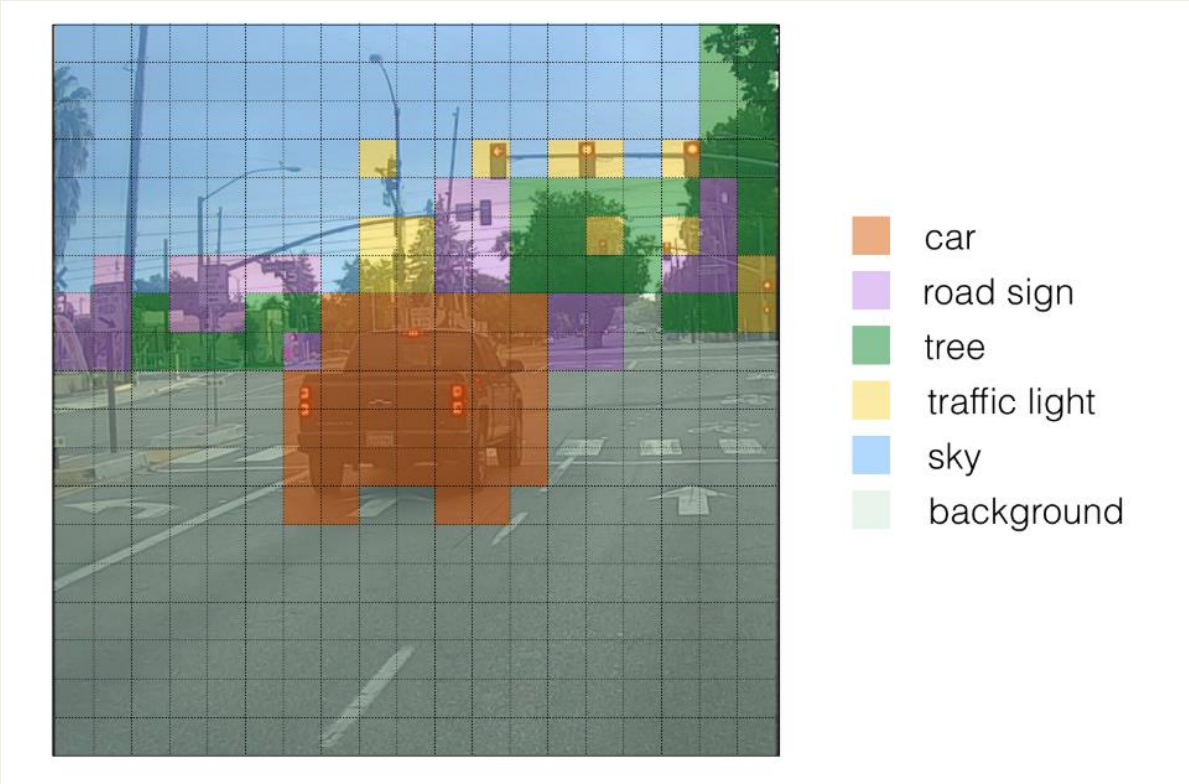
OBJECT RECOGNITION

(Supervised and Unsupervised)

Varsha Nagarajan, Kevin Alvarez, Darshil Patel, Ashwin Risbood

Why Object Recognition?



