High Level Computer Vision

Project Introduction | SS 2021

31.05.2021 - Farzaneh Rezaeianaran

Slide Credit: Rakshith Shetty

Logistics

- Start date: 31.05
- Project Proposal Phase
  - 5+1 slides + max. 2 page report, due on Friday 11.06, 23:59 – task - goals - method - dataset - evaluation - references
  - Offline feedback through CMS
- Interim report
  - Slides + max. 2 page report, due on Friday 02.07, 23:59 – progress report / problems encountered / feedback
  - Offline feedback through CMS
- Final presentation (Will be graded)
  - Slides due on Friday 23.07, 23:59 (preliminary date) - Progress and presentation evaluation
  - Talk (15+5min)* - The exact information regarding final presentation will be announced later.
- Written report submitted on 02.08, (23:59)(preliminary date) (Will be graded)
  - Report evaluation

Project → Research

- Choose a dataset and task:
  - **Datasets**: Caltech4, Caltech101, Buffy Stickmen, HOI, UKBench, MPII Human Pose, ImageNet, COCO, CelebA etc.
  - **Tasks**: object detection/localization, person identification, gender recognition, scene classification, image captioning, visual question answering, image generation etc.
- What is the hypothesis you want to answer ?
- Conduct the experiments to test your hypothesis, present the analysis of your results
  - Necessary simplifications are OK (e.g. additional annotations)
  - Can you think of a new twist to the method?

Project → Application

- Application:
    - Apply computer vision techniques to a real-world problem.
    - New/interesting application of techniques learned in the course.
- Model
    - Build a new model/algorithm or a new variant of existing models for an existing computer vision task.
- Apply your methods to the task, present the analysis of your results.

Proposal Slides Structure

- Slide 1 – Task and motivation
  - Task statement and definitions
  - Motivation
  - Related work
- Slide 2 – Goals
  - Precisely, what do you want to achieve by end of the project
    - Eg. Implement method x on task k, compare it to method z and so on
  - What you want to have completed by the mid-term
    - Setup code for tasks x,y,z
    - Collect data
    - Setup baselines

Proposal Slides Structure

- Slide 3 – Methods
    - What is the primary models you will use
        - GANs, segmentation model with architecture X,
        - Provide exact references you will use
    - What tools/ code is already available, that you will use.
    - Related work
- Slide 4 - Data
    - What datasets you are going to use/collect and why
    - What simplifications if any you will perform

Proposal Slides Structure

- Slide 5 – Evaluation
    - How is your method going to be evaluated. What metrics are suitable?
    - Automatic metrics like accuracy, mAP, ...
    - User study
    - Public Leaderboards
    - Or your own method of evaluation
- Slide 6-X References
    - List your references for related work/ datasets/ code/ tools
    - Put the directly related work only
- Short Report Based on slides

Report Structure for Final Report

- Title
- Abstract
- Introduction
- Related work
- Proposed method explained
- Experimental results
- Conclusions and Future work
- References
- Reports to indicate assignments of each group member
- Honor Code: clearly cite your sources in your code and your report.

