Research Institute for Future Media Computing
未来媒体技术与研究所
Institute of Computer Vision
计算机视觉研究所

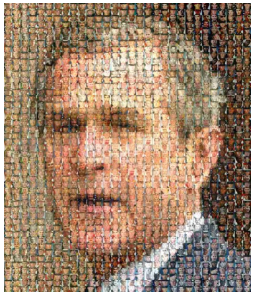
多媒体检索
Image Retrieval

Background

- Necessity of **retrieval**
  - *Information is of no use, unless you can actually access it.*

[from the TREC homepage: trec.nist.gov]

- Why do we need image retrieval?
  - "A Picture is worth thousand words"
  - Not everything can be described in text
  - Not everything is described in text



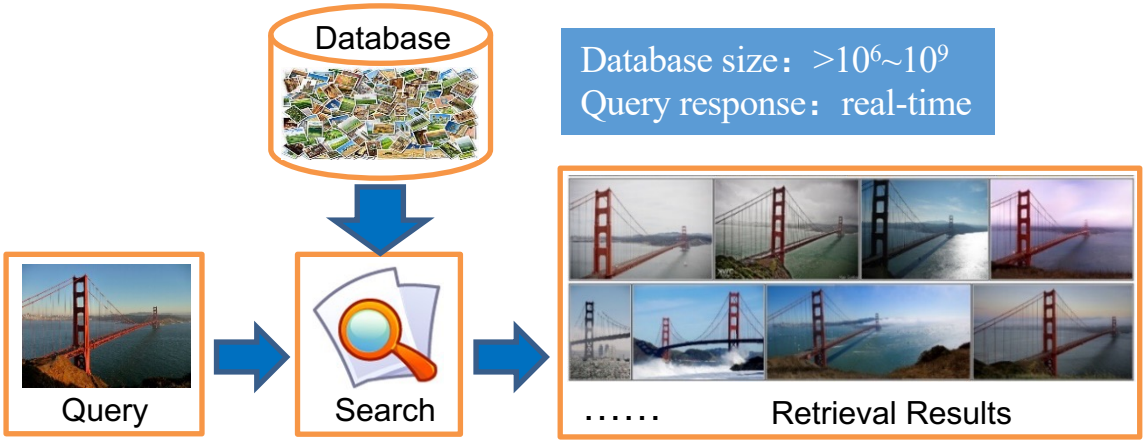
Background


Google