AUTONOMOUS DRIVING

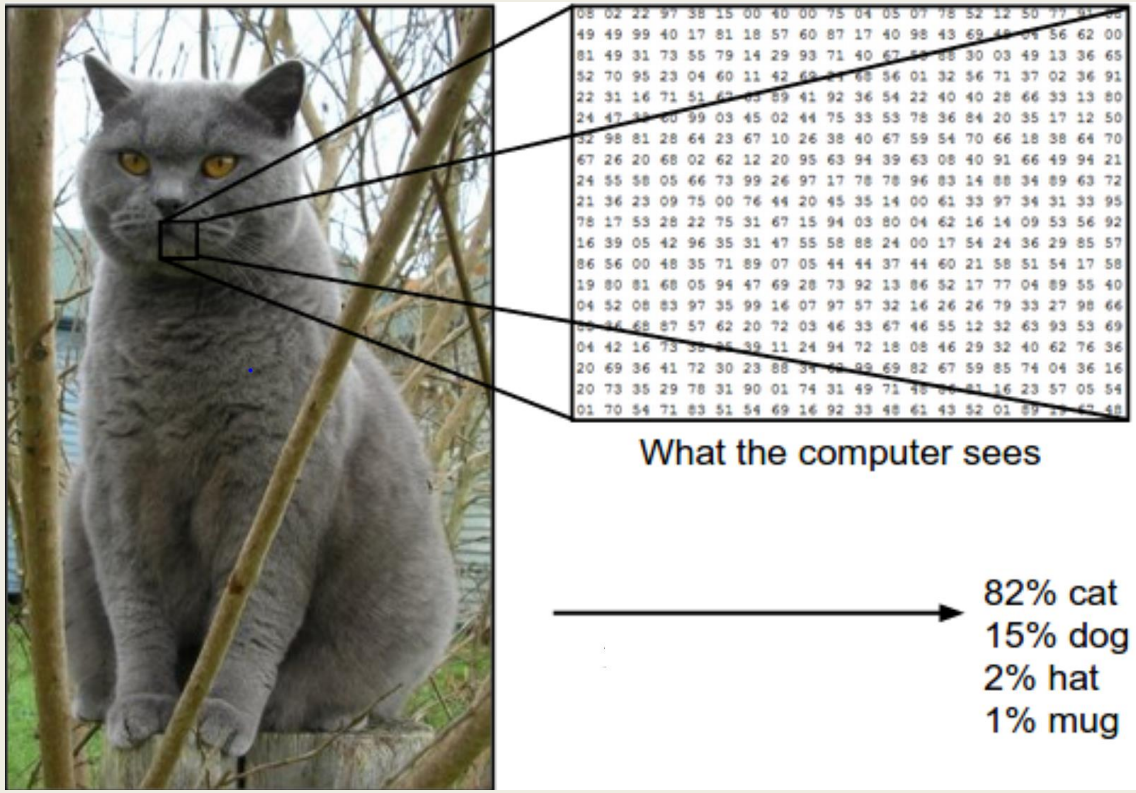
DATASET: CALTECH256 | CIFAR-10/100

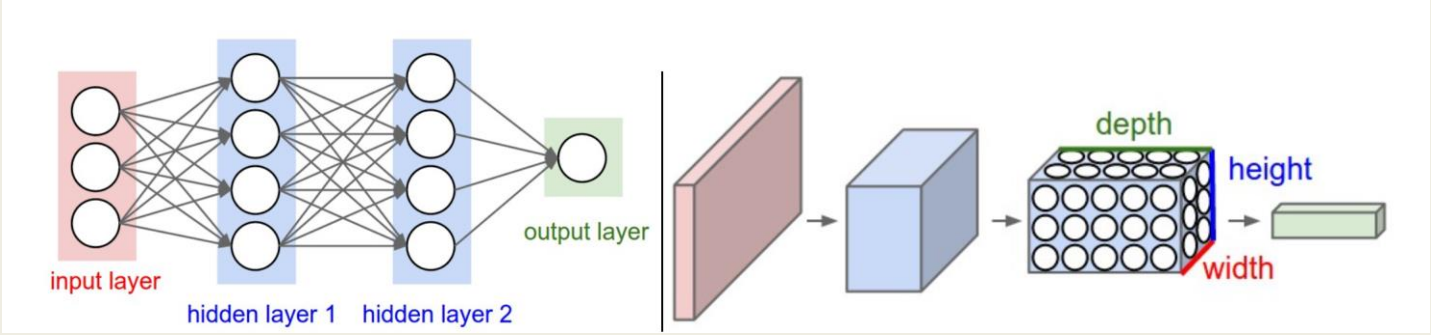
- **Caltech**: Train, Validation and Test set splits with probabilities of 0.6, 0.15 and 0.25 resulted in 17803 train images, 4665 validation images and 7322 test images
- **CIFAR-10:** 50000 train images (25% validation split), 10000 test images.
- **CIFAR-100:** 500 training images and 100 testing images per class. (25% validation split)

