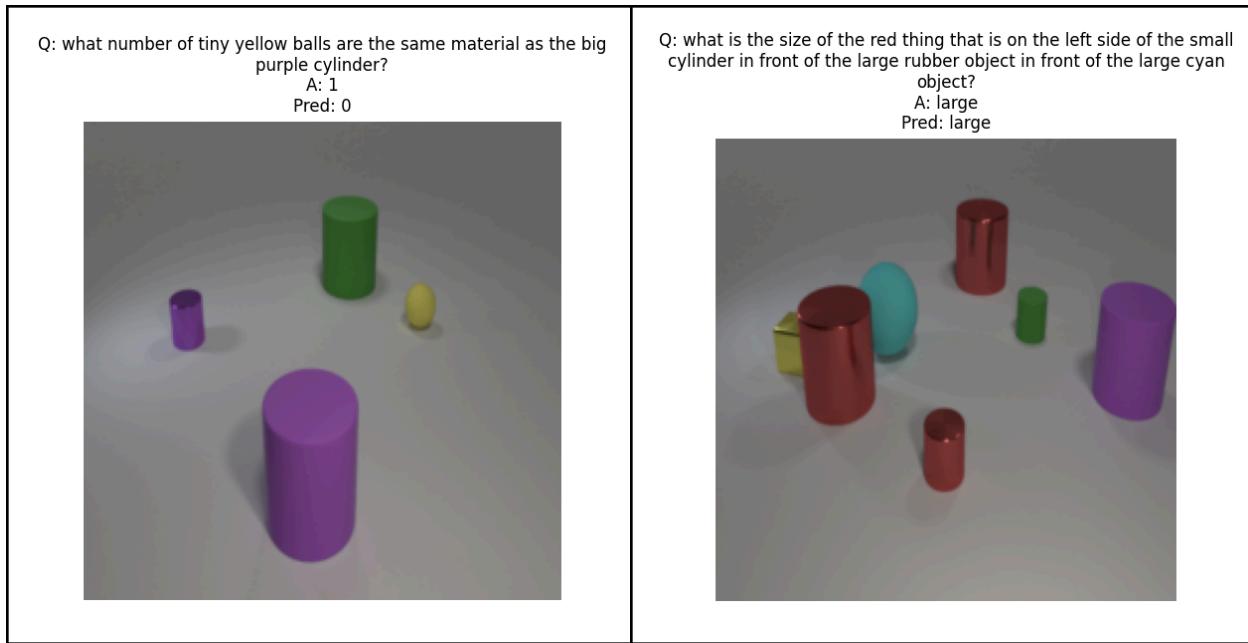
A: brown
Pred: yellow



Q: are any yellow cylinders visible?