在 Google 上搜索，或者输入一个网址
Bai du 百度
百度一下


Bai du 图片
热门搜索:
壁纸
图腾
头像
高清动漫

Background

- Content based image retrieval



- Potential applications of content-based image retrieval

Image Retrieval

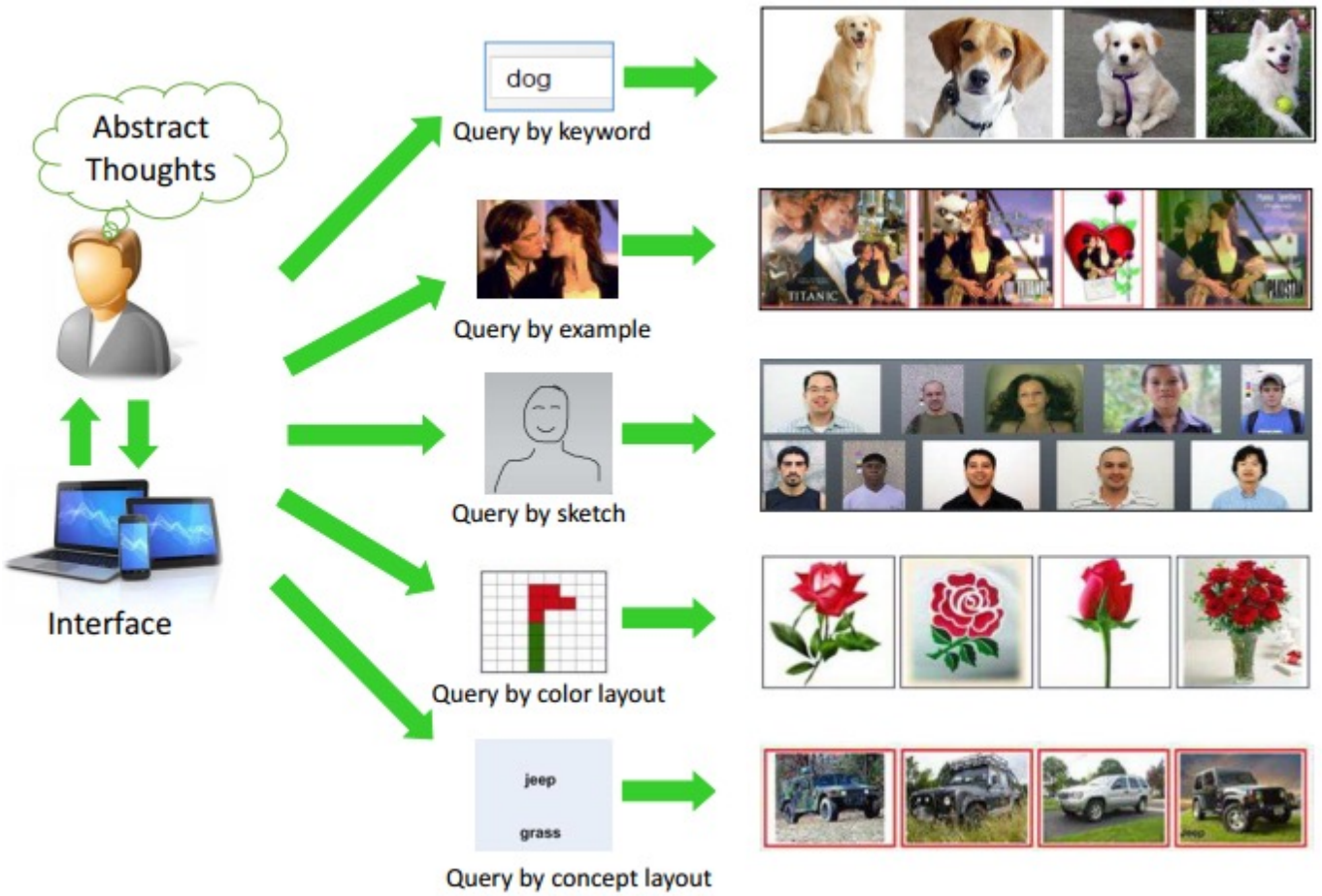


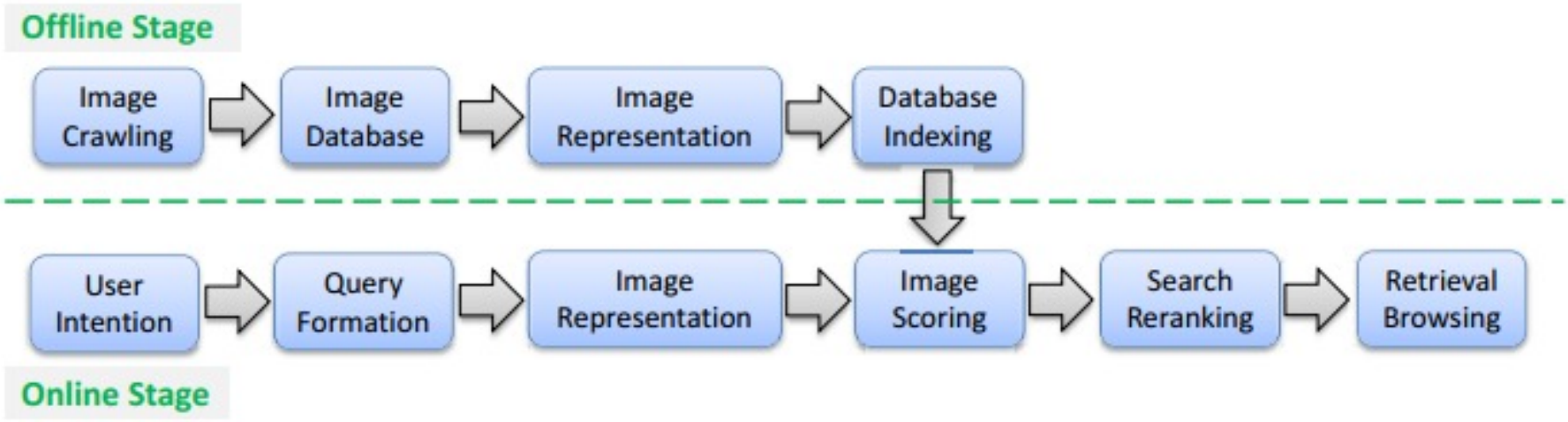
Illustration of different query schemes with the corresponding retrieval results

Why is Image Retrieval Hard ?

- What is the topic of this image
- What are right keywords to index image
- What words would you use to retrieve this image ?
- The ***Semantic Gap***