Convolutional Neural Networks

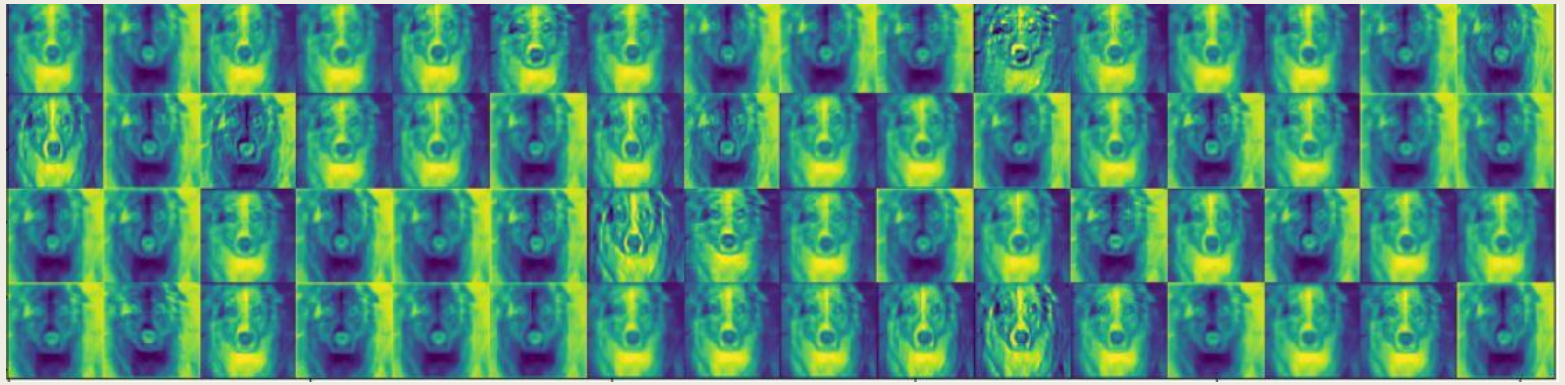
> Works well with images

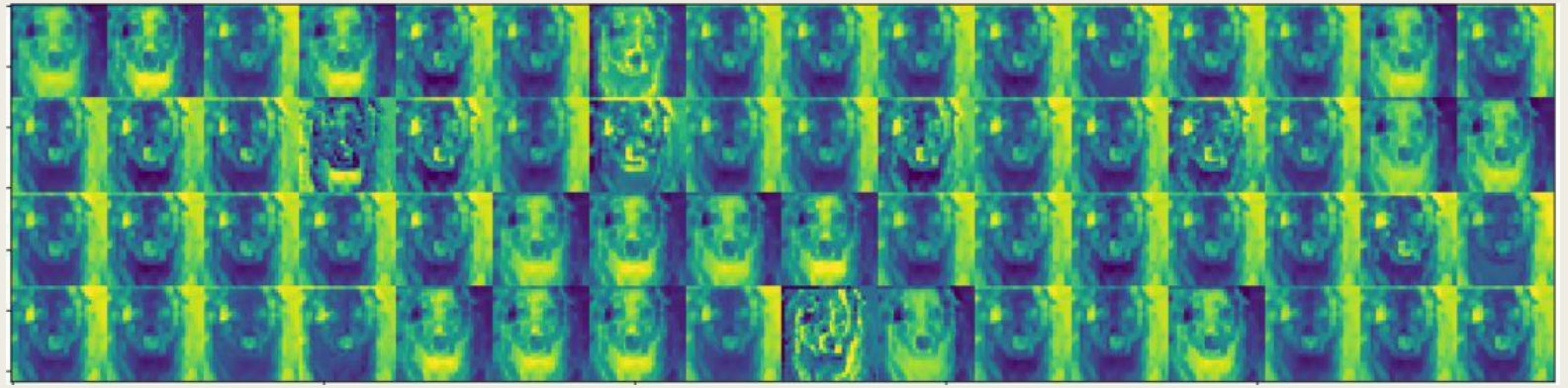
> Less parameters to learn

