

# Feature representation and learning

---

Dr. Tushar Sandhan

# Contents

- Different types of data features
- How computer perception work?
- What is feature representation
  - Computer Vision (CV) Features
  - Audio Features
  - Natural Language Processing (NLP) Features

- Feature Representation

- Input layer
- Convolutional layer
- ReLu
- Softmax
- Pooling layers
- Flatten layer

CNN

probability vector

final feature representation

end to end. =

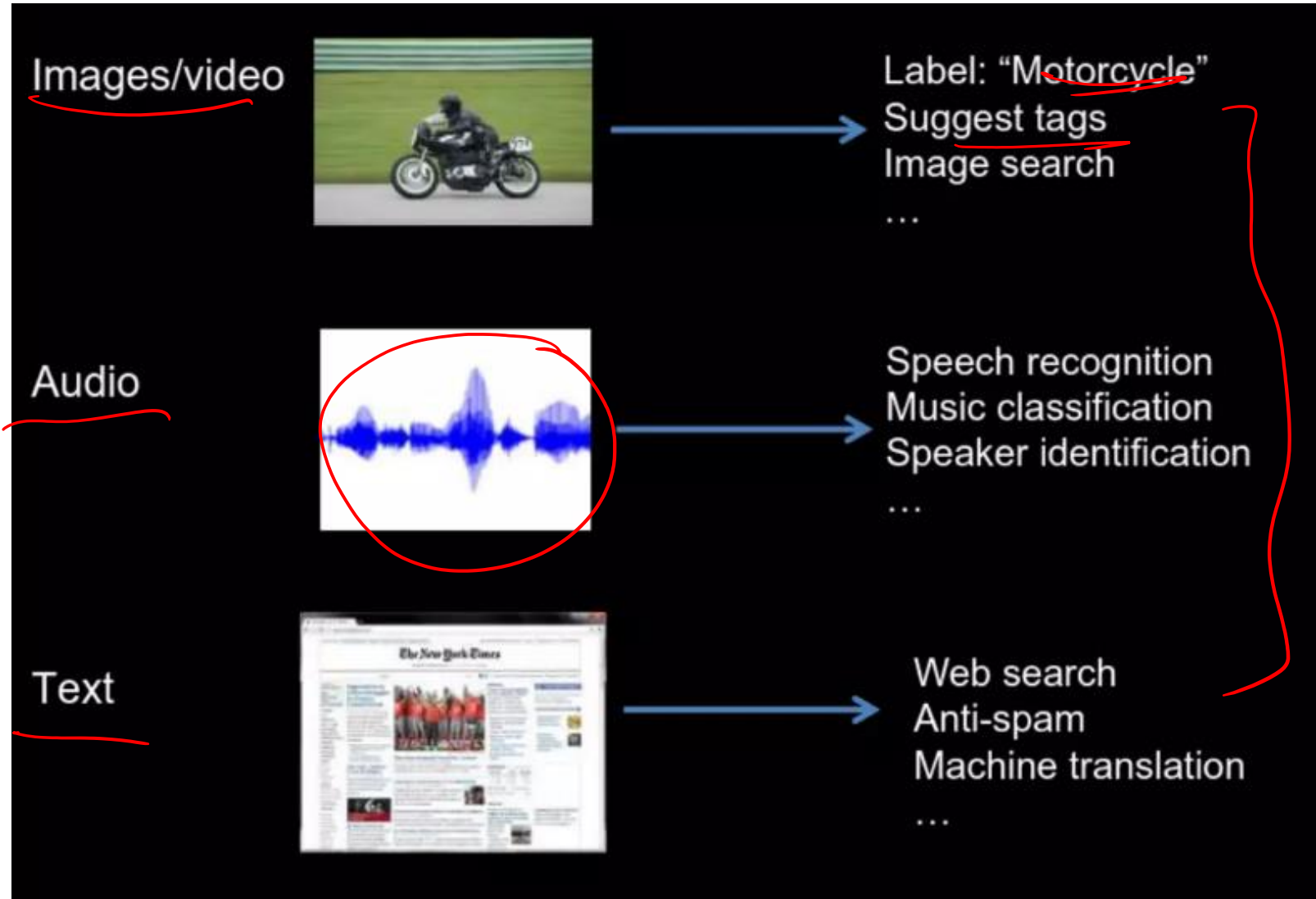
flatten → N.N.  
(classifier or regressor)