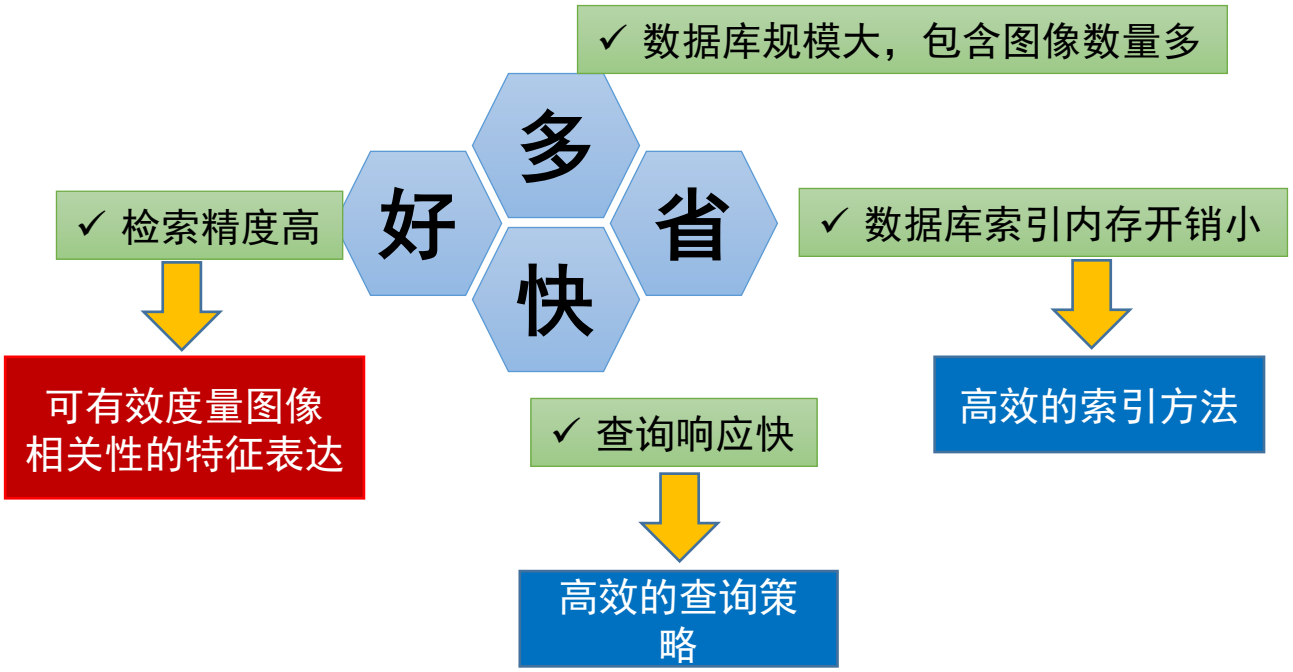
- A picture is worth a thousand words
- The meaning of an image is highly individual and subjective

Framework



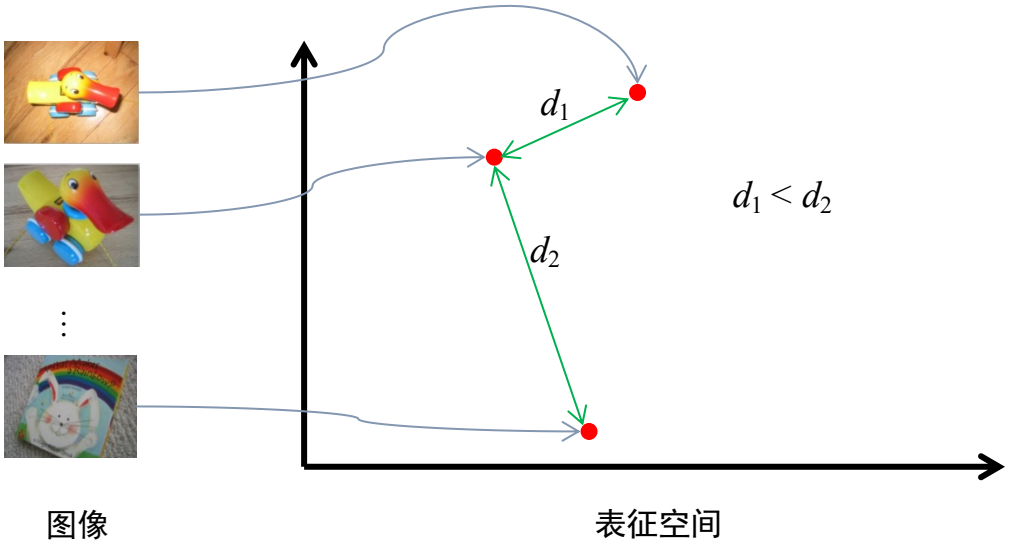
The general framework of content-based image retrieval