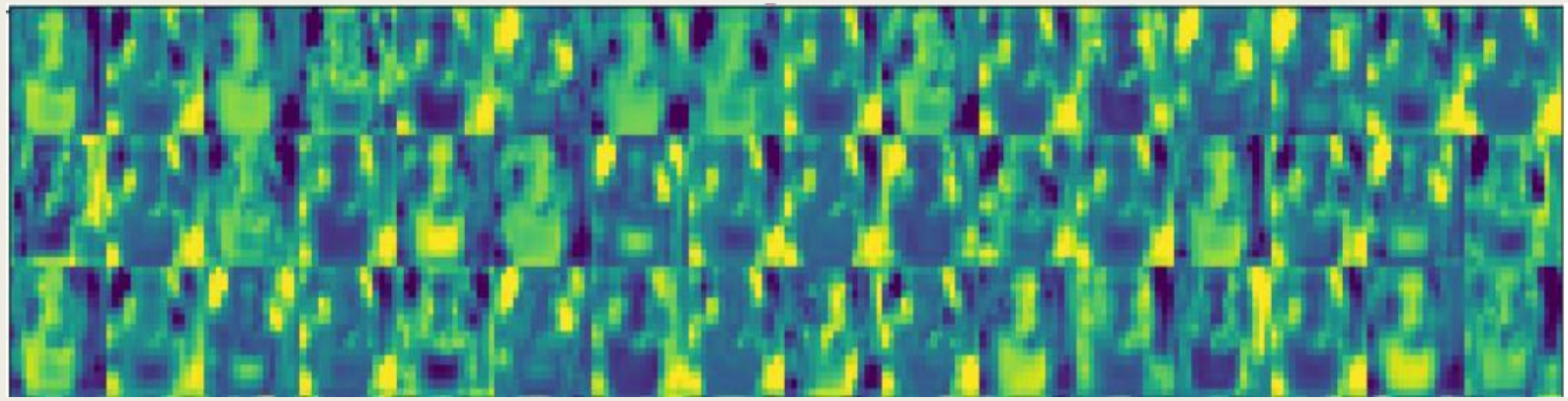




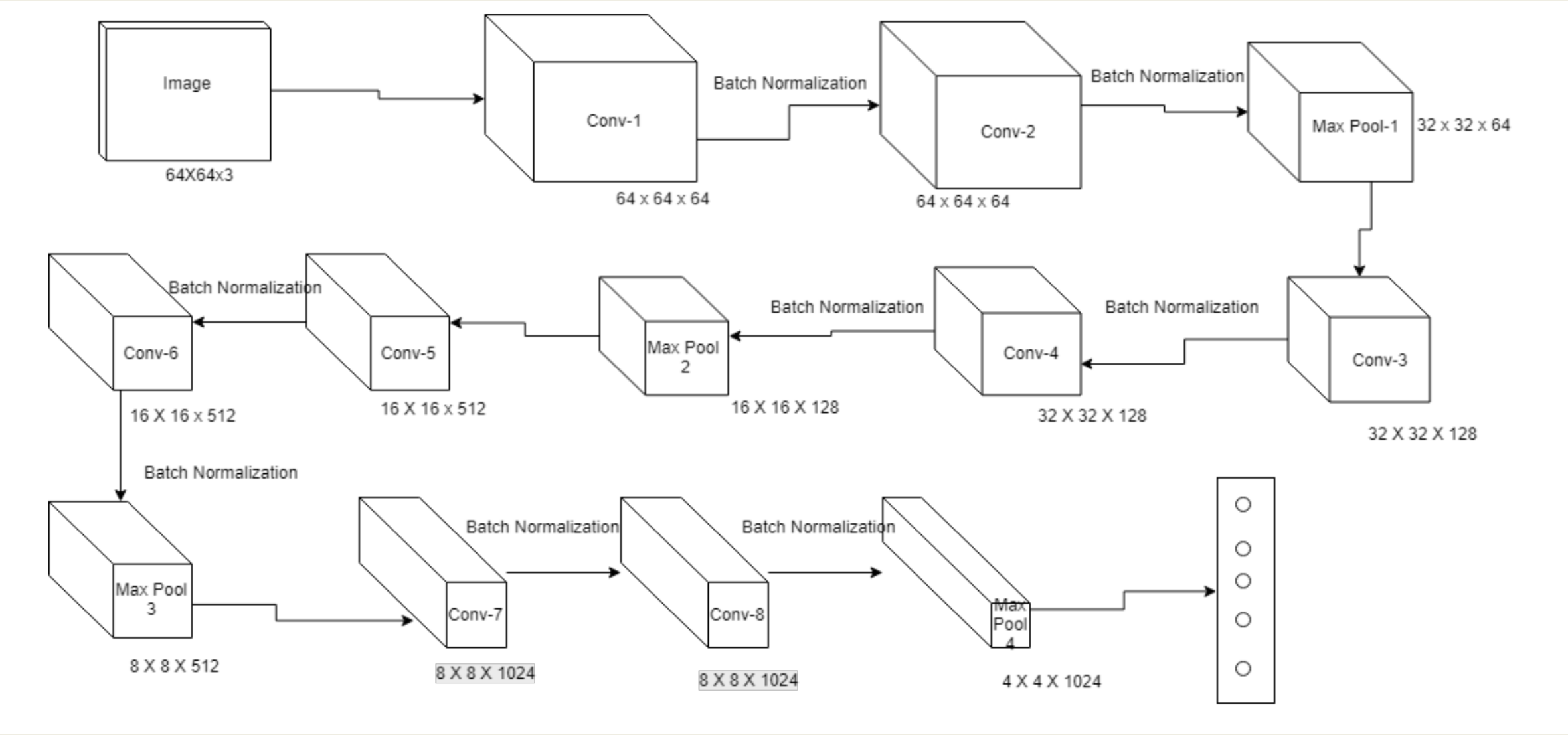
Visualization of Layers





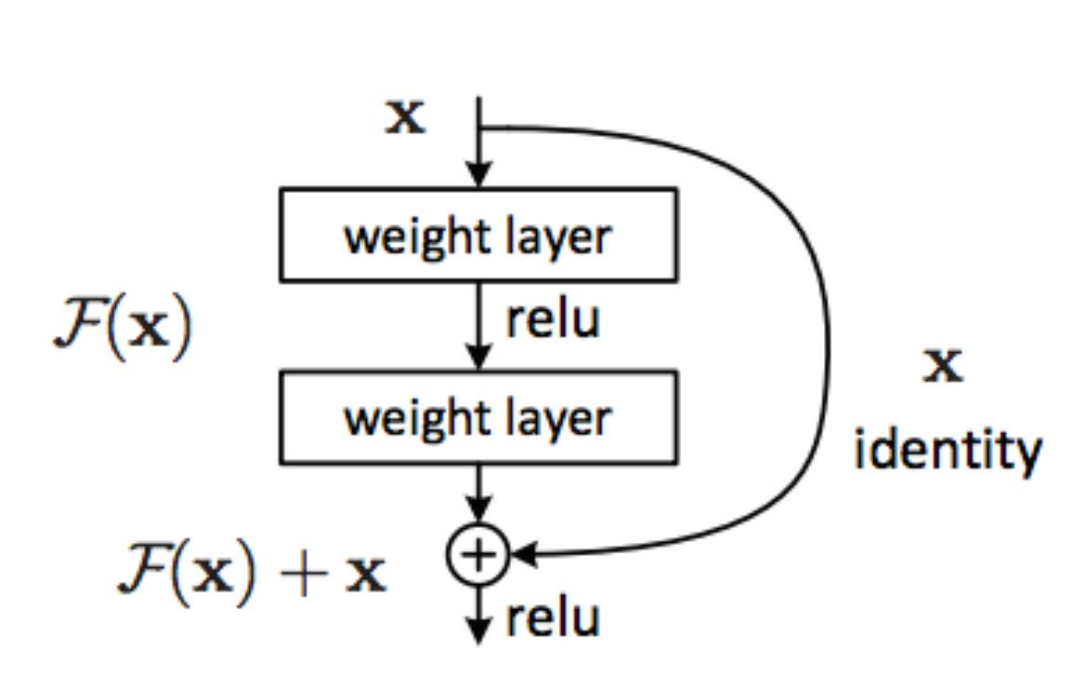


Method 1 – Plain Classifier Model

