Final Project Topics

CS6550 Computer Vision

Brought to you by CVLabs

List of Final Topics

Image Classification

Image Saliency Detection

Image Segmentation (Learning Based)

Image Restoration

Object Detection

Facial Expression Recognition

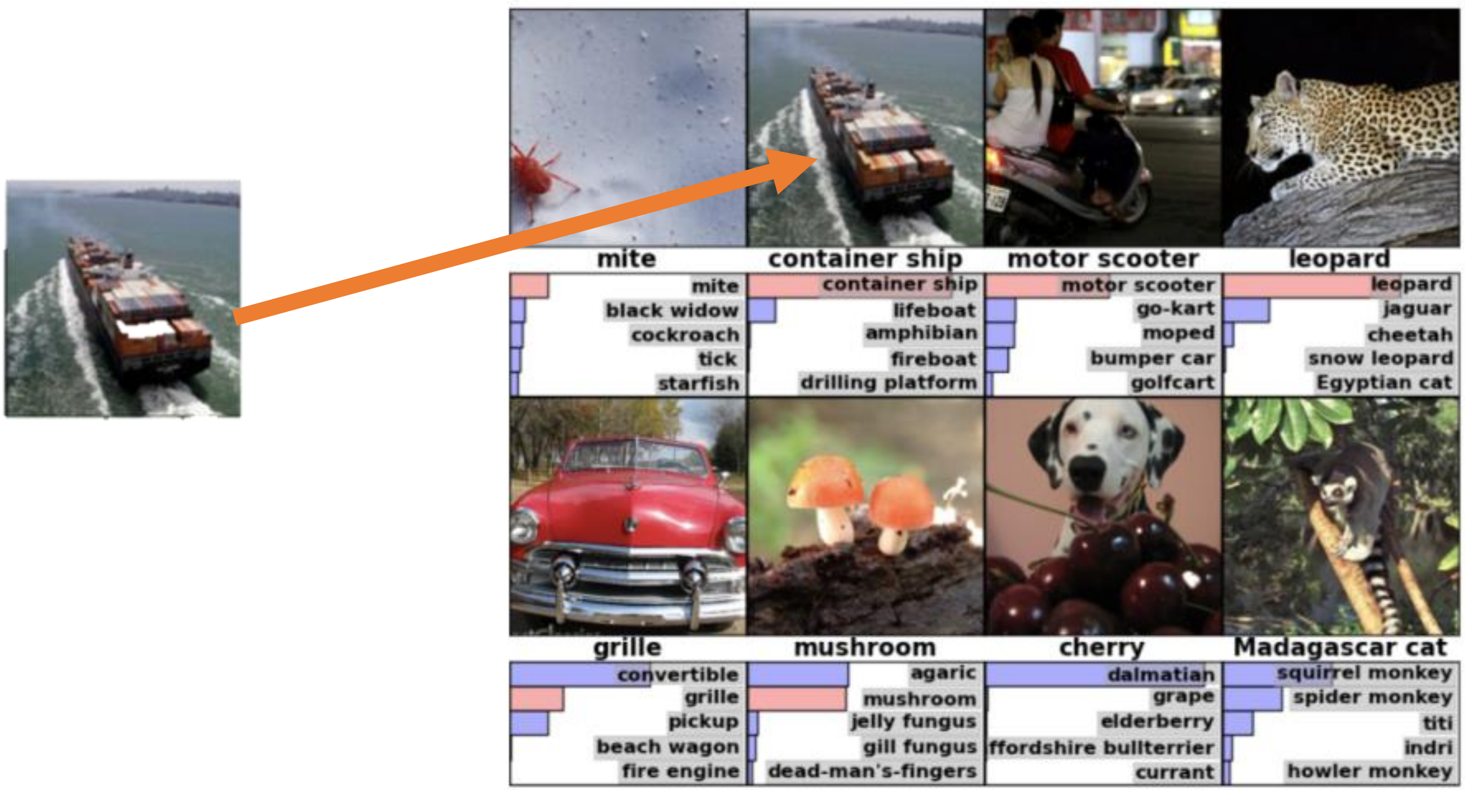
Gesture Recognition

Stereo Matching

Image Classification

Image Classification

- Categorize or detect text in the picture.



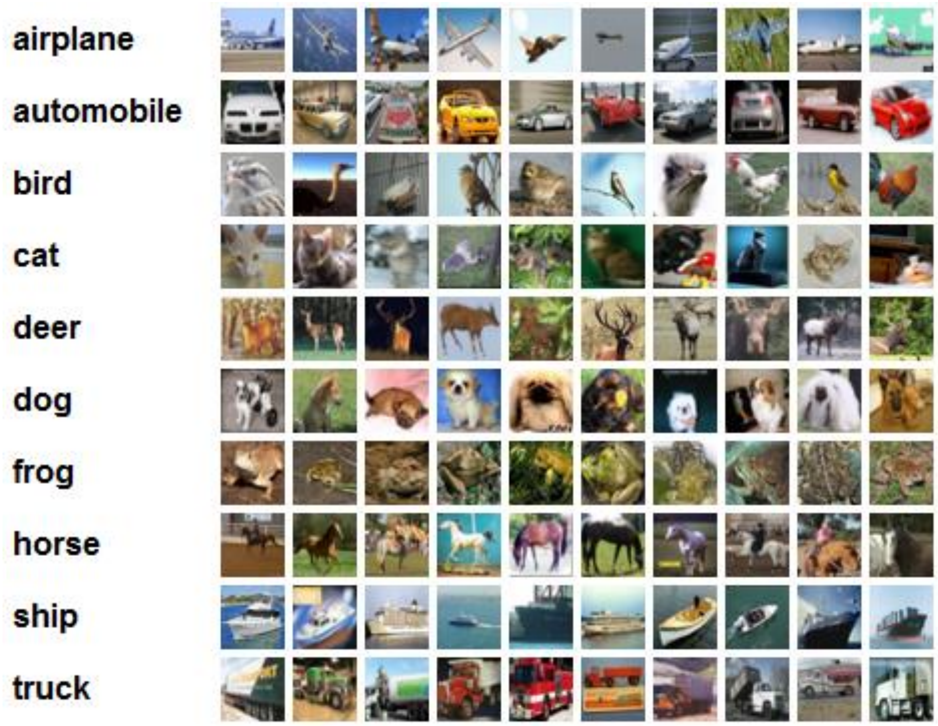
References: https://blog.acolyer.org/2016/04/20/imagenet-classification-with-deep-convolutional-neural-networks/

Image Classification

- How to do “Classification” ?
  1. Preprocessing
  2. Feature extraction
  3. Modeling
     - Description of each class in mathematical form
  4. Classification
     - The classifier divides the feature space into class regions

Image Classification

- Datasets
  - ImageNet : http://image-net.org/
  - CIFAR : https://www.cs.toronto.edu/~kriz/cifar.html



References: https://www.cs.toronto.edu/~kriz/cifar.html

Image Classification

- References
  - https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf
  - http://vision.cse.psu.edu/seminars/talks/2009/random_tff/bosch07a.pdf
  - http://cs.utsa.edu/~qitian/seminar/Spring11/02_18_11/ECCV10.pdf

Image Saliency Detection

Image Saliency Detection

- Finding salient objects in the source image.

