



Datasets for Image Classification and Benchmarks

Dr. Mongkol Ekpanyapong



Datasets

- MNIST
- CIFAR-10
- SMILES
- Kaggle:Dogs vs. Cats

Datasets

- Flowers-17
- CALTECH-101
- Adience
- ImageNet
- Kaggle: Facial Expression Recognition Challenge

Datasets

- Indoor CVPR
- Stanford Cars
- LISA Traffic Signs
- Flickr Material Database
- DeepFashion

A decorative graphic in the top left corner consisting of a blue square above a blue and white circuit-like pattern.

Challenges

- ImageNet
- Face Recognition Vendor Test (FRVT) by the Intelligence Advanced Research Projects Activity (IARPA) and the US National Institute of Standards and Technology (NIST)
- NVIDIA AI City Challenge



MNIST

- Stands for modified National Institute of Standards and Technology released in 2004
- The goal of this dataset is to correctly classify the handwritten digits 0-9 (28x28 pixels)
- It is like “Hello, World” in machine learning
- It consists of 60,000 training images and 10,000 testing images. Pixel intensities are in the range [0,255]

Three small decorative images at the bottom left: a colorful abstract pattern, a solid yellow square, and a circuit board.

4	1	5	7	1	3	3	6	4	8	1	9	7	6	3	6	9	3	0	6
4	7	7	8	1	3	7	2	4	6	4	3	2	8	6	1	4	3	0	9
1	7	7	6	5	8	6	0	0	3	9	5	4	1	5	7	2	3	2	1
3	5	2	5	7	3	2	9	7	1	6	9	4	6	2	3	2	4	1	9

CIFAR-10

- It is considered another standard benchmark data for image classification (released in 2009)
- CIFAR-10 consists of 60,000 of 32x32x3 (RGB) (= 3072 features) images
- CIFAR-10 consists of 10 classes including airplanes, automobiles, birds, cats, deer, dogs, frogs, horses, ships, and trucks

airplane



automobile



bird



cat



deer



dog



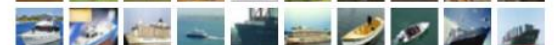
frog



horse



ship



truck



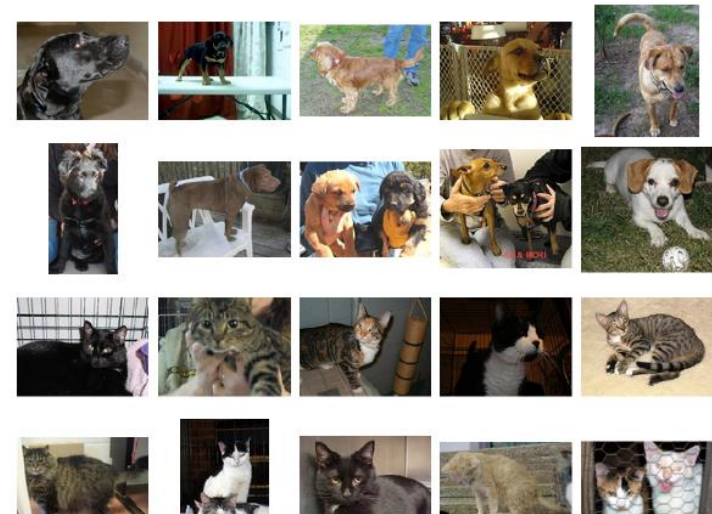
SMILES

- The SMILES dataset (smiling or not smiling) consists of 13,165 grayscale images with the size of 64x64.
- The early days of machine learning is using only cropped data with minimum background



Kaggle: Dogs vs. cats

- Dogs vs. Cats challenge is part of a Kaggle competition for machine learning algorithms
- The dataset is provided in 2013 from Kaggle company
- A total of 25,000 RGB images are provided



Flowers-17

- Flowers-17 dataset is a 17 category dataset with 80 images per class
- It is considered a challenging dataset due to the dramatic changes in scale, view-point angles, background clutter, light condition, and intra-class variation
- It also has 80 images per class



CALTECH-101

- It is introduced in 2004
- The data is a popular benchmark for object detection
- The dataset of 8,677 images includes 101 categories
- The datasets exhibits heavy class imbalances (more images for some categories than others)



