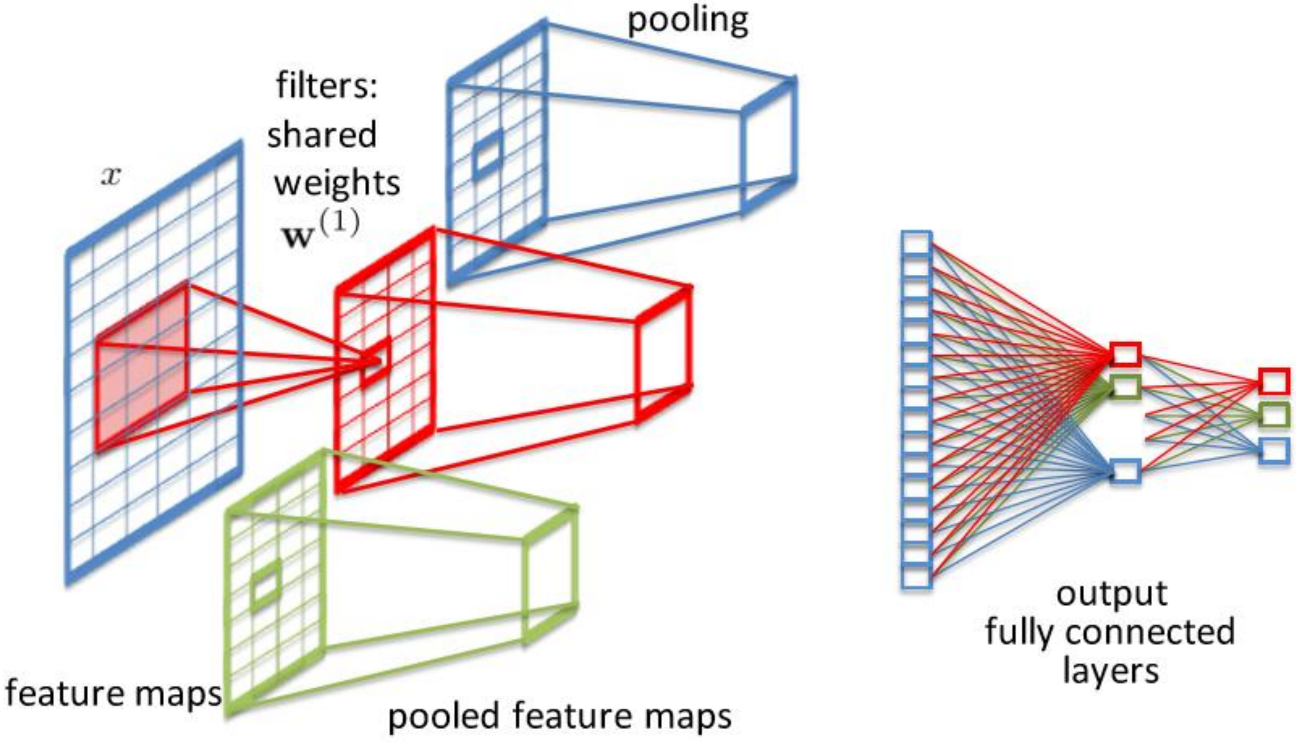
CNN FOR CBIR

RECAP:
WHAT IS WRONG WITH CONVENTIONAL FEATURES?

- ❖ Semantic Gap
- ❖ Not adaptive to learning spatial
- ❖ Only targeted features extraction

RECAP: THE SOLUTION?

- Convolutional neural networks (CNN)



BRIEF HISTORY: CNN

Yann LeCun, Professor of Computer Science
The Courant Institute of Mathematical Sciences
New York University
Room 1220, 715 Broadway, New York, NY 10003, USA.
(212)998-3283 yann@cs.nyu.edu

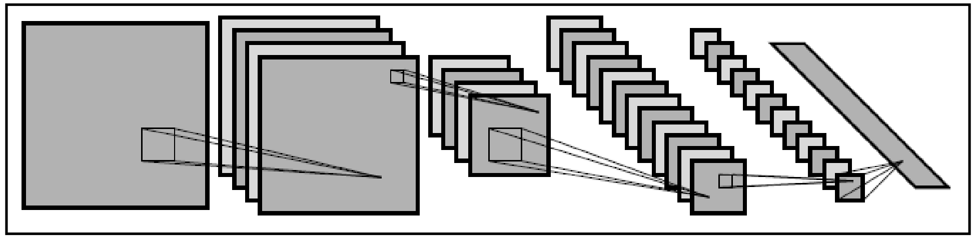
- In 1995, Yann LeCun and Yoshua Bengio introduced the concept of convolutional neural networks.

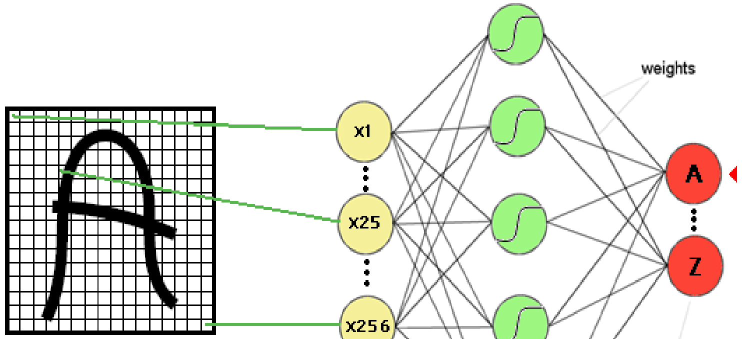
CONVOLUTIONAL NEURAL NETWORK

- CNN's Were neuro-biologically motivated by the findings of locally sensitive and orientation-selective nerve cells in the visual cortex.
- They designed a network structure that implicitly extracts relevant features.
- Convolutional Neural Networks are a special kind of multi-layer neural networks.
- CNN is a feed-forward network that can extract topological properties from an image.
- Like almost every other neural networks they are trained with a version of the back-propagation algorithm.
- Convolutional Neural Networks are designed to recognize visual patterns directly from pixel images with minimal preprocessing.
- They can recognize patterns with extreme variability (such as handwritten characters).

CONVOLUTION LAYER

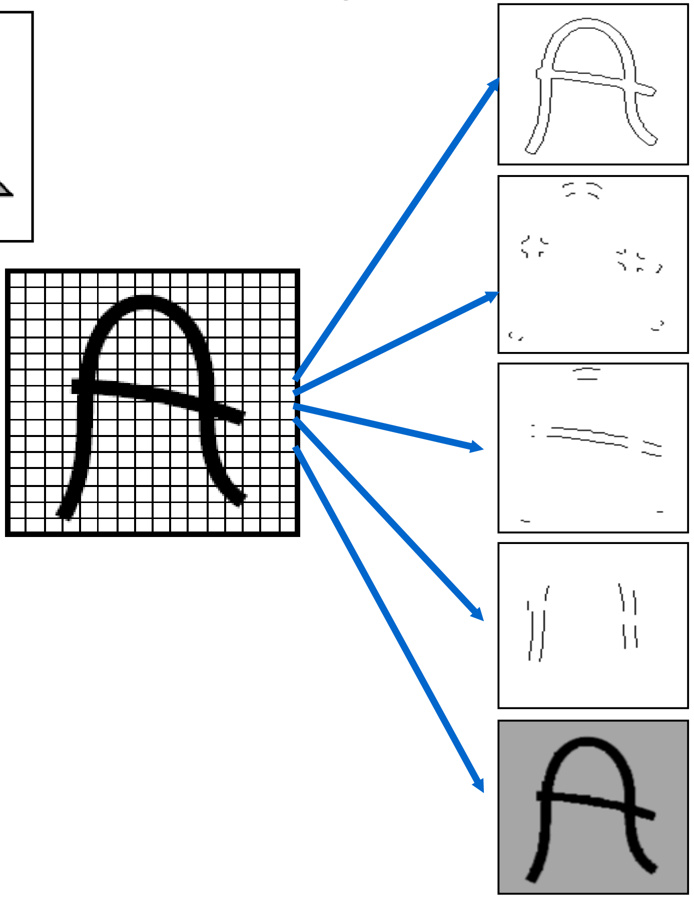
- Detect the same feature at different positions in the input image.



