CTE: Intro to Machine learning and Deep Learning

Computer Vision

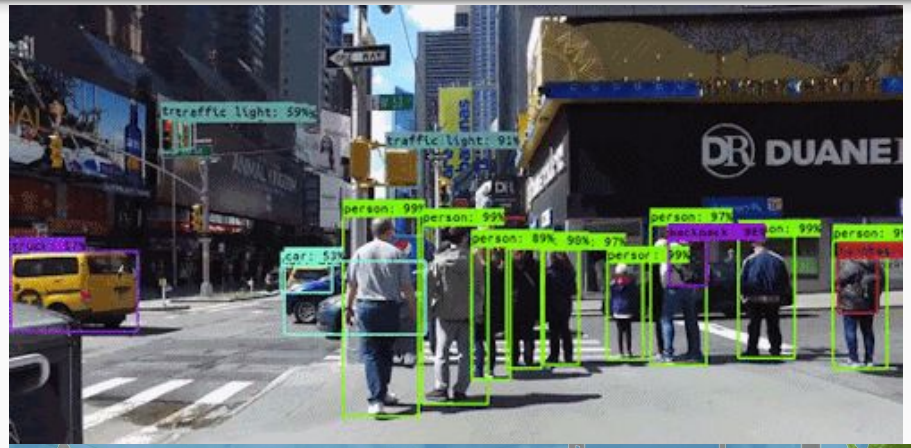
Contents -

- Applications of CV
- Famous architectures
- Transfer Learning
- Image Augmentation
- Pooling techniques and other techniques (Reading Material*)

Computer Vision -

Part of Deep Learning that deals with the visual Data - like Images and Videos

Applications of Computer Vision



Application of this field

Current CV techniques have many applications like -

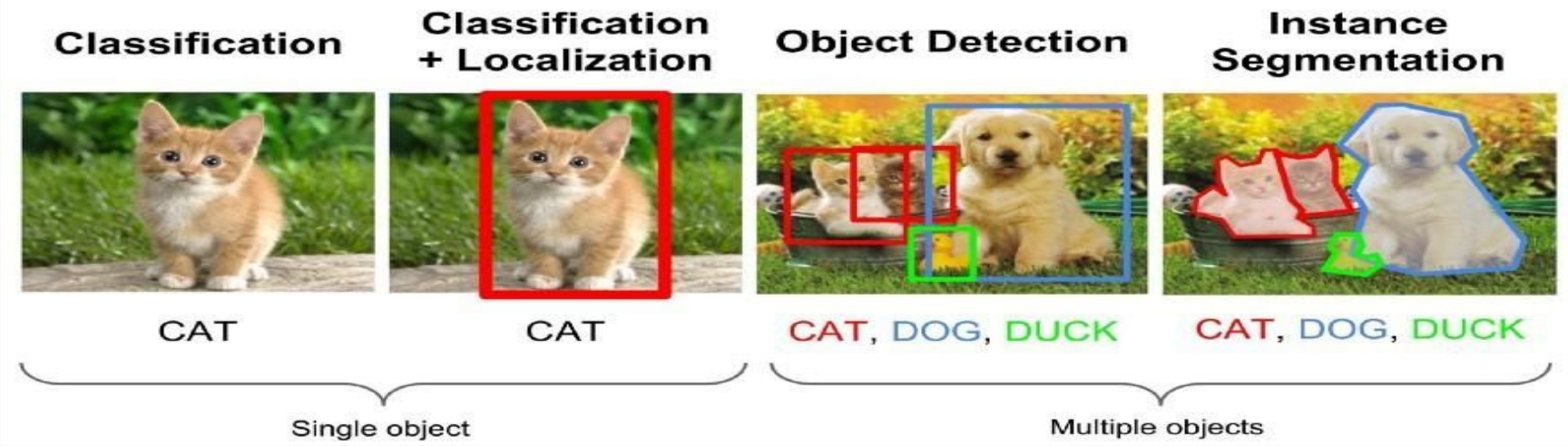
Segmentation (used for self-driving cars)

Detection (used for surveillance)

Image/Video Captioning (could be used to help blind for letting them know about the surroundings)

Key point / Pose estimation (think this application by yourself :))

Applications of Computer Vision



Difference between various applications

Classification is to classify a particular image into various classes (like dog/cat etc.)