Problems with Image Retrieval



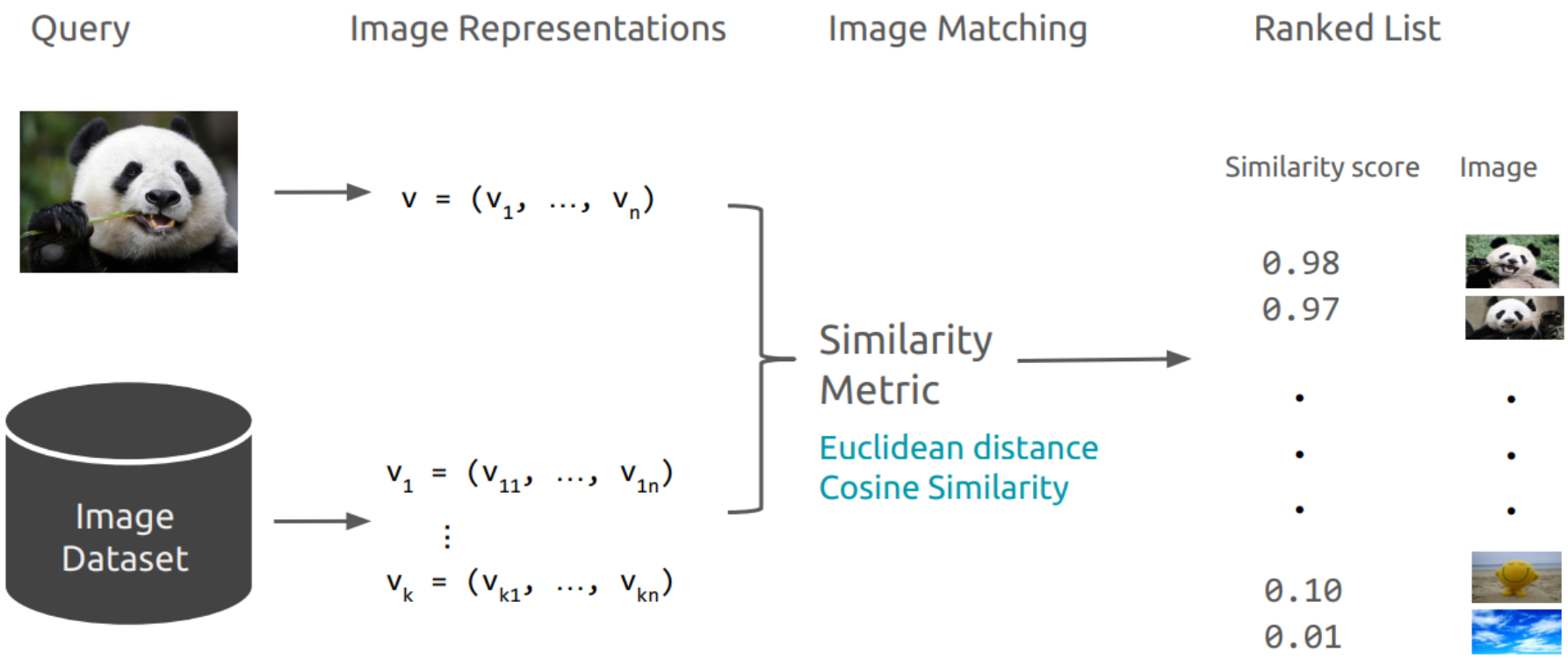
Basic Problems

- 图像检索基本问题之一：如何计算图像间的内容相关性?
  - **图像表征**：非结构化图像数据的结构化表达
    - SIFT + BoW/VLAD/FV
    - 神经网络中间层
  - **相似性度量：基于图像表征的相关性计算**