A: no
Pred: no

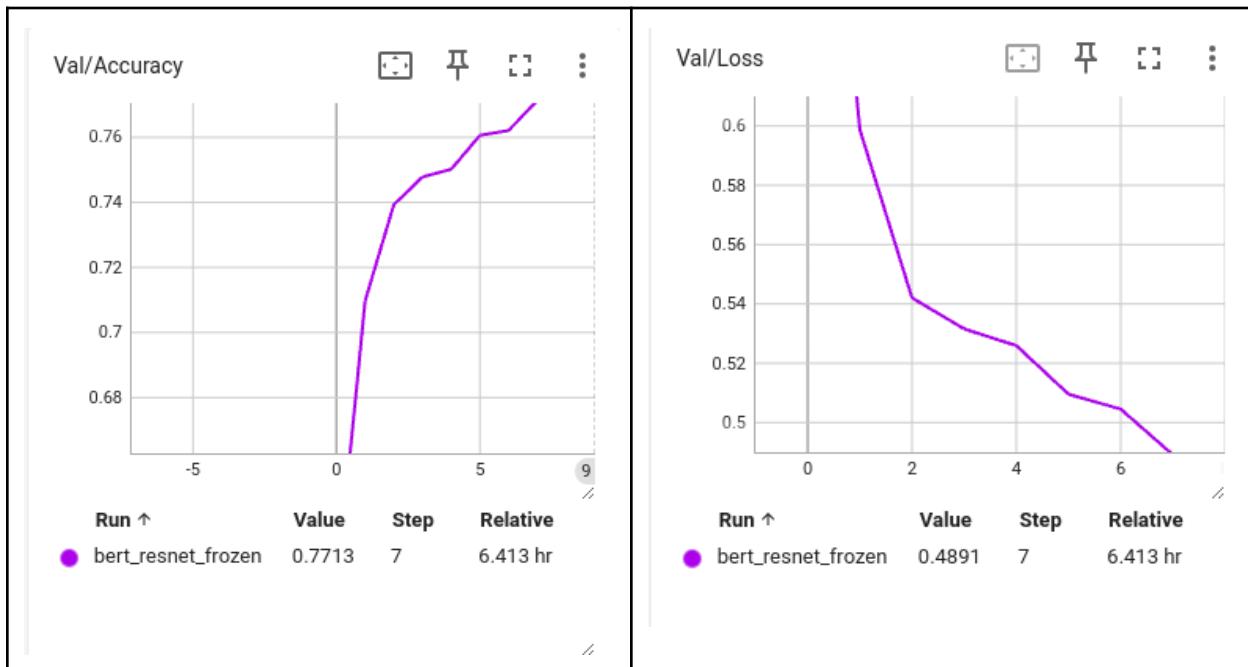


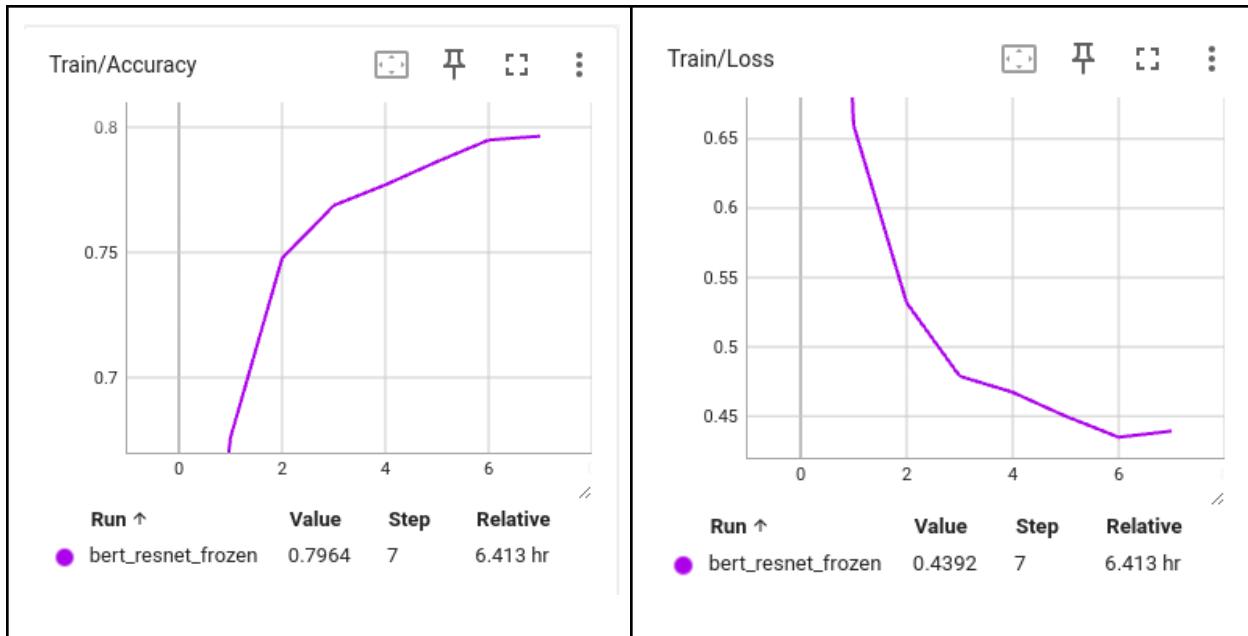


e) Additionally shown 5 images which were misclassified (above). Based on these examples
 Model looks to be struggling with identifying materials.