Localization is detecting an object and drawing a bounding box over it

Detection can be thought of as localisation + classification

Segmentation is when you label each pixel of the image and classify that pixel to what category/object it belongs to

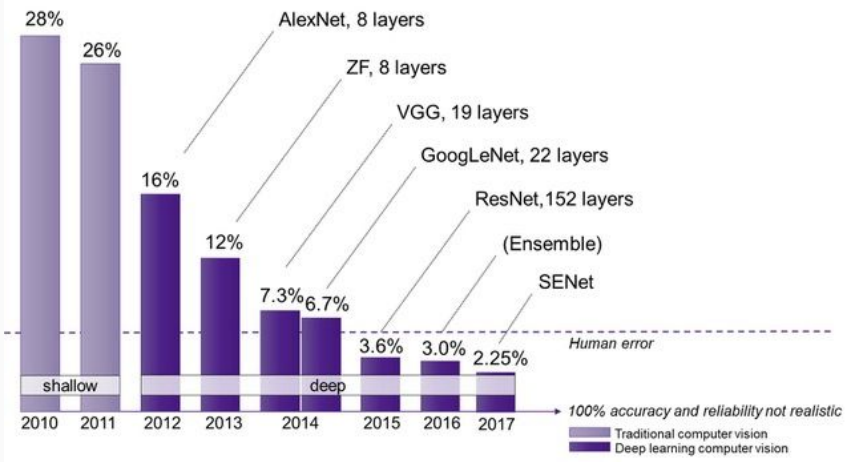
Some famous architectures

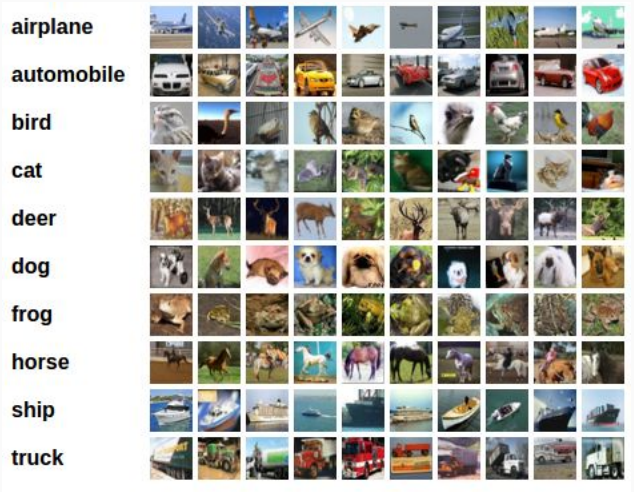
Datasets!

-ImageNet (14 Million Images)

Cifar 10/100 (10/100 classes for classification)

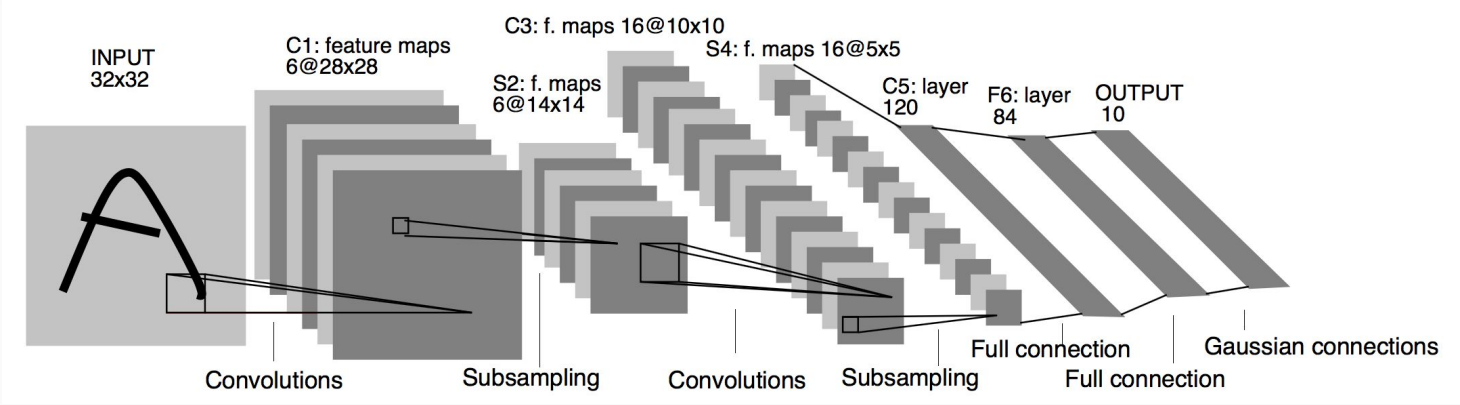
Fashion MNIST (easy)