Based on Hand-Crafted Features

- The retrieval pipeline

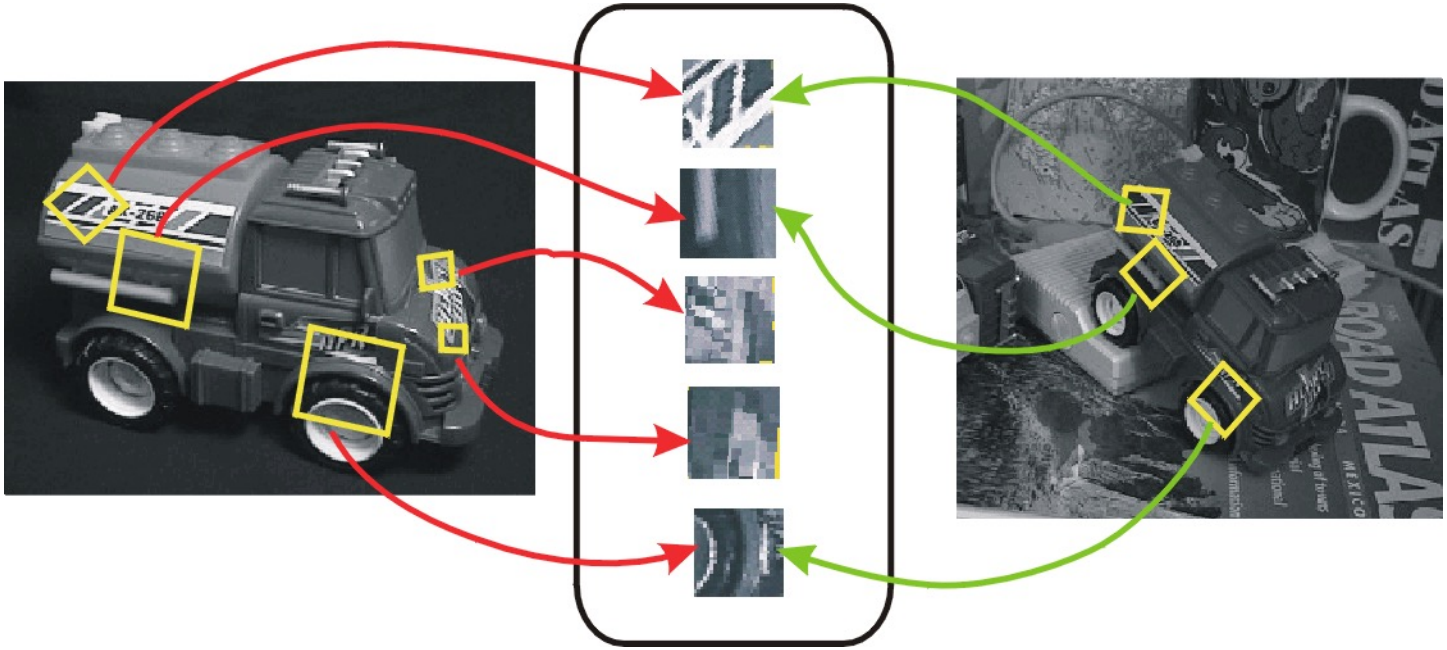


Global Features

- Color (histograms, gridded layout, wavelets)
- Texture (Laws, Gabor filters, LBP, polarity)
- Shape (What preprocessing must occur to get shape?)
- Objects and their Relationships

Local Features

- Find features that are invariant to transformations
  - geometric invariance: translation, rotation, scale
  - photometric invariance: brightness, exposure, …



Local Features

- SIFT
- LBP
- SURF
- BRISK
- And so on

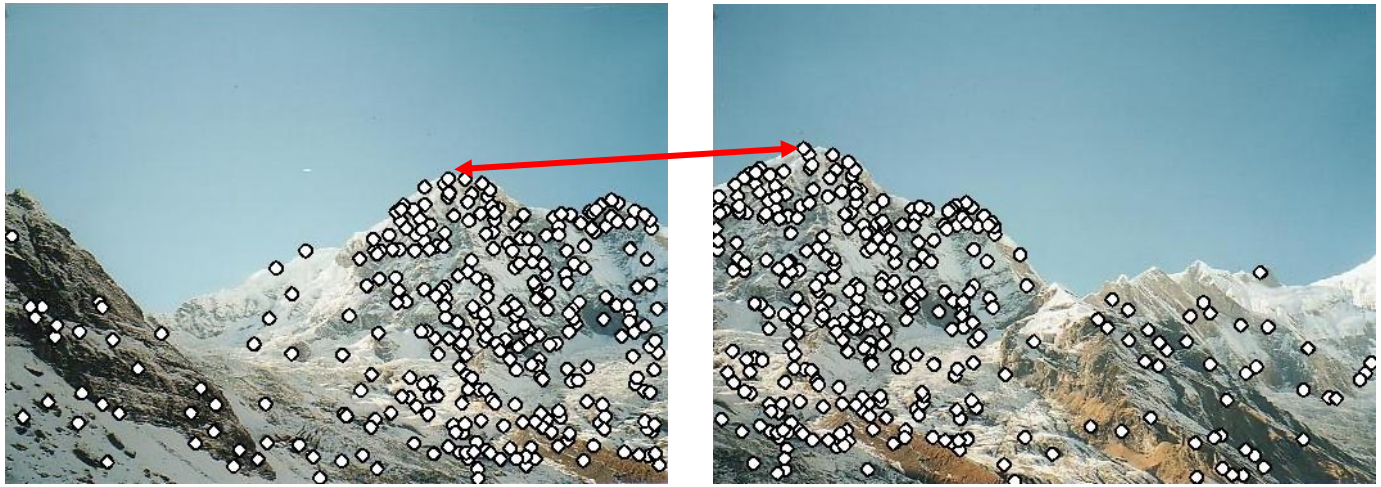


Image Retrieval & Image Classification

Query: This chair

Image Retrieval

Results from dataset ranked by similarity to the query

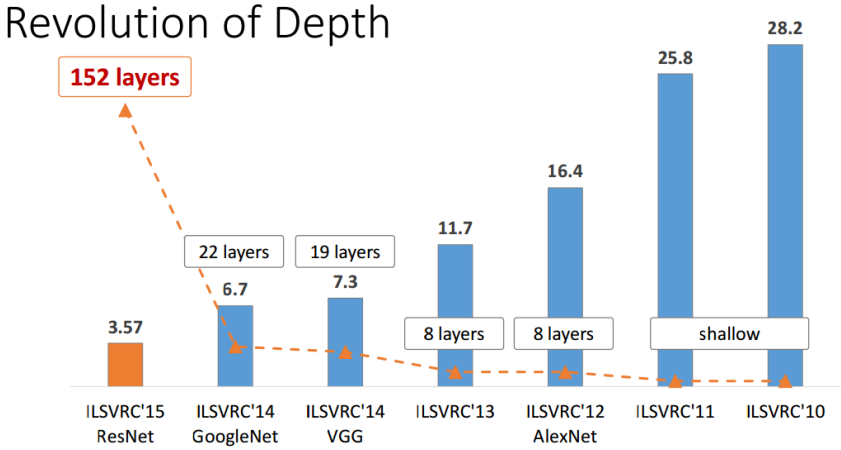
Query: This chair

Image Classification

Results from dataset classified as “chair”

