# Question and Model Prediction

Model Used: ViLT  
Question Asked: How many cars are there?  
Prediction: 2

# Prediction Details

2 0.96228  
3 0.02562  
1 0.00970  
4 0.00160  
5 0.00029

# User Settings

Camera Effect:

# Original Image



# Model Visualizations

## Combined Attention Patches.png



## Token 0 - '[CLS]'.png



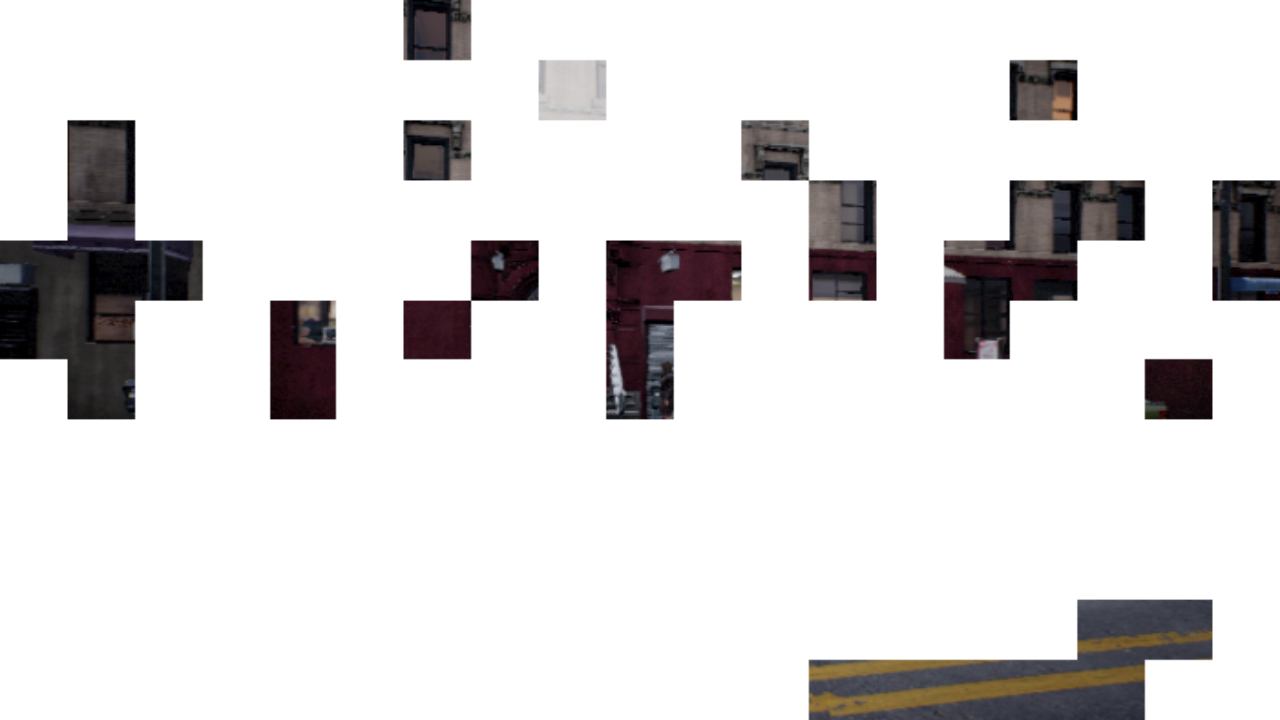
## Token 1 - 'how'.png



## Token 2 - 'many'.png



## Token 3 - 'cars'.png



## Token 4 - 'are'.png



## Token 5 - 'there'.png



## Token 6 - '[SYMBOL - Question Mark]'.png



## Token 7 - '[SEP]'.png