Famous architectures in CV

- Le-Net
- Alex Net
- ResNet
- Inception



Le-Net (Basic)

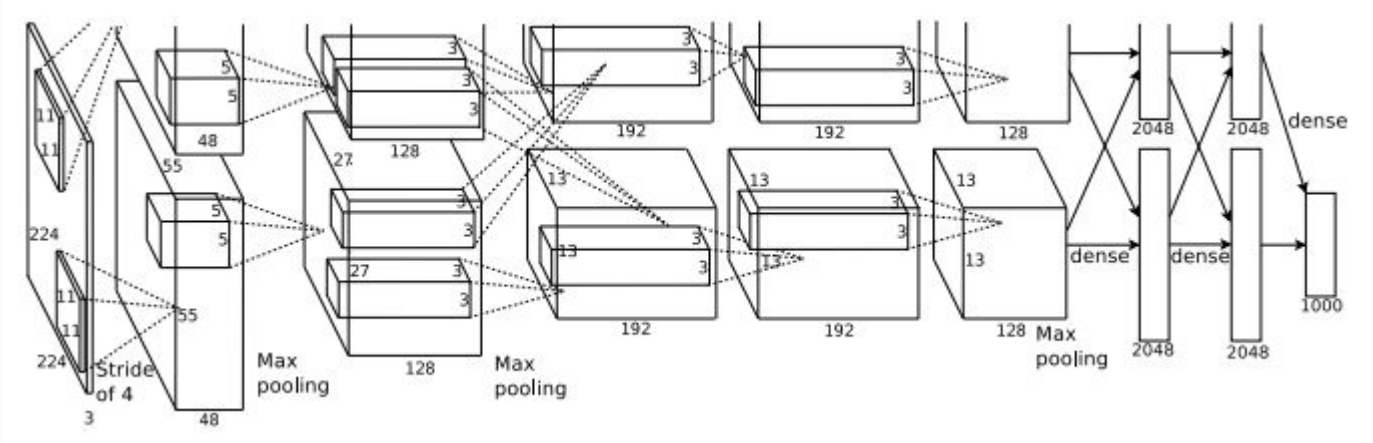
- 7-level convolutional network
- Input - 32x32 pixel greyscale

Famous architectures in CV

Alex-Net

more filters per layer

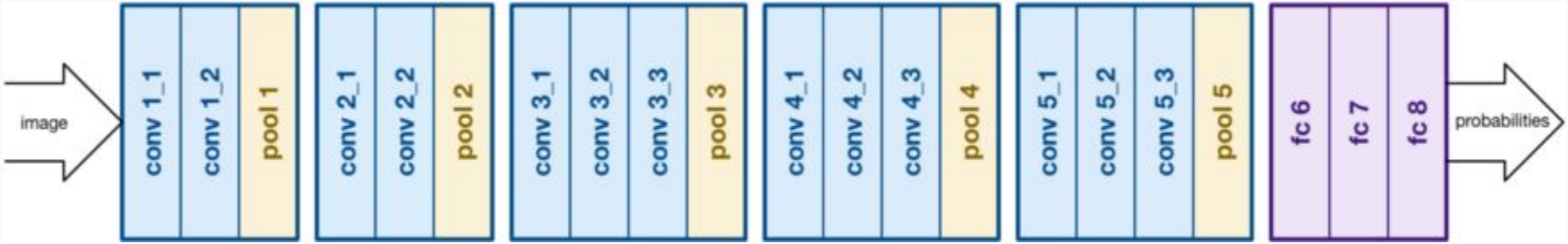
parameters-
max pooling, dropout,
data augmentation,
ReLU activations,
SGD with momentum



AlexNet was trained for 6 days simultaneously on two Nvidia Geforce GTX 580 GPUs

Famous architectures in CV

VGG -16



VGGNet consists of 16 convolutional layers and is very appealing because of its very uniform architecture.