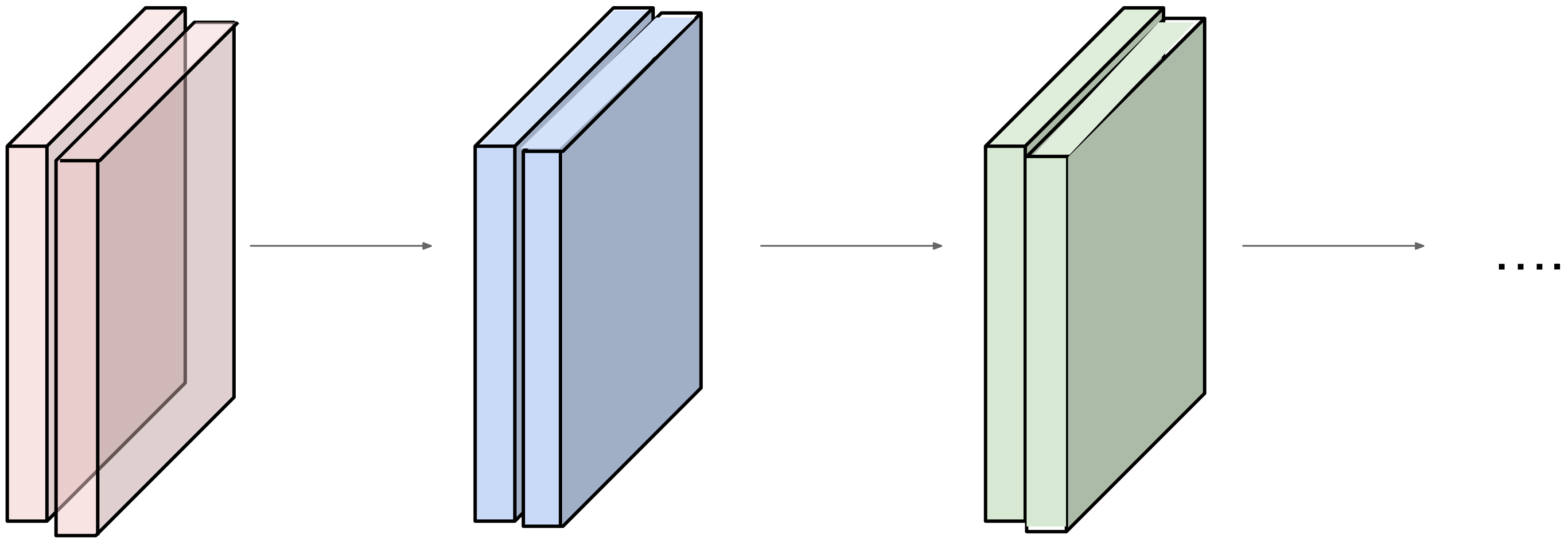


Input Image
16 * 16

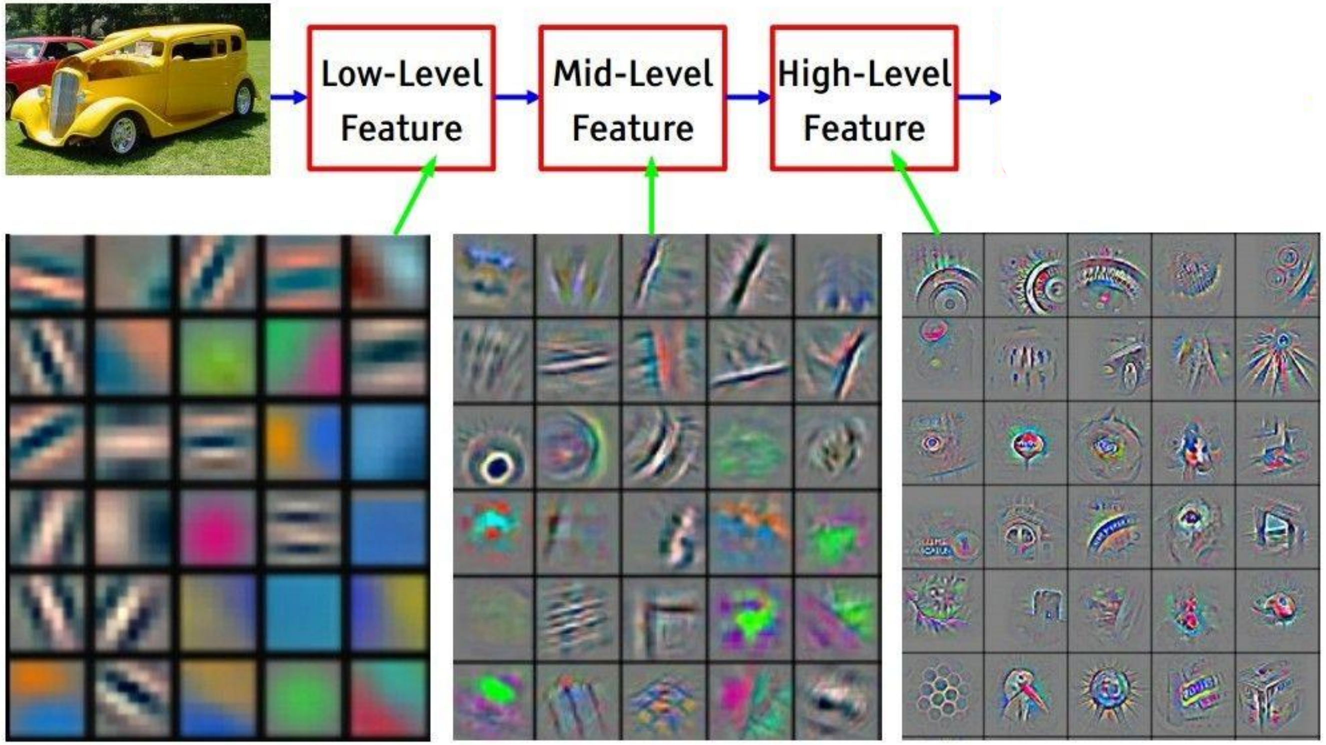
features



- **Preview:** ConvNet is a sequence of Convolutional Layers, interspersed with activation functions