Adience

- It is constructed by Edinger et al. in 2014
- It is used to study age, and gender classification
- A total of 26,580 images are included with ages ranging from 0-60



ImageNet

- ImageNet is a project aimed at labelling and categorizing images into 22,000 categories
- There are more than 14 million images in the project (more than 1,000 images per set)
- ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is the de facto standard for computer vision classification algorithms
- ILSVRC has 1,000 images per class



ImageNet

- ILSVRC has 1.2 million images with 1,000 object classes for training, 50,000 for validation and 100,000 for testing
- In Computer Vision, when you refer to ImageNet, they usually means this ILSVRC
- ImageNet is put together by Prof. Fei-Fei Li from Stanford University
From 2010-2017



ImageNet

- Currently, the world largest dataset

<https://www.image-net.org/>

IMAGENET

14,197,122 images, 21841 synsets indexed

[Home](#) [Download](#) [Challenges](#) [About](#)

Not logged in. [Login](#) | [Signup](#)

ImageNet is an image database organized according to the **WordNet** hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. The project has been **instrumental** in advancing computer vision and deep learning research. The data is available for free to researchers for non-commercial use.

Mar 11 2021. ImageNet website update.

© 2020 Stanford Vision Lab, Stanford University, Princeton University imagenet.help.desk@gmail.com Copyright infringement

ImageNet in Research

IMGENET on Google Scholar

4,386
Citations

[Imagenet: A large-scale hierarchical image database](#)

[J Deng, W Dong, R Socher, L J Li, K Li...](#) - Computer Vision and ..., 2009 - [ieeexplore.ieee.org](#)

Abstract: The explosion of image data on the Internet has the potential to foster more sophisticated and robust models and algorithms to index, retrieve, organize and interact with images and multimedia data. But exactly how such data can be harnessed and organized

[Cited by 4386](#) [Related articles](#) [All 30 versions](#) [Cite](#) [Save](#)

2,847
Citations

[Imagenet large scale visual recognition challenge](#)

[O Russakovsky, J Deng, H Su, J Krause...](#) - International Journal of ..., 2015 - Springer

Abstract The **ImageNet** Large Scale Visual Recognition Challenge is a benchmark in object category classification and detection on hundreds of object categories and millions of images. The challenge has been run annually from 2010 to present, attracting participation

[Cited by 2847](#) [Related articles](#) [All 17 versions](#) [Cite](#) [Save](#)

...and many more.

https://www.image-net.org/static_files/files/imagenet_ilsvrc2017_v1.0.pdf

Comparing with other data set

SUN, 131K

[Xiao et al. '10]

LabelMe, 37K

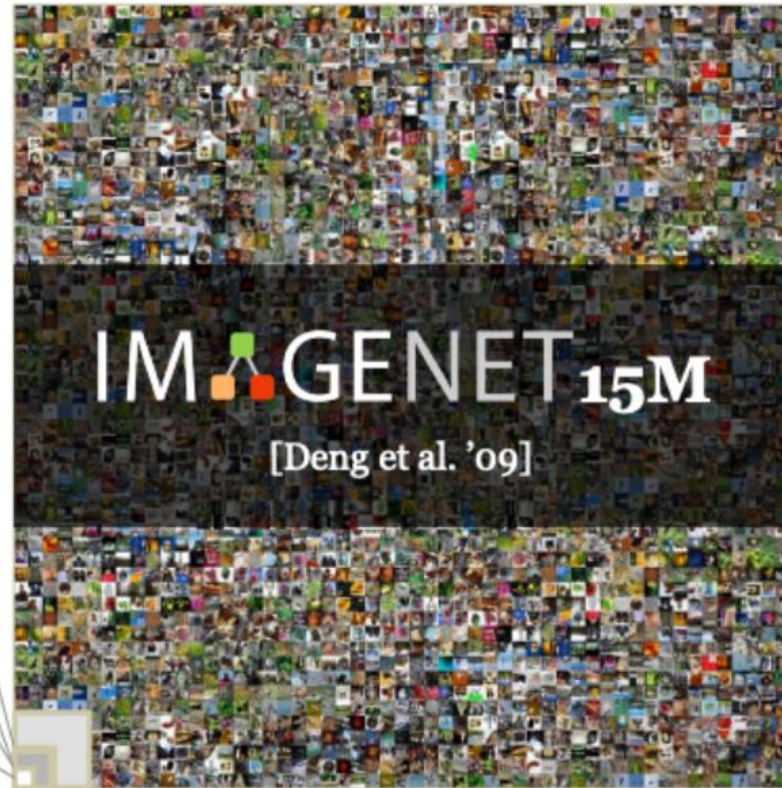
[Russell et al. '07]