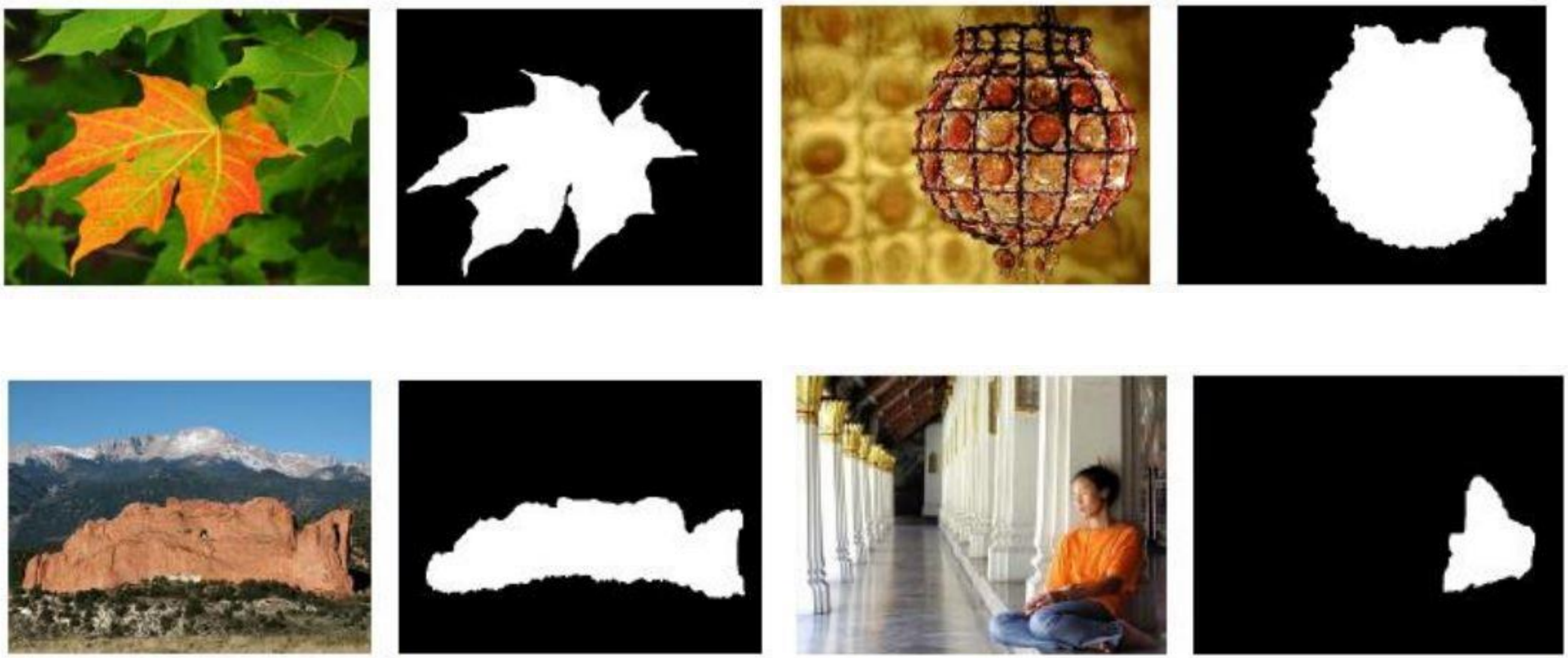


Image Saliency Detection

- Contrast Based Filter Saliency
  - Abstraction : decomposes an image into compact represented by their mean color.
  - Uniqueness : regions which stand out from other regions should be labeled more salient.
  - Distribution : foreground objects are generally more compact, thus we measure the spatial distribution.

(a) Source image. (b) Abstraction. (c) Uniqueness. (d) Distribution. (e) Saliency.

References: https://graphics.ethz.ch/~perazzif/saliency_filters/

Image Saliency Detection

- Datasets
  - https://graphics.ethz.ch/~perazzif/saliency_filters/files/SF_maps.zip
  - http://saliency.mit.edu/datasets.html


Image Saliency Detection

- References
  - https://graphics.ethz.ch/~perazzif/saliency_filters/files/saliency_filters_cvpr_2012.pdf
  - http://www.cse.cuhk.edu.hk/leojia/projects/hsaliency/papers/hsaliency.pdf
  - https://arxiv.org/pdf/1505.01173v1.pdf

