

# Visualization vs. Interpretability

## 1. “CNN” **Model** Summary

June 11th, 2025

Lab 1. CAM (Self-defined, sequential)

Model: "sequential"

Dense  
units = 10

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 28, 28, 16)	160
max_pooling2d (MaxPooling2D)	(None, 14, 14, 16)	0
conv2d_1 (Conv2D)	(None, 14, 14, 32)	4,640
max_pooling2d_1 (MaxPooling2D)	(None, 7, 7, 32)	0
conv2d_2 (Conv2D)	(None, 7, 7, 64)	18,496
max_pooling2d_2 (MaxPooling2D)	(None, 3, 3, 64)	0
conv2d_3 (Conv2D)	(None, 3, 3, 128)	73,856
global_average_pooling2d (GlobalAveragePooling2D)	(None, 128)	0
dense (Dense)	(None, 10)	1,290

Total params: 98,442 (384.54 KB)  
Trainable params: 98,442 (384.54 KB)  
Non-trainable params: 0 (0.00 B)

Model: "functional\_9"

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	(None, 28, 28, 1)	0	-
conv2d (Conv2D)	(None, 28, 28, 16)	160	input_layer[0][0... input_layer[0][0]
max_pooling2d (MaxPooling2D)	(None, 14, 14, 16)	0	conv2d[6][0], conv2d[8][0]
conv2d_1 (Conv2D)	(None, 14, 14, 32)	4,640	max_pooling2d[5]... max_pooling2d[7]...
max_pooling2d_1 (MaxPooling2D)	(None, 7, 7, 32)	0	conv2d_1[4][0], conv2d_1[6][0]
conv2d_2 (Conv2D)	(None, 7, 7, 64)	18,496	max_pooling2d_1[... max_pooling2d_1[...
max_pooling2d_2 (MaxPooling2D)	(None, 3, 3, 64)	0	conv2d_2[2][0], conv2d_2[4][0]
conv2d_3 (Conv2D)	(None, 3, 3, 128)	73,856	max_pooling2d_2[... max_pooling2d_2[...
global_average_poo... (GlobalAveragePool...	(None, 128)	0	conv2d_3[2][0]
dense (Dense)	(None, 10)	1,290	global_average_p...

Total params: 98,442 (384.54 KB)  
Trainable params: 98,442 (384.54 KB)  
Non-trainable params: 0 (0.00 B)

Loss l: sparse\_categorical\_crossentropy

Label format: Integer (sparse)

1D array of shape (batch\_size, )

Lab 2. CAM (Self-defined, sequential)

Dense  
units = 1

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 300, 300, 16)	448
max_pooling2d (MaxPooling2D)	(None, 150, 150, 16)	0
conv2d_1 (Conv2D)	(None, 150, 150, 32)	4,640
max_pooling2d_1 (MaxPooling2D)	(None, 75, 75, 32)	0
conv2d_2 (Conv2D)	(None, 75, 75, 64)	18,496
max_pooling2d_2 (MaxPooling2D)	(None, 37, 37, 64)	0
conv2d_3 (Conv2D)	(None, 37, 37, 128)	73,856
global_average_pooling2d (GlobalAveragePooling2D)	(None, 128)	0
dense (Dense)	(None, 1)	129

Total params: 97,569 (381.13 KB)  
Trainable params: 97,569 (381.13 KB)  
Non-trainable params: 0 (0.00 B)

Model: "functional\_9"

