

# Attempts to Interpret a Neural Network

# Attempts to Interpret a Neural Network

We can

- **Understand the model architecture**
- Visualize the filters / weights
- Extract the output of intermediate neurons / layers
- Locate important parts of the image according to the model

# Attempt 1 - Understanding the Model Architecture

Model: "sequential\_1"

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 218, 218, 32)	4736
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 32)	0
conv2d_2 (Conv2D)	(None, 48, 48, 32)	50208
max_pooling2d_2 (MaxPooling2D)	(None, 12, 12, 32)	0
conv2d_3 (Conv2D)	(None, 6, 6, 32)	50208
max_pooling2d_3 (MaxPooling2D)	(None, 1, 1, 32)	0
flatten_1 (Flatten)	(None, 32)	0
preds (Dense)	(None, 1)	33
Total params: 105,185		
Trainable params: 105,185		
Non-trainable params: 0		

# Attempt 1 - Understanding the Model Architecture

Model: "sequential\_1"

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 218, 218, 32)	4736
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 32)	0
conv2d_2 (Conv2D)	(None, 48, 48, 32)	50208
max_pooling2d_2 (MaxPooling2D)	(None, 12, 12, 32)	0
conv2d_3 (Conv2D)	(None, 6, 6, 32)	50208
max_pooling2d_3 (MaxPooling2D)	(None, 1, 1, 32)	0
flatten_1 (Flatten)	(None, 32)	0
preds (Dense)	(None, 1)	33
Total params: 105,185		
Trainable params: 105,185		
Non-trainable params: 0		

# Attempt 1 - Understanding the Model Architecture

Model: "sequential\_1"

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 218, 218, 32)	4736
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 32)	0
conv2d_2 (Conv2D)	(None, 48, 48, 32)	50208
max_pooling2d_2 (MaxPooling2D)	(None, 12, 12, 32)	0
conv2d_3 (Conv2D)	(None, 6, 6, 32)	50208
max_pooling2d_3 (MaxPooling2D)	(None, 1, 1, 32)	0
flatten_1 (Flatten)	(None, 32)	0
preds (Dense)	(None, 1)	33
Total params: 105,185		
Trainable params: 105,185		
Non-trainable params: 0		

# Attempt 1 - Understanding the Model Architecture

Model: "sequential\_1"

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 218, 218, 32)	4736
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 32)	0
conv2d_2 (Conv2D)	(None, 48, 48, 32)	50208
max_pooling2d_2 (MaxPooling2D)	(None, 12, 12, 32)	0
conv2d_3 (Conv2D)	(None, 6, 6, 32)	50208
max_pooling2d_3 (MaxPooling2D)	(None, 1, 1, 32)	0
flatten_1 (Flatten)	(None, 32)	0
preds (Dense)	(None, 1)	33
Total params: 105,185		
Trainable params: 105,185		
Non-trainable params: 0		

# Attempt 1 - Understanding the Model Architecture

Model: "sequential\_1"

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 218, 218, 32)	4736
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 32)	0
conv2d_2 (Conv2D)	(None, 48, 48, 32)	50208
max_pooling2d_2 (MaxPooling2D)	(None, 12, 12, 32)	0
conv2d_3 (Conv2D)	(None, 6, 6, 32)	50208
max_pooling2d_3 (MaxPooling2D)	(None, 1, 1, 32)	0
flatten_1 (Flatten)	(None, 32)	0
preds (Dense)	(None, 1)	33

Total params: 105,185  
Trainable params: 105,185  
Non-trainable params: 0

# Attempt 1 - Understanding the Model Architecture

Model: "sequential\_1"

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 218, 218, 32)	4736
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 32)	0
conv2d_2 (Conv2D)	(None, 48, 48, 32)	50208
max_pooling2d_2 (MaxPooling2D)	(None, 12, 12, 32)	0
conv2d_3 (Conv2D)	(None, 6, 6, 32)	50208
max_pooling2d_3 (MaxPooling2D)	(None, 1, 1, 32)	0
flatten_1 (Flatten)	(None, 32)	0
preds (Dense)	(None, 1)	33

Total params: 105,185  
Trainable params: 105,185  
Non-trainable params: 0





Thank You