2. Bert Embedding Layer:

- a. Same hyperparameters as above, we have **frozen** the ResNet model since introducing the embedding layer only affects the output of the decoder.
- b. Plots for Training and Validation Loss and Accuracy





c. --- Evaluating on Test Set A ---

```
{"loss": 0.49355383441289263, "metrics": {"accuracy": 0.7708533333333333, "precision": 0.6594424064642846, "recall": 0.6301100304404832, "f1": 0.6369727380130688}}
```

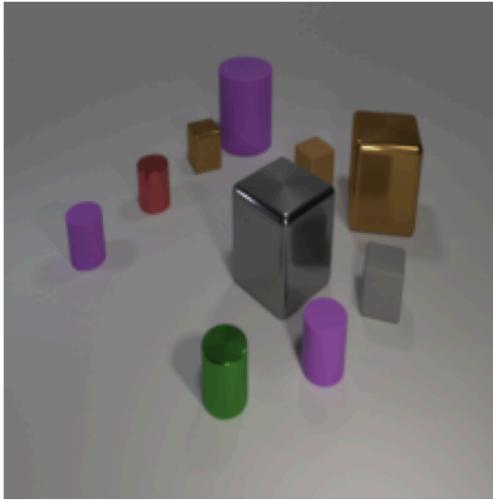
d. **Visualization:** Images, Question, Answer, Predictions

Incorrect	Correct
<p>Q: what number of things are either big cyan shiny things or large things that are in front of the large metallic thing?</p> <p>A: 1 Pred: 2</p>	<p>Q: how many metal objects are either cyan cylinders or tiny purple things?</p> <p>A: 1 Pred: 1</p>

Q: how many blocks are either large objects or gray rubber objects?

A: 3

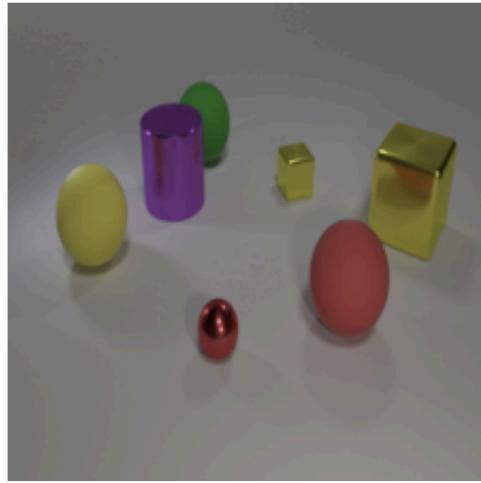
Pred: 2



Q: the large matte thing that is the same color as the tiny sphere is what shape?

A: sphere

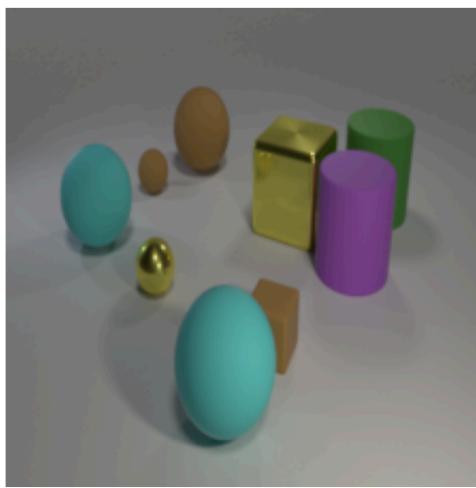
Pred: sphere



Q: there is a large thing that is the same color as the tiny metallic sphere ; what is its material?

