Module 1 Project: Computer Vision Application

Machine Learning

Fall 2019

Motivation

Computer vision and image classification tools are rapidly reshaping how we live our lives. Machine learning approaches have driven recent progress in an array of technologies that have the potential to realize huge positive impacts on our world. These tools have become increasingly easy to apply to masses of existing data, however, these tools and their real-world application do not always have the results that their creators intend. To achieve better outcomes, it is important to deeply consider the potential implications and limitations throughout the application design and implementation process. This project aims to help you to begin to develop the skills to understand, implement, and critically evaluate machine learning systems.

Goal-Setting and Customization

You will be considering your own learning goals for this project, as well as the learning goals of your teammate(s) (if applicable). You can customize this project to support your own learning, and we are happy to help you shape the project to support your goals and challenge yourself.

Project Description

In this project, you will contemplate a potential application for computer vision (machine learning of images) and you will implement a machine learning algorithm on a data set in support of this application. As part of this project, you will:

- Document the important considerations for your application (e.g., what data are available for training, how well would the algorithm need to work to create value, how could the algorithm and application be tested (beyond the testing you choose to implement), what are the implications of this application in the world, what are the stakeholders, what are the risks).

- Build and iterate on a computer vision model as a step toward this application. We don't expect you to build the entire application, just to work toward implementing an early version of an algorithm that could be used for this application. It's likely that you will use a convolutional neural network, but not required. You will build a test multiple versions of your model.

- Visualize key aspects of your data/model. Include at least two visualizations in your report, but you'll probably find it valuable to visualize various aspects of the data and model (this is helpful in sanity checking and can give good insight into how the model is working).

- Test and evaluate your model. This will likely include accuracy on training and test sets. Depending on your application, you may also want to collect your own photos, find other images online, and/or manipulate images from your original test set (adding noise, shifting/flipping images, etc). You will also evaluate the effectiveness of this model for your specific application and describe the limitations of your current model and data.
- Document your final analysis pipeline for transparency. (A simplified version that does not need to include every parameter, visualization, or tweak that you tried.)

Resources

- Interactive visualization for considerations in your application: https://www.cdt.info/ddtool/. The background for this visualization is here: https://cdt.org/issue/privacy-data/digital-decisions/
- FAT ML principles for accountable algorithms (from the first day of class): https://www.fatml.org/resources/principles-for-accountable-algorithms
- Suggested questions to consider about your application in Appendix A.
- Suggested data sets in Appendix B and here

Appendix A: Some questions to consider

source

- What do you want your model to be able to do?
- How can you imagine your model (or an extension of it) being used in the real world? Feel free to get creative.
- If those ideas came true, who might they affect? In what ways?
- What measures would your model's real-world implementers need to take to ensure its effects live up to your intentions?
- What pitfalls might your model fall into, and what could you quantitatively measure to avoid those?
- Why was the dataset you used to train your model created?
- In what other ways could the same data be used? How do you feel about those possibilities?
- How was that dataset assembled? From where was the data sourced? Who or what labeled it? If there are any elements of this process you think were either particularly well done or problematic, how so?

- If your dataset contains information about people, to what degree did those people have agency over their inclusion? Do you feel that matters in this case? Why or why not?
- Skimming through your dataset, does anything stand out to you about representation in its contents?
- Do you feel the potential use cases for this dataset justify it being created and published? Why or why not?

Appendix B: Computer Vision Datasets

PyTorch built-in datasets

- CIFAR-10: 60,000 labeled 32x32 color images of 10 classes of objects
- CIFAR-100: 60,000 labeled 32x32 color images of 100 fine classes of objects, also grouped into 20 course superclasses
- EMNIST: 800,000 labeled 28x28 grayscale images of handwritten digits and letters (uppercase and lowercase) that expand the MNIST dataset
- Fashion-MNIST: 70,000 labeled 28x28 grayscale images of 10 classes of clothing articles
- ImageNet 2012: 1,331,167 labeled color images of tens of thousands of classes of nouns in the WordNet hierarchy
- KMNIST: 70,000 labeled 28x28 grayscale images of 10 classes of handwritten Japanese Hiragana characters
- LSUN: 708,564 labeled large-scale color images of 10 classes of scenes/settings
- MNIST: 70,000 labeled 28x28 grayscale images of 10 classes of handwritten digits
- SVHN: Color images of house numbers (addresses) obtained from Google Street View w/ labeled bounding boxes around individual digits

TensorFlow built-in

- Cats and Dogs: 25,000 labeled images of cats and dogs
- CelebA: 202,599 total face images of 10,177 identities w/ facial landmarks, aligned & cropped images, bounding boxes, and binary attribute labels
- CIFAR-10: 60,000 labeled 32x32 color images of 10 classes of objects
- CIFAR-100: 60,000 labeled 32x32 color images of 100 fine classes, also grouped into 20 course superclasses
- CIFAR-10-C: Images from CIFAR-10 manipulated using 15 common corruptions at 5 levels of severity each