Save them →

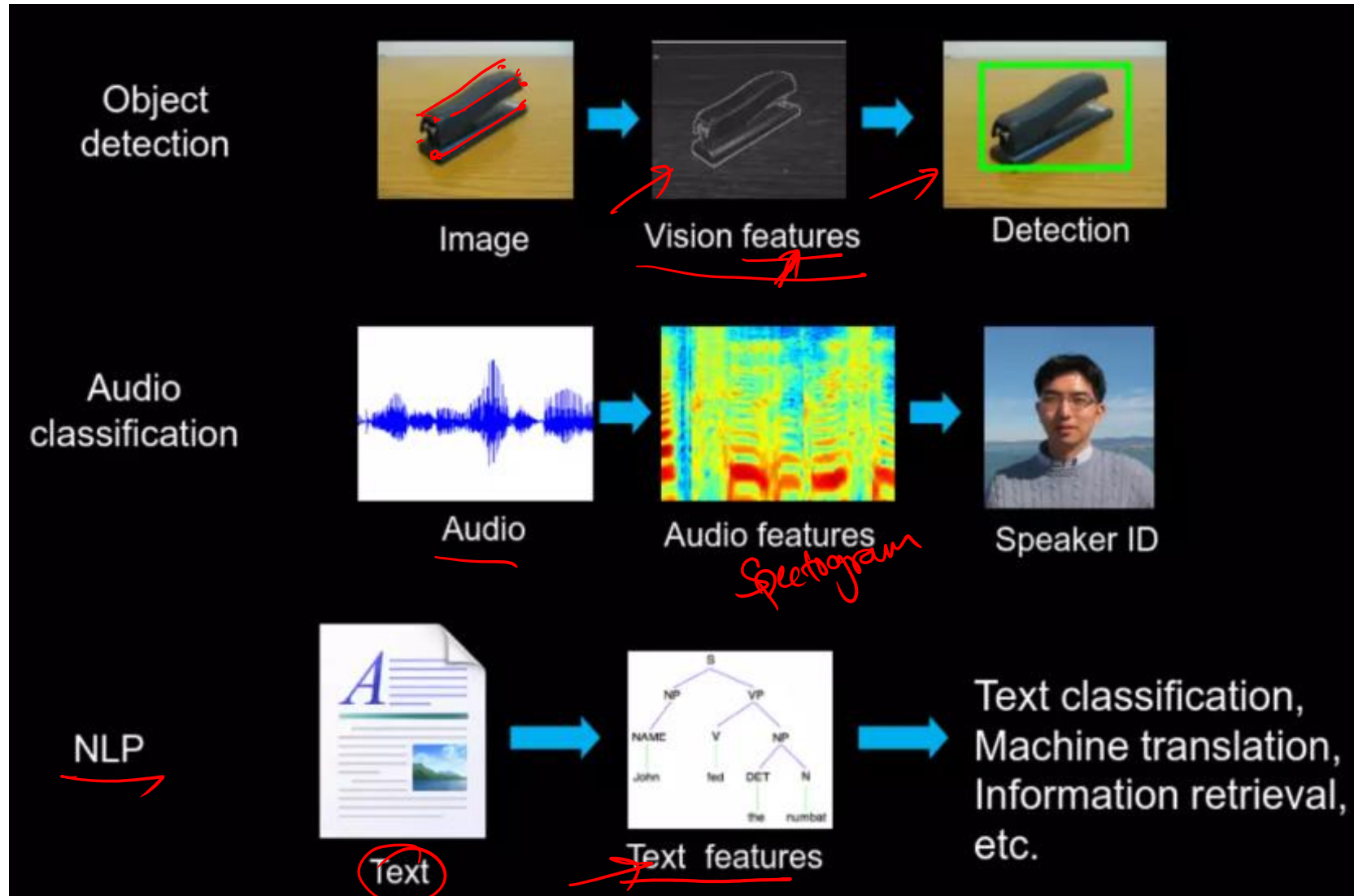
Use any M.L.