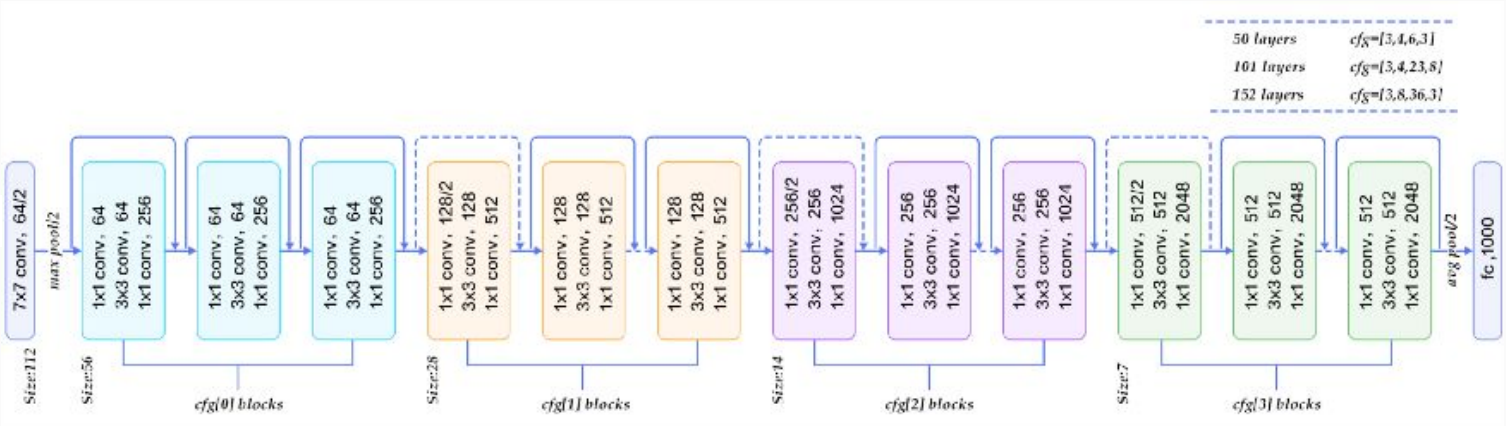
VGG takes in a 224x224 pixel RGB image.

Lesser filter size so possibility of more layers.

Famous architectures in CV

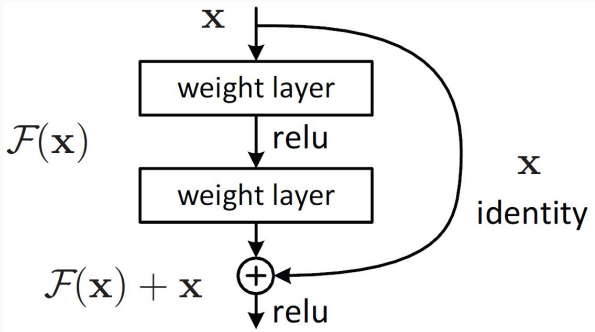
Res-Net

"skip connections"



Why ResNet is better -

- **More layers is better**
- **Identity function helps not to perform bad when layers increase**

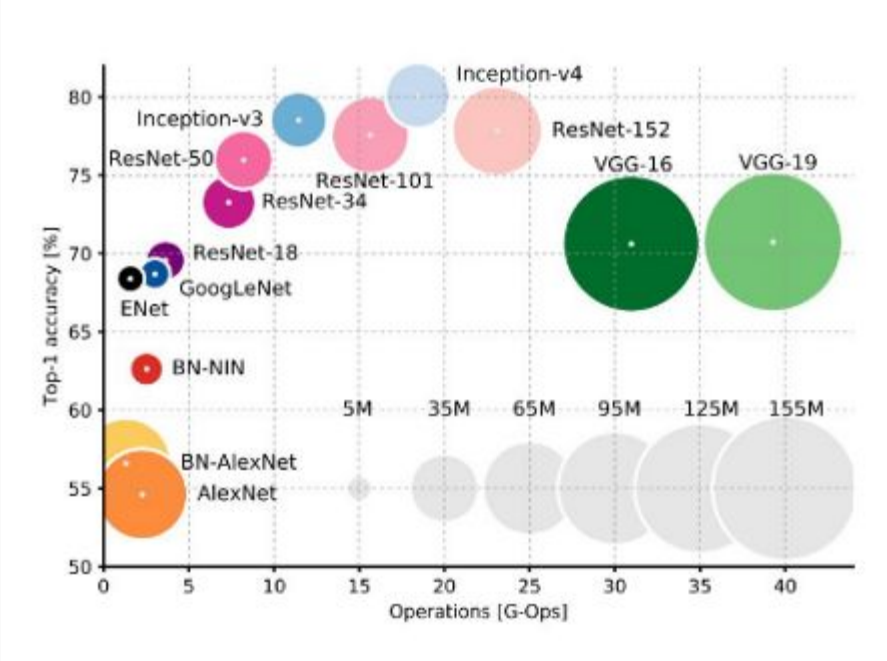


Also Have a look at Inception and Xception

Famous Architecture

Other Structures to have a look at (MUST)