Conferences

- CVPR
- ICCV
- ECCV
- NeurIPS
- ICLR
- BMVC
- ACCV
- GCPR

Datasets

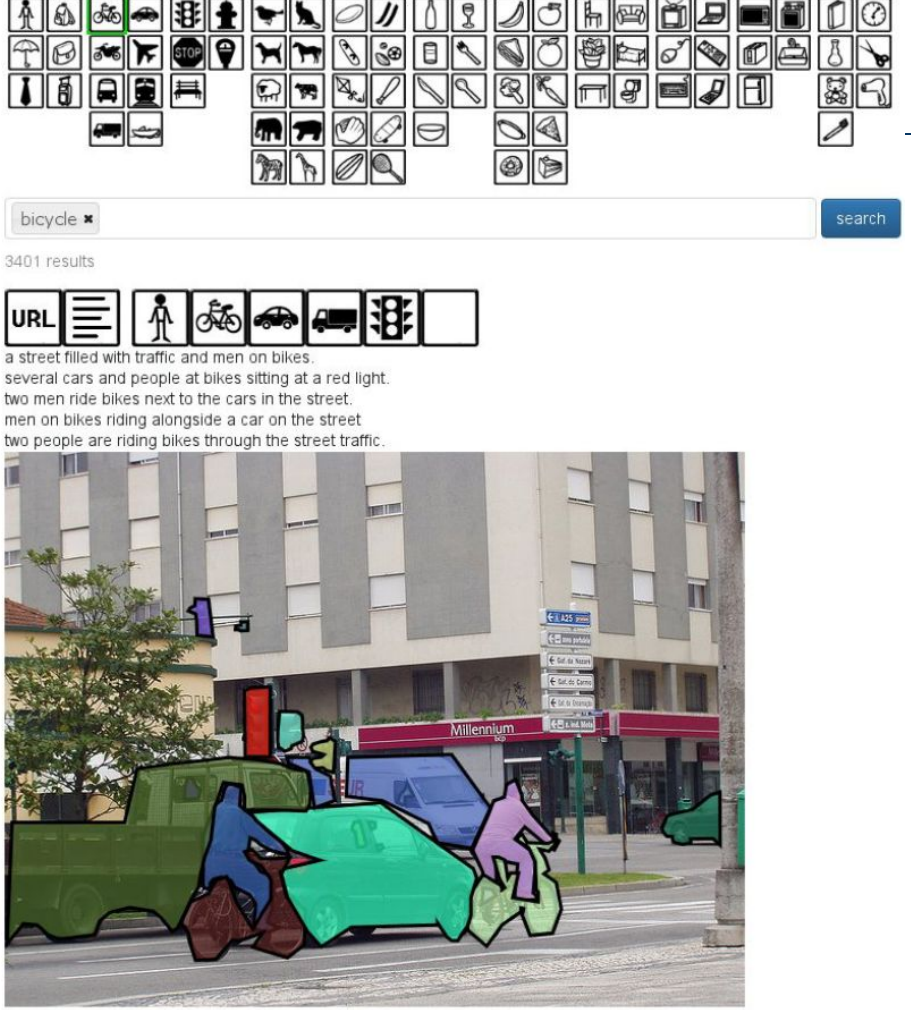
- Database of datasets
    - https://riemenschneider.hayko.at/vision/dataset/
    - http://www.cvpapers.com/datasets.html
- Data is key for any deep learning based model
- Think carefully about what data you use/ collect while choosing your task.
    - Fully supervised
    - Semi/Weakly supervised
    - Synthetic data