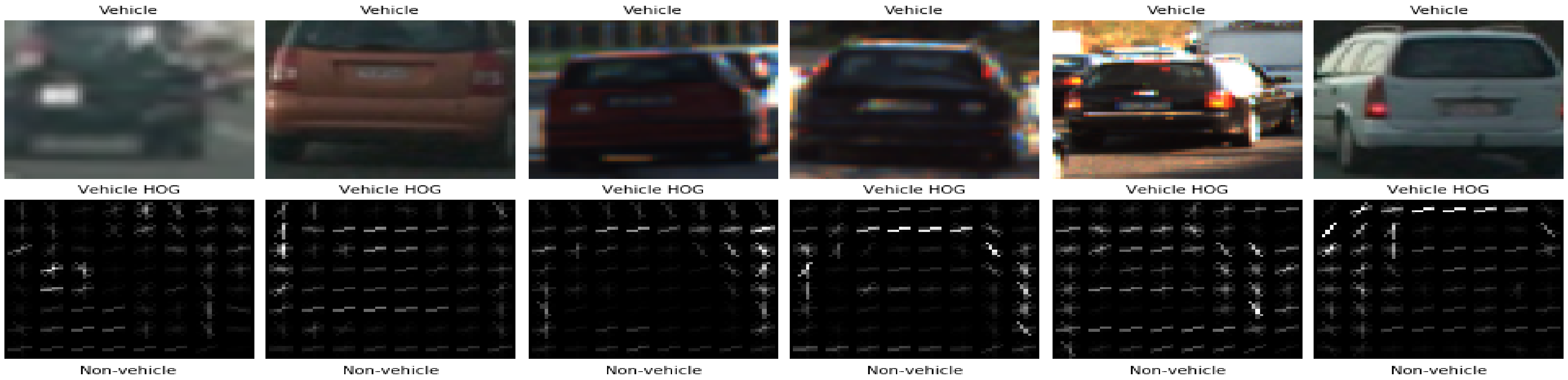


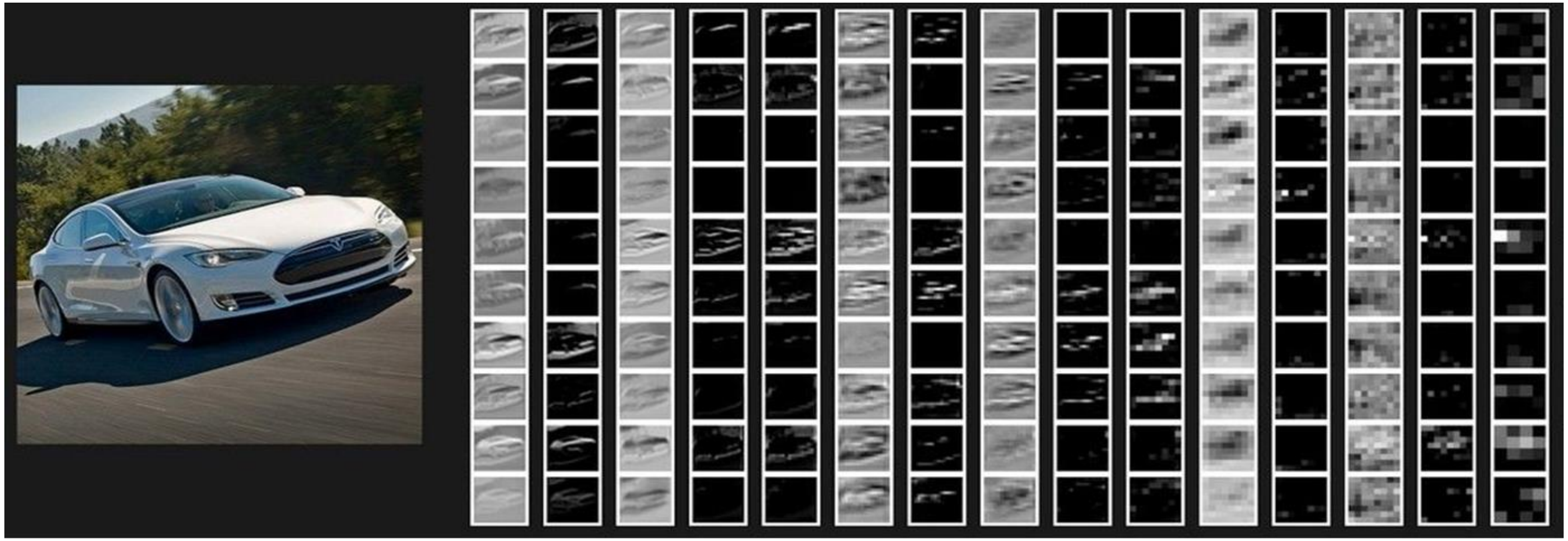
- **Preview**

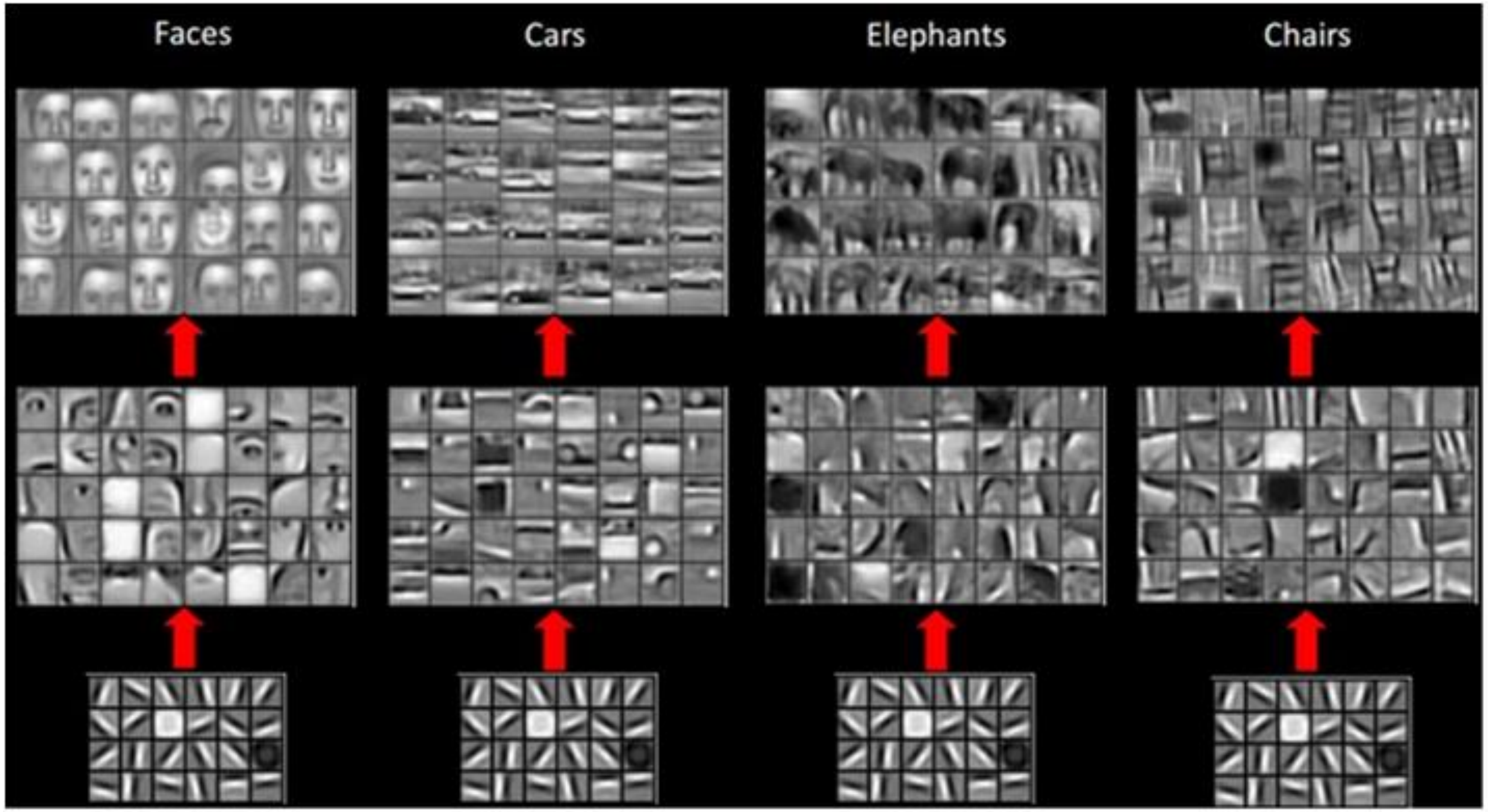


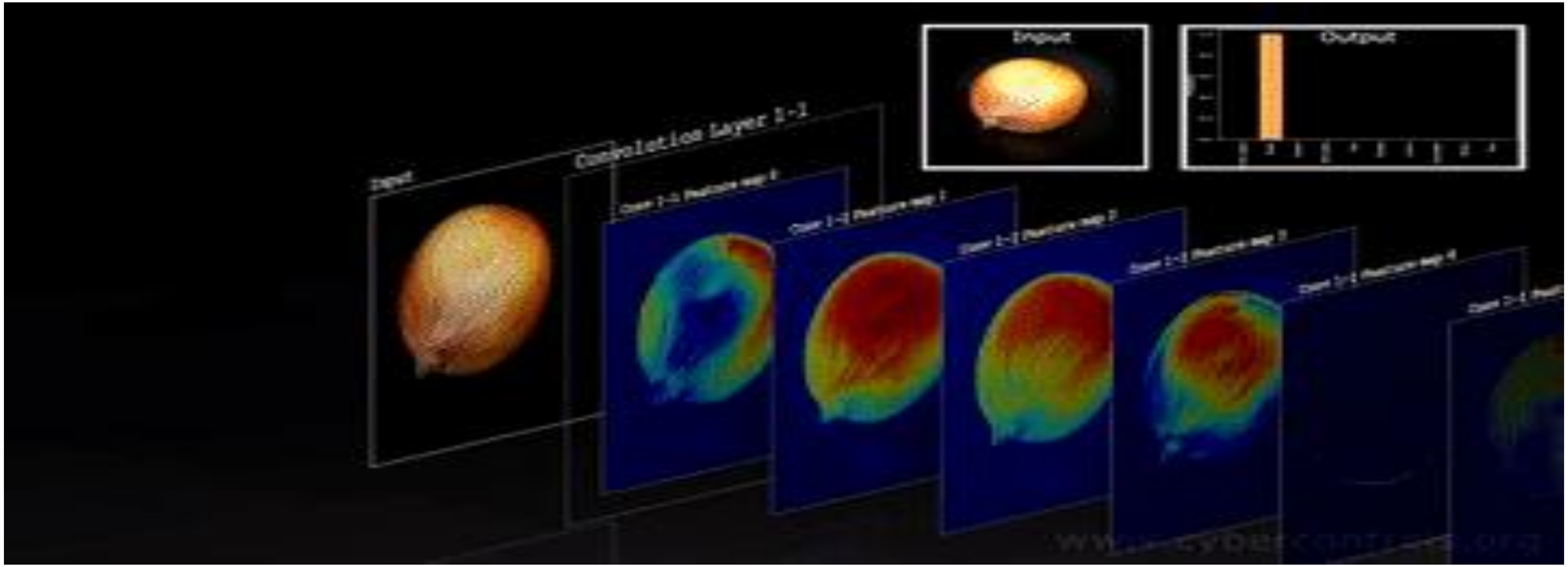
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

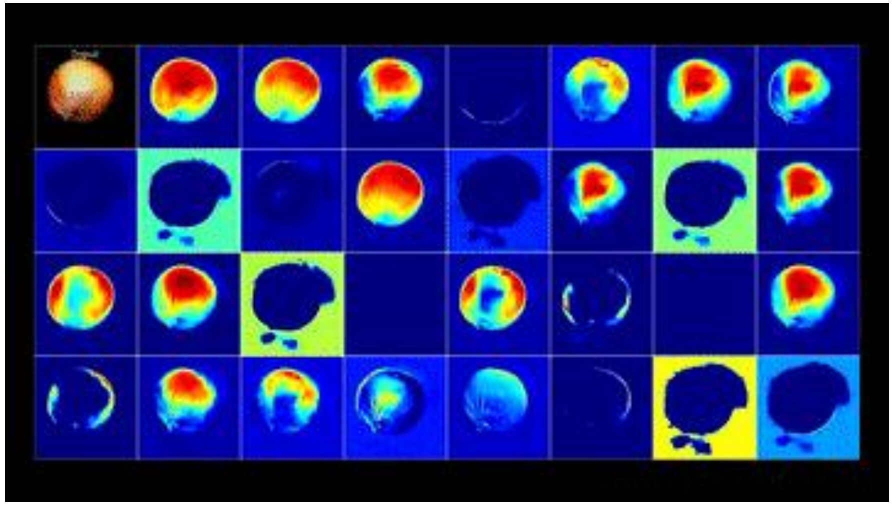
CONVENTIONAL VS CONVOLUTION FEATURES

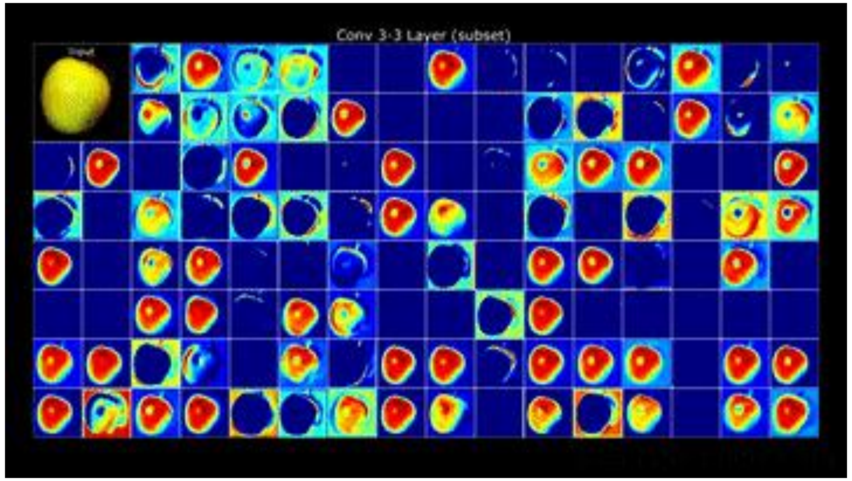


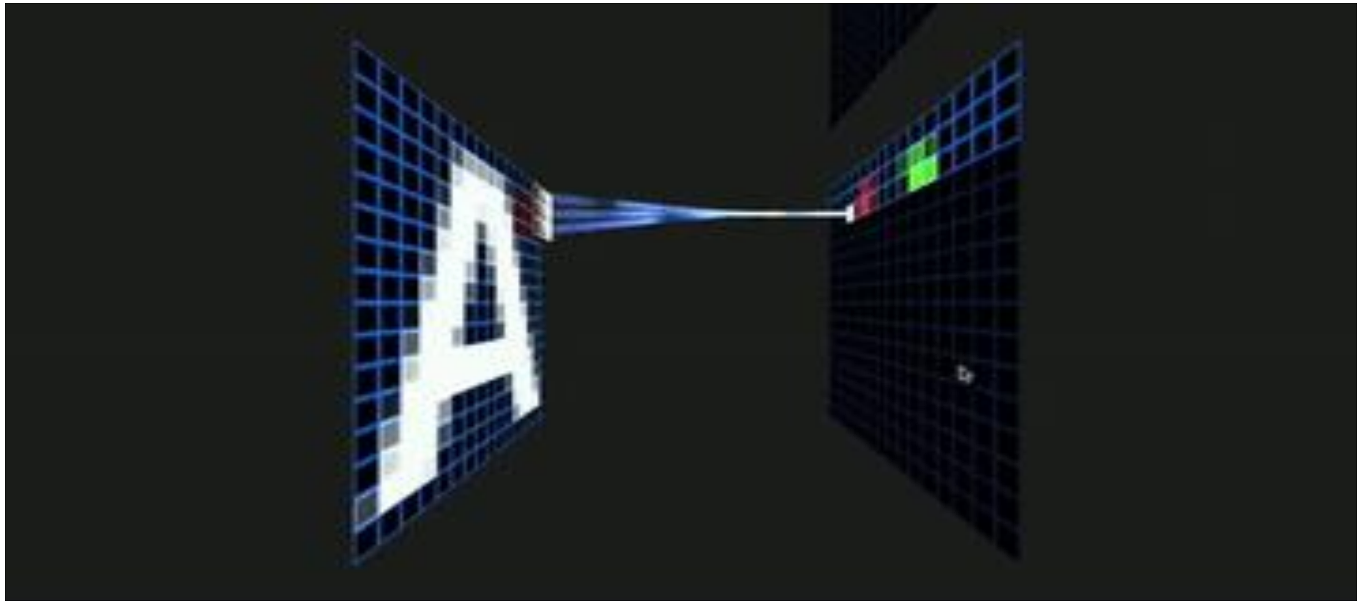


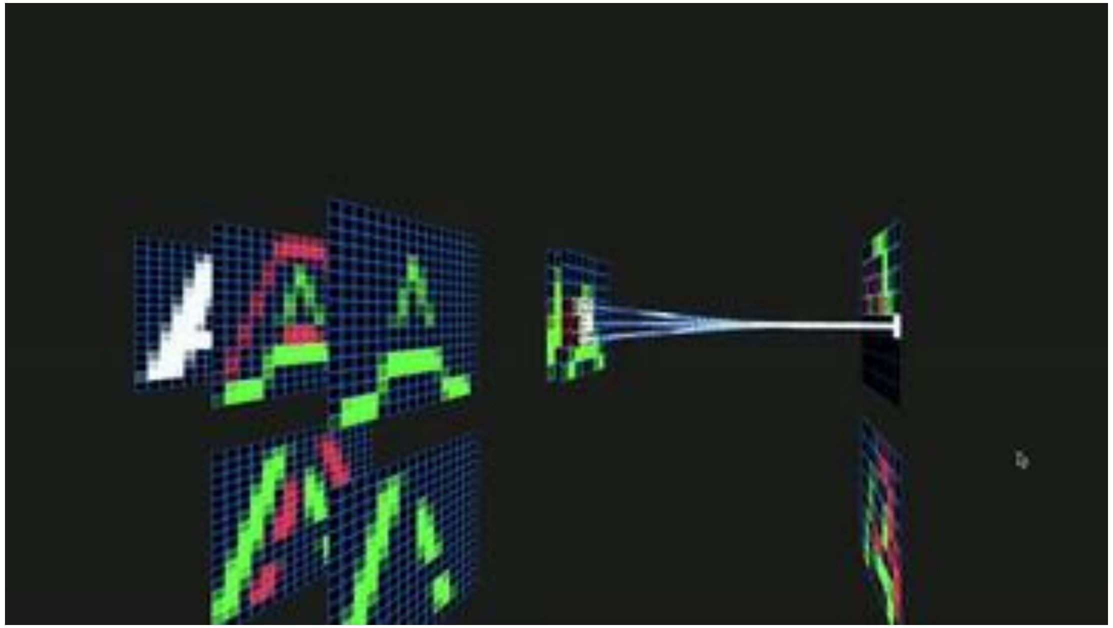

Faces
Cars
Elephants
Chairs


Input
Output
Input
Convolution Layer 1-1




Conv 3-3 Layer (subset)