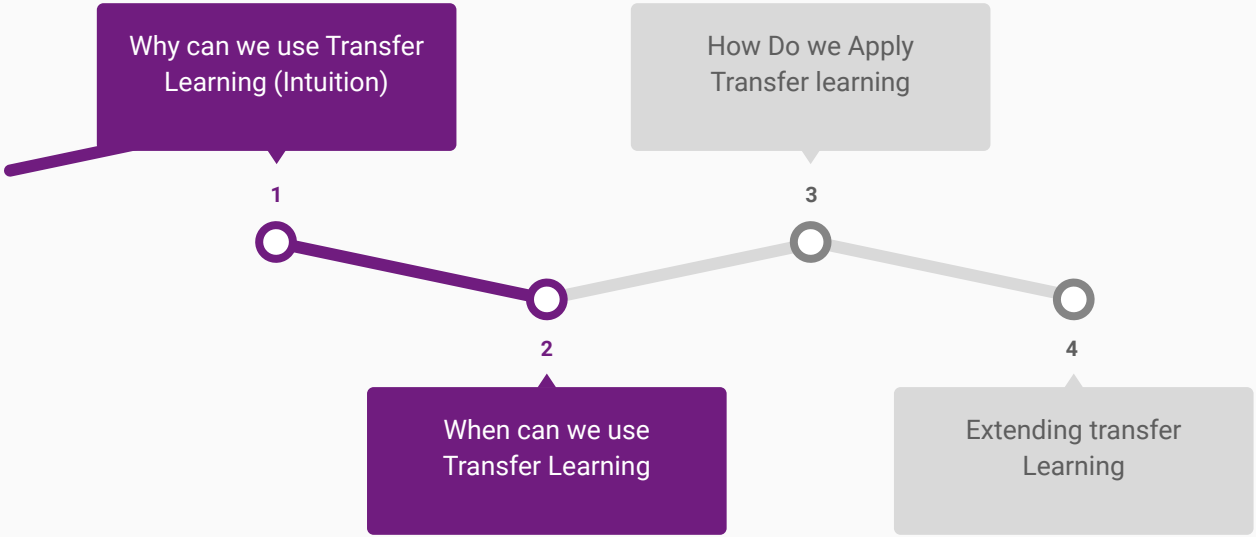
- Inception
- Xception



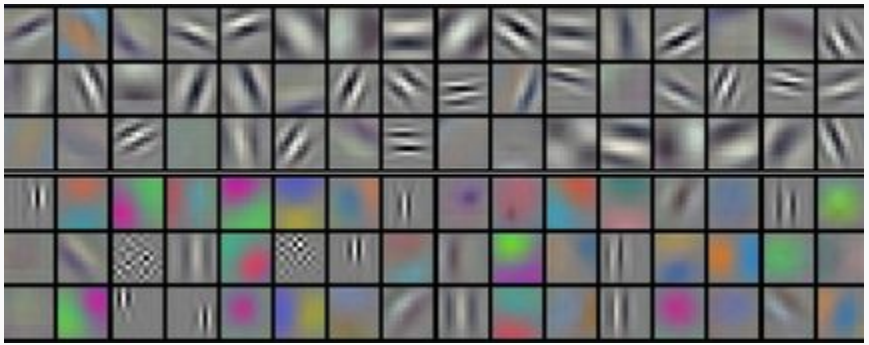
Transfer Learning

Transfer- Learning

The knowledge of a model trained on a large dataset (eg - ImageNet) can be transferred/applied to another model that trains on similar but less dataset.