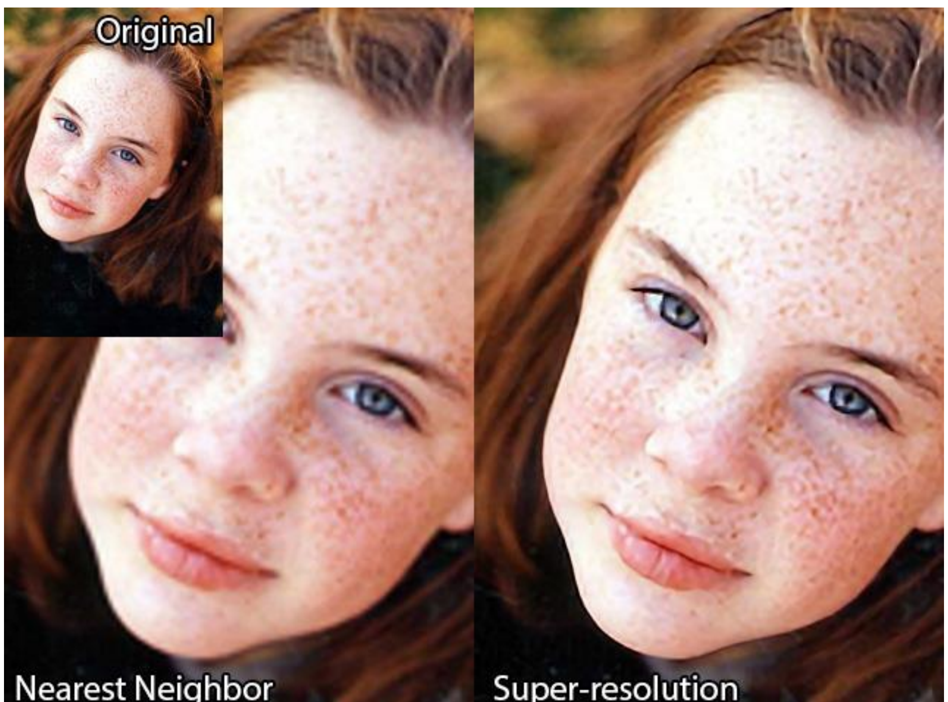
Image Restoration

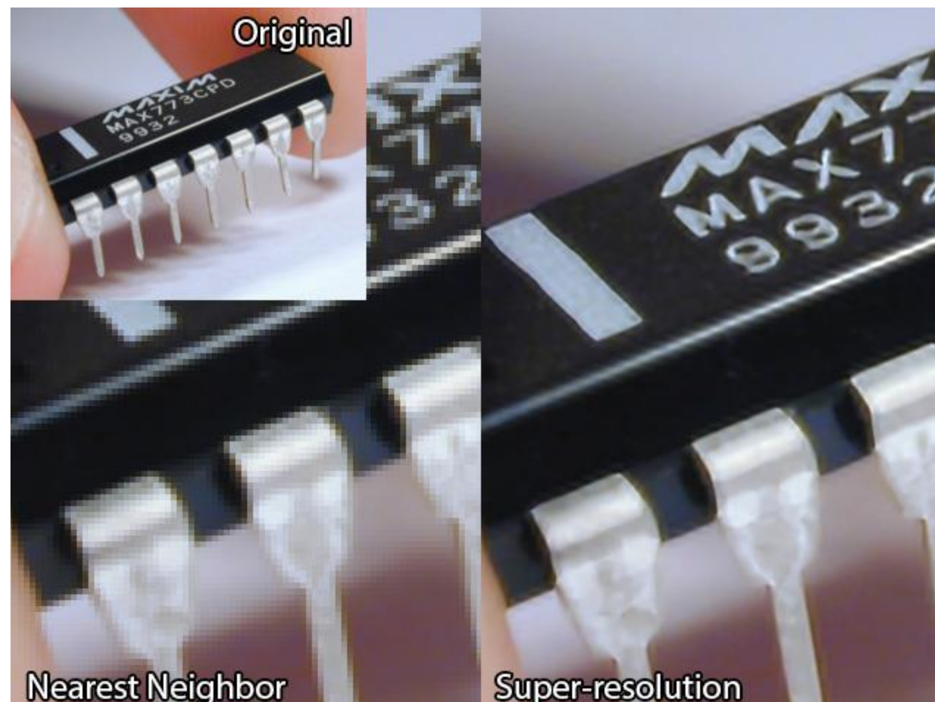
Super-Resolution

Inpainting / Completion

Image Restoration — Super-Resolution

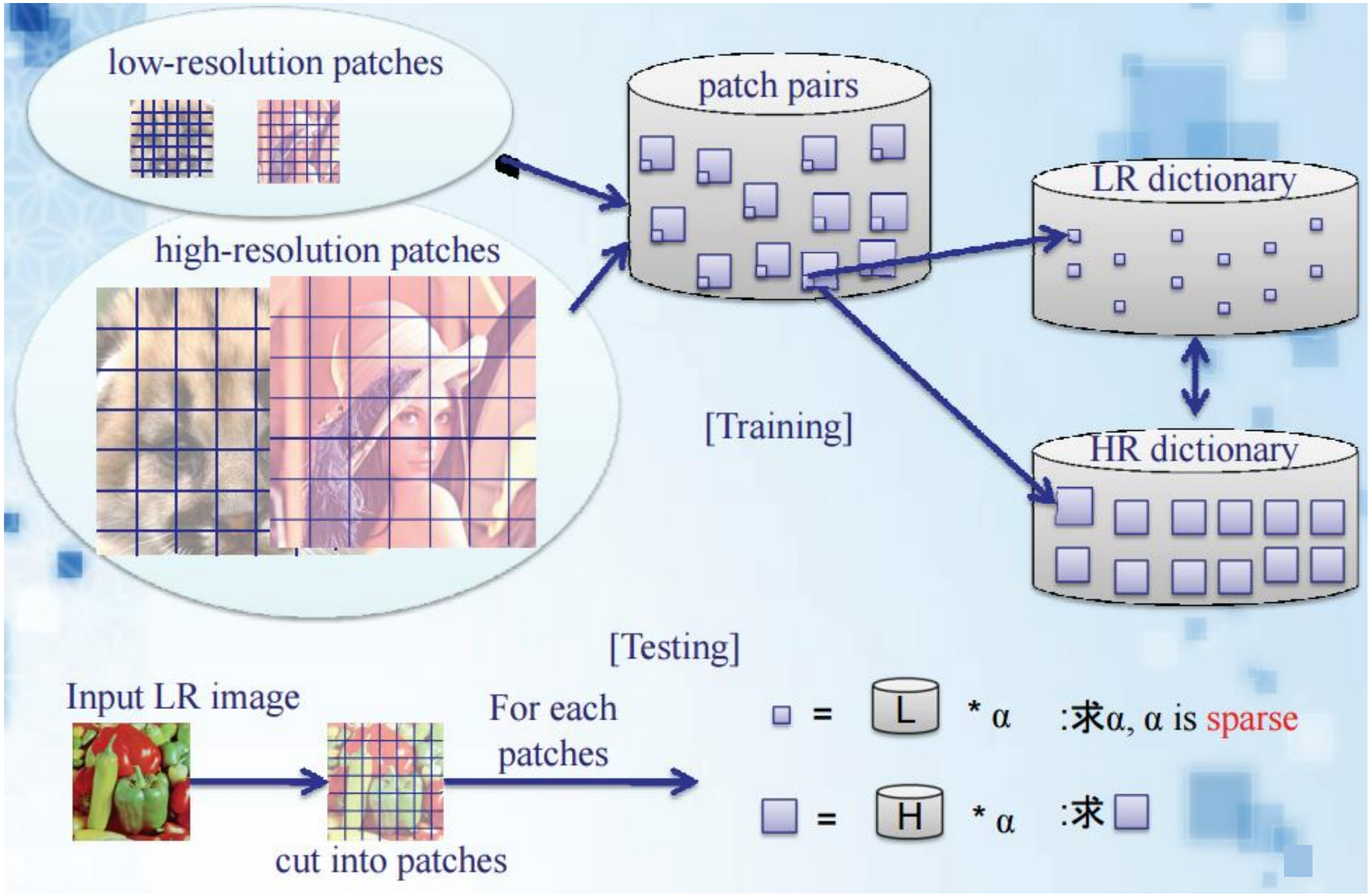
- What is Super-Resolution ?
  - Upscale the image to desire size (x2, x4, x8…)
- Possible solutions
  - Interpolations: Nearest Neighbor / Bicubic / Bilinear
  - Learning-based: Sparse Representation / Deep Learning



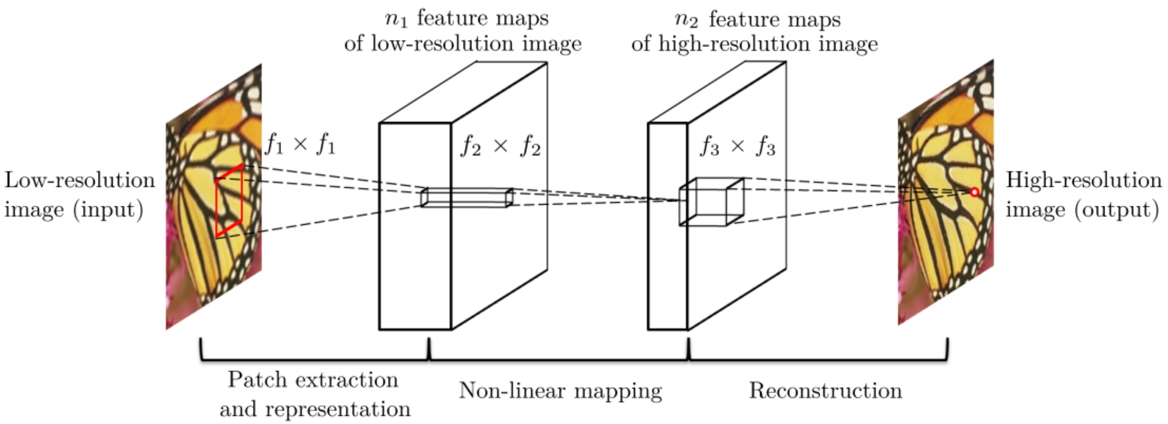


- Sparse Representation [1]
  1. Training: Learned HR-LR pair dictionaries
  2. Testing: Solve the sparse coding
  3. Reconstruct HR image

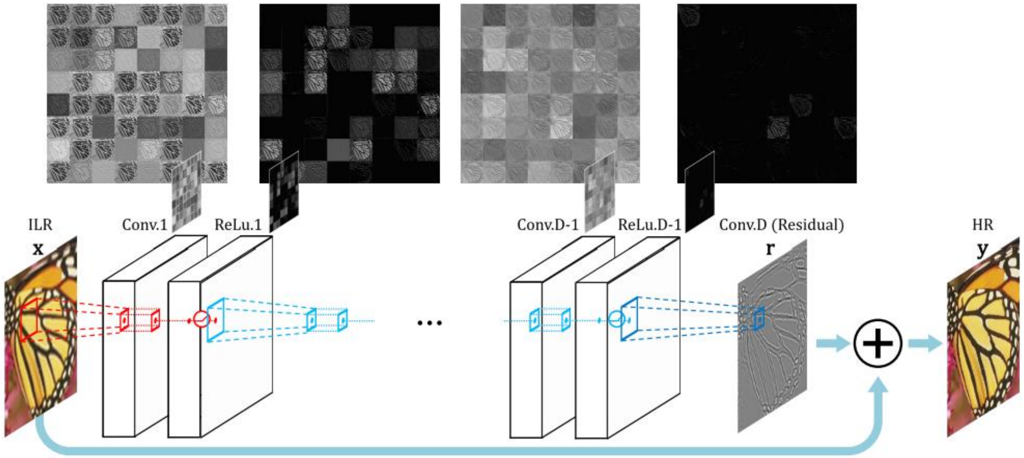
References
[1] Image Super-Resolution via Sparse Representation



- Deep Learning
  - SRCNN [2] [Image Super-Resolution Using Deep Convolutional Networks]



