CSC420
Detecting The Avengers

Frank Karunaratna & Michael Ly

Problem Definition

- Detecting faces in an image or frames of a video
- Classifying the faces in an image
- Classifying the image (global classifier)
- Detecting cuts in a video from the frames

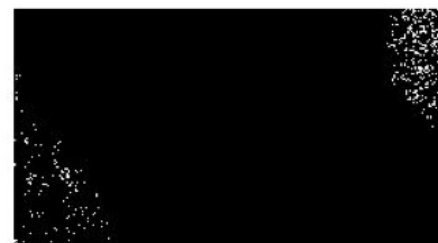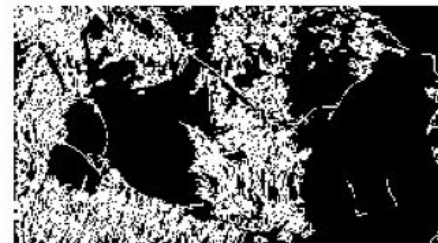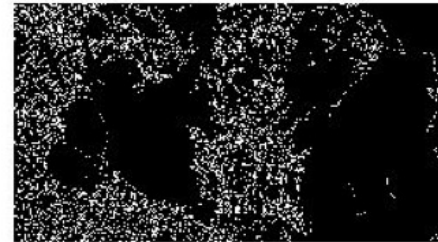# Face Classification Method

- Developed a CNN classifier to classify faces
- Found sufficient loss function
  - Cross Entropy Loss
- Data augmentation
  - Flips
  - Zooms
  - Filters
- Low learning rate of 0.0001

```
Net(
  (ConvLayer1): Sequential(
    (0): Conv2d(1, 8, kernel_size=(3, 3), stride=(1, 1))
    (1): Conv2d(8, 16, kernel_size=(3, 3), stride=(1, 1))
    (2): ReLU()
    (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (ConvLayer2): Sequential(
    (0): Conv2d(16, 32, kernel_size=(5, 5), stride=(1, 1))
    (1): Conv2d(32, 32, kernel_size=(3, 3), stride=(1, 1))
    (2): ReLU()
    (3): MaxPool2d(kernel_size=4, stride=4, padding=0, dilation=1, ceil_mode=False)
  )
  (ConvLayer3): Sequential(
    (0): Conv2d(32, 64, kernel_size=(3, 3), stride=(1, 1))
    (1): Conv2d(64, 64, kernel_size=(5, 5), stride=(1, 1))
    (2): ReLU()
    (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (ConvLayer4): Sequential(
    (