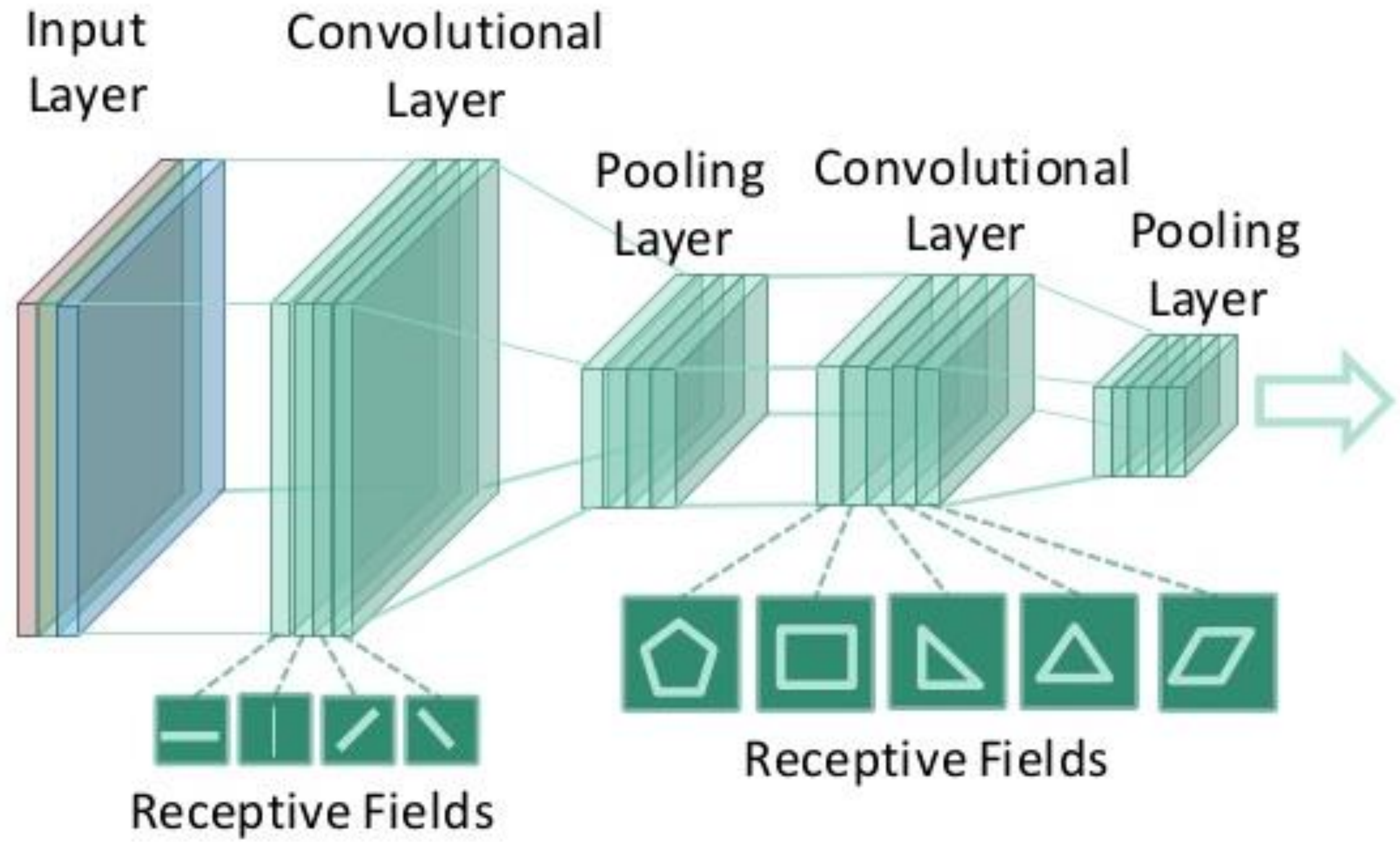




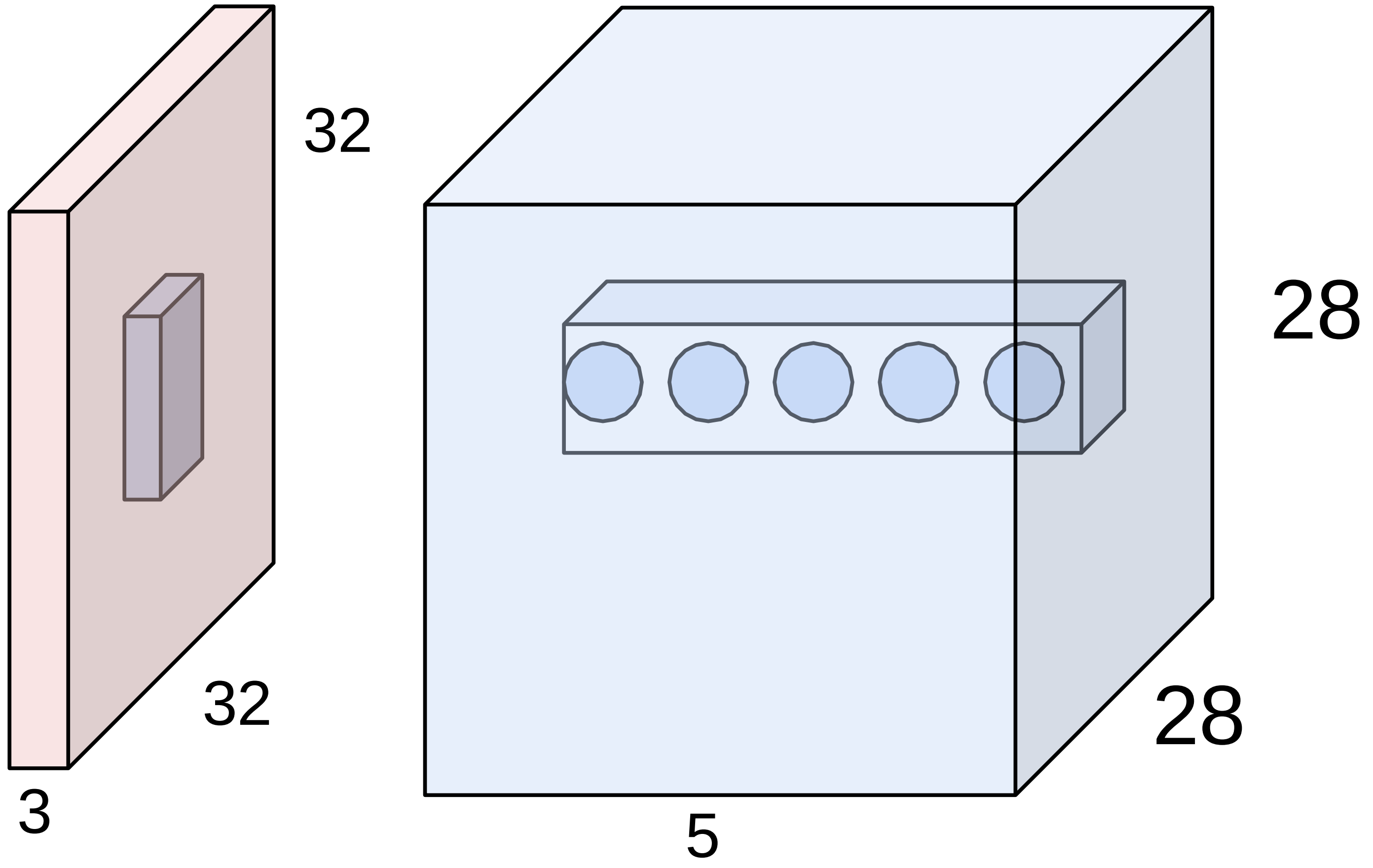
Summary. To summarize, the Conv Layer:

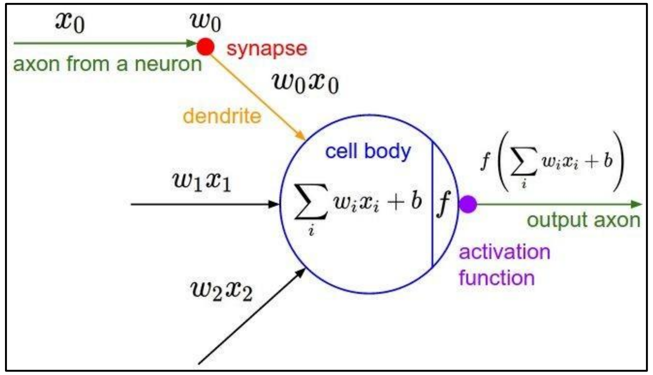
- Accepts a volume of size $W_1 \times H_1 \times D_1$
- Requires four hyperparameters:
  - Number of filters K,
  - their spatial extent F,
  - the stride S,
  - the amount of zero padding P.
- Produces a volume of size $W_2 \times H_2 \times D_2$ where:
  - $W_2 = (W_1 - F + 2P)/S + 1$
  - $H_2 = (H_1 - F + 2P)/S + 1$ (i.e. width and height are computed equally by symmetry)
  - $D_2 = K$
- With parameter sharing, it introduces $F \cdot F \cdot D_1$ weights per filter, for a total of $(F \cdot F \cdot D_1) \cdot K$ weights and K biases.
- In the output volume, the d-th depth slice (of size $W_2 \times H_2$) is the result of performing a valid convolution of the d-th filter over the input volume with a stride of S, and then offset by d-th bias.