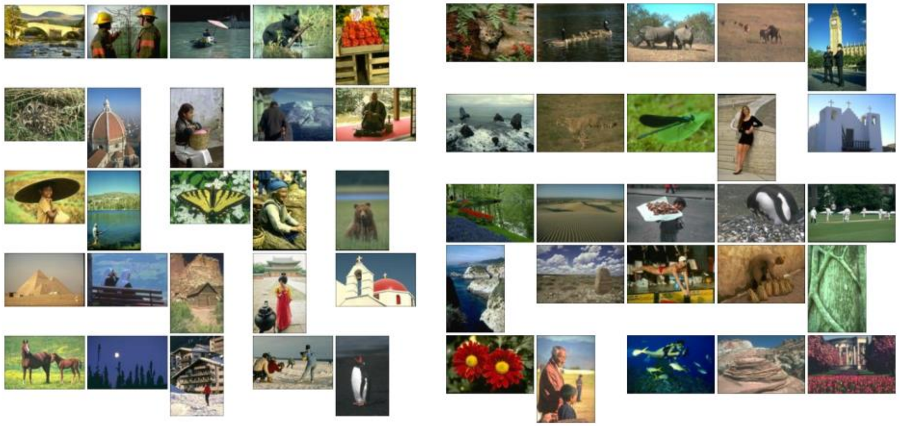
  - VDSR [3] [Accurate Image Super-Resolution Using Very Deep Convolutional Networks]



Super-Resolution

- Training Datasets

The Berkeley Segmentation Dataset and Benchmark (BSD 200)



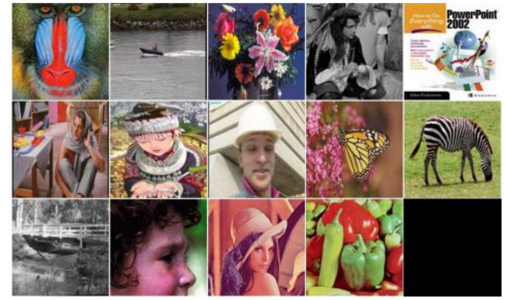
Yang 91 Images

ImageNet (Not Suggested)

Image Restoration | Super-Resolution

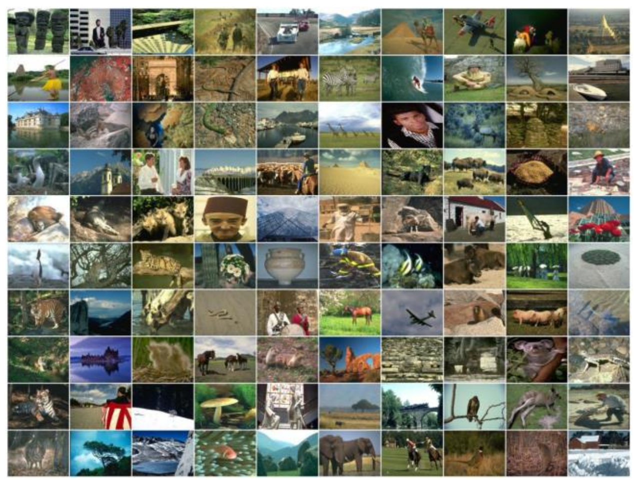
- Testing Datasets

Set 5



Set 14

Urban 100



BSD 100

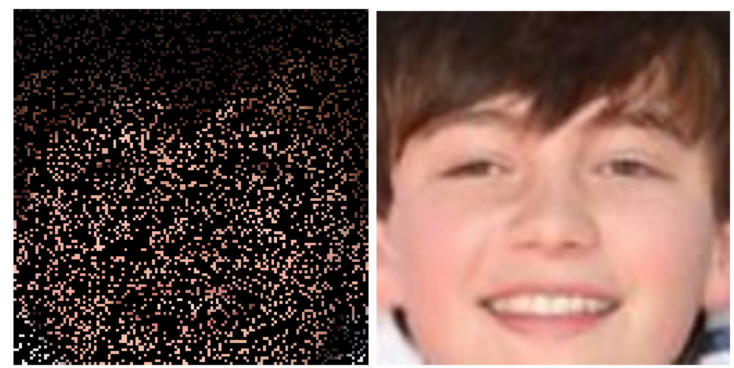
Sun-Hays 80

GitHub Resources:

- **Comparison with the [state-of-the-art datasets]()**
- **[Super-Resolution-Benchmarks]()**

Image Restoration — Inpainting / Completion

- What is Inpainting or Completion ?
  - To fill or complete the lost or unwanted regions in the images
- Related solutions
  - Vision-based (Patch Inpainting) / Numerical Optimization (TNNR)
  - MRF (Field of Experts) / Autoencoders / Deep Learning



- Autoencoders / Deep Learning
  - Conditional GANs
  - Context Encoder [1] [Context Encoders: Feature Learning by Inpainting]

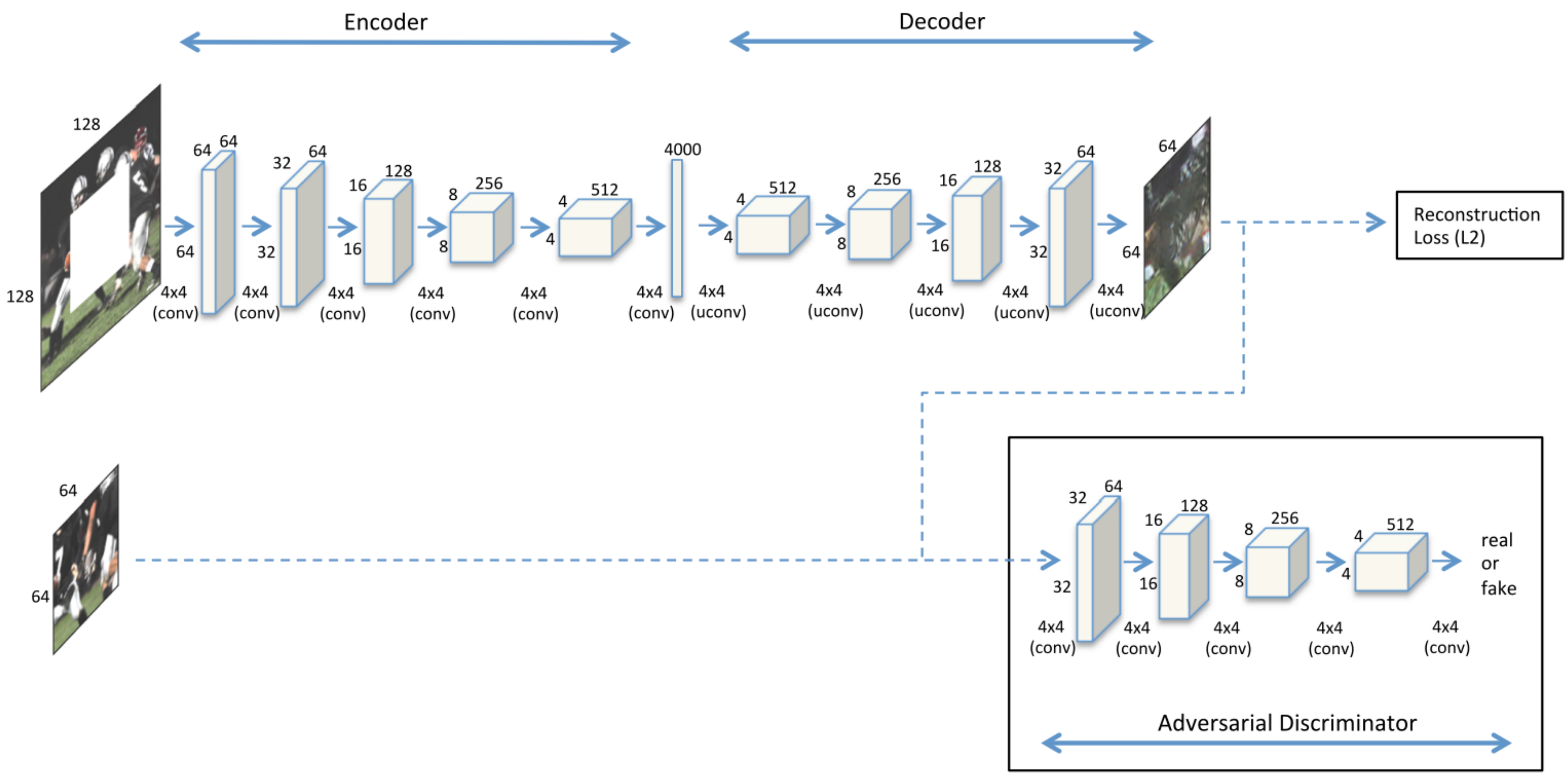
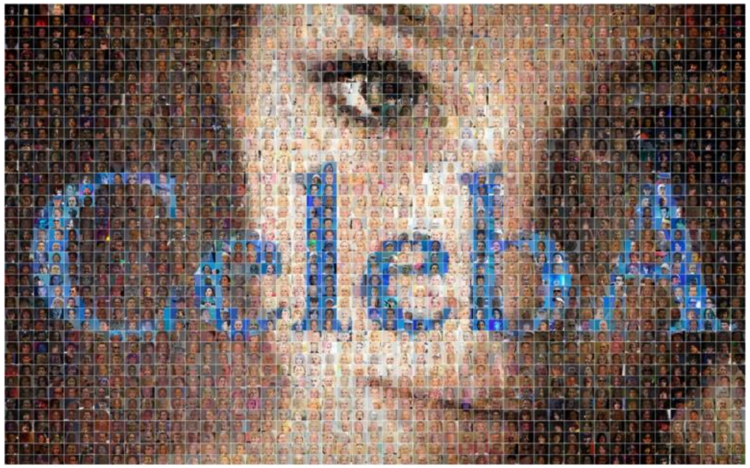