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	(None, 300, 300, 3)	0	-
conv2d (Conv2D)	(None, 300, 300, 16)	448	input_layer[0][0... input_layer[0][0]
max_pooling2d (MaxPooling2D)	(None, 150, 150, 16)	0	conv2d[6][0], conv2d[8][0]
conv2d_1 (Conv2D)	(None, 150, 150, 32)	4,640	max_pooling2d[5]... max_pooling2d[7]...
max_pooling2d_1 (MaxPooling2D)	(None, 75, 75, 32)	0	conv2d_1[4][0], conv2d_1[6][0]
conv2d_2 (Conv2D)	(None, 75, 75, 64)	18,496	max_pooling2d_1[... max_pooling2d_1[...
max_pooling2d_2 (MaxPooling2D)	(None, 37, 37, 64)	0	conv2d_2[2][0], conv2d_2[4][0]
conv2d_3 (Conv2D)	(None, 37, 37, 128)	73,856	max_pooling2d_2[... max_pooling2d_2[...
global_average_pooling2d (GlobalAveragePooling2D)	(None, 128)	0	conv2d_3[2][0]
dense (Dense)	(None, 1)	129	global_average_p...

Total params: 97,569 (381.13 KB)  
Trainable params: 97,569 (381.13 KB)  
Non-trainable params: 0 (0.00 B)

Loss III: binary\_crossentropy

Label format: multi-hot encoded

2D array of shape (batch\_size, num\_classes)



# Lab 3. Saliency (Self-defined, sequential)

Dense units = 2

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 300, 300, 16)	448
max_pooling2d (MaxPooling2D)	(None, 150, 150, 16)	0
conv2d_1 (Conv2D)	(None, 150, 150, 32)	4,640
max_pooling2d_1 (MaxPooling2D)	(None, 75, 75, 32)	0
conv2d_2 (Conv2D)	(None, 75, 75, 64)	18,496
max_pooling2d_2 (MaxPooling2D)	(None, 37, 37, 64)	0
conv2d_3 (Conv2D)	(None, 37, 37, 128)	73,856
global_average_pooling2d (GlobalAveragePooling2D)	(None, 128)	0
dense (Dense)	(None, 2)	258

Total params: 97,698 (381.63 KB)  
Trainable params: 97,698 (381.63 KB)  
Non-trainable params: 0 (0.00 B)

Assignment use this model too~

Loss II: categorical\_crossentropy

Label format: One-hot encoded  
2D array of shape (batch\_size, num\_classes)

# Lab 4. GradCAM (VGG16)

Model: "functional"

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1,792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36,928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73,856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147,584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295,168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590,080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1,180,160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0

block5_conv1 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 512)	0
dense (Dense)	(None, 2)	1,026

Total params: 14,715,714 (56.14 MB)

Trainable params: 7,080,450 (27.01 MB)

Non-trainable params: 7,635,264 (29.13 MB)

Loss l: `sparse_categorical_crossentropy`

Label format: Integer (sparse)

1D array of shape (batch\_size, )

# Visualization vs. Interpretability

2. **Review.** Label format vs. Last Dense Layer # of Units  
Label format vs. **Loss**

June 11th, 2025

# Loss - 1. `sparse_categorical_crossentropy`

- You have **single-label** classification
- Labels are given as **integer class indices** (not one-hot)
- Predicted probabilities: [ 0.1, 0.2, 0.7 ]

$$\text{loss} = -\log(p_{\text{true class}})$$

# Loss - 2. categorical\_crossentropy

- You have **single-label** classification
- Labels are **one-hot encoded**
- One-hot label: [ 0 , 0 , 1 ]
- Predicted probabilities: [ 0 . 1 , 0 . 2 , 0 . 7 ]

$$\text{loss} = - \sum_{i=1}^N y_i \log(p_i)$$



# Loss - 3. binary\_crossentropy

- You have **multi-label classification** (multi-hot labels)
- Each class is treated as a **separate binary problem** (present or not)
- Multi-hot label: [ 1 , 0 , 1 ]
- Prediction: [0.9, 0.2, 0.8]

$$\text{loss} = - \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

# Loss vs. Label Dimension

## Summary Table

Task Type	Label Format	Label Shape	Loss Function	
Single-label	Integer	<code>(batch_size, )</code>	<code>sparse_categorical_crossentropy</code>	
Single-label	One-hot	<code>(batch_size, num_classes)</code>	<code>categorical_crossentropy</code>	
Multi-label	Multi-hot	<code>(batch_size, num_classes)</code>	<code>binary_crossentropy</code>	