Based on Deep Learning

- 深度学习在计算机视觉领域取得巨大成功
  - ImageNet Grand Challenge



- 深度学习：时势造英雄
  - 大规模图像视频数据
  - 强大的计算能力：GPU/TPU

IMAGENET

flickr

60亿张图片，2011年

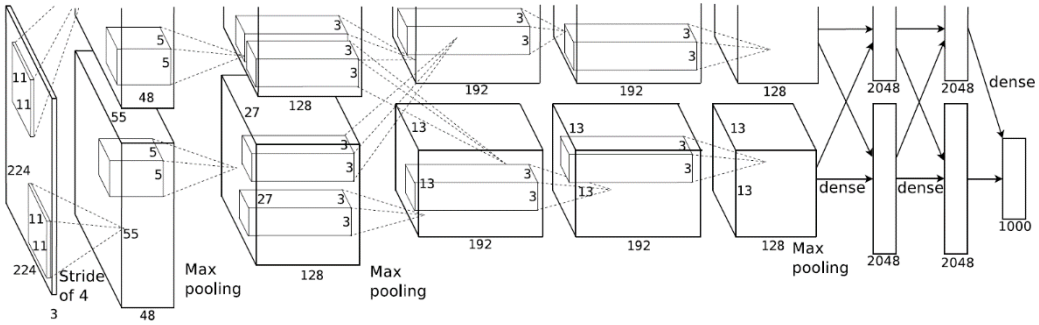
facebook

2500亿张图片，2013年

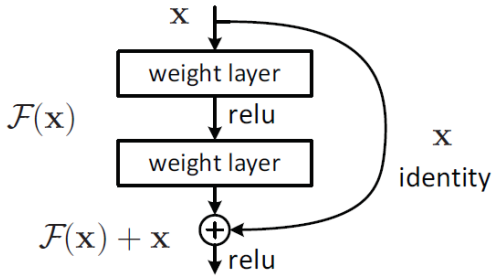
NVIDIA

研究背景：深度学习

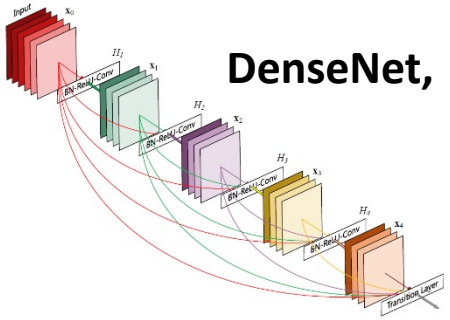