Image Restoration — Inpainting / Completions

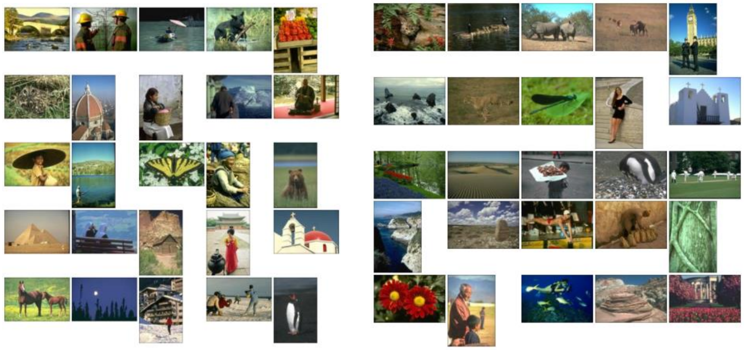
- Datasets

[TUM-Image Inpainting Database]()



[CelebA Datasets]()

[Google Street View Data Set]()



[The Berkeley Segmentation Dataset and Benchmark]() (BSD 200)

Image Segmentation

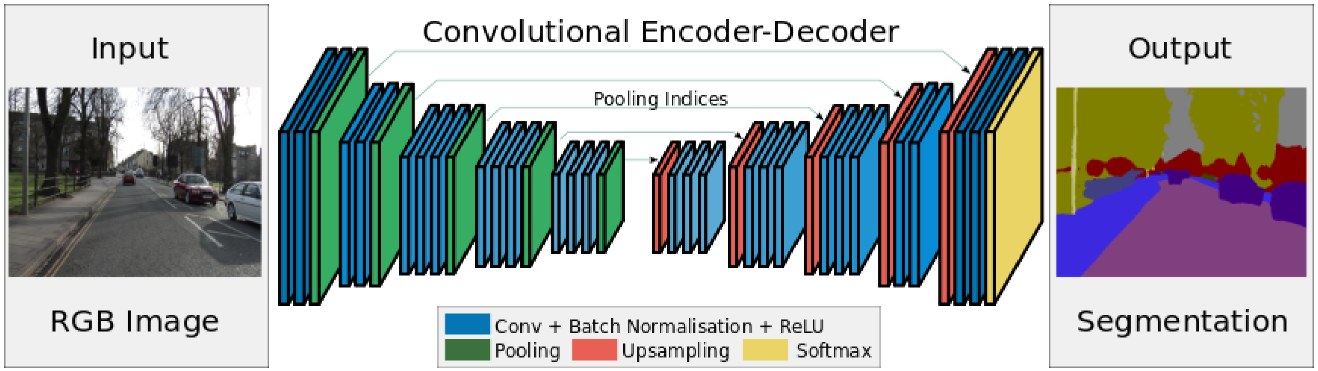
Focusing on Learning-based methods

Image Segmentation

- What is Image Segmentation ?
  - the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels).
  - The goal of segmentation is to simplify and/or change the representation of an image into something that is **more meaningfu**l and **easier to analyze**.
- Related solutions
  - Grouping Super-Pixels / Kmeans
  - Deep Learning

Image Segmentation

- Deep Learning Methods: Directly generate segmentation results
    - SegNet [1]



- Fully Convolutional Networks for Semantic Segmentation [2]

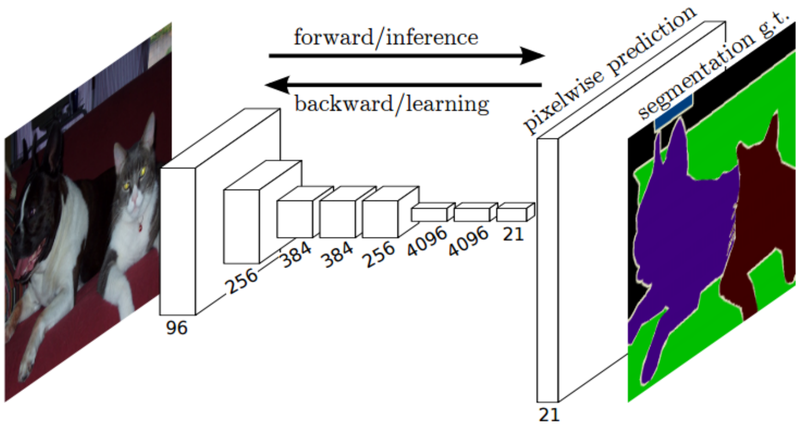
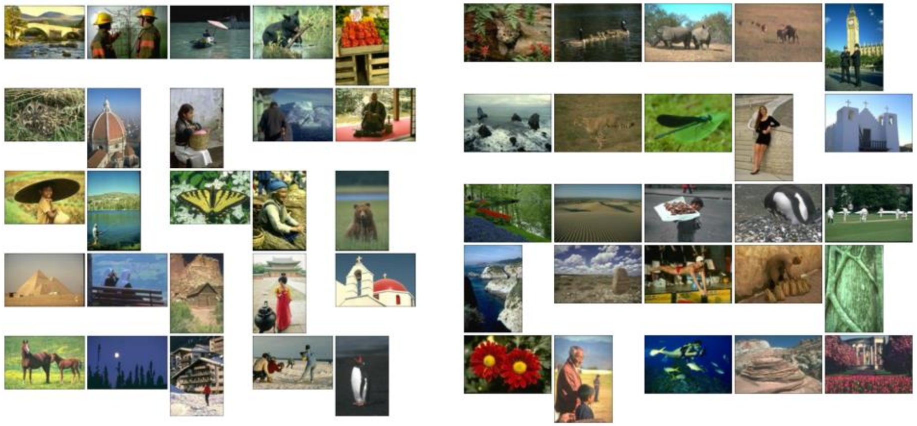


Image Segmentation

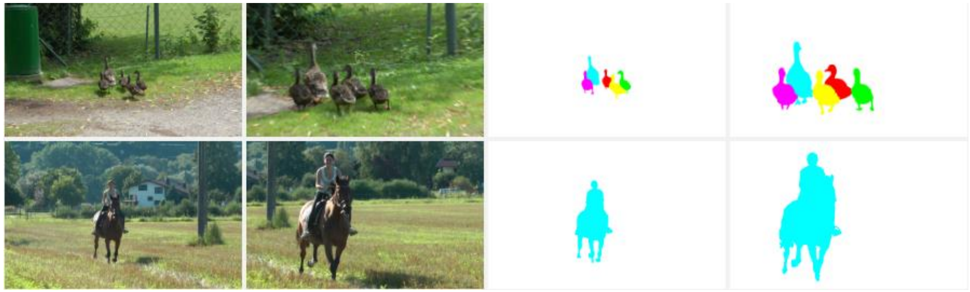
- Datasets

The Berkeley Segmentation Dataset and Benchmark (BSD 200)



Visual Object Classes Challenge 2012

Freiburg-Berkeley Motion Segmentation Dataset (FBMS-59)