VISUAL PERCEPTION OF COMPUTER



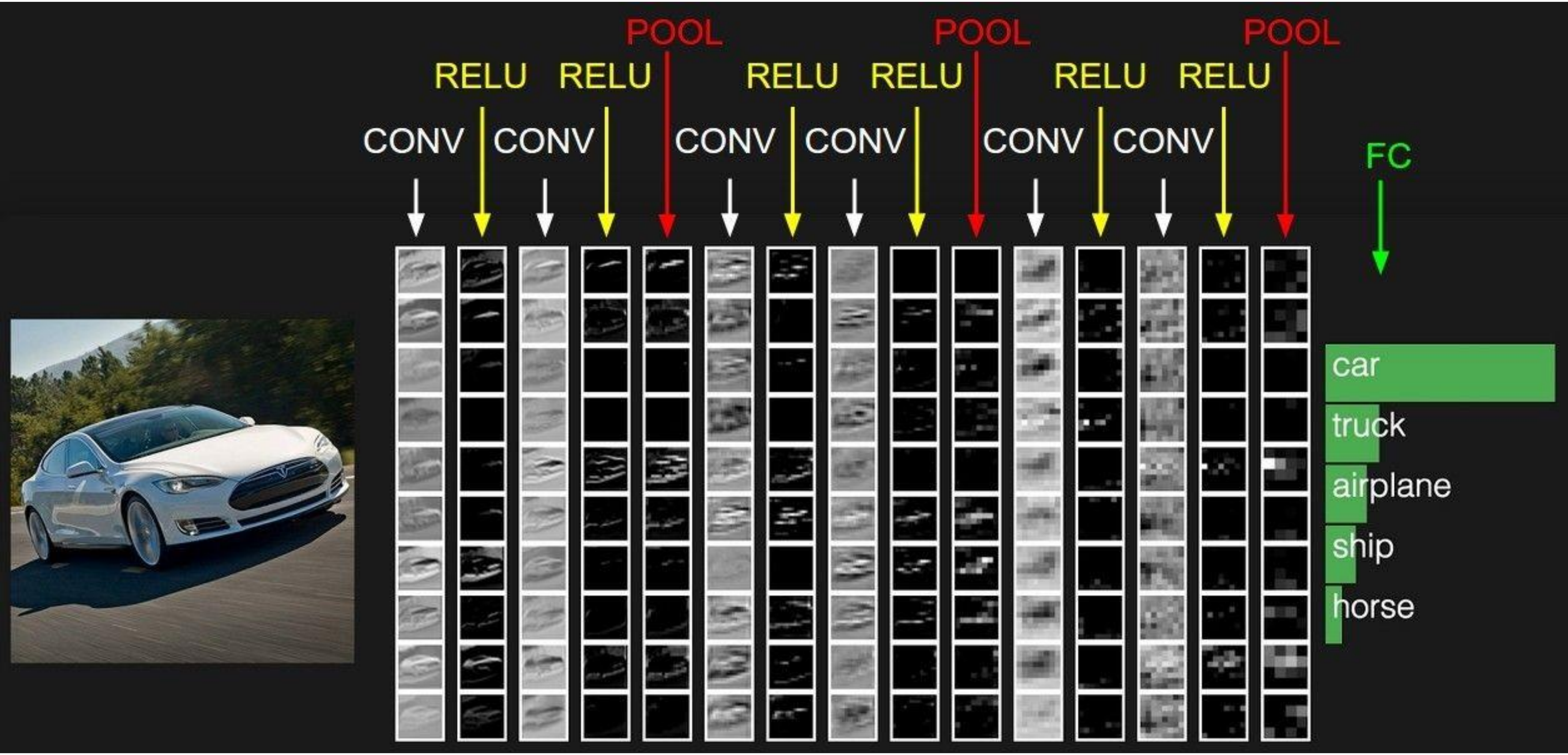
THE BRAIN/NEURON VIEW OF CONV LAYER





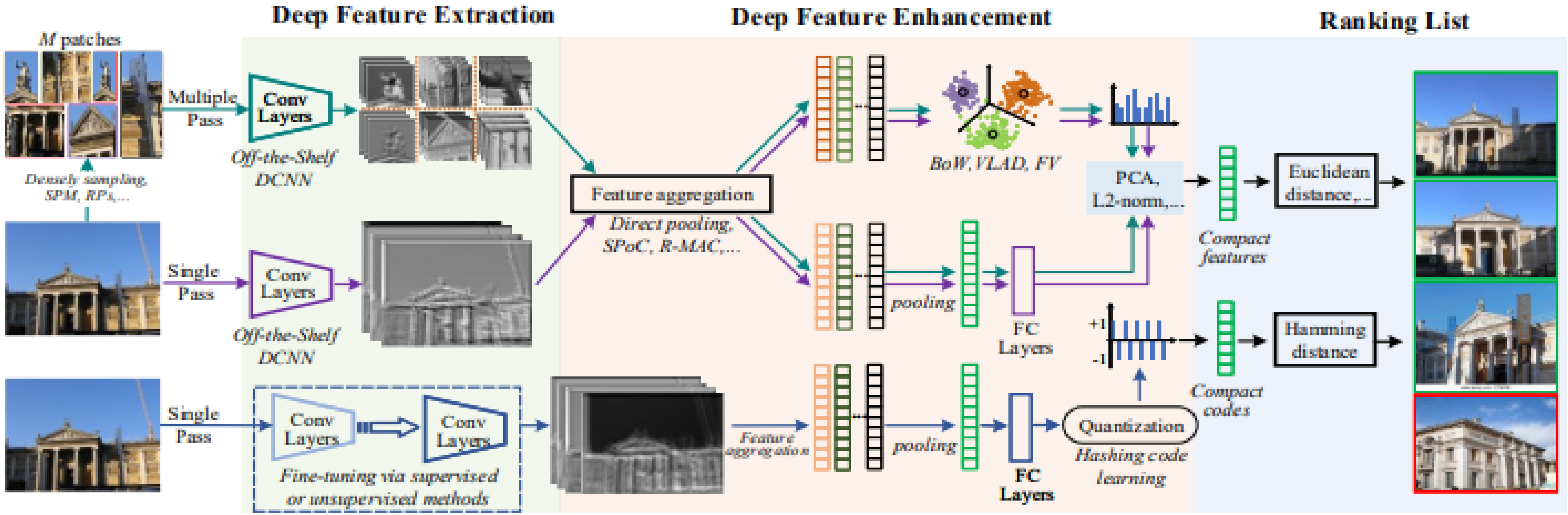
E.g. with 5 filters,
CONV layer consists of neurons arranged in a 3D grid (28x28x5)

There will be 5 different neurons all looking at the same region in the input volume two more layers to go: POOL/FC



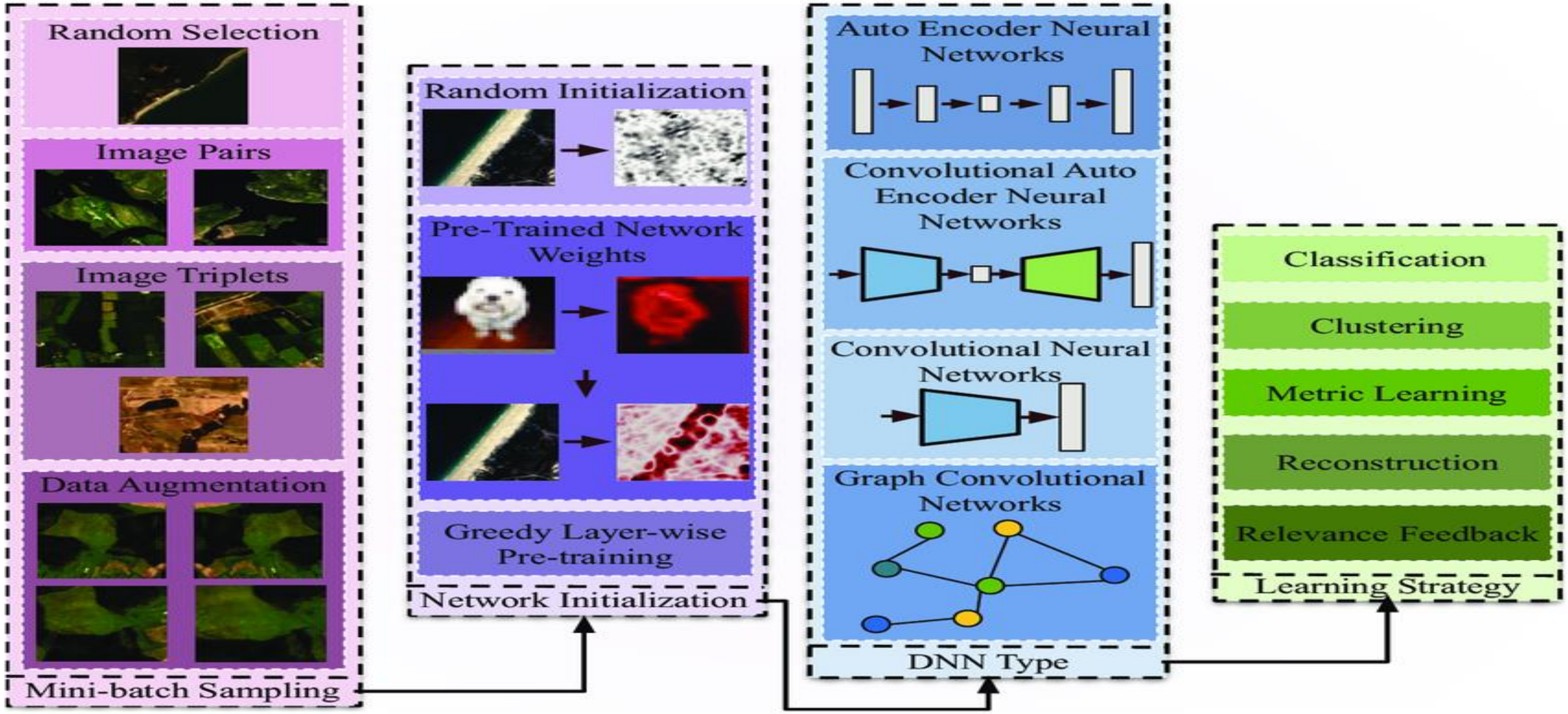
CONVOLUTION IN CBIR

- CNNs automatically learn hierarchical feature representations from raw input images.
- They capture both low-level and high-level features, making them effective for recognizing objects and patterns.
- CNNs extract compact and discriminative image representations.
- These representations serve as effective features for content-based image retrieval.



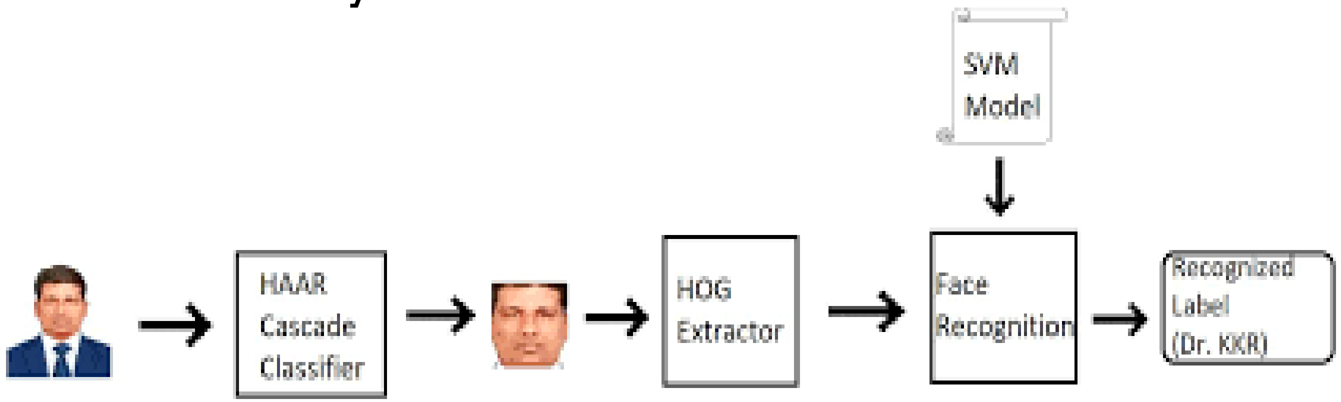
DIFFERENT STRATEGIES CONSIDERED WITHIN DL BASED CBIR SYSTEMS.