Why can we use transfer learning



Layer visualization of AlexNet Model

What do different layers learn-

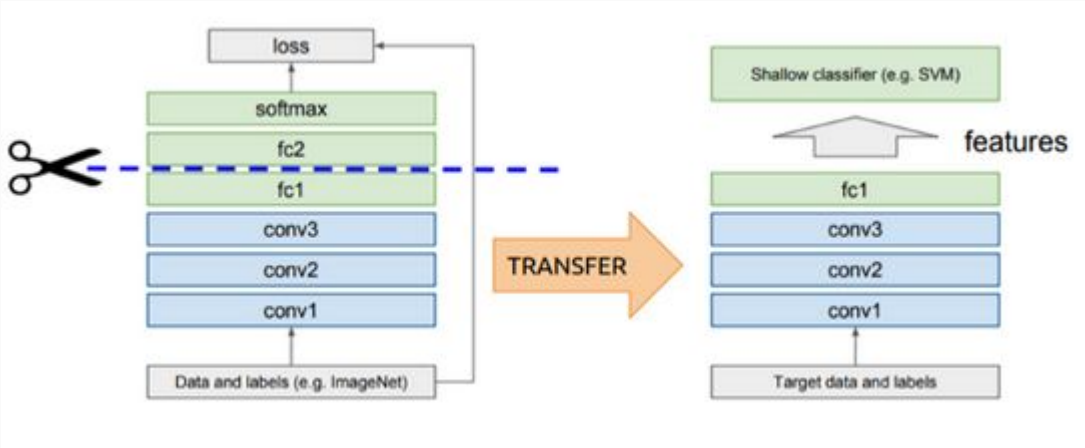
- Initial layers - Generic features (lines,edges)

- Deep layers - Specific features (color, pattern)

Why can we use transfer learning

All models trained on a particular type of dataset learn the similar kind of features.

So why not use the weights of model trained on larger dataset to substitute the generic layer weights of the model training on smaller dataset

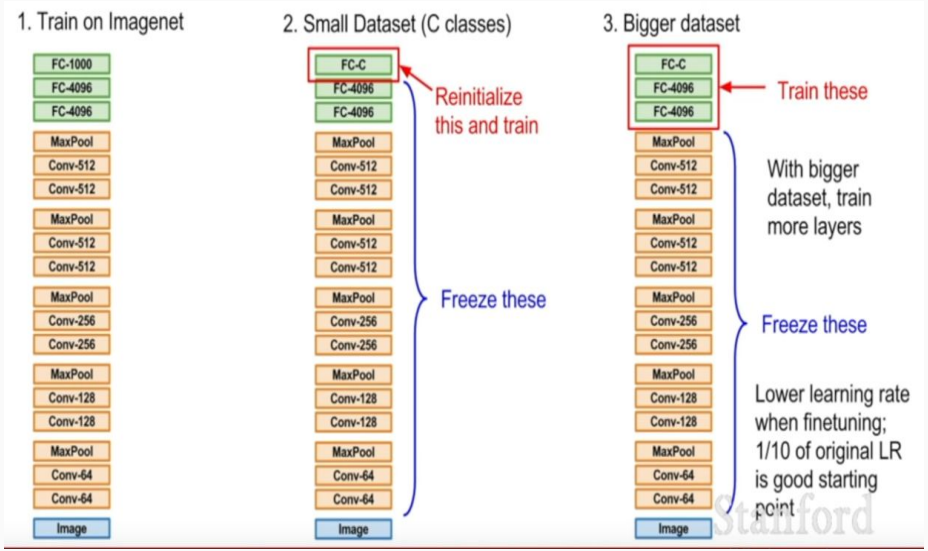


When can we use transfer learning

- Task A and Task B have similar input (Images/Audio/..)
- We have more Dataset for Task A than Task B.
- Low level Features of A can be helpful for learning B

Numbers/Alphabet type of Dataset	MNIST (60,000)	Omniglot (1623)
Animal/Face/Object	VGG, Alexnet , CIFAR-10/100	Any custom model/ other dataset

How can we Apply Transfer Learning

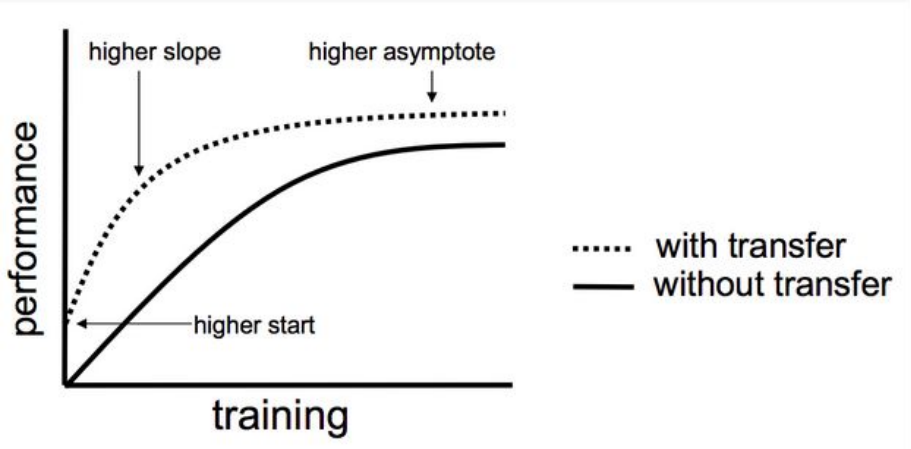


- Freezing the layers - Keeping the weights constant (LR=0)
- Model can be thought of as 2 separate models , One that gives the output which is same for the same image(constant weights) which has many layers (almost all) -**Pretrained part**
- Final classifier model - Which takes the input from initial model and train to give the required output - **Fine Tuning Part**