inp data represented as features  
imp. characteristics of data

# Different types of data features



# How computer perception work?

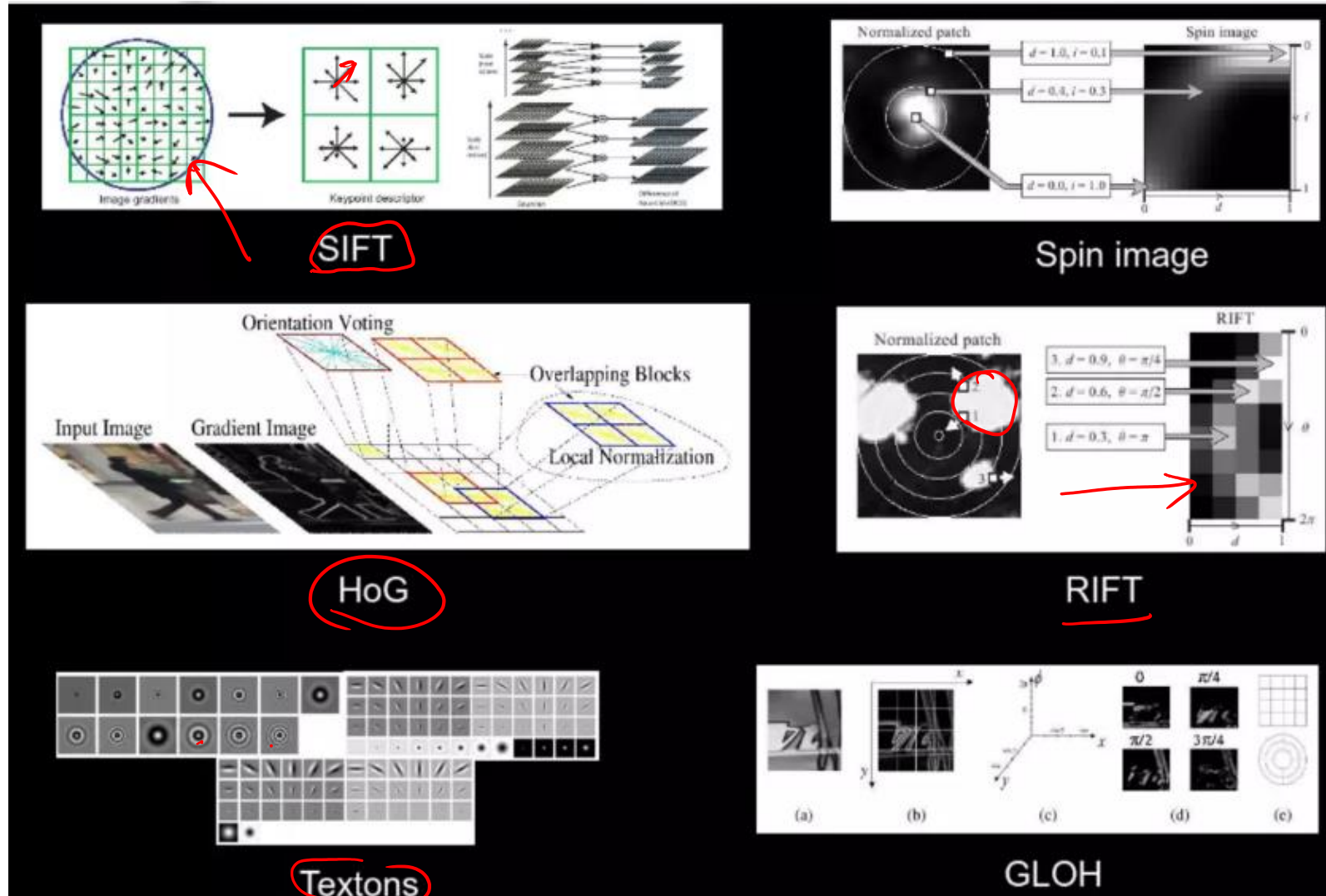


# What is feature representation

---

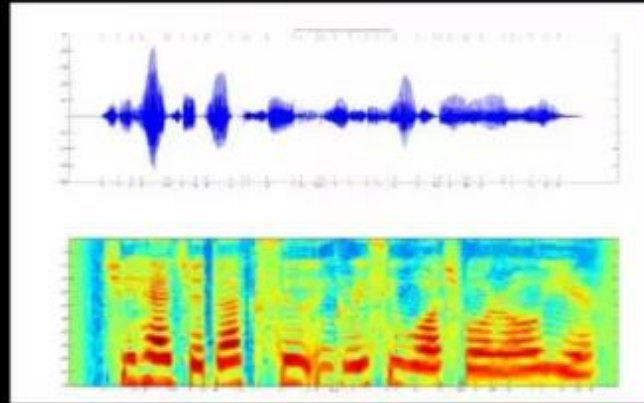
- In a deep learning architecture, the output of each intermediate layer can be viewed as a representation of the original input data. ✓
- Each level uses the representation produced by previous level as input, and produces new representations as output, which is then fed to higher levels. ✓

# Feature Representation: CV Features

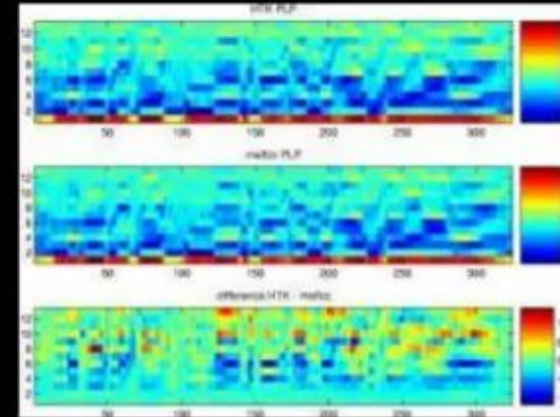




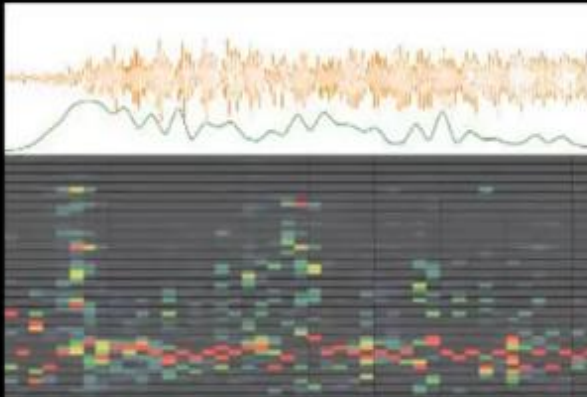
# Feature Representation: Audio Features



