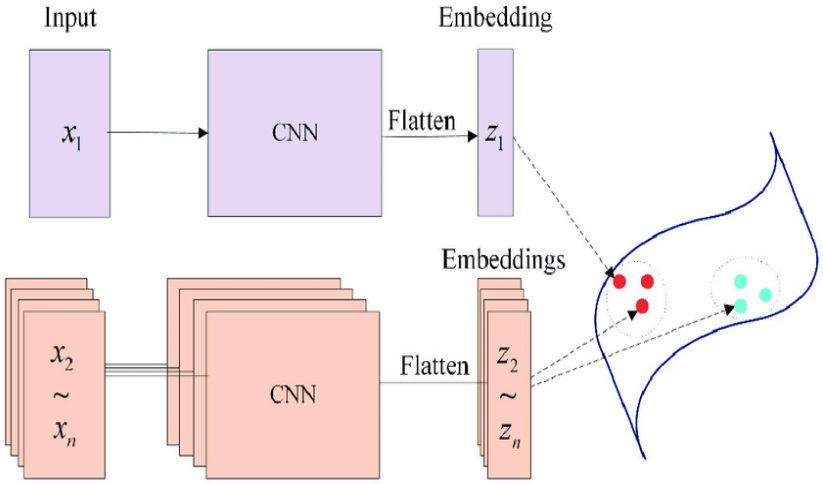
What is Image similarity?

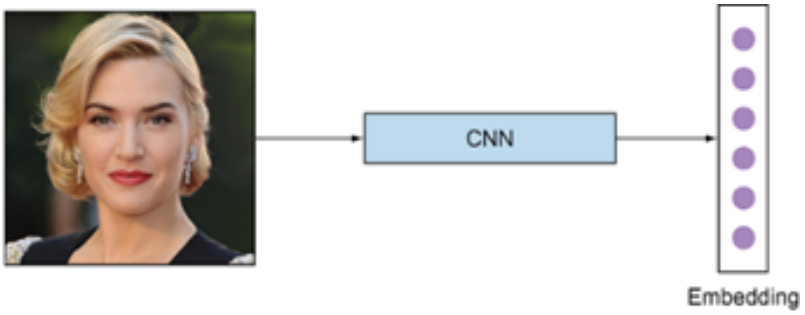
- Image similarity is the measure of how similar two images are.
- In other words, it quantifies the degree of similarity between intensity patterns in two images.
- Finding Image similarity also called reverse image search

Architecture of Image Similarity



What is an Embedding?

- An image embedding is a lower-dimensional representation of the image.
- In other words, it is a dense vector representation of the image that can be used for many tasks such as classification.
- A convolutional neural network (CNN) can be used to create the image embedding.



Drawbacks of VGG and AlexNet

VGG:

- A large number of parameters: VGG has a large number of parameters, making it prone to overfitting and memory-intensive.

- Computationally expensive: The use of small 3x3 filters results in a large number of operations and can be computationally expensive.

AlexNet:

- Limited representational capacity: AlexNet has a relatively shallow architecture with limited representational capacity, which can limit its ability to learn complex features and representations.

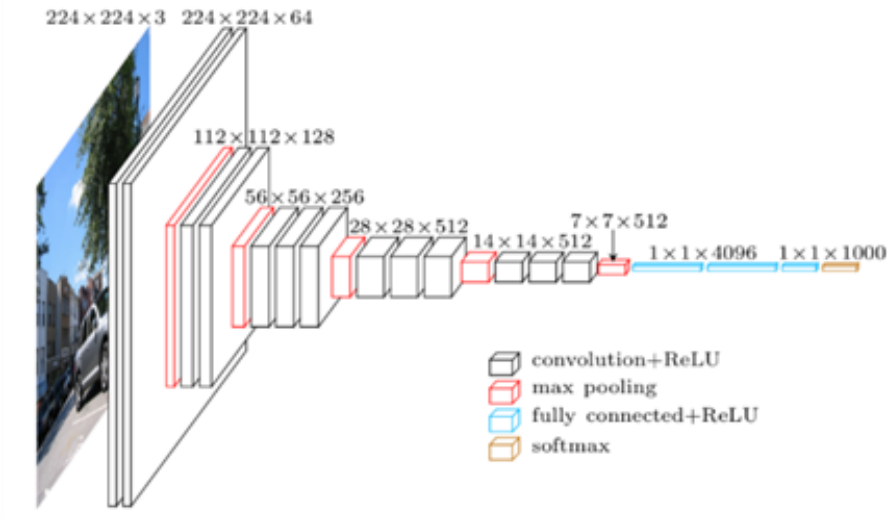
- Over-reliance on data augmentation: AlexNet was heavily reliant on data augmentation techniques, such as image flipping and color shifting, which can be time-consuming and may not always be appropriate.

Why do we need Deeper Architecture?

- Deeper architectures in Convolutional Neural Networks (CNNs) are needed to improve the performance of the network by allowing it to learn more complex representations of the data

- A deeper network has more layers, and each layer can learn to extract a different and more abstract representation of the data.

- This allows the network to learn increasingly complex features, such as shapes, textures, and patterns, that are important for image classification and other computer vision tasks.



How ResNet and Inception better than AlexNet and VGG?

ResNet:

- **Residual connections:** ResNet introduced the concept of residual connections, which allow the network to learn residual representations instead of direct representations.

- This helps prevent vanishing gradients and enables the network to train deeper models without degradation in performance.

- Improved representational capacity: ResNet has a much deeper architecture than VGG and AlexNet, which allows it to learn more complex representations.

Inception:

- **Improved efficiency:** The Inception architecture uses multiple parallel branches with different filter sizes and concatenates the results, which allows the network to learn multiple scales of features efficiently.