- Colorectal Histology: 5,000 labeled 150x150 color histological images of 8 tissue types of human colorectal cancer
- CBIS-DDSM: 2,620 labeled scanned film mammography studies of normal, benign, and malignant cases of human breast cancer
- Diabetic Retinopathy Detection: 88,712 labeled images of eyes with diabetic retinopathy at none, mild, moderate, severe, and proliferative levels of severity
- EMNIST: 800,000 labeled 28x28 grayscale images of handwritten digits and letters (uppercase and lowercase) that expand the MNIST dataset
- Fashion-MNIST: 70,000 labeled 28x28 grayscale images of 10 classes of clothing articles
- Horses or Humans: 1,283 labeled 300x300 color images of cgi humans and horses
- ImageNet 2012: 1,331,167 labeled color images of tens of thousands of classes of nouns in the WordNet hierarchy
- ImageNet-C 2012: Images from ImageNet 2012 manipulated using 12 common corruptions at 5 levels of severity each
- KMNIST: 70,000 labeled 28x28 grayscale images of 10 classes of handwritten Japanese Hiragana characters
- LSUN: 708,564 labeled large-scale color images of 10 classes of scenes/settings
- MNIST: 70,000 labeled 28x28 grayscale images of 10 classes of handwritten digits
- MNIST-C: Images from MNIST manipulated using 15 common corruptions at 5 levels of severity each
- Omniglot: 38,300 labeled grayscale images of 1,623 classes of handwritten characters from 50 alphabets w/ stroke data in [x,y,t] coordinates where time (t) is in milliseconds
- Oxford Flowers: 8,189 labeled color images of 102 classes of flowers
- Oxford-IIIT Pet: 7,349 labeled color images of 37 classes of pet breeds w/ tight bounding boxes around the heads and pixel-level foreground-background-boundary segmentation
- PatchCamelyon: 327,680 labeled 96x96 color images of histopathologic lymph node scans where metastatic tissue is either present or absent
- PetFinder: 72,776 color images of dogs and cats w/ many attribute labels
- Quick, Draw! Bitmap: 50,426,266 labeled 28x28 grayscale images of 345 classes of drawn objects contributed by Quick, Draw! players
- NWPU-RESISC45: 31,500 labeled 256x256 color images of 45 classes of scenes/settings
- Rock Paper Scissors: 2,892 labeled 300x300 color images of cgi hands in rock, paper, and scissors gestures

- small NORB: 48,600 labeled 96x96 color images of 5 classes of toys
- Sun397: 108,753 labeled color images of 397 classes of scene/settings
- SVHN-cropped: 600,000 labeled 32x32 color images of digits of house numbers (addresses) obtained from Google Street View
- TF Flowers: 3,670 labeled color images of 5 classes of flowers
- UC Merced Land Use: 2,100 labeled 256x256 color aerial images of 21 classes of land uses collected from USGS National Map Urban Area Imagery

Other datasets

- Bark-101: 2,594 labeled color images of 101 classes of tree bark
- Caltech 101: 9,146 labeled color images of 101 classes of objects
- Caltech 256: 30,607 labeled color images of 256 classes of objects
- CARS: 604 labeled color images with either cars or no cars
- CAS-PEAL-R1: 30,900 labeled grayscale images of 1,040 faces (exclusively Chinese)
- CCPD: 300,000 labeled color images of license plates w/ bounding boxes, license plate numbers, and several dimension labels
- CINIC-10: A drop-in replacement for CIFAR-10 with 270,000 images collected by downsampling images from ImageNet
- CRACK: 1,428 labeled 299x299 color images with either cracks or no cracks
- CyberExtruder Ultimate Face Matching Dataset: 10,205 labeled 600x600 color images of 1000 faces scraped from the internet
- DFW: 11,157 labeled color images of 1000 subjects' faces, including examples of that subject attempting to obfuscate their face and of other individuals attempting to impersonate that subject
- DiF: 1,000,000 color images of diverse faces w/ dimension labels for objective physical properties
- FERET: 14,126 labeled 384x286 color images of 1,119 faces
- Food-101: 101,000 labeled 512x512 color images of 101 classes of food
- Food-475: 247,636 labeled color images of 475 classes of food
- Food-524: 247,636 labeled color images of 524 classes of food
- HistAerial: 4,900,000 labeled grayscale aerial images of 7 classes of ground use
- IMDb: 460,723 color images of celebrities w/ bounding boxes around faces and attribute labels including DOB, name, binary gender (literally)

- IP102: 75,000 labeled color images of 102 classes of insects
- LAG: 11,760 labeled color images of suspicious or negative glaucoma samples
- LFW: 13,233 labeled 250x250 color images of 5,749 faces
- LFW-a: Images from LFW that have been aligned using a commercial face alignment software
- LFWcrop: Images from LFW that have been cropped to only include the contents of facial bounding boxes
- ML-Images: 17,698,491 labeled color images of 11,166 classes of objects
- Multi-PIE: 750,000 labeled color images of 337 faces
- PIE: 41,368 labeled color images of 68 faces
- QMNIST: An extension of the MNIST data set to include 50,000 additional test images
- Real Rain: 29,500 labeled color image pairs, one with and one without rain
- ROOMS: 10,029 labeled (out of 20,001 total) color images of 6 classes of room types
- SCface: 4,160 labeled color images of 130 faces taken from 6 security cameras of varying quality in either visible light or infrared mode
- SoF: 42,592 labeled color images of 112 faces (exclusively glasses-wearing)
- Stanford Dogs: 20,580 labeled color images of 120 breeds of dogs w/ bounding boxes
- VERI-Wild: 416,314 labeled color images of 40,671 vehicles