A: metal

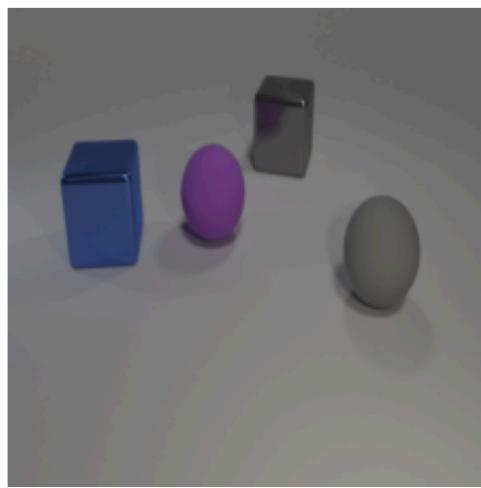
Pred: rubber



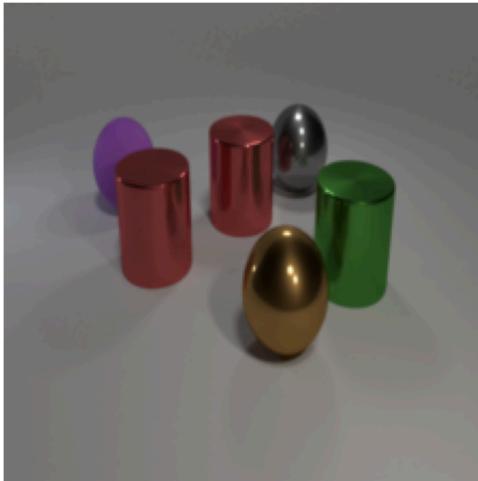
Q: what number of matte things are either big gray objects or purple balls?

A: 2

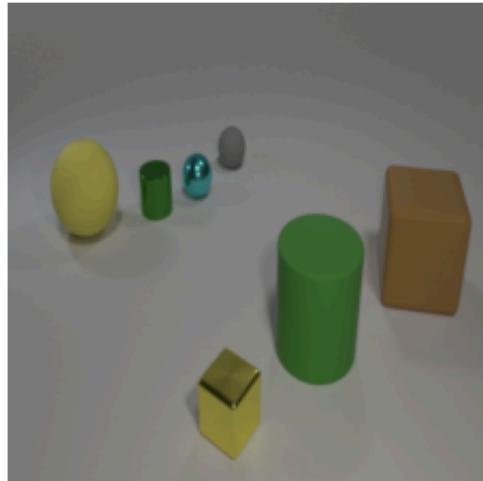
Pred: 2



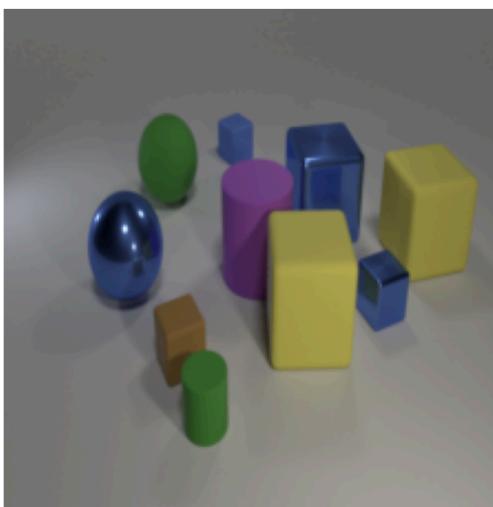
Q: there is a large metallic sphere on the left side of the large gray thing ; what color is it?
A: brown
Pred: gray