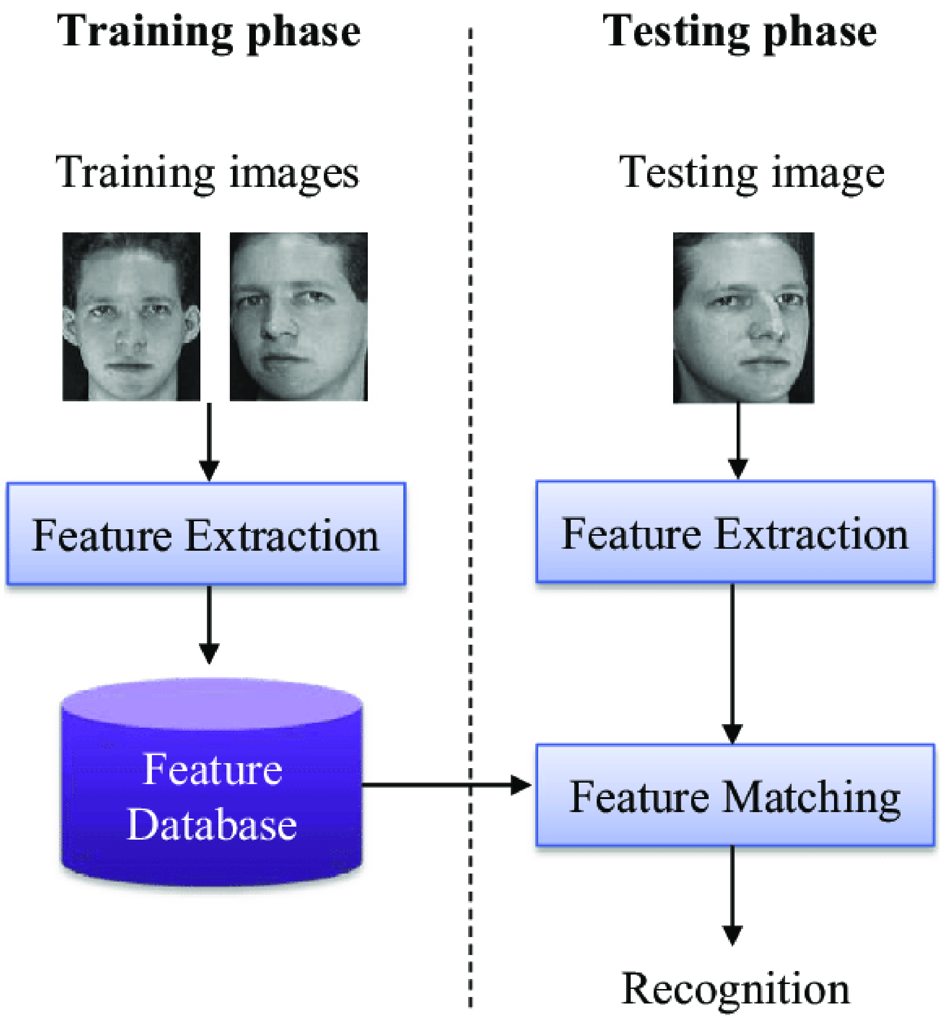
❖ combining deep learning techniques, attention mechanisms, and domain-specific adaptations enhances CBIR systems' effectiveness.



CASE STUDY: FACE-RECOGNITION

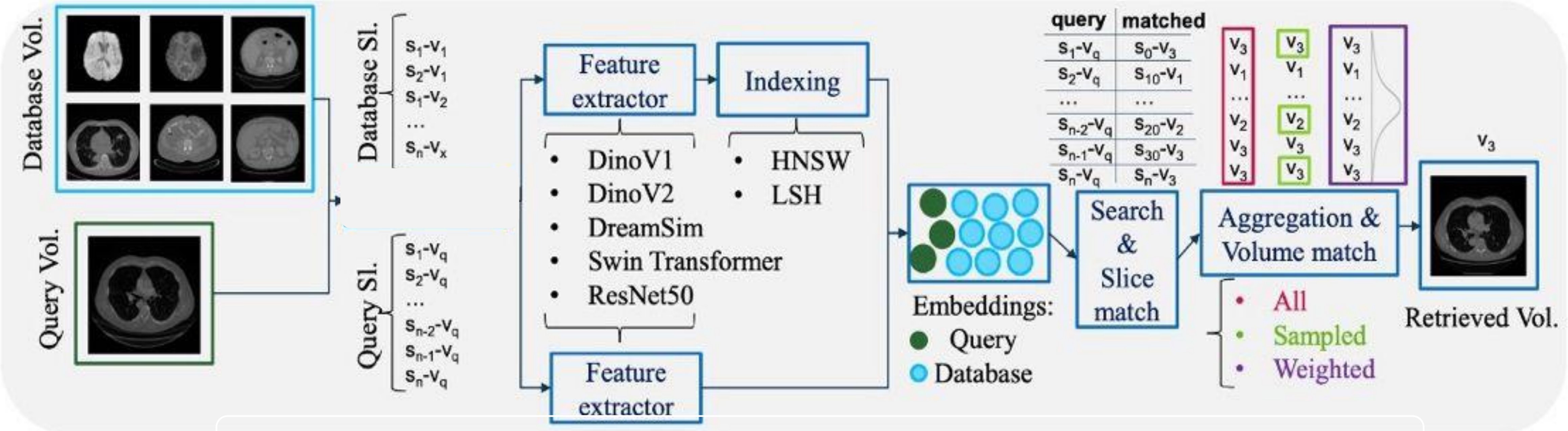
- Face recognition, also known as facial recognition, identifies or verifies individuals based on their facial features.
- It involves capturing an image of a person's face and comparing it to a database of known faces to determine their identity.





CASE STUDY: MEDICAL IMAGE RETRIEVAL

- ❖ CBIR plays a crucial role in medical imaging, and leveraging pretrained models can enhance retrieval performance.
- ❖ The transferability of features from natural images to medical tasks is an active area of research.
- ❖ CBIR assists in medical image analysis, such as retrieving similar X-rays, MRIs, or histopathology images based on visual content.
- ❖ It aids in diagnosing diseases, tracking treatment progress, and identifying anomalies

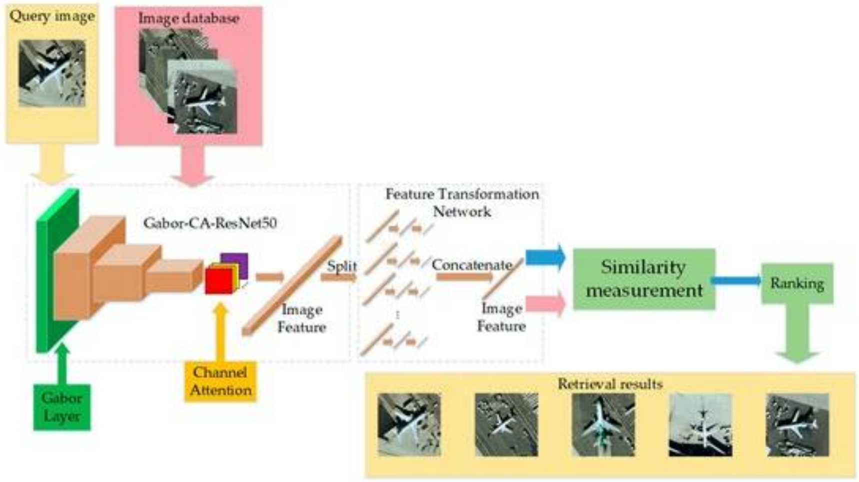


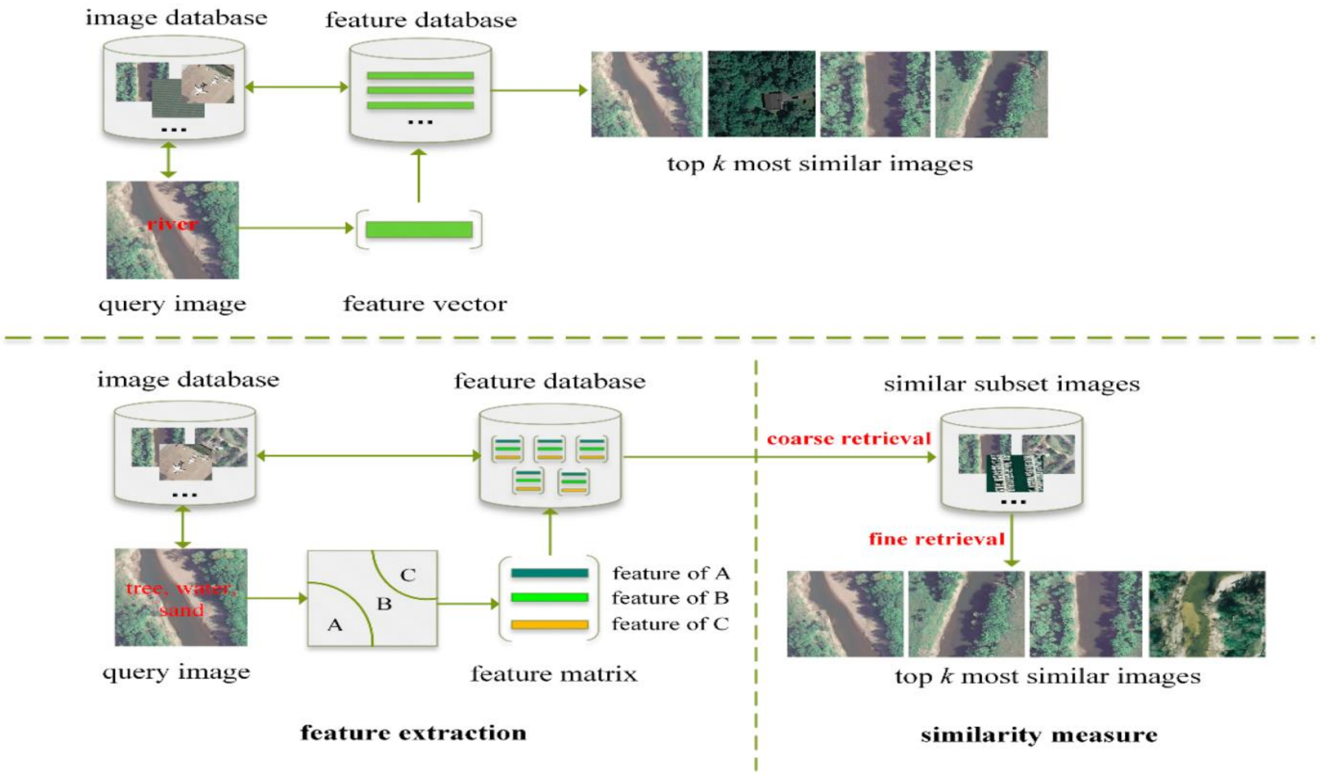
Medical Image Retrieval Using Pretrained Embeddings [1]

CASE STUDY: REMOTE SENSING

- ❖ Change Detection:
  - ❖ CBIR helps identify changes between satellite images captured at different times.
- ❖ Disaster Monitoring and Damage Assessment

- ❖ Urban Planning and Infrastructure Management
- ❖ CBIR techniques enhance remote sensing applications by intelligently retrieving relevant images for analysis, monitoring, and decision-makin