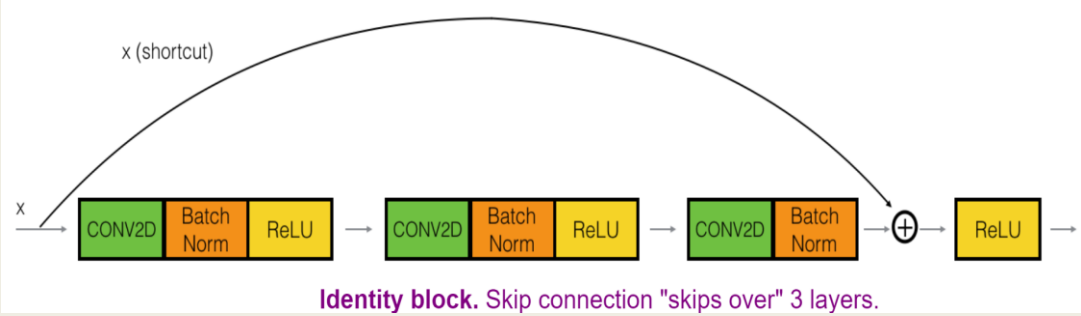


Method 2 – Residual Networks (ResNet50)

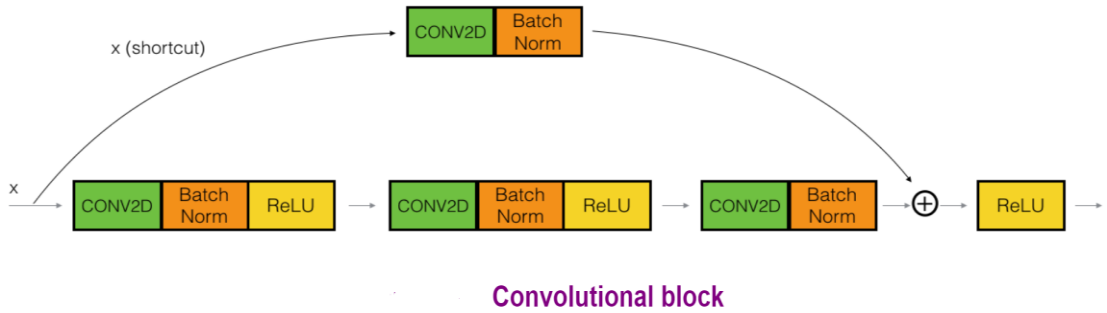
- Vanishing gradient problem (batch normalization and normalized initialization