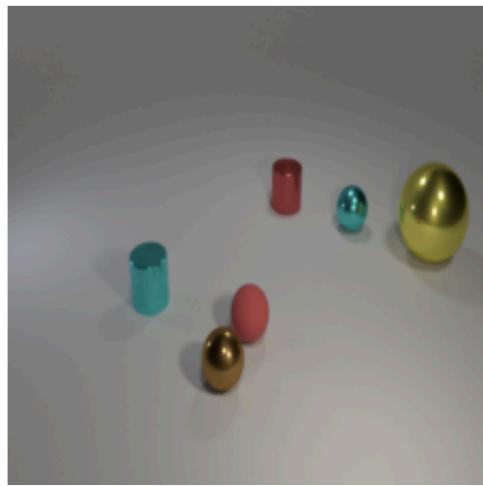
Q: do the cylinder in front of the brown matte block and the small metal cylinder have the same color?
A: yes
Pred: yes



Q: how many small objects are rubber cylinders or metallic things?
A: 2
Pred: 3



Q: there is a rubber thing that is the same size as the brown metal thing ; what color is it?
A: red
Pred: red



- e. Questions with more than 9 shapes that are packed closely(3,5) are hard to answer.

Improvement Due to Two Strategies

c) Focal Loss Improvement

b) Bert Embedding Improvement

b)Bert Embedding Improvement

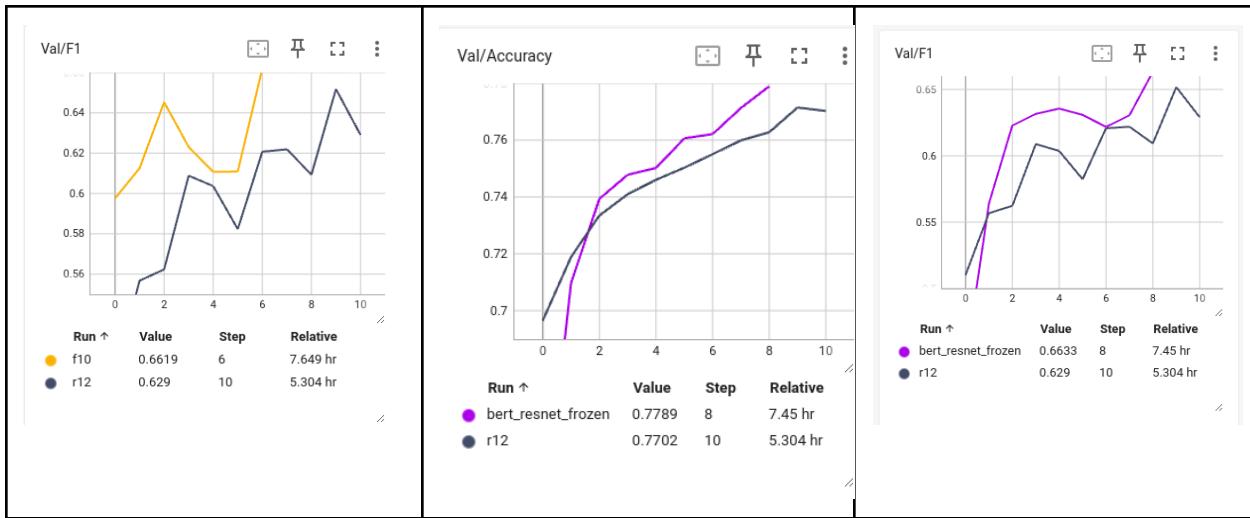


Figure Illustrating the Improvement of F1 metric when using Focal Loss, similarly we show the improvement of accuracy and F1 of the model when fine tuned with Bert Embedding.