Object Detection

Object Detection

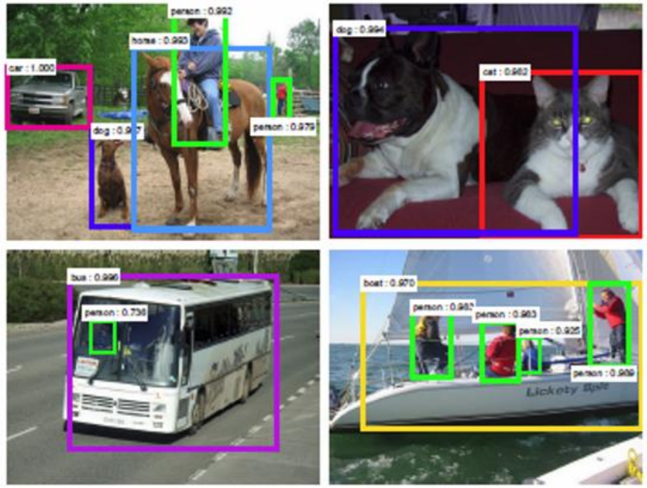
- Problem

–Given an image, find all objects and mark them up with bounding boxes and categories

- Sub-problems

–Object proposal

–Image Classification

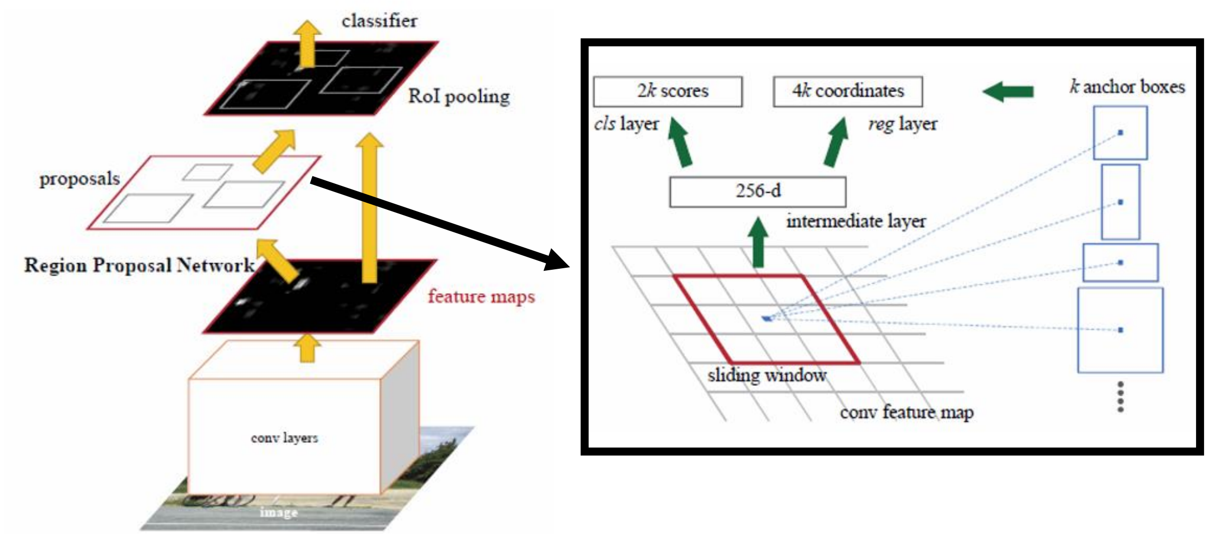


Example of detection results

Basic framework of object detection (FER)

Object Detection

- Framework of Faster R-CNN
  - Deep learning model
  - Shared CNN
  - Region Proposal Network
  - RoI pooling



Faster R-CNN, NIPS2015

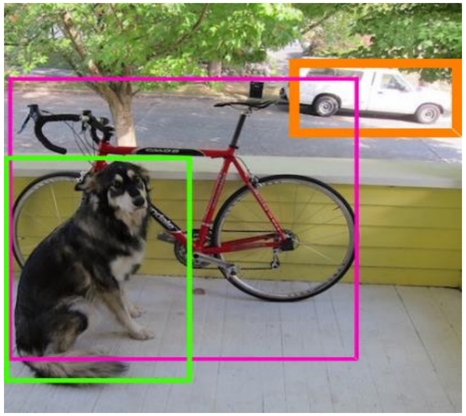
Object Detection

- Evaluation dataset
  - MIT Street Scenes

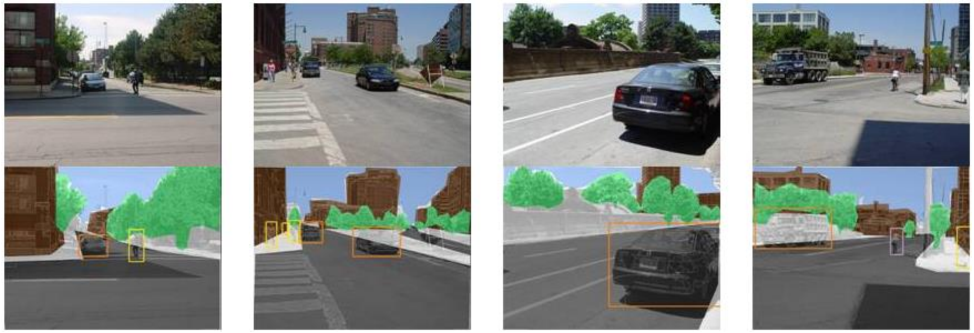
    http://cbcl.mit.edu/software-datasets/streetscenes/
  - PASCAL VOC dataset

    http://host.robots.ox.ac.uk/pascal/VOC/
  - KITTI dataset

    http://www.xrce.xerox.com/Our-Research/Computer-Vision/Proxy-Virtual-Worlds



Example of PASCAL VOC



Example of MIT Street Scenes

Example of KITTI

Object Detection

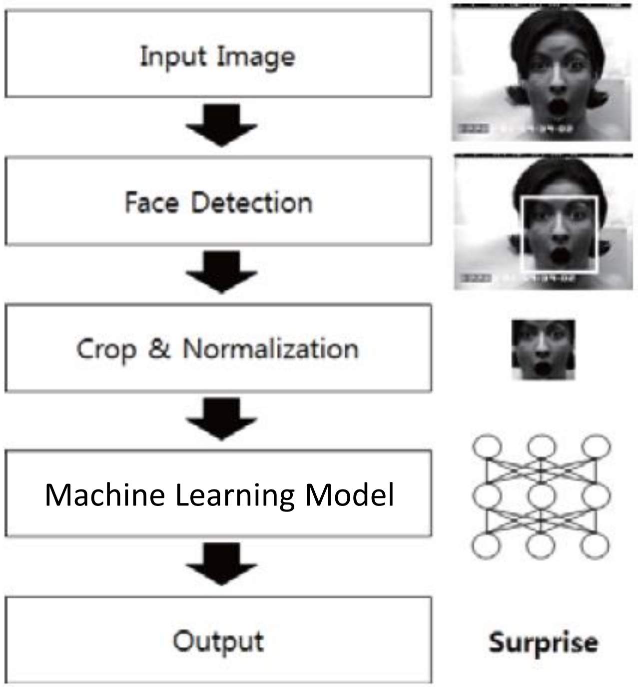
- Reference
  - Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *arXiv preprint arXiv:1506.02640* (2015).
  - Girshick, Ross. "Fast r-cnn." *Proceedings of the IEEE International Conference on Computer Vision*. 2015
  - Kye-Hyeon Kim, et al. "PVANET: Deep but Lightweight Neural Networks for Real-time Object Detection" arXiv:1608.08021.
  - Dai, Jifeng, et al. "R-FCN: Object Detection via Region-based Fully Convolutional Networks." *arXiv preprint arXiv:1605.06409* (2016).