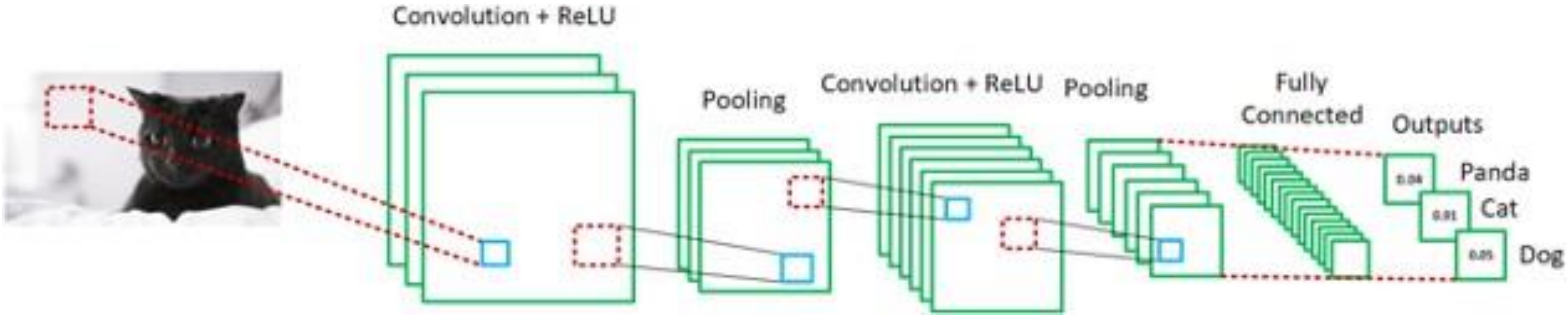




TRANSFER LEARNING - 2 MAJOR TYPES

Feature Extractor

Fine Tuning



TRANSFER LEARNING

- **Transfer learning** or inductive transfer is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem.
- For example, knowledge gained while learning to recognize cars could apply when trying to recognize trucks. This area of research bears some relation to the long history of psychological literature on transfer of learning, although formal ties between the two fields are limited.

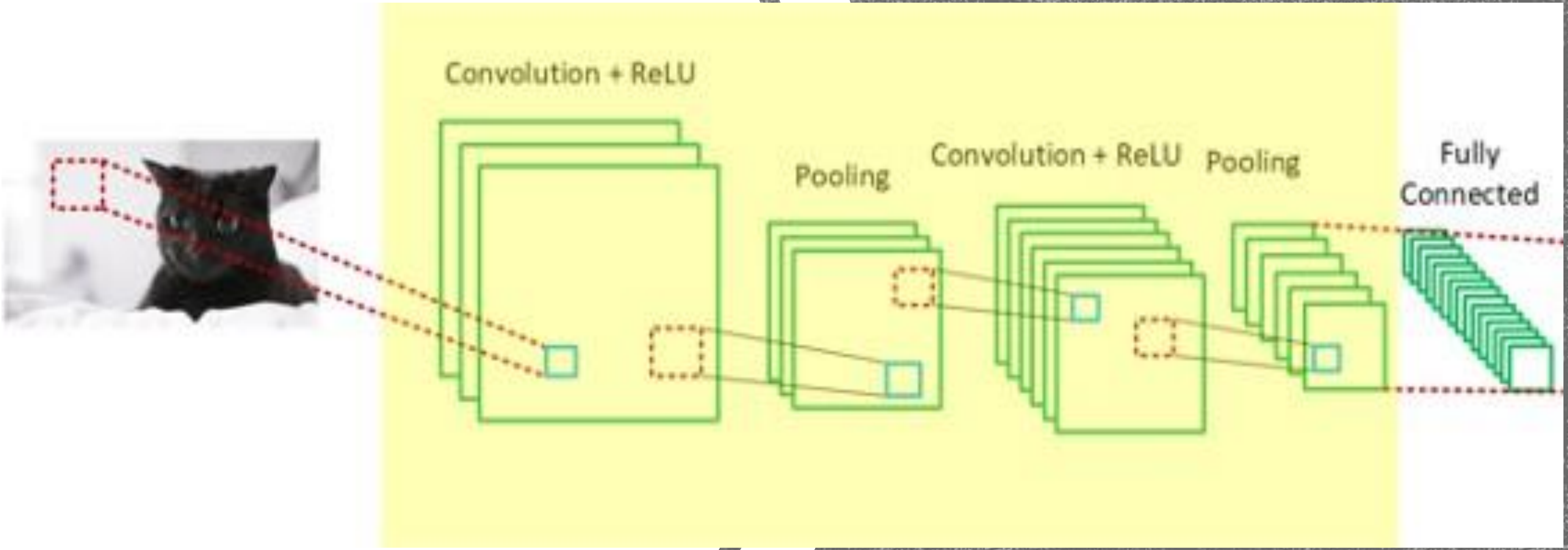
TRANSFER LEARNING

- The three major Transfer Learning scenarios look as follows:

1. **ConvNet as fixed feature extractor**.

- Take a ConvNet pretrained on ImageNet, remove the last fully-connected layer (this layer's outputs are the 1000 class scores for a different task like ImageNet), then treat the rest of the ConvNet as a fixed feature extractor for the new dataset.

1. CONVNET AS FIXED FEATURE EXTRACTOR.

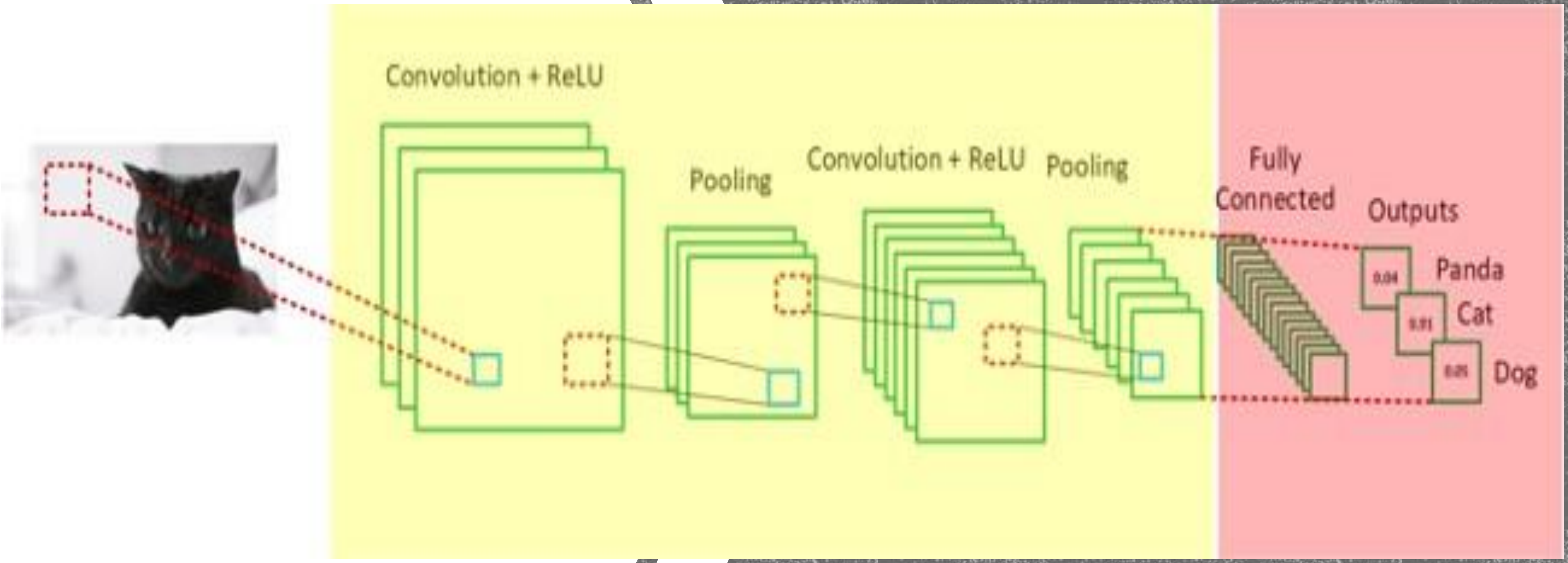


TRANSFER LEARNING

2. Fine-tuning the ConvNet.

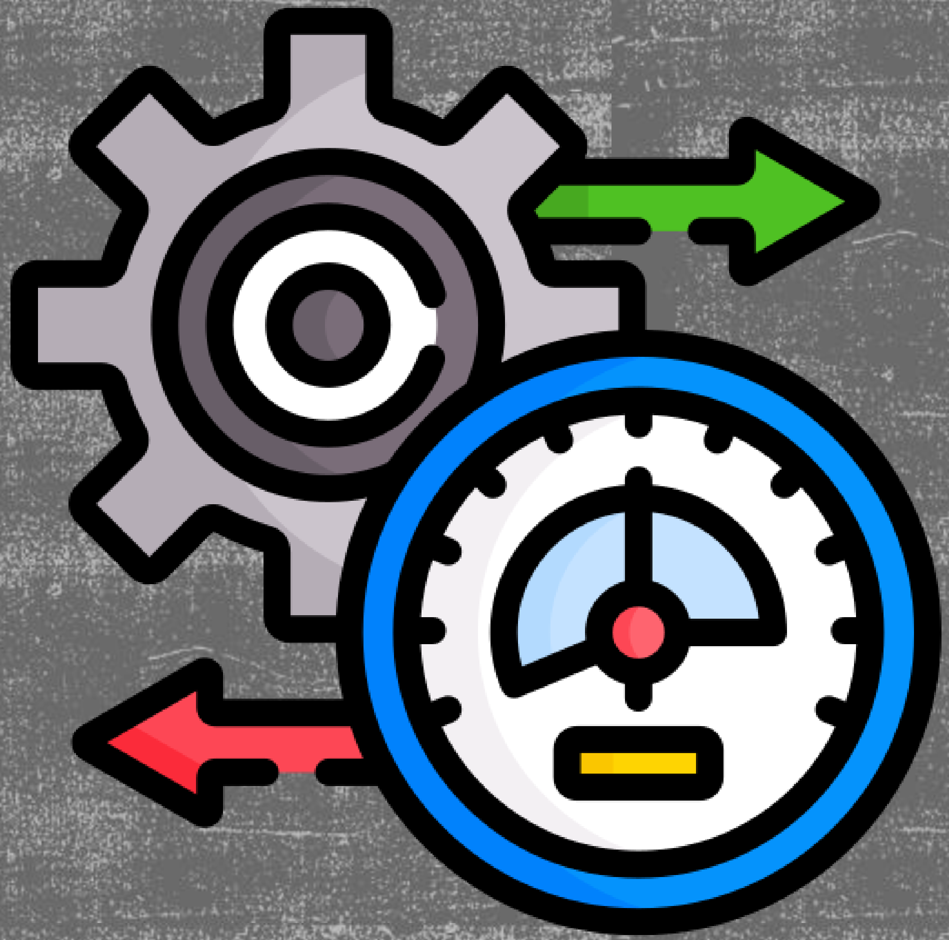
- The second strategy is to not only replace and retrain the classifier on top of the ConvNet on the new dataset, but to also fine-tune the weights of the pretrained network by continuing the backpropagation. It is possible to fine-tune all the layers of the ConvNet, or it's possible to keep some of the earlier layers fixed (due to overfitting concerns) and only fine-tune some higher-level portion of the network.