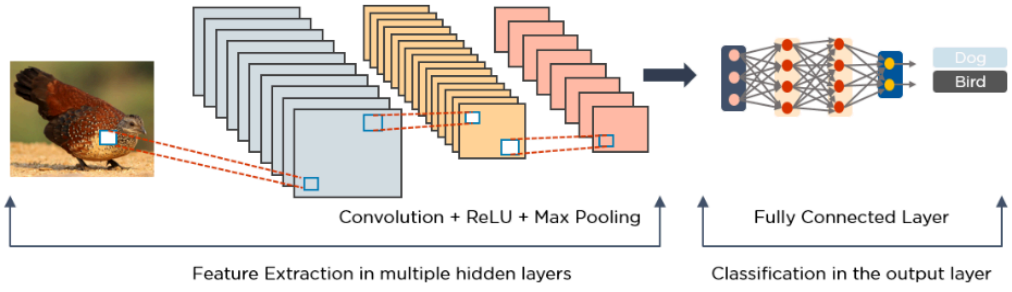
- Reduced number of parameters: Inception uses multiple smaller filters instead of a few large filters, which reduces the number of parameters and reduces the risk of overfitting.

Why do we need a Multi-level Feature Extractor?

- We use multiple levels of Feature Extraction in CNN because At lower levels, the network extracts simple features such as edges and corners, which are combined to form higher-level features such as textures and shapes.

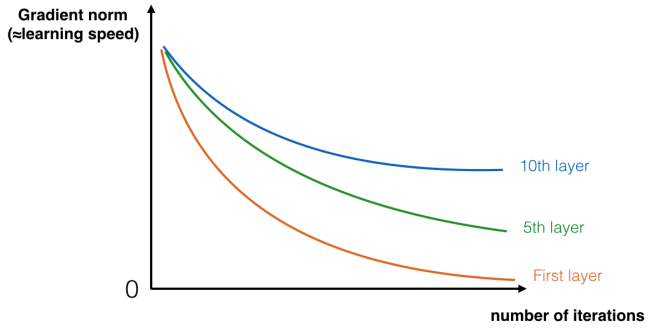
- These higher-level features are then combined to form even higher-level representations such as object parts and eventually complete objects.

- By using multiple levels of feature extraction, the network can learn to recognize objects and their relationships from a wide range of viewpoints and scales.



Vanishing Gradient

The speed of learning decreases very rapidly for the early layers as the network trains



Residual Blocks:

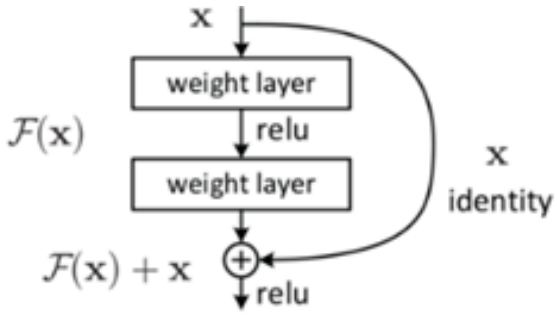
1. Identity Block/skip connections

- Skip connections in Convolutional Neural Networks (CNNs) are connections that bypass one or more layers in the network, allowing the gradients to flow more easily through the network during training.

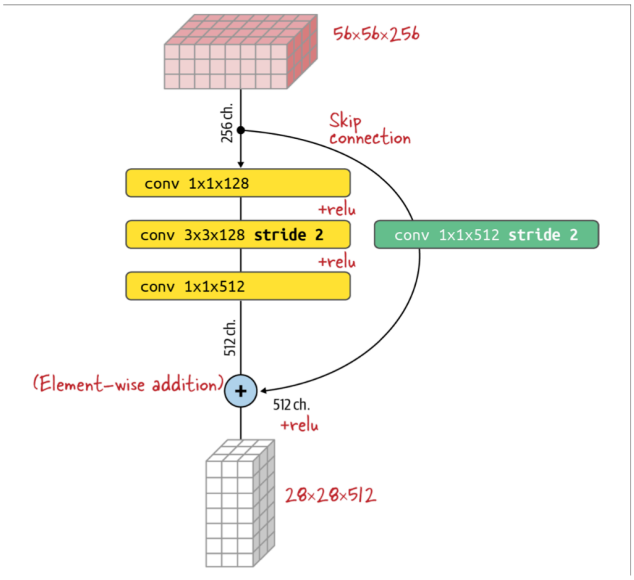
- The idea behind skip connections is to mitigate the vanishing gradient problem, which can occur in very deep networks and cause the gradients to become very small, making training difficult.

- Skip connections allow the activations from earlier layers to be combined with the activations from later layers, allowing the network to maintain a stronger gradient signal throughout the network.

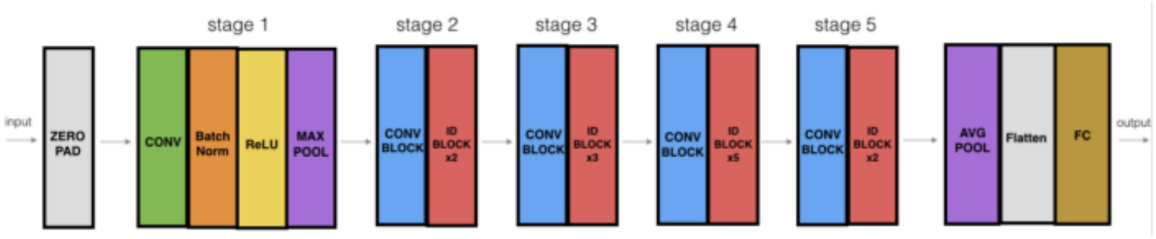
- This can help to improve the training process and allow the network to learn more complex features.



2. Convolutional Block



ResNet50 Architecture



Architecture of Inception module