- Degradation problem
- Easier to learn an identity function



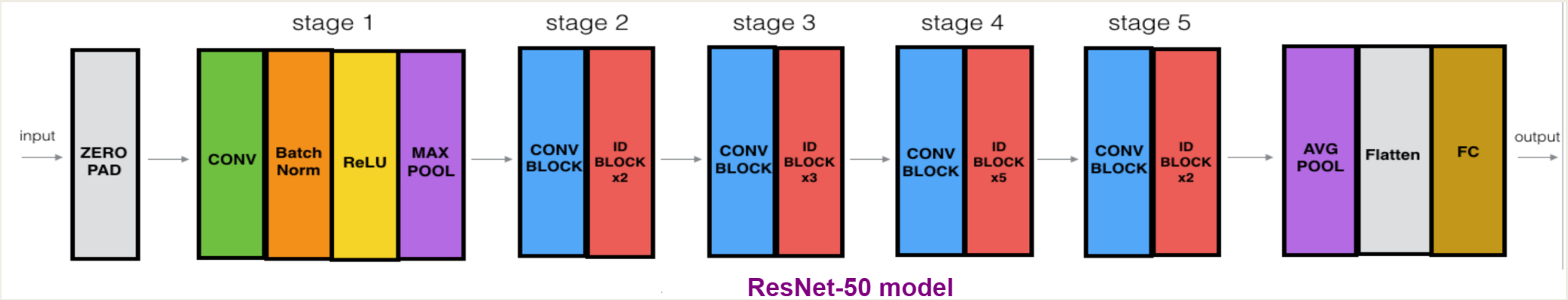
Our ResNet Model



Identity block. Skip connection "skips over" 3 layers.