Datasets

- Pascal VOC:
  - Object detection
  - Segmentation

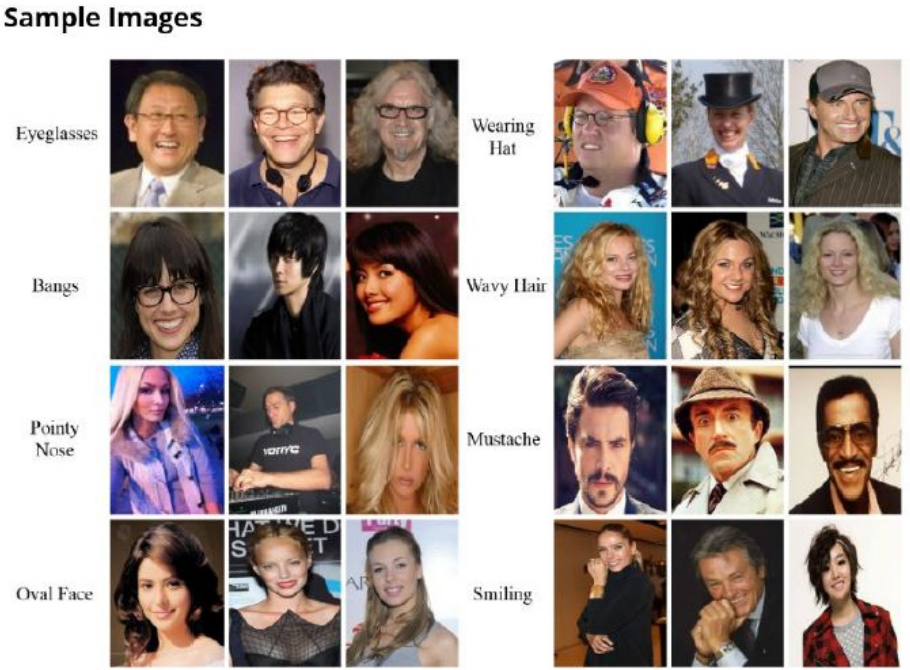
MS - COCO

- Common objects in Common Context
- Tons of annotation
  - Bounding Boxes
  - Instance Segmentations
  - Keypoints for humans
  - Panoptic segmentation
  - Image captions



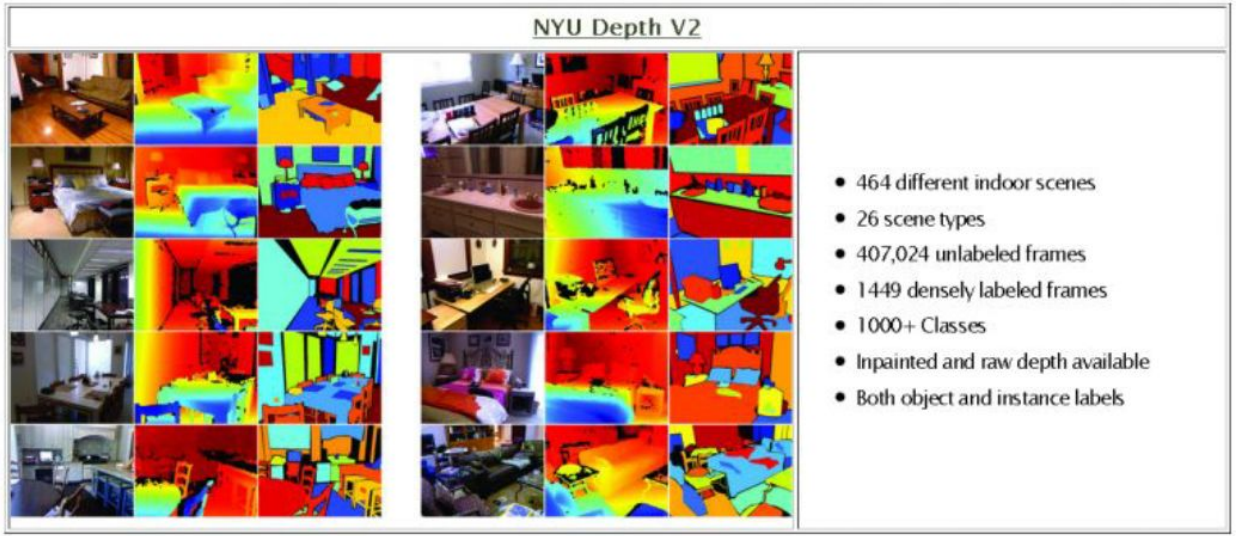
CelebA dataset

- 200k images
- 5 landmark locations and 40 binary attributes
- Cropped and centered version
- Commonly used in image generation and manipulation research



CelebA dataset

- RGB-D Indoor Scenes Dataset (http://cs.nyu.edu/~silberman/datasets/)
    - Scene classification
    - Object detection, recognition, segmentation



More Datasets

- ImageNet
- SUN Database
- Places Database
- MPII Human Pose
- Open Images https://storage.googleapis.com/openimages/web/index.html
- Labeled Faces in the Wild
- ….