PASCAL VOC, 30K

[Everingham et al. '06-'12]

Caltech101, 9K

[Fei-Fei, Fergus, Perona, '03]



https://www.image-net.org/static_files/files/imagenet_ilsvrc2017_v1.0.pdf

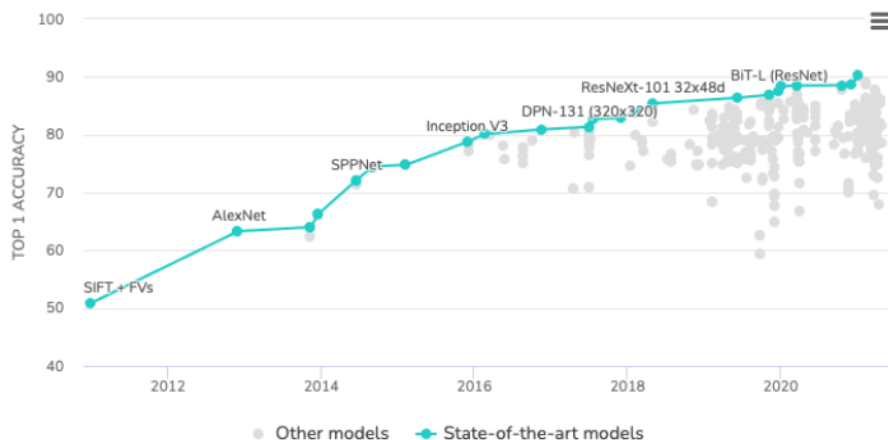


ImageNet – Current State of the Art

Image Classification on ImageNet

Leaderboard

Dataset



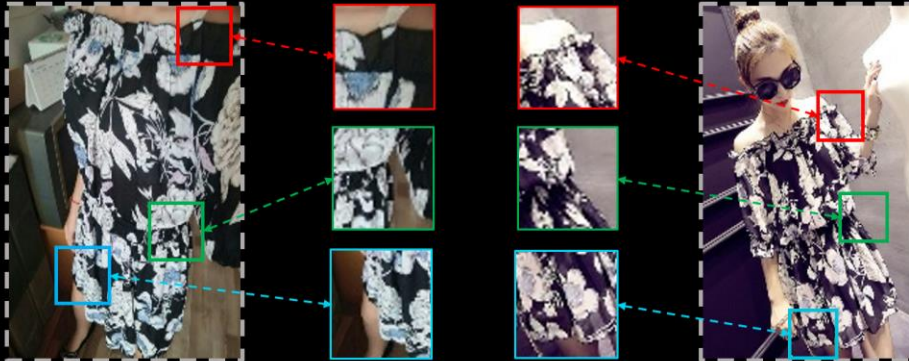
<https://paperswithcode.com/sota/image-classification-on-imagenet>

RANK	MODEL	TOP 1 ACCURACY	TOP 5 ACCURACY	NUMBER OF PARAMS	EXTRA TRAINING DATA	PAPER
1	Meta Pseudo Labels (EfficientNet-L2)	90.2%	98.8%	480M	✓	Meta Pseudo Labels
2	Meta Pseudo Labels (EfficientNet-B6-Wide)	90%	98.7%	390M	✓	Meta Pseudo Labels
3	NFNet-F4+	89.2%		527M	✓	High-Performance Large-Scale Image Recognition Without Normalization
4	ALIGN (EfficientNet-L2)	88.64%	98.67%	480M	✓	Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision
5	EfficientNet-L2-475 (SAM)	88.61%		480M	✓	Sharpness-Aware Minimization for Efficiently Improving Generalization



DeepFashion

- The dataset is proposed in CVPR 2016
- It consists of 800,000 fashion images
- The dataset is labelled with 50 categories