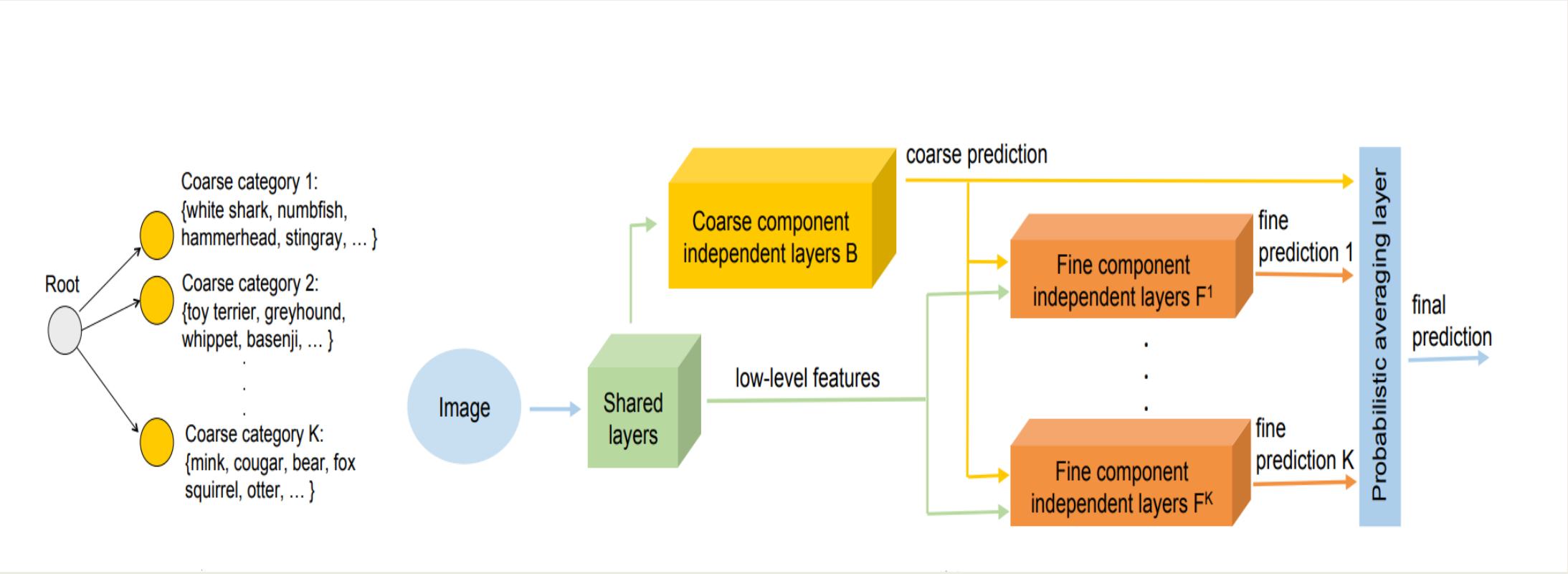


Convolutional block



ResNet-50 model

Spectral Clustering + Hierarchical Deep CNN



RESULTS

Model	Caltech256			CIFAR-10		CIFAR-100	
	Classes	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
Plain Classifier	256	15.58%	30.79%	67.34%	96.98%	37.56%	68.40%
ResNet50	256	8.50%	22.65%	68.91%	96.72%	26.66%	54.94%
HD-CNN	256	10.17%*	-	-	-	42.84%	-
Random Guess	256	0.8%	1.25%	0.2%	1.04%	0.4%	0.98%