2. FINE-TUNING THE CONVNET



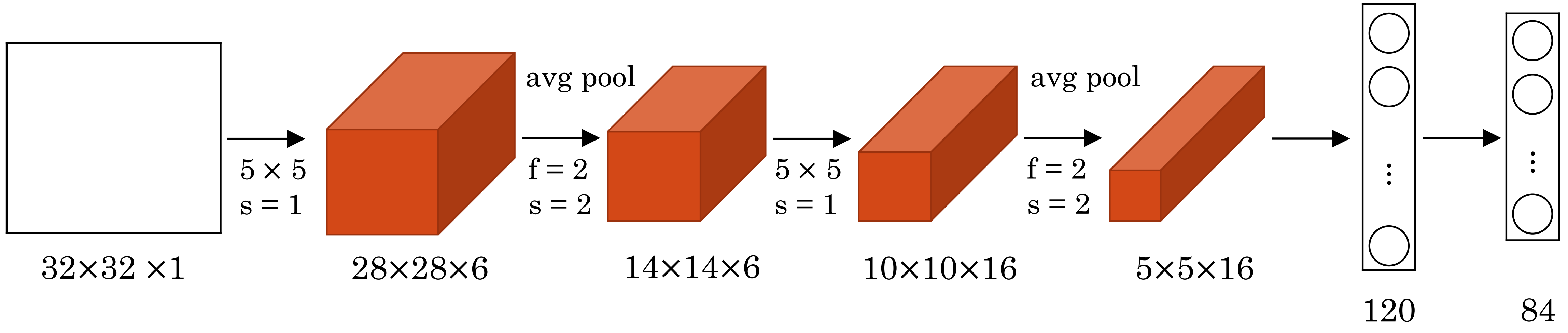
WHY FINE TUNE FOR RETRIEVAL

- Fine-tuning deep learning models for retrieval ensures that the features extracted are better aligned with the specific task, leading to improved performance.
- Designing your own fully connected layers, you can ensure that the semantics align with your retrieval task.
- Reducing Dimensionality:
  Use fewer neurons in the fully connected layers
  If you want smaller feature vectors.

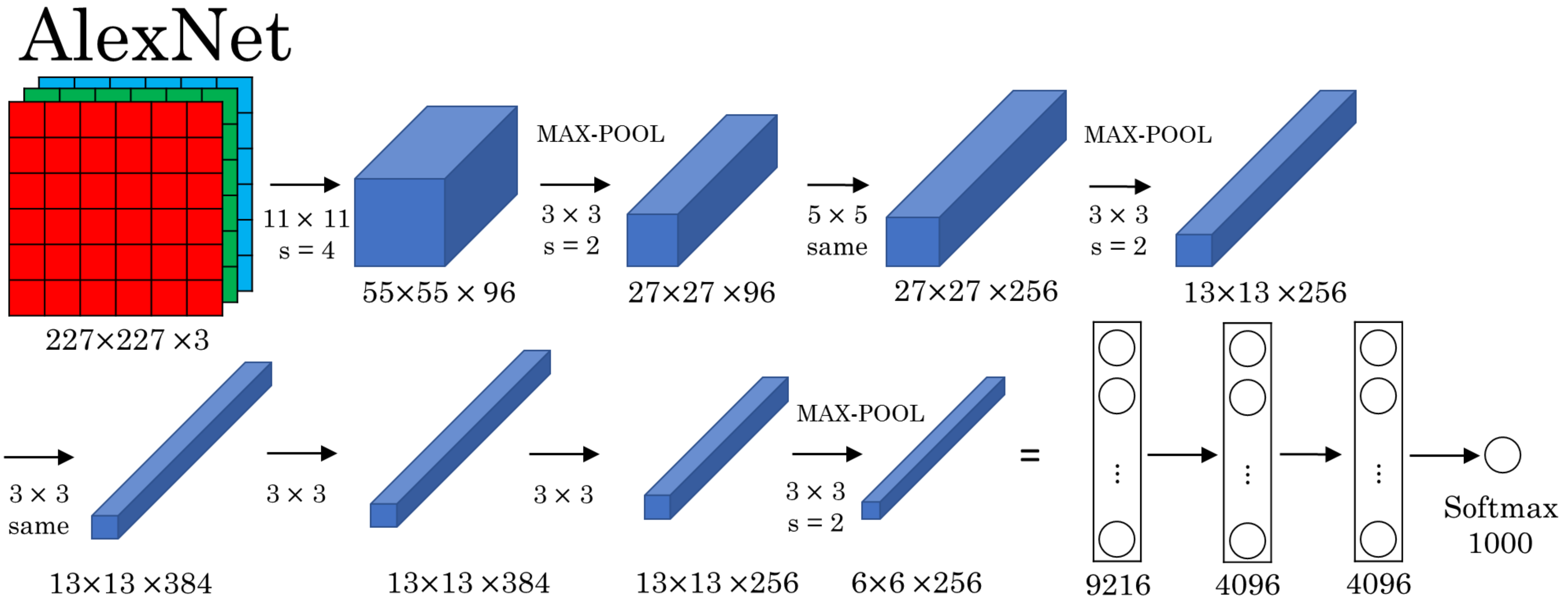


PLAIN NETWORKS

- Simple CNN architectures where connections among the layers follow simple straight-forward paths
- E.g. LeNet-5



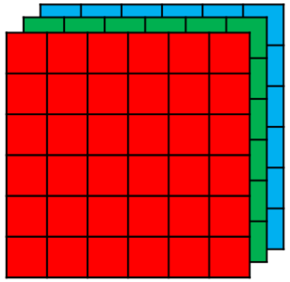
PLAIN NETWORKS



THE LAST PLAIN NETWORKS

VGG-16/19

CONV = 3×3 filter, s = 1, same   MAX-POOL = 2×2 , s = 2

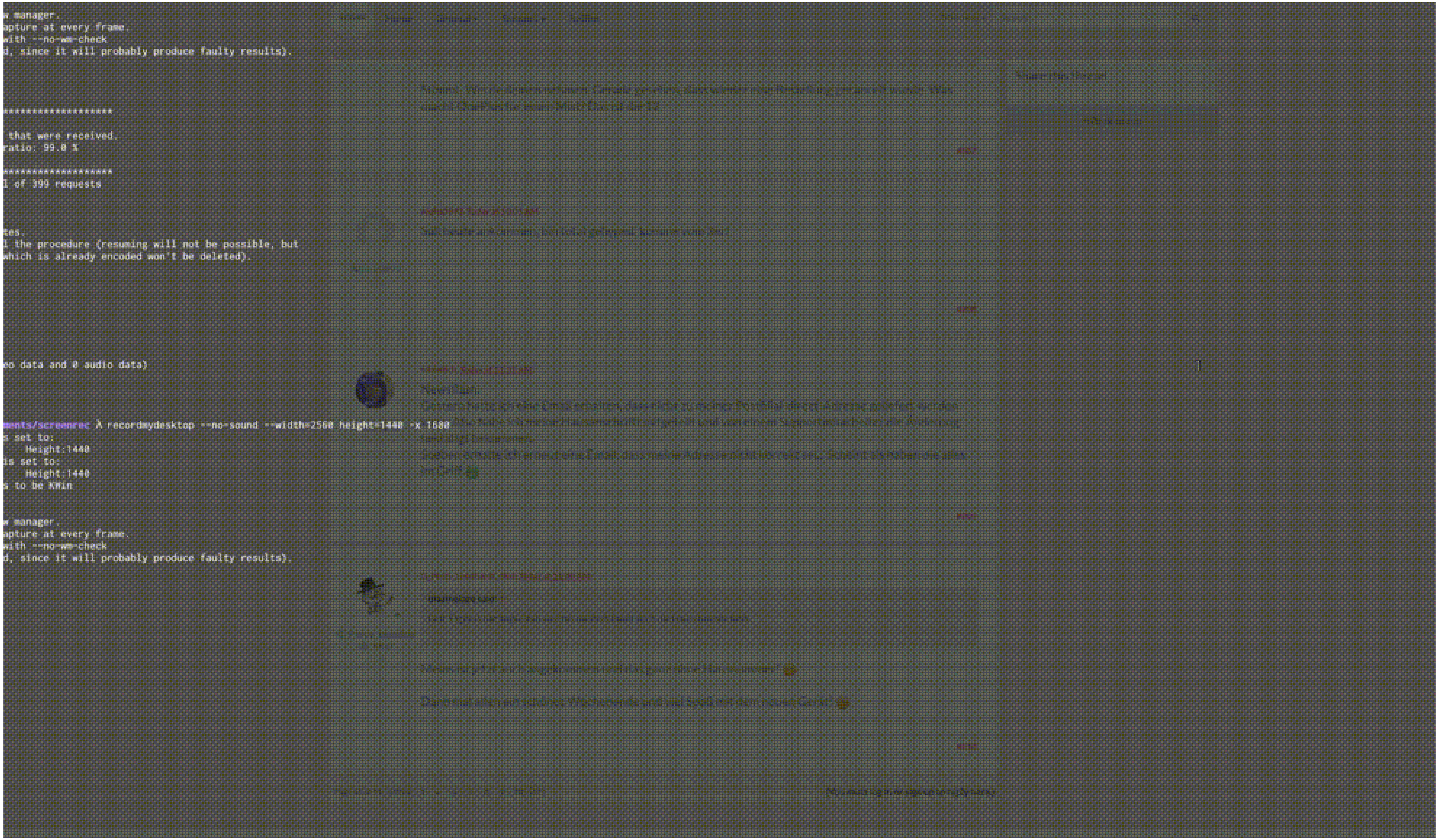


224×224 ×3

→ [CONV 64] ×2 → 224×224×64 → POOL → 112×112 ×64 → [CONV 128] ×2 → 112×112 ×128 → POOL → 56×56 ×128

→ [CONV 256] ×3 → 56×56 ×256 → POOL → 28×28 ×256 → [CONV 512] × 3 → 28×28 ×512 → POOL → 14×14×512

→ [CONV 512] ×3 → 14×14 ×512 → POOL → 7×7×512 → FC 4096 → FC 4096 → Softmax 1000


w manager.
apture at every frame.
with --no-wm-check
d, since it will probably produce faulty results).
****
that were received.
ratio: 99.0 %
****
1 of 399 requests
tes.
1 the procedure (resuming will not be possible, but
which is already encoded won't be deleted).
eo data and 0 audio data)
ments/screenrec λ recordmydesktop --no-sound --width=2560 height=1440 -x 1600
s set to:
Height:1440
is set to:
Height:1440
s to be KWin
w manager.
apture at every frame.
with --no-wm-check
d, since it will probably produce faulty results).

REFERENCES

Medical Image Retrieval Using Pretrained Embeddings
F. K. Jush, T. Truong, S. Vogler and M. Lenga
arXiv preprint arXiv:2311.13547 2023

A Content-Based Remote Sensing Image Change Information Retrieval Model

http://ruder.io/optimizing-gradient-descent/