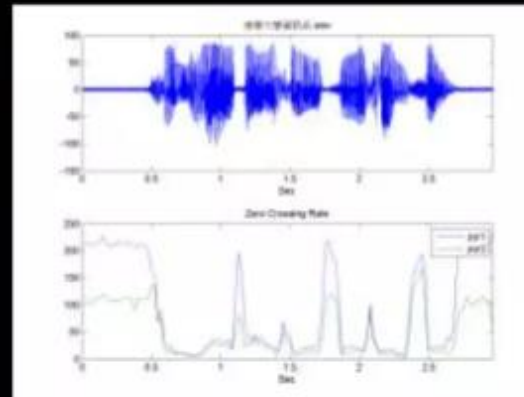
Spectrogram



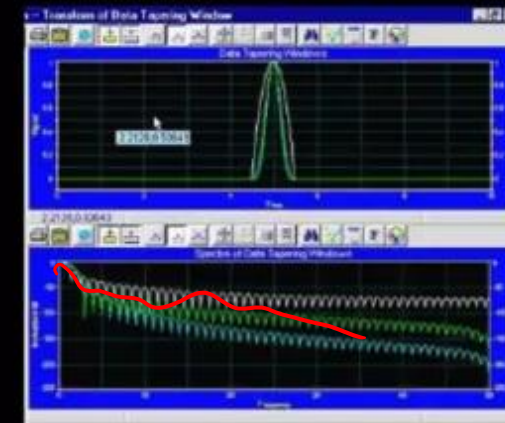
MFCC



Flux

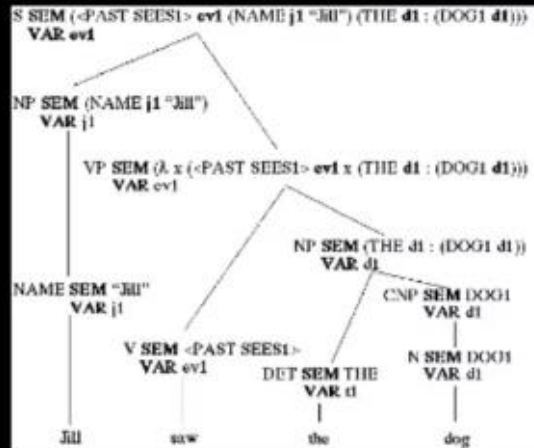


ZCR



Rolloff

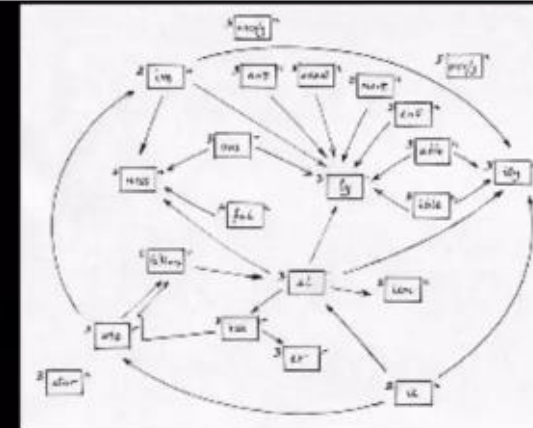
# Feature Representation: NLP Features



Parser features

```
<DOC>
<DOCID> ws94.008.0212 </DOCID>
<DOCHD> 940413-0062 </DOCHD>
<CHL> Who's News:
$ Burns Fry Ltd. </REL>
<DD> 04/13/94 </DD>
<SO> WALL STREET JOURNAL (J), PAGE B10 </SO>
<CO> MEK </CO>
<IND> SECURITIES (SCH) </IND>
<TXT>
<p>
  Burns Fry Ltd. (Burns Fry) -- Donald Wright, 46 years old, was
  named executive vice president and director of fixed income at this
  brokerage firm. Mr. Wright resigned as president of Merrill Lynch
  Canada Inc., a unit of Merrill Lynch & Co., to succeed Mark
  Bassard, 49, who left Burns Fry last month. A Merrill Lynch
  spokesman said it hasn't named a successor to Mr. Wright, who is
  expected to begin his new position by the end of the month.
</p>
</TXT>
</DOC>
```

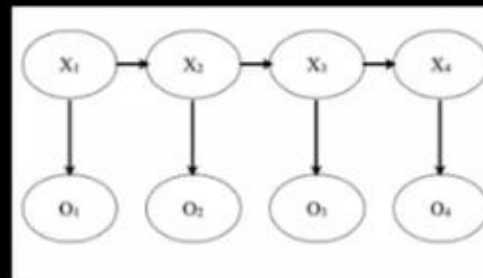
NER/SRL



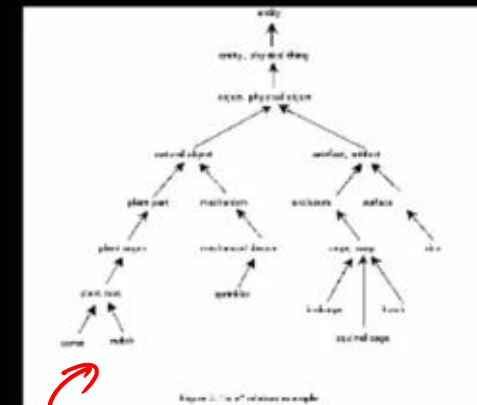
Stemming

His father, Nick Begich, won an election posthumously, only they didn't know for sure that it was posthumous because his plane just disappeared. It still hasn't turned up. It's why locators are now required in all US planes.