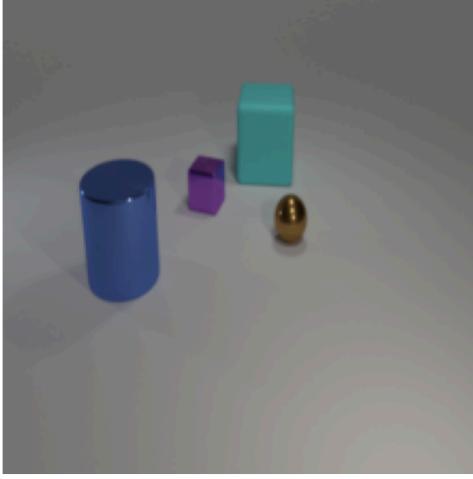
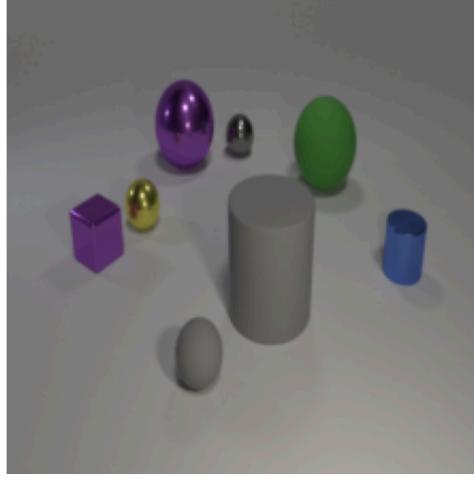
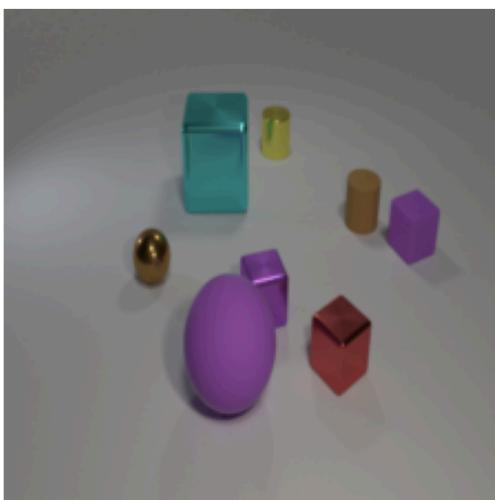
Checkpoint at step 6 was used as a starting point to train both these models.

Question 11: Zero Shot Evaluation

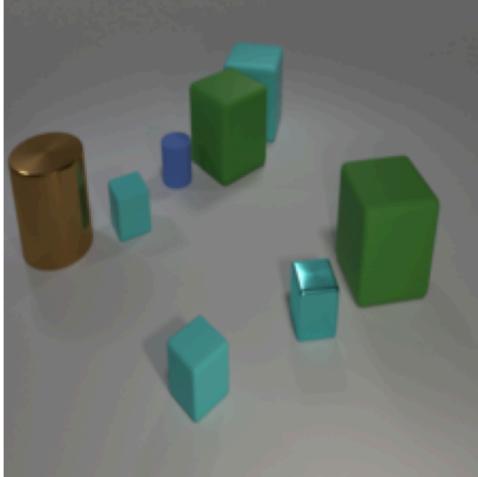
Best Model Metrics

Dataset	Avg Loss	Accuracy	Precision	Recall	F1-score
Validation	0.4899	0.7703	0.6512	0.6213	0.6291
Test	0.4954	0.7707	0.6800	0.6258	0.6350
TestB	1.6795	0.6556	0.4894	0.4718	0.4765
Train (for eval comparison)	0.4139	0.8068	0.6698	0.6416	0.6507

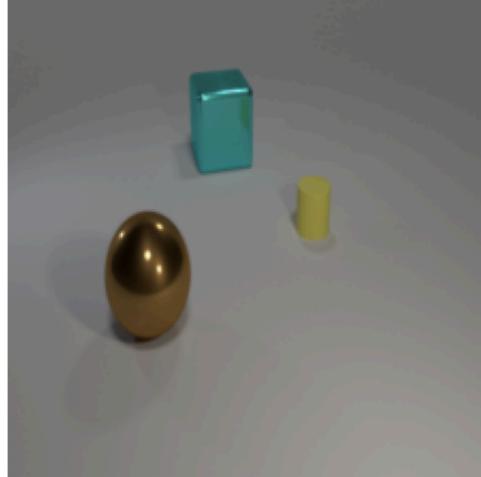
- a) Precision, Recall and F1 suffer heavily when doing evaluation on Test B in the above table.
- b) The table shows the sample predictions. Observations:
 - i) In the first figure, the object corresponding to the correct answer is a cube. In the original dataset cubes were brown, thus it seems that model is conflating brown with cube leading to wrong answer(brown instead of purple).
 - ii) Similarly in the third example, the model is conflating the colors cyan and blue with cylinder identity.