- 面向图像分类的深度学习模型



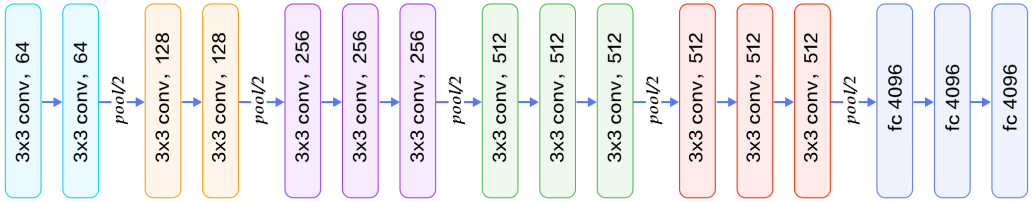
AlexNet, 2012



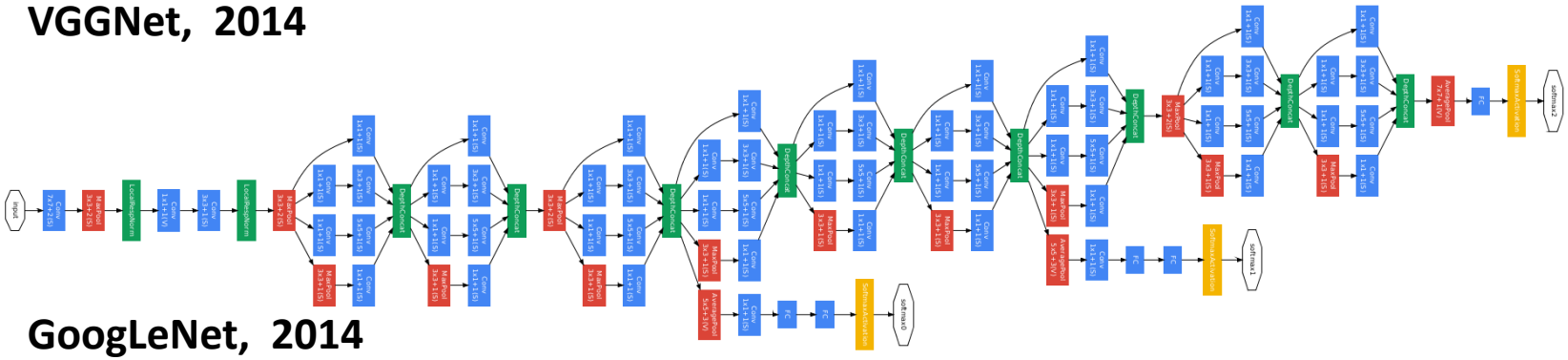
ResNet, 2015



DenseNet, 2017



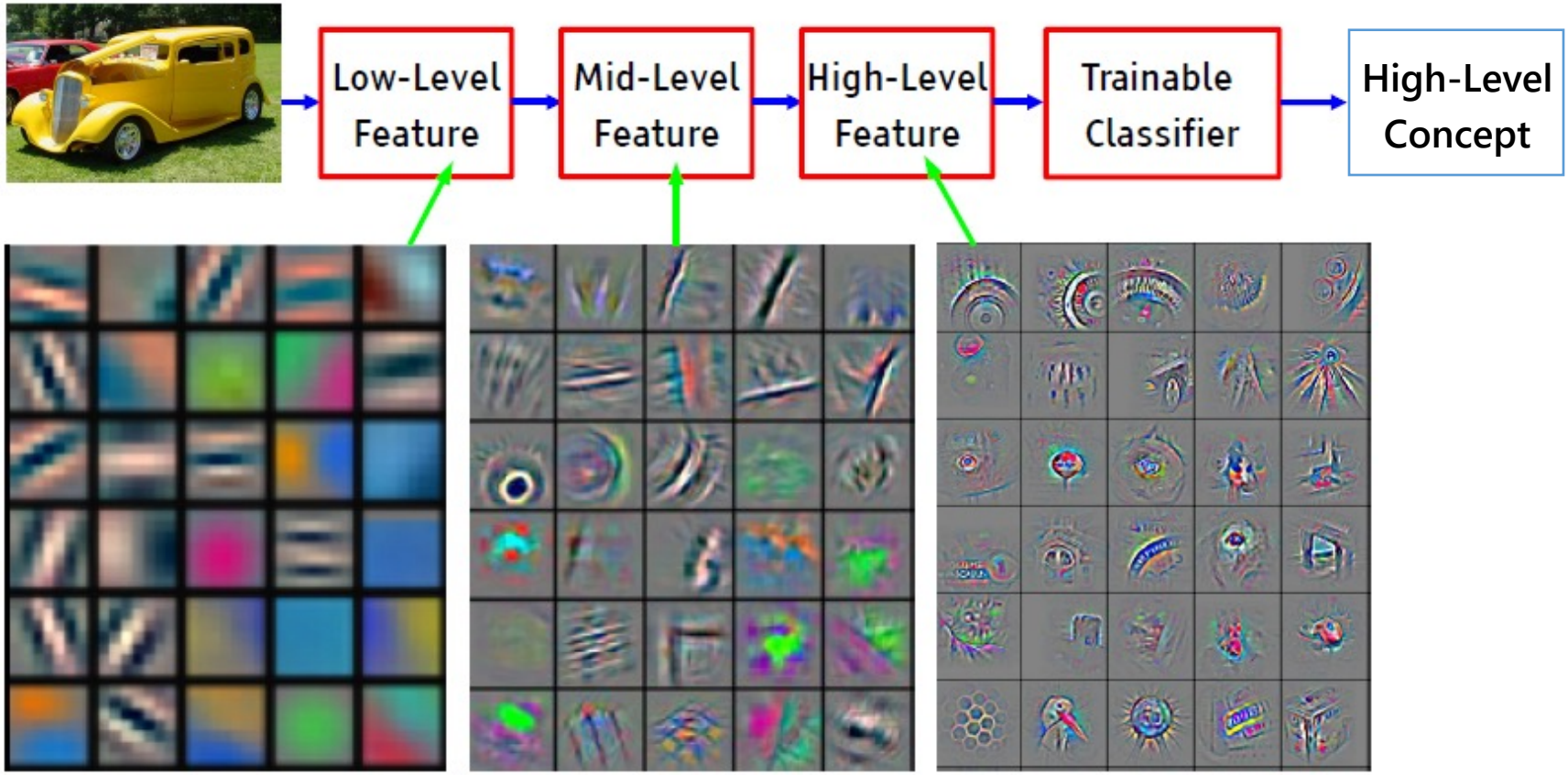
VGGNet, 2014



GoogLeNet, 2014

研究背景：深度学习

- 深度学习本质：层次化的表征学习



[Courtesy of Yann Le Cun]