Anaphora



POS tagging

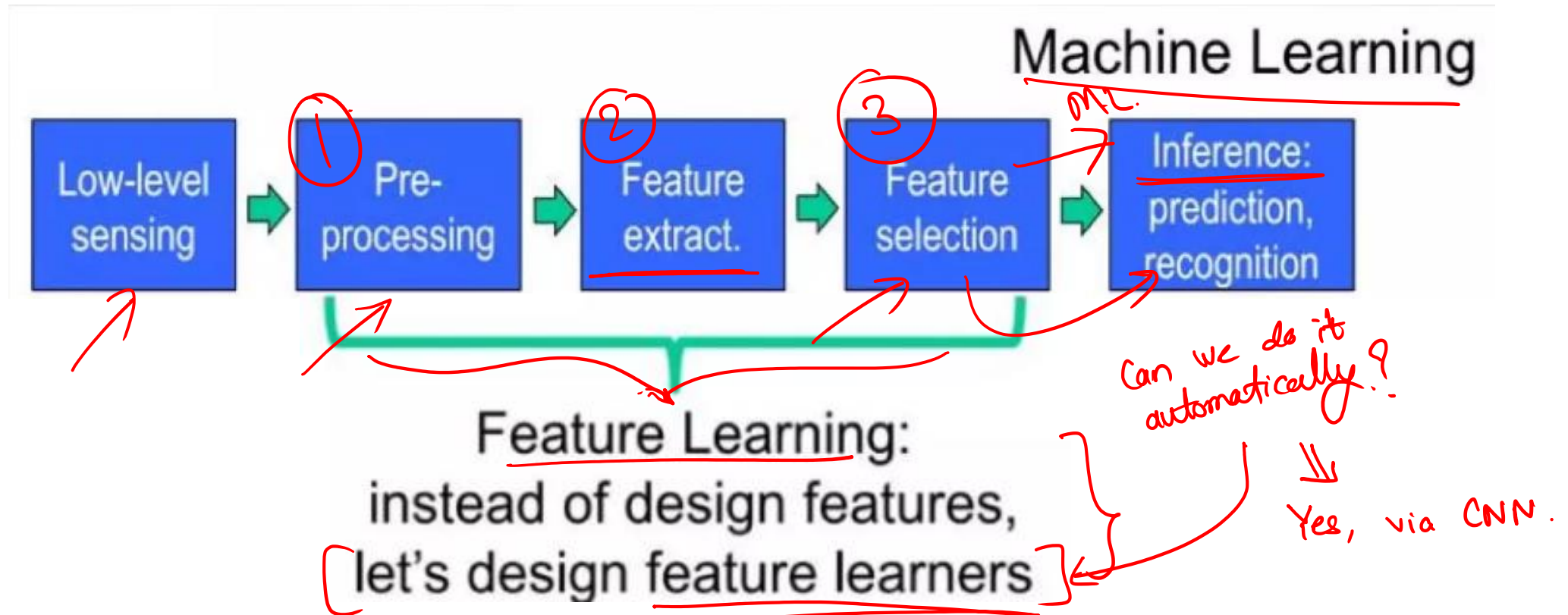


WordNet features



# Feature Representation

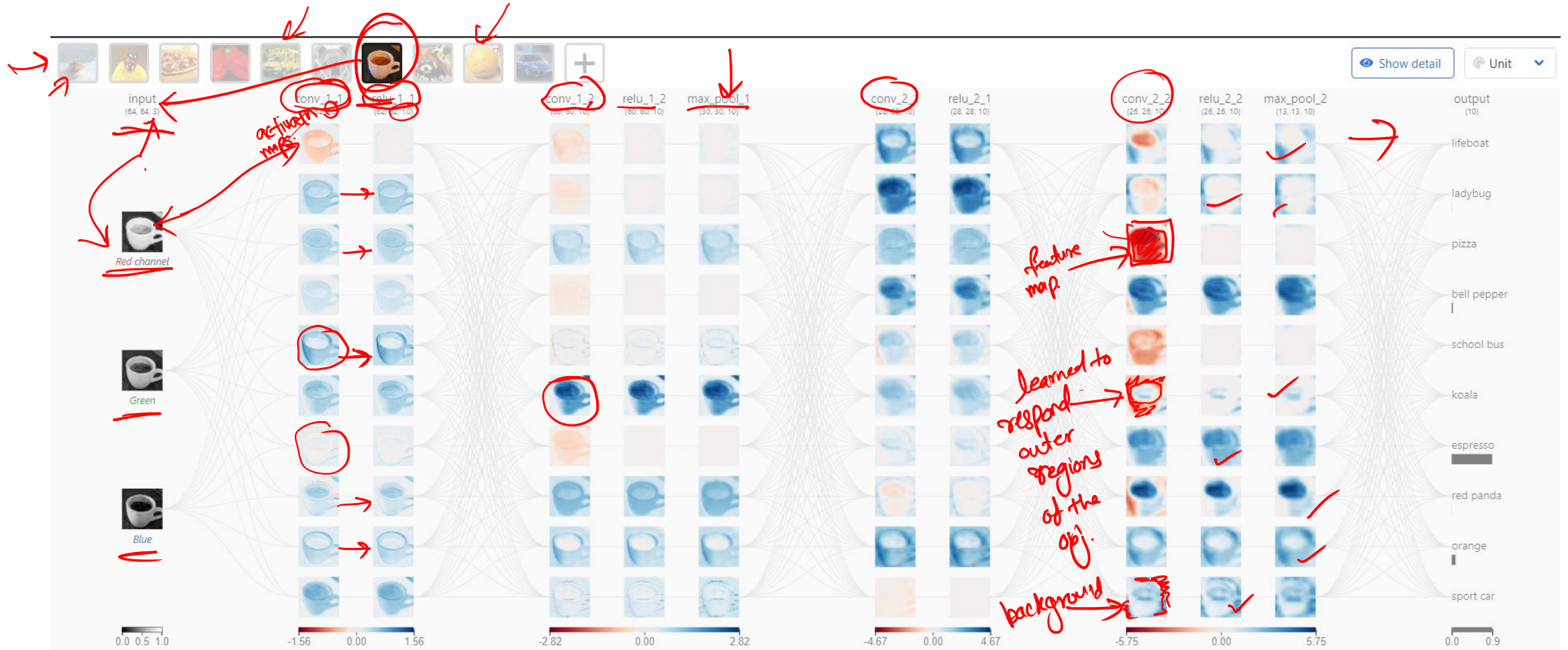
- Certainly, Coming up with features is difficult, time-consuming and requires expert knowledge.
- A lot of time is spent tuning the features which are often hand-crafted.