- Tools
  - Caffe http://caffe.berkeleyvision.org/
  - Tensorflow https://www.tensorflow.org/
  - Torch http://torch.ch/

Facial Expression Recognition

Facial Expression Recognition

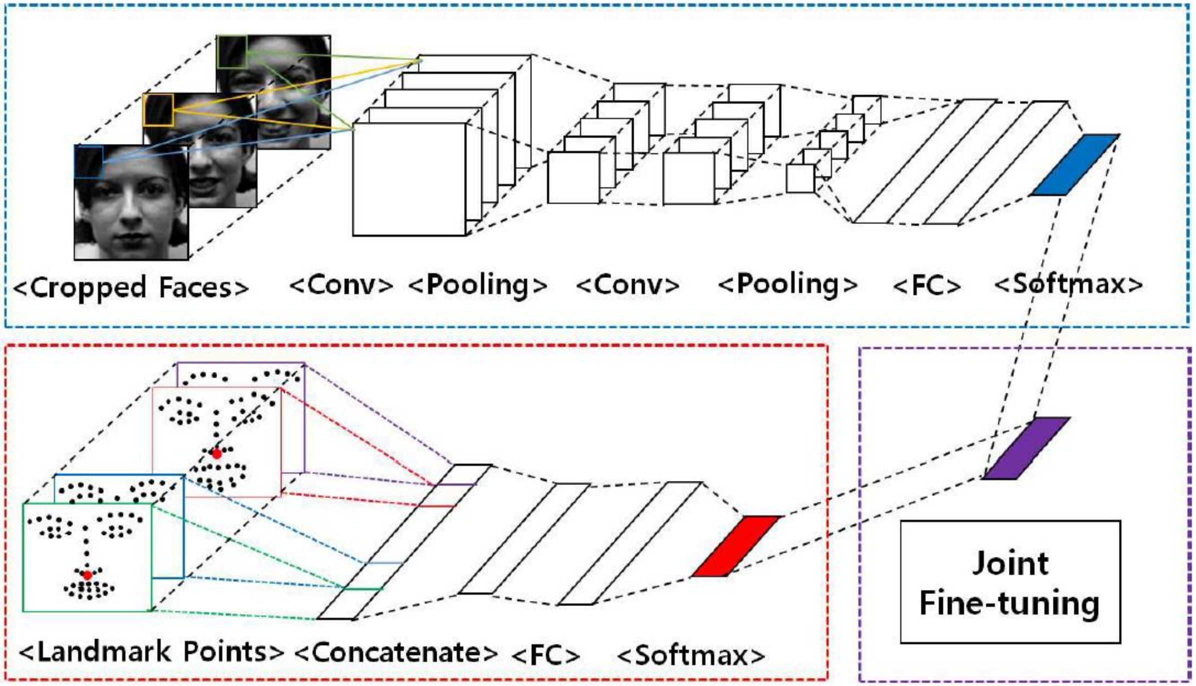
- Problem
  - Predict the emotion category (Anger, Disgust, Fear, Happiness, Sadness or Surprise) from a still-image or an image sequence
- Approaches
  - Image based framework
  - Sequence based framework
- Preprocess
  - Face detection
  - Face alignment



Basic framework of facial expression recognition (FER)

Facial Expression Recognition

- Framework of image sequence approach
  - Deep learning model
  - Appearance feature(CNN)
  - Geometry feature(NN)
  - Joint fine-tuning



Joint Fine-Tuning Method, ICCV2015

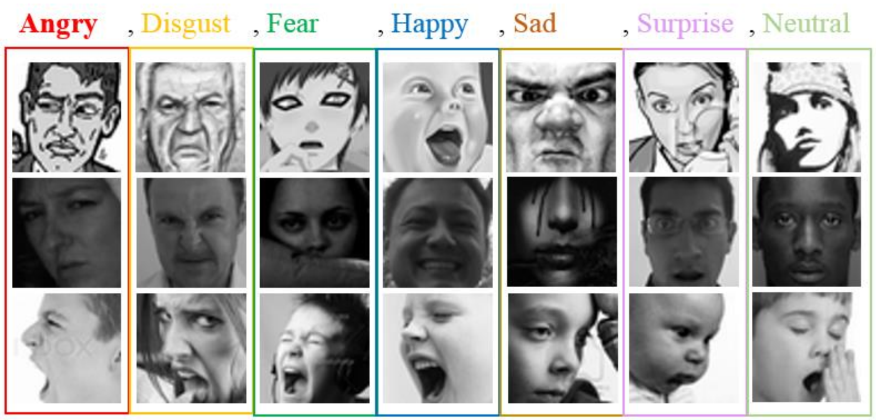
Facial Expression Recognition

- Evaluation dataset
  - JAFFE database

    http://www.kasrl.org/jaffe.html
  - The CK+ database

    http://www.consortium.ri.cmu.edu/ckagree/
  - FER2013 dataset

    https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data



Example of the FER2013 dataset

Example of the JAFFE dataset

Example of the CK+ dataset

Facial Expression Recognition

- References

– Jung, Heechul, et al. "Joint fine-tuning in deep neural networks for facial expression recognition." *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

– Zhao, Xiangyun, et al. "Peak-piloted deep network for facial expression recognition." *European Conference on Computer Vision*. Springer International Publishing, 2016.

– Fabian Benitez-Quiroz, C., Ramprakash Srinivasan, and Aleix M. Martinez. "EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

– Liu, Ping, et al. "Facial expression recognition via a boosted deep belief network." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.

- Tools

– Dlib C++ Library http://dlib.net/

– OpenCV http://opencv.org/

Gesture Recognition

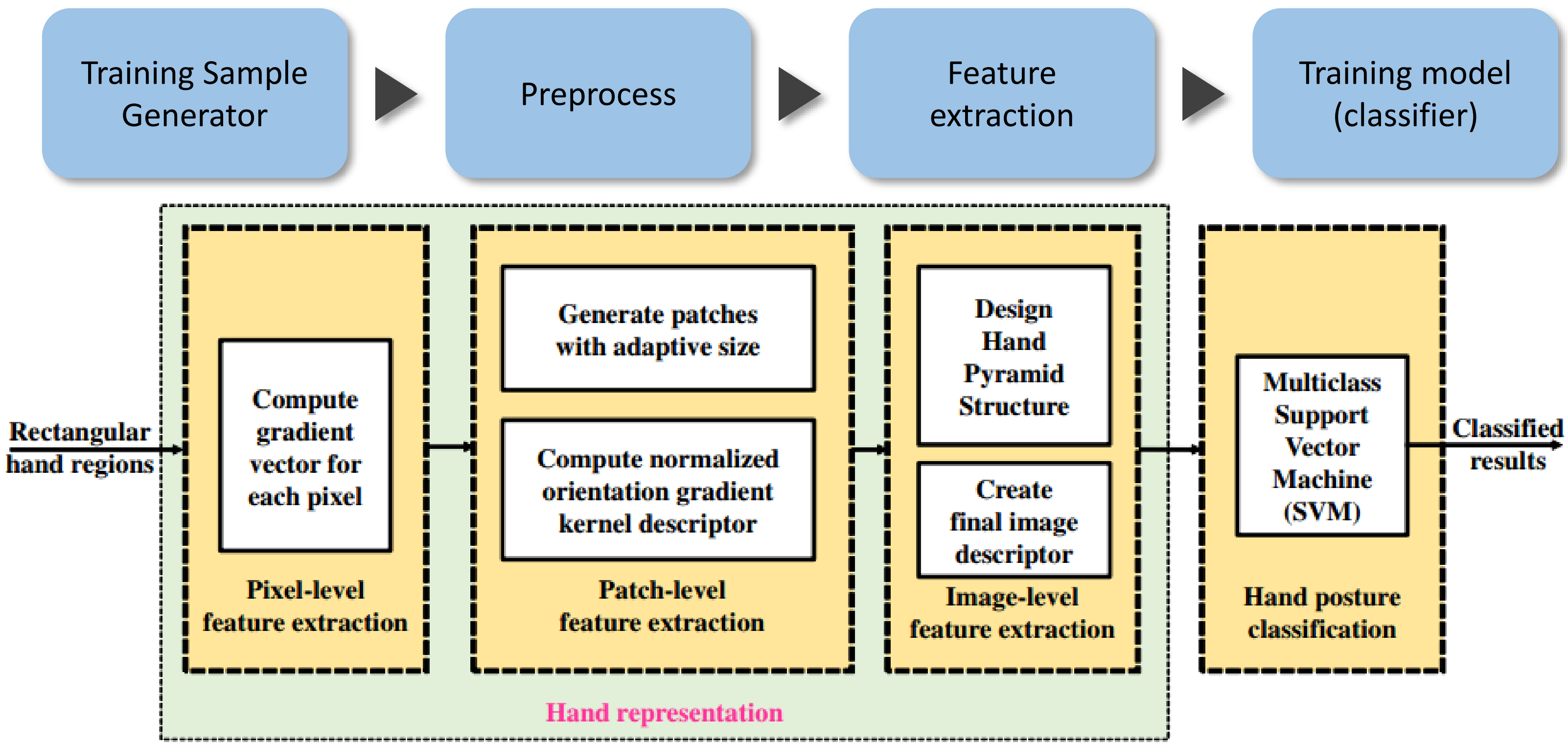
Gesture Recognition

- Problem description
  - Given depth or RGB images with a specific hand pose
  - The goal is to solve a classification problem that seeks the pre-defined gesture classes which is most similar to the input data