Kaggle: Facial Expression Recognition Challenge

- Kaggle FER challenge is to identify emotion
- It has a total of 35,888 images with these categories:
 - Angry
 - Disgust
 - Fear
 - Happy
 - Sad
 - Surprise
 - Neutral



Indoor CVPR

- It consists of number of indoor scenes including stores, houses, leisure spaces, working areas, and public spaces
- It is made available in 2009



Stanford Cars

- This is Cars Dataset put together by Stanford in 2013
- It consists of 16,185 images of 196 classes of cars
- The information includes vehicle make, model, or manufacturer year



LISA Traffic Sign

- It was put by Mogelmose et al. in 2012
- The dataset consists of 47 different United States traffic sign types including stop signs, pedestrian crossing sign.



Flickr Material Database

- It is available in 2014
- It consists of 100 images per category and 10 categories
- It consists of:
fabric, foliage,
glass, lather,
metal, paper,
plastic, stone,
water, wood

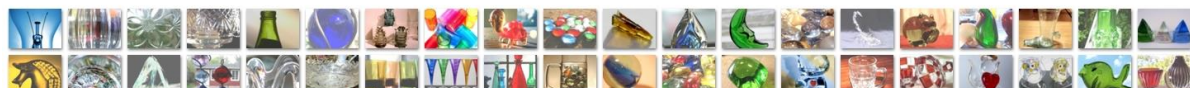
Fabric



Foliage

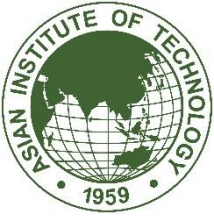


Glass



FRVT Challenge

- FRVT challenge is the event once 3-4 years
- In 2014, NEC won the first prize
- In 2018, Yitu technology ranked first place with 95.5 % accuracy subjected to false match rate of 10 million (based on 10 billion samples)
- In 2021, IDEMIA from France ranked first place with



NVIDIA AI City challenge

- First time in 2017
- Estimating vehicle speed
- Detect anomalies caused by crashes
- Multi-sensor tracking



COCO Dataset

COCO is a large-scale object detection segmentation and captioning dataset:

- Object Segmentation
- Recognition in context
- 330k images
- 1.5 Million objects
- 80 object categories
- 5 captions per image

Dataset examples

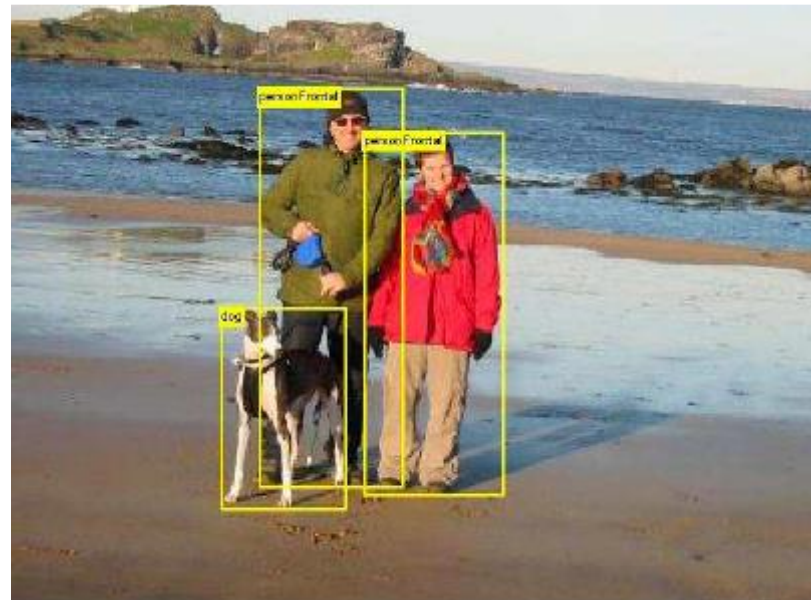


Pascal VOC

The PASCAL Visual Object Classes Challenge

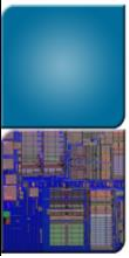
2012 challenges:

- 20 classes, 11,530 images containing 27,450 objects



Kaggle

- Website that post many challenge regularly including sample data



Questions?