# Feature Representation

## ■ CNN Explainer

ResNet 50



# Feature Representation

---

## Input Layer

- The input layer (leftmost layer) represents the input image into the CNN. It uses RGB images as input, the input layer has three channels, corresponding to the red, green, and blue channels respectively.

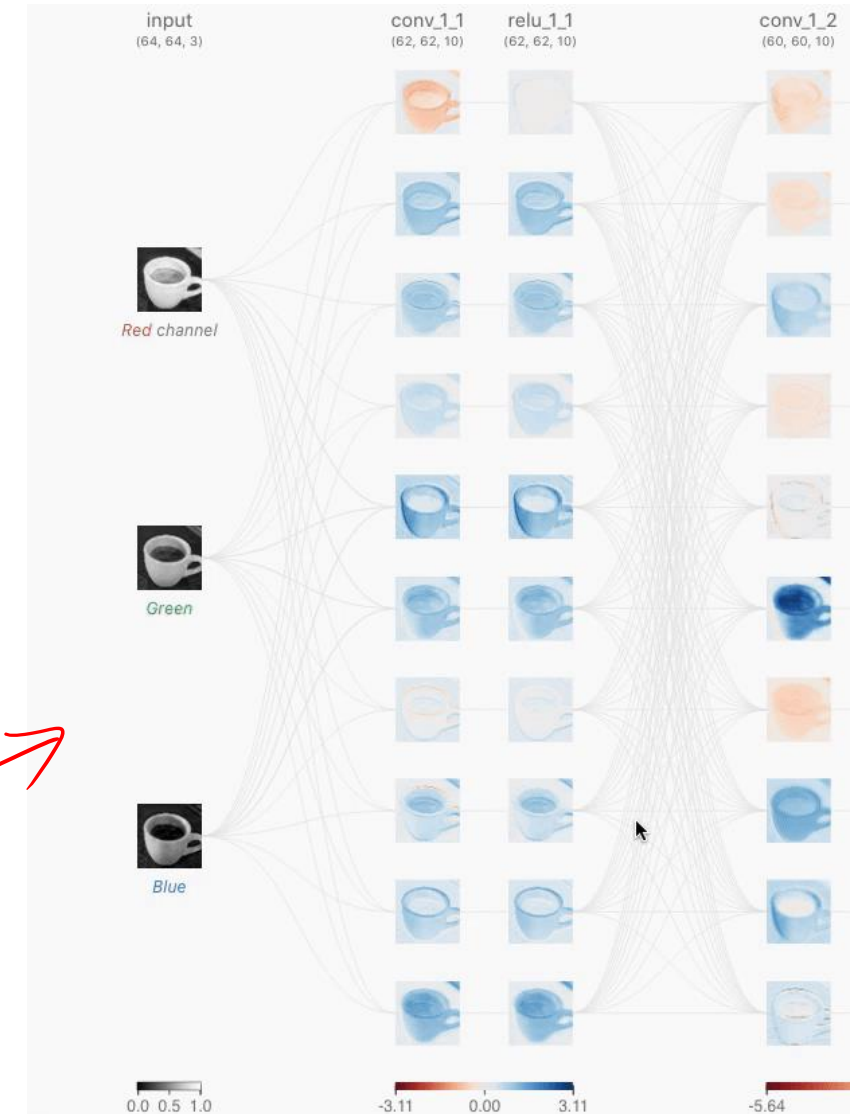
## Convolutional Layers

- The convolutional layers are the foundation of CNN, as they contain the learned kernels (weights), which extract features that distinguish different images from one another—this is what we want for classification
- The convolutional neuron performs an elementwise dot product with a unique kernel and the output of the previous layer's corresponding neuron. This will yield as many intermediate results as there are unique kernels. The convolutional neuron is the result of all of the intermediate results summed together with the learned bias.