How can we Apply Transfer Learning

	Similar Dataset	Different Dataset
Very little Dataset available(Dataset you have)	Use Linear classifier on top of the model	Problem!
Large amount of dataset Available	Try to fine Tune few more layers	Fine Tune large amount of layers (Not great use of TL)

Performance of Transfer learning



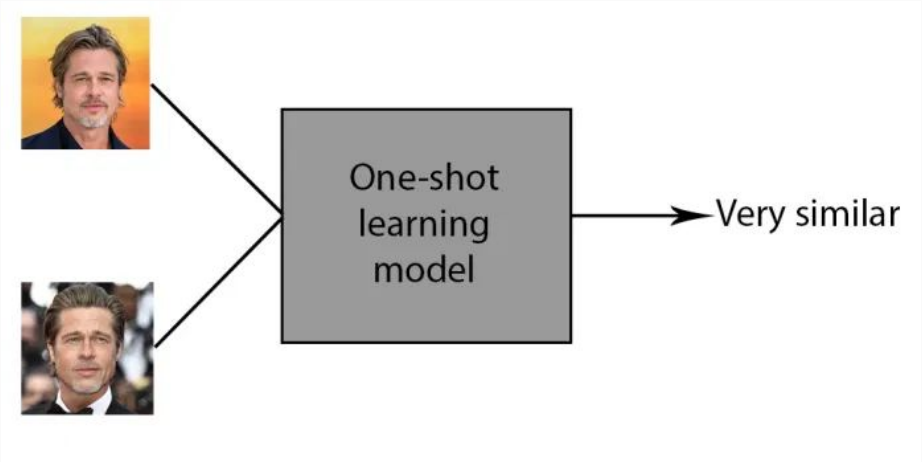
- The performance increases very fast as most of the initial layers have been thoroughly learned.
- Saturates at much higher level

Extending Transfer learning

Few shot learning like One shot and Zero shot learning use Transfer learning model as their base.

One shot learning , is an extension of Transfer learning where we have even smaller amount of dataset (2-5 images per class).

Applications in Face recognition!