  - Sensor-based approach
  - Vision-based approach
    - Model-based approach
    - Appearance-based approach

Gesture Recognition

- Framework of hand posture recognition
  - Components
    - Training images
    - Testing images
    - Model



Gesture Recognition

- Datasets
  - Multimedia Technology and Telecommunications Laboratory, University of Padova
    - http://lttm.dei.unipd.it/downloads/gesture/
  - ICVL Hand Posture Dataset
    - http://www.iis.ee.ic.ac.uk/~dtang/hand.html
  - IEEE Computer Society Workshop on Observing and understanding hands in action (HANDS 2015)
    - http://www.ics.uci.edu/~jsupanci/HANDS-2015/#

Gesture Recognition

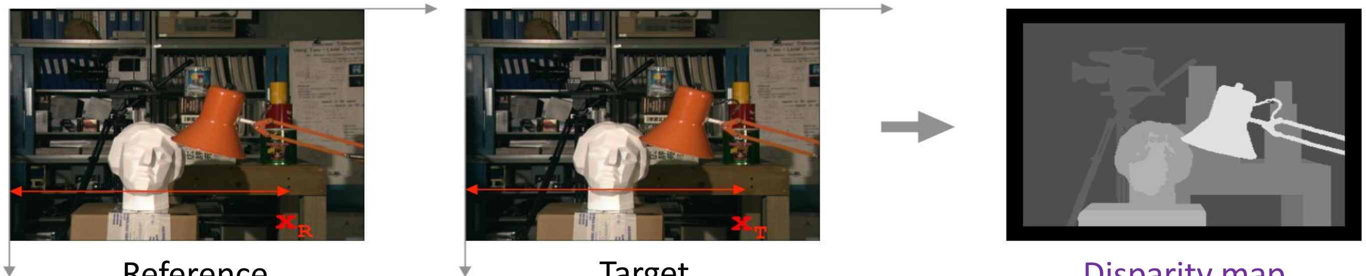
- References

  - G. Marin, F. Dominio, P. Zanuttigh, "Hand gesture recognition with Leap Motion and Kinect devices", IEEE International Conference on Image Processing (ICIP), Paris, France, 2014
  - G. Marin, F. Dominio, P. Zanuttigh, "Hand Gesture Recognition with Jointly Calibrated Leap Motion and Depth Sensor", Multimedia Tools and Applications, 2015
  - D. Tang, H.J. Chang*, A. Tejani*, T-K. Kim
    Latent Regression Forest: Structured Estimation of 3D Hand Posture, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)
  - D. Tang, T.H. Yu and T-K. Kim
    Real-time Articulated Hand Pose Estimation using Semi-supervised Transductive Regression Forests, Proc. of IEEE Int. Conf. on Computer Vision (ICCV), Sydney, Australia, 2013
  - Van-Toi NGUYEN et al,A New Hand Representation Based on Kernels for Hand Posture Recognition

Stereo Matching

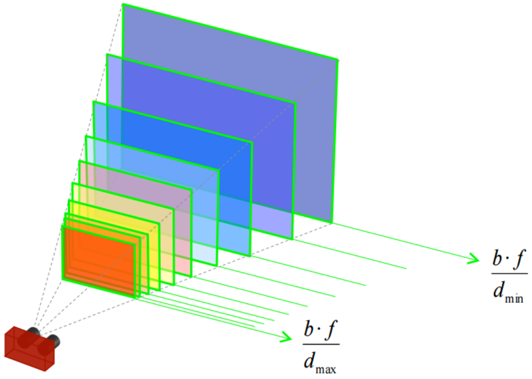
Stereo Matching

- Problem description
  - It is a technique aimed at inferring **depth** from two or more cameras.
  - The **disparity** is the difference between the $\mathbf{x}$ coordinate of two corresponding points.
    - It is typically encoded with greyscale image (closer points are brighter).



Reference Target Disparity map

- Depth measured by a stereo vision system is discretized into parallel planes (one for each disparity value)



Stereo Matching

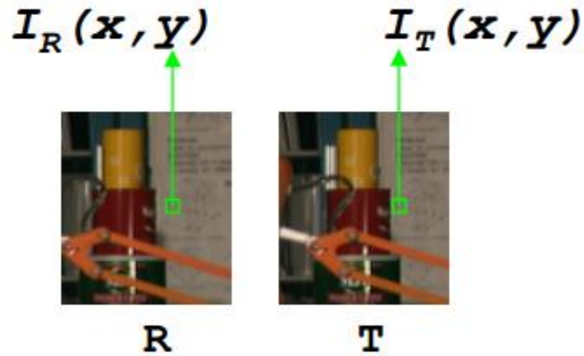
- According to [1] most stereo algorithms perform these steps:
    1. Matching cost computation
    2. Cost aggregation
    3. Disparity computation/optimization
    4. Disparity refinement
- Local algorithms
- Global algorithms

[1] System flowD. Scharstein and R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms Int. Jour. Computer Vision, 47(1/2/3):7–42, 2002

Stereo Matching

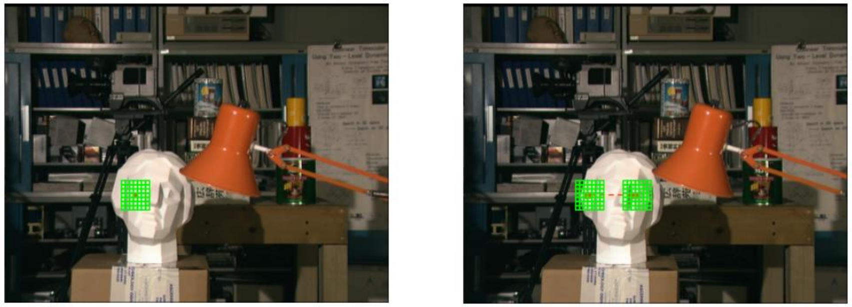
- Matching cost computation
  - Pixel-based matching costs
    - E.g. squared differences

$$E(x, y, d) = (I_R(x, y) - I_T(x + d, y))^2$$



- Cost aggregation
  - Sum of squared differences (SSD)
    - E.g. squared differences

$$\sum_{(x,y) \in w} (I_R(x, y) - I_T(x + d, y))^2$$



Reference Target

Stereo Matching

- Disparity computation/optimization
  - Energy function $E(d) = E_{data}(d) + E_{smooth}(d)$
  - Relevant approaches are:
    - Graph Cuts
    - Belief Propagation
- Disparity refinement
  - Raw disparity maps computed by correspondence algorithms contain outliers that must be identified and corrected
    - Sub-pixel interpolation
    - Image filtering techniques
    - Bidirectional Matching

Stereo Matching

- Evaluation & Datasets
  - Middlebury Stereo Datasets :
    - http://vision.middlebury.edu/stereo/data/
  - MPI Sintel Datasets :
    - http://sintel.is.tue.mpg.de/stereo

Stereo Matching

- References
  - Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér. Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. PAMI 2008.
  - X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang. On building an accurate stereo matching system on graphics hardware. GPUCV 2011.
  - Z. Wang and Z. Zheng. A region based stereo matching algorithm using cooperative optimization. CVPR 2008.
  - D. Scharstein and R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms
  - http://vision.deis.unibo.it/~smatt/Seminars/StereoVision.pdf
  - H. Ha. et al High-quality Depth from Uncalibrated Small Motion Clip

End

You can still choose other interesting topics