# Feature Representation

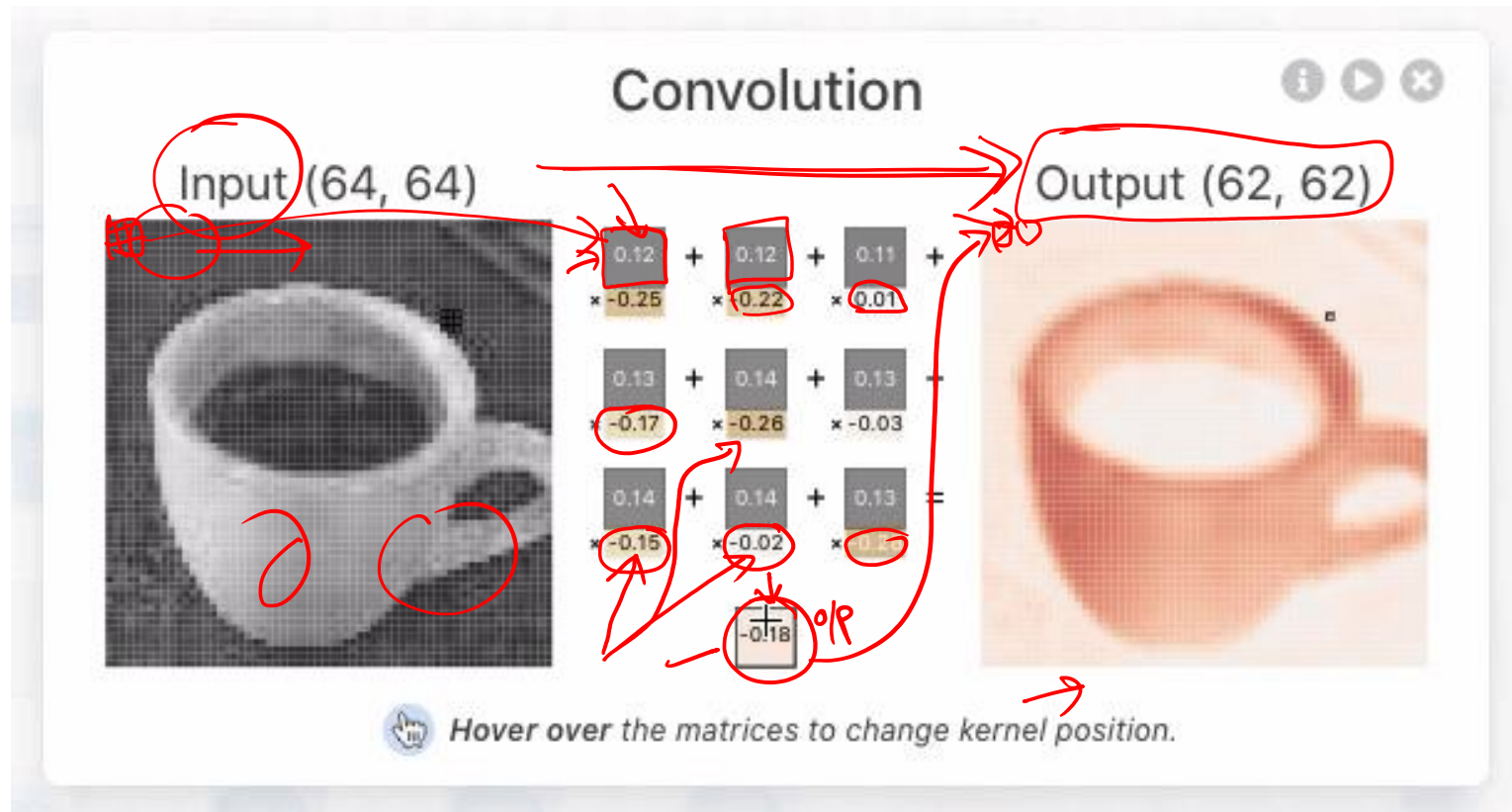
## Convolutional Layer

- The convolutional neuron performs an elementwise dot product with a unique kernel and the output of the previous layer's corresponding neuron. This will yield as many intermediate results as there are unique kernels. The convolutional neuron is the result of all of the intermediate results summed together with the learned bias.
- As you hover over the activation map of the topmost node from the first convolutional layer, you can see that 3 kernels were applied to yield this activation map.
- After clicking this activation map, you can see the convolution operation occurring with each unique kernel.



# Feature Representation

- The size of these kernels is a hyper-parameter specified by the designers of the network architecture. In order to produce the output of the convolutional neuron (activation map), we must perform an elementwise dot product with the output of the previous layer and the unique kernel learned by the network.

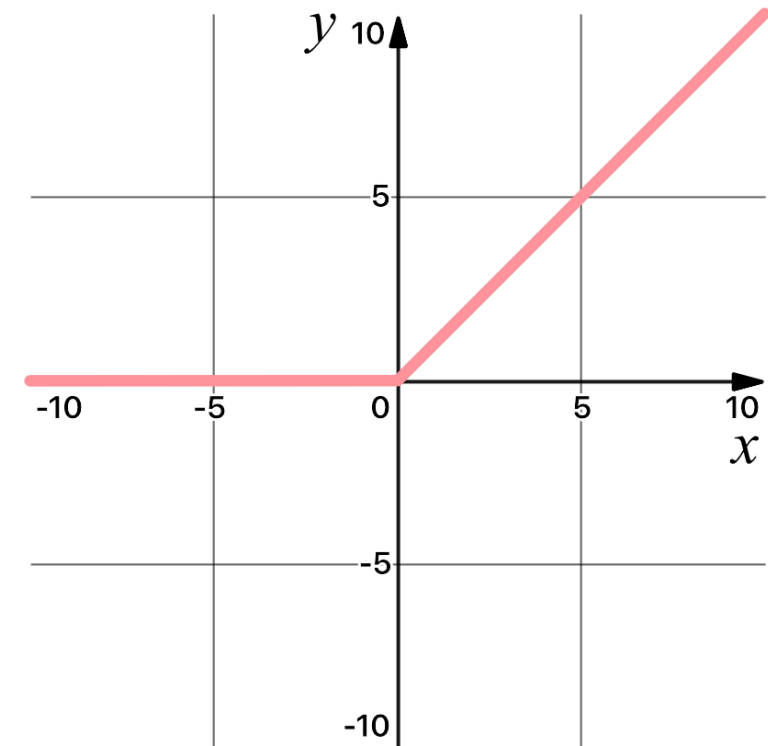




# Feature Representation

## ReLU

- The ReLU activation function is specifically used as a non-linear activation function, as opposed to other non-linear functions such as Sigmoid because it has been empirically observed that CNNs using ReLU are faster to train than their counterparts.
- This activation function is applied elementwise on every value from the input tensor. For example, if applied ReLU on the value 2.24, the result would be 2.24, since 2.24 is larger than 0.