One - Shot learning

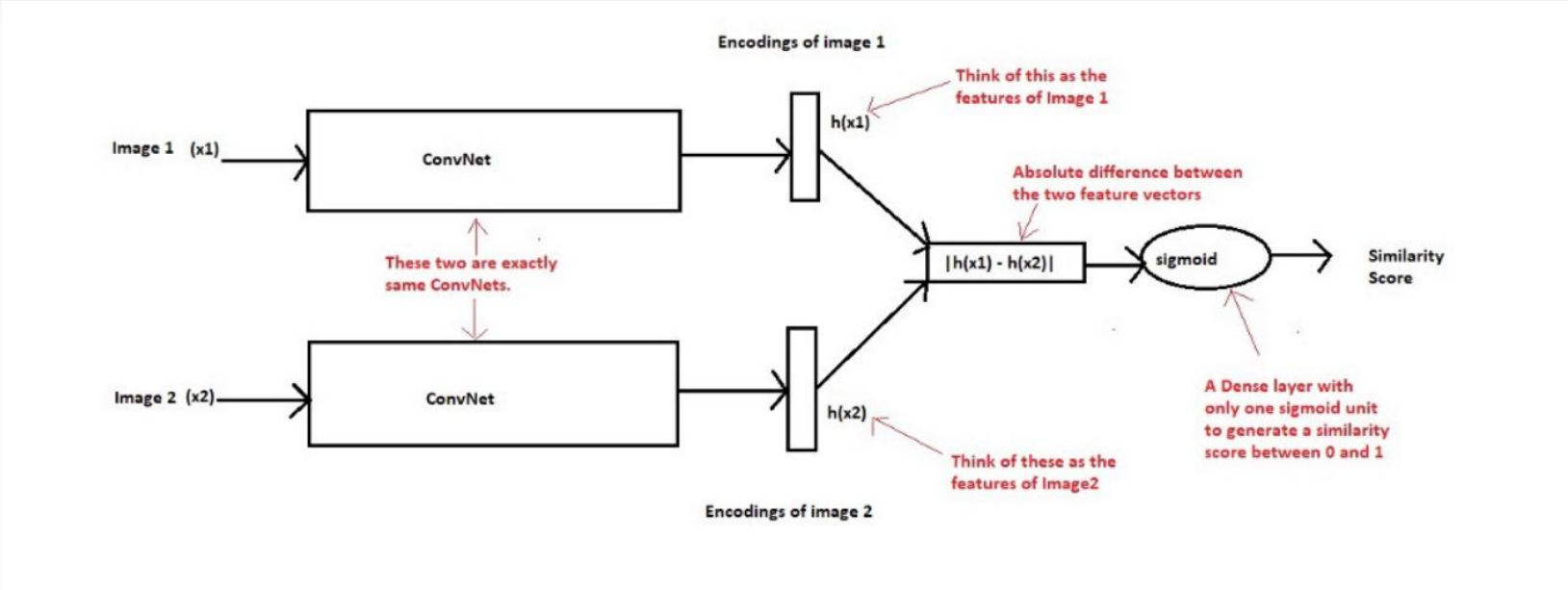


Image Augmentation

Image Augmentation

Another technique if we have less dataset.

Idea - Flipped photo of a Cat is still a cat

Types of image augmentation -

- Flipping
- Random Rotation
- Zoom
- Randomized contrast and brightening

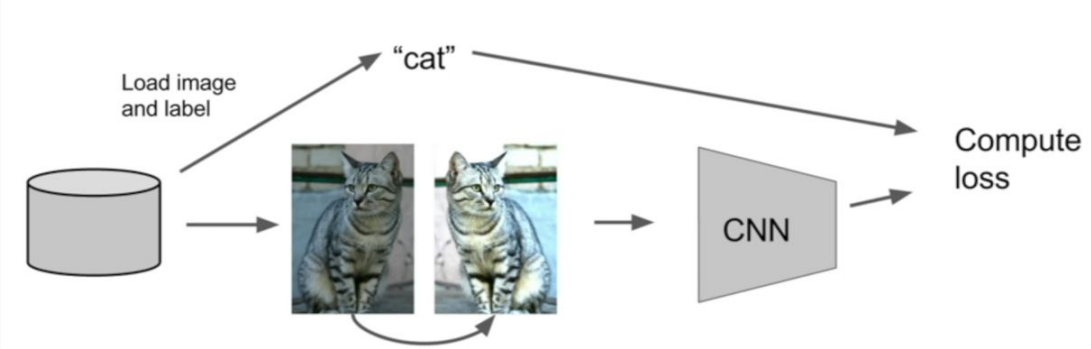


Image Augmentation

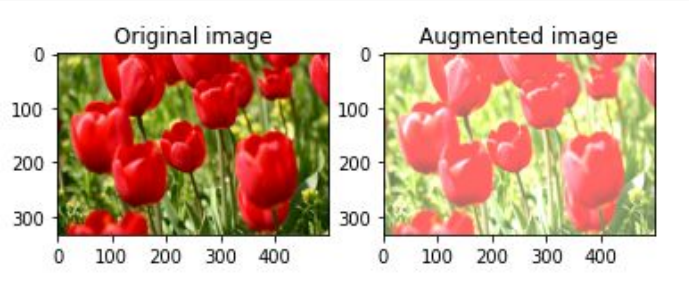
Flipping - Randomly flipping the images , options for Both horizontal and vertical.

Cropping - Center crop , random crop

Rotation-Randomly rotating images by any angle

Brightening

```
# define pytorch transforms
transform = transforms.Compose([
     transforms.ToPILImage(),
     transforms.Resize((300, 300)),
     transforms.CenterCrop((100, 100)),
     transforms.RandomCrop((80, 80)),
     transforms.RandomHorizontalFlip(p=0.5),
     transforms.RandomRotation(degrees=(-90, 90)),
     transforms.RandomVerticalFlip(p=0.5),
     transforms.ToTensor(),
     transforms.Normalize((0.5, 0.5, 0.5), (0.5, 0.5, 0.5)),
     ])
```



When shouldn’t we use a particular Augmentation

When flipping might change the meaning of image - MNIST dataset.

When random crops / Zoom might lead to uncertain images.

Here Image augmentation decreases the performance of the model