Incorrect	Correct
<p>Q: there is another small thing that is the same shape as the matte thing ; what color is it? A: purple Pred: brown</p> 	<p>Q: how many small blue shiny objects are in front of the large rubber thing that is in front of the small metal cylinder? A: 0 Pred: 0</p> 
<p>Q: how many red things are the same material as the red block? A: 0 Pred: 1</p> 	<p>Q: are the cylinder and the cyan object to the left of the red object made of the same material? A: no Pred: no</p> 

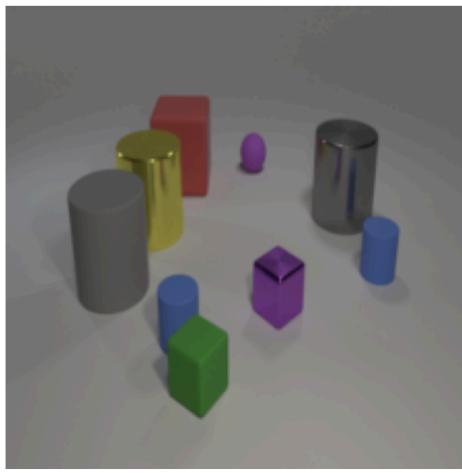
Q: is there anything else that has the same color as the matte cylinder?
A: no
Pred: yes



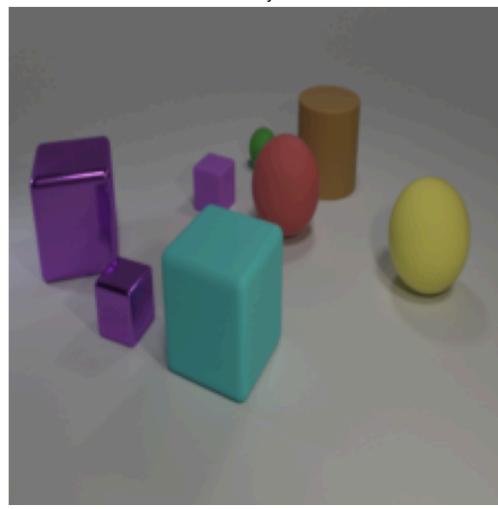
Q: are there any other brown objects that have the same shape as the brown thing?
A: no
Pred: no



Q: is the number of blue objects that are in front of the purple sphere the same as the number of green cubes on the left side of the large red cube?
A: no
Pred: yes



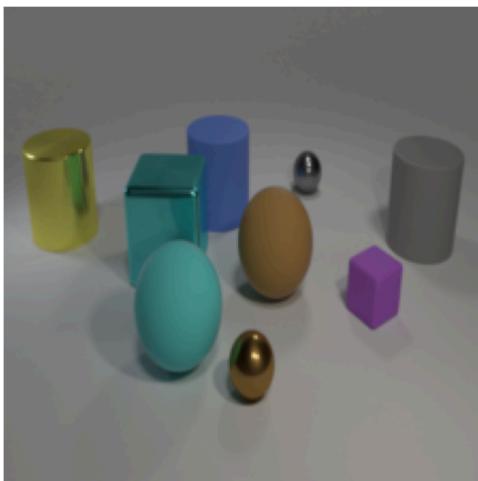
Q: do the brown cylinder and the red ball have the same material?
A: yes
Pred: yes



Q: how many blue matte objects are the same shape as the yellow metallic object?

A: 1

Pred: 0



Q: the metallic thing that is the same color as the rubber cylinder is what shape?

A: cylinder

Pred: cylinder