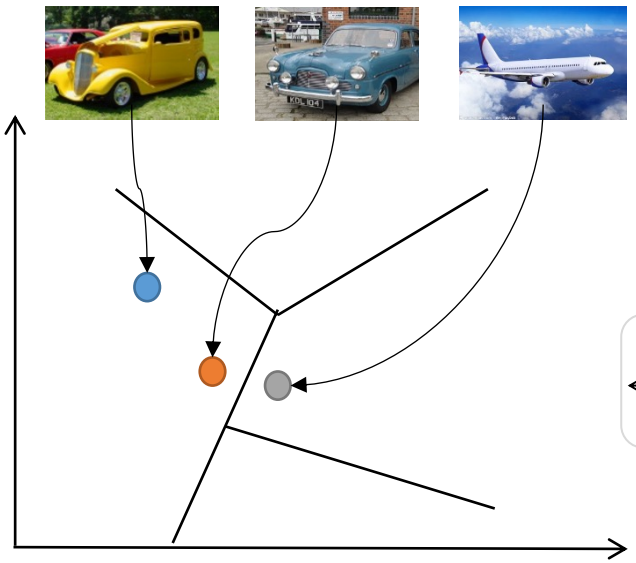
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

研究背景：深度学习

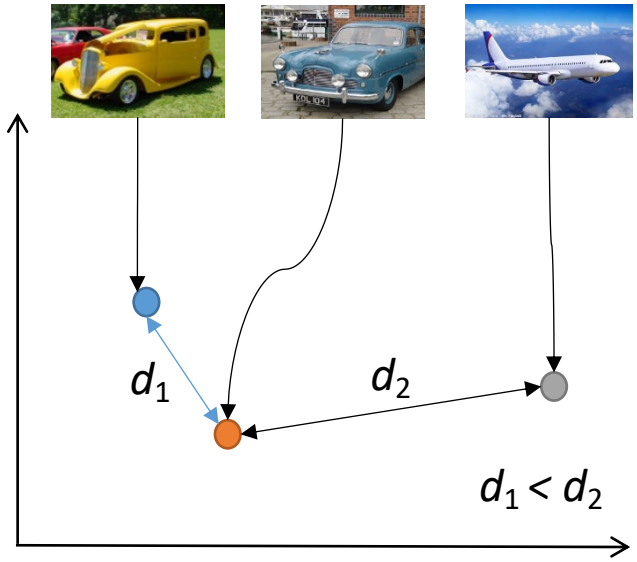
- 图像分类 vs. 图像检索

Low-Level Feature → Mid-Level Feature → High-Level Feature → Trainable Classifier → High-Level Concept

Feature Extraction　　Classification



已知　图像类别　未定义



图像分类：特征空间划分　　图像检索：相关度度量

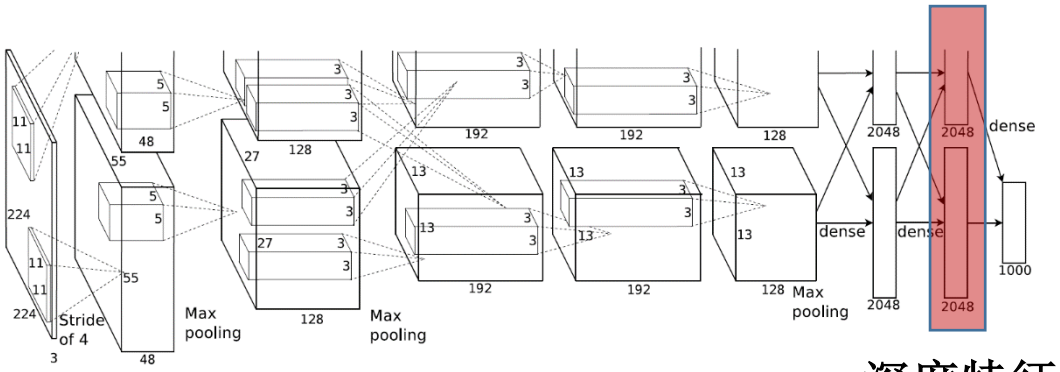
如何将深度学习用于图像检索?

- 关键：大规模的标注的训练数据

Low-Level Feature → Mid-Level Feature → High-Level Feature → Trainable Classifier → High-Level Concept

Feature Extraction　　Classification

Option 1:
- 利用图像分类任务进行预训练
- 基于上述网络提取特征，用于检索



深度特征

Option 2:
- 面向检索任务，设计监督信息
- fine-tune预训练好的CNN模型