Deep Learning Projects (CSE 641)

General Guidelines for All Projects

1. Decide your project and **submit a proposal by 11:59PM, Wednesday, 6th March, 2019**.
2. We will have **in-class presentation on 26th March, 2019**. You will be required to present your model and training strategy.
3. The **final presentations will be on 1st May, 2019**, for which you will be submitting a short report on what you have done (with code, evaluation strategy, accuracy and individual contributions).

This year's project recommendations:

1. Traffic Monitoring for Road Safety

a. The 2018 NVIDIA AI City Challenge, CVPR 2018

2. Visual Animal Biometrics[1]

a. Identifying Individual Monkeys/Zebras/Whale Sharks

3. Analytics for Video Surveillance and Forensics

a. Visual Q&A for Image Retrieval

b. Real-Time Object Tracking (Single Camera or Multi-Camera Tracking)

c. Object Detection in Compressed Domain Videos

Reference:

Visual Q&A in Compressed Domain Videos?

Fast Object Detection in Compressed Video, ArXiv Jan 2019

Compressed Video Action Recognition, CVPR 2018

4. Unsupervised Domain Adaptation for Visual Object Detection

Reference: Domain Adaptive Faster-RCNN, CVPR 2018

5. Automatic Speech Recognition on Mobile Phones

6. Natural Speech Synthesis / Voice Conversion

a. Change of Speaker

b. Change of Emotion

7. Text based Conversational Agent

a. Understanding long-term context in a conversation

8. Detection of Satire in Headlines

Reference: West and Horvitz, 'Reverse-Engineering Satire, or "Paper on Computational Humor Accepted Despite Making Serious Advances"', AAAI 2019

[1] Meet me regarding datasets if you are interested in this project.

Last Year's (DL-2018) Projects

1. Deep Learning for Mobile Platforms

OBJECTIVE

Current works in the area of object recognition and semantic segmentation rely on expensive computational resources like distributed CPU clusters and high end GPUs. Today's mobile platforms have become powerful enough to run traditional Machine Learning algorithms. Your task is to port an existing object detection framework (eg. SSD) to a mobile platform (preferably android). Evaluation will be done on the basis of the time taken for detecting objects, number of objects correctly detected, class confidence score and RAM usage. You may use different model compression techniques for benchmarking.

DATASET

You can use any publicly available dataset for training. The testing images would be a small set of images containing many objects taken from the Internet.

REFERENCES

1. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg "SSD: Single Shot MultiBox Detector", ECCV 2016.
2. https://developer.android.com/studio/index.html
3. https://www.slideshare.net/anirudhkoul/squeezing-deep-learning-intomobile-phones

2. Low Resolution Face Recognition

OBJECTIVE

To train a deep learning model that works with low-resolution face recognition. For this purpose, you can use the Labeled Faces in the Wild dataset and synthetically generate low-resolution images for training and testing your model.

DATASET

Labeled Faces in the Wild (LFW) with synthetically downsampled images. As is standard procedure, train and test splits should have different identities.

EVALUATION

Report accuracy and cumulative match characteristic (CMC) scores for different factors of subsampling. For the original resolution, the accuracy should meet the state-of-the-art (acc_{orig}) and for a lower-resolution image with a subsampling factor of *k*, the accuracy should reduce gracefully as $acc_{orig} – k$*x. Your goal would be to achieve the smallest x. Report accuracies for k= 1, 2 and 3.

REFERENCES

1. Erik Learned-Miller, Gary B. Huang, Aruni RoyChowdhury, Haoxiang Li, and Gang Hua. **Labeled Faces in the Wild: A Survey.**

3. Wildlife Detection and Recognition in Camera Trap Images

OBJECTIVE

The use of camera trap images for monitoring animals in the wild is on the rise. This results in a huge number of images, which have to be classified on the basis of species. Use the provided dataset to train an animal species detector.

Given an image, perform the following tasks:

1. Draw bounding boxes around the animals present in the camera trap images.
2. For each bounding box, classify the species of the animal present in it.
3. Report accuracies for the test set provided.

DATASET

Use the dataset with 20 species found in India.

REFERENCES:

1. Stefan Schneider∗ , Graham W. Taylor† , Stefan C. Kremer∗, “Deep Learning Object Detection Methods for Ecological Camera Trap Data”, ArXiv, March 2018.
2. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg “SSD: Single Shot MultiBox Detector”, ECCV 2016.
3. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”

4. Domain Adaptation for Image Segmentation

OBJECTIVE

Training a CNN based Semantic Segmentation model requires ground-truth annotations at the pixel level. Such annotations are readily available in synthetic datasets generated from game engines like Unity, Unreal, etc. However, upon training DNN models on synthetic data and testing it on real data, there is a significant domain shift. In this project, your goal is to use state-of-the-art unsupervised domain adaptation techniques for performing well on real-world images.

DATASET

Train on Synthetic dataset like Synthia and test it on Camvid

Download dashcam videos for Indian roads (e.g., from youtube) and show qualitative results on a test set.

REFERENCES

1. Sankaranarayanan et al. Learning from Synthetic Data: Addressing Domain Shift for Semantic Segmentation
2. Hoffman et al. FCNs in the wild: Pixel-level adversarial and constraint-based adaptation
3. Bousmalis et al. Unsupervised Pixel–Level Domain Adaptation with Generative Adversarial Networks