Dataset

- Or capture your own
    - Digital camera, mobile phone, Google glasses..
    - Microsoft Kinect
    - Web / Google image search

For a good project

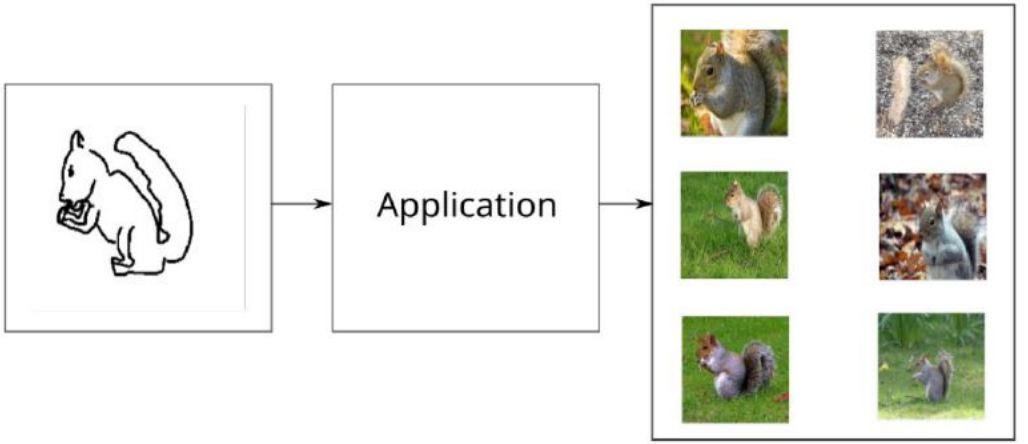
- 5 W's
    - What? (a problem)
    - Why? (motivation)
    - How? (proposed strategy)
    - Where? (dataset and benchmark)
    - Who? (team assignments)
- It is recommended
    - Baseline
- It is desired.. your considerations on
    - Influence of parameter and dataset choice
    - Results: what is expected and what is surprising.. not just numbers!
    - Observations must be substantiated by results or references

Example Projects

- Gender/ Age Recognition
- Object recognition or detection with the Kinect (RGB + Depth)
- Image retrieval for 3D objects
- Object retrieval in videos / on a mobile phone
- Person identification
- Image and video segmentation
- Detection and segmentation
- Tracking
- Vision and Language tasks (captioning, question answering, explanations)
- Image generation tasks (GANs, conditional generation, style transfer)

Previous year projects

Sketch based image retrieval



- Train models for embedding sketch and images
- Learn to keep the embeddings for matching items close
- Nearest neighbor search to retrieve matching images in test time.

Reconstruct images from sparse version

- Studied if it is possible to reconstruct images from sparse version
- What kind of sampling works best ?
- What kind of architecture works best ?

Tumor segmentation

- Limited training data available
- Transfer learning from larger brain tumor segmentation dataset to smaller lungs dataset
- Shows better performance than training from scratch

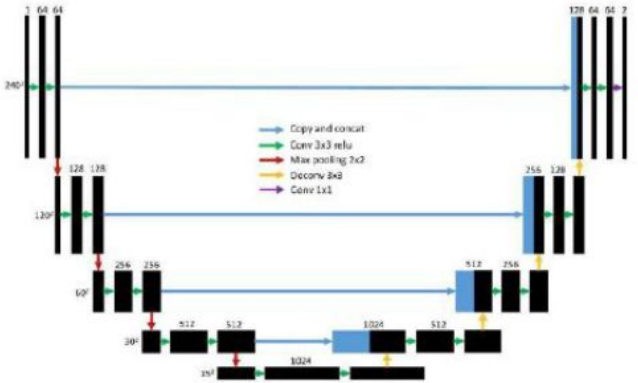
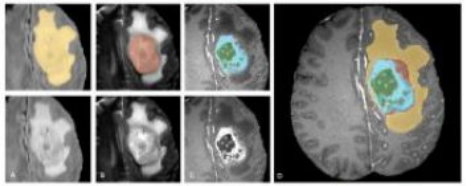


Figure 1. The U-Net architecture has a 240x240x4 input and the output is a semantic label for each pixel



Manga colorization

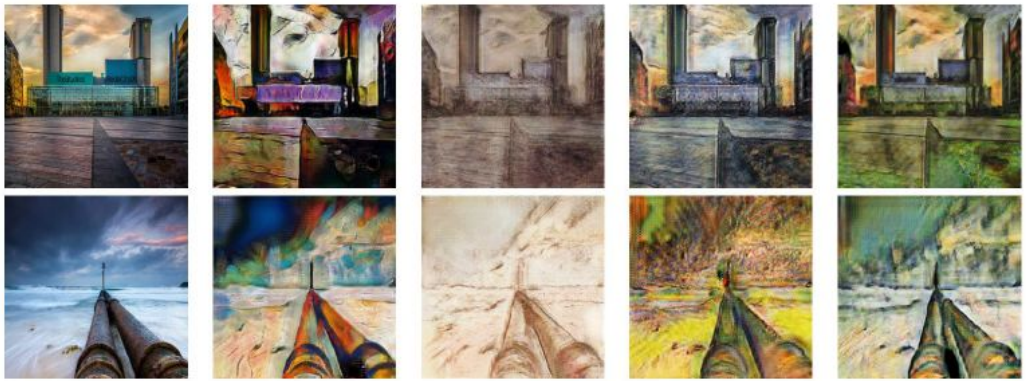
- They collected the dataset by scraping the web and pre-processing to extract paired data

- GAN based generator
  - Comparison to simple L1 L2 baselines
  - Different color-spaces
  - Comparison across monochrome and binary settings

Figure 1. Manga Colorization: from monochrome or grayscale to colored images

Painting Style Transfer

- Conditional GAN based architecture
- Single generator to switch to different styles based on input condition
- Quantitative evaluation using classifier and user study

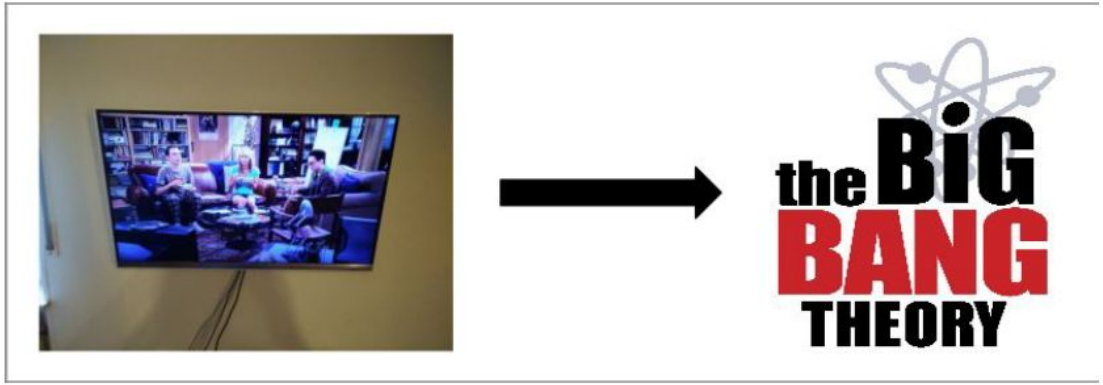


Face beauty filter

- Based on people rating of beauty on a bunch of photos
- Try to create a version of the image which maximizes this score.
- Very subjective!

TV series classification

- Classify tv series from short smartphone videos
- CNN frame level classification (designed based on related work)
- Collected own dataset !!
- Synthetic data augmentation



More project ideas

- Look at student projects done here
  - http://cs231n.stanford.edu/2017/reports.html
  - http://cs231n.stanford.edu/2016/reports.html
  - http://cs231n.stanford.edu/2015/reports.html
  - http://cs229.stanford.edu/proj2019aut/
  - http://cs229.stanford.edu/proj2018/
  - http://cs229.stanford.edu/proj2019spr/

Any Question?