Background - Clustering

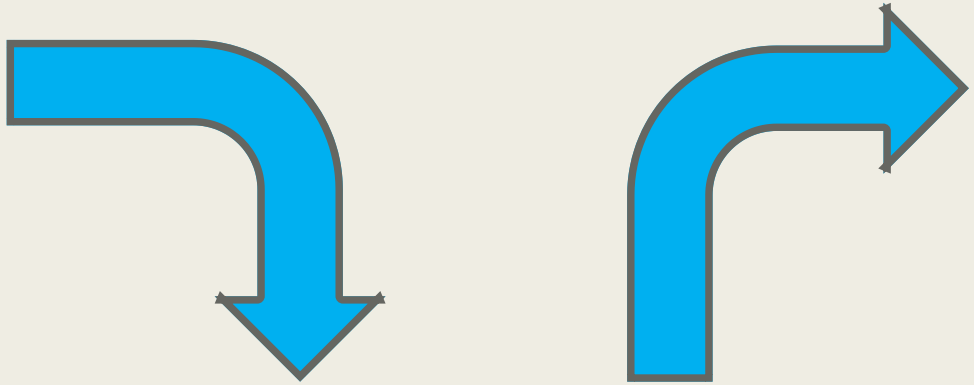
Goal: learn object categories without supervision

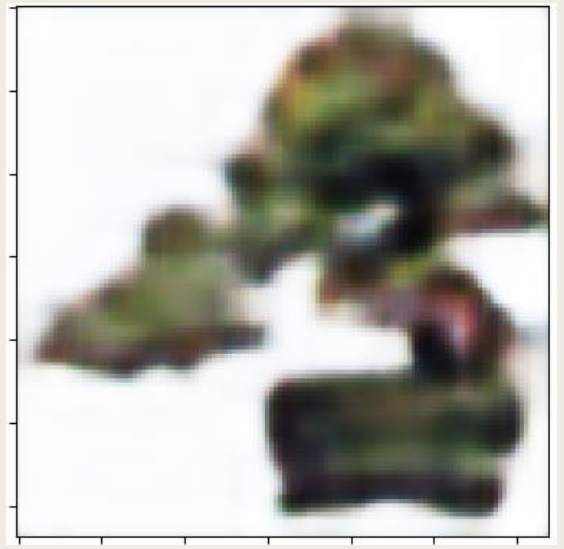
The effectiveness of any clustering algorithm is highly dependent on the representation of the objects.

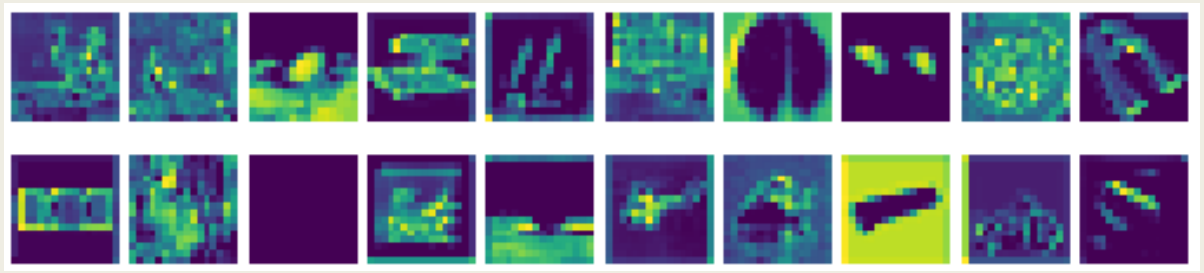
Trouble with high dimensional, noisy data

Images – high dimensional and noisy

Method 1 – Convolutional Autoencoder







Method 2 – t-Distributed Stochastic Neighbor Embedding

- Uses t-distribution compared Gaussian used by SNE to calculate similarity
- Dimensionality Reduction by preserving local structure in latent space

RESULTS

Feature Extraction	Purity	Entropy
CAE	**0.143**	**5.45**
t-SNE	0.119	5.61
Raw Pixels	0.125	5.66

Conclusion

- Poor image quality
- Very less training and validation samples
- High-dimensional data caused computational complexity issues with K-means
- The main factor in the effectiveness of the clustering seemed to be the quality of the image representation.