# Feature Representation

---

## Softmax

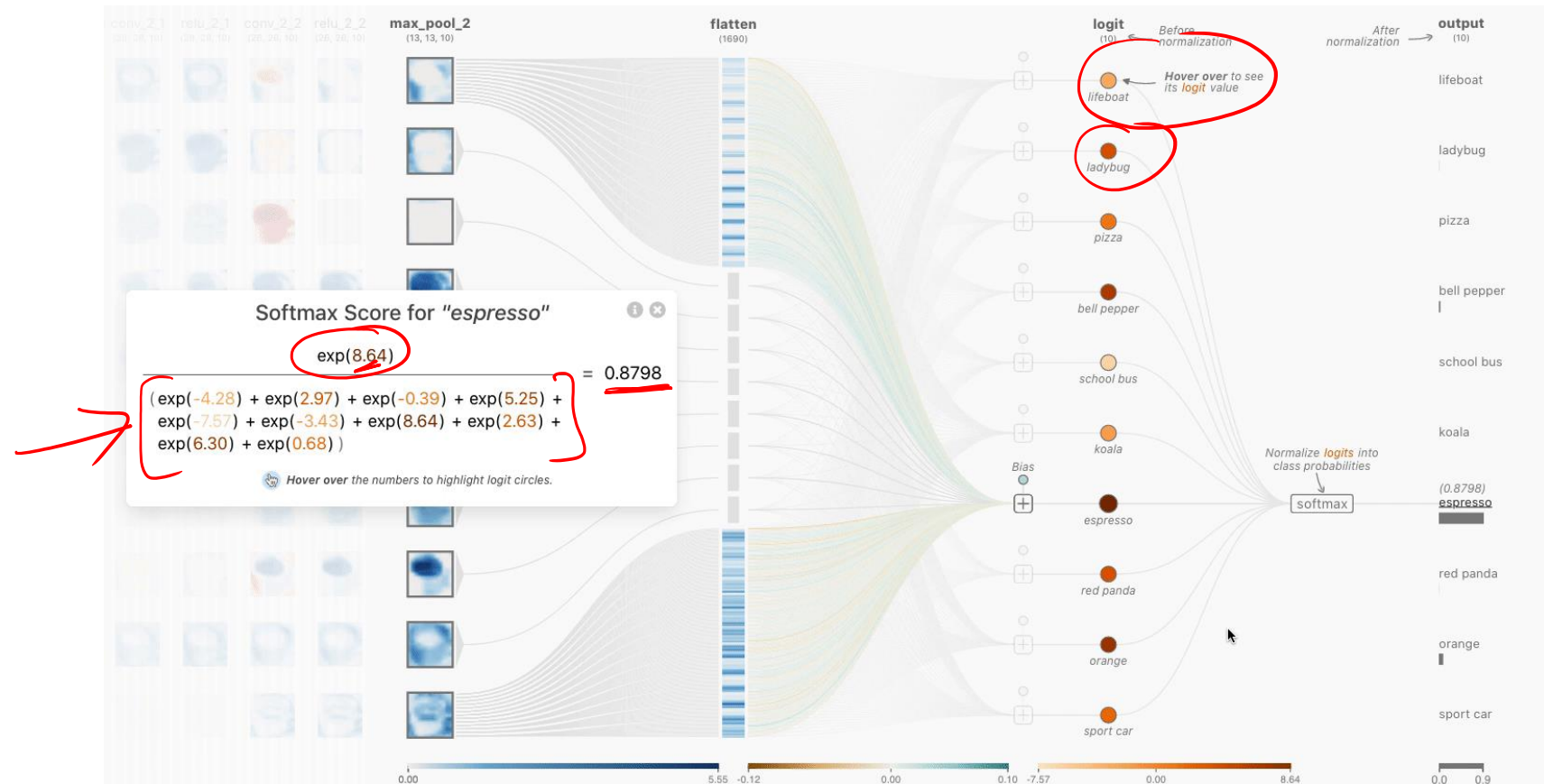
- A softmax operation serves a key purpose: making sure the CNN outputs sum to 1. Because of this, softmax operations are useful to scale model outputs into probabilities.

$$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$$

- For a visual indication of the impact of each logit (unscaled scalar value), they are encoded using a light orange → dark orange color scale. After passing through the softmax function, each class now corresponds to an appropriate probability.

# Feature Representation

- The Softmax Interactive Formula View allows a user to interact with both the color encoded logits and formula to understand how the prediction scores after the flatten layer are normalized to yield classification scores.



# Feature Representation

---

## ✓ Pooling Layers

- There are many types of pooling layers in different CNN architectures, but they all have the purpose of gradually decreasing the spatial extent of the network, which reduces the parameters and overall computation of the network. The type of pooling used in the previous architecture is Max-Pooling.

## ✓ Flatten Layer

- This layer converts a three-dimensional layer in the network into a one-dimensional vector to fit the input of a fully-connected layer for classification. For example, a  $5 \times 5 \times 2$  tensor would be converted into a vector of size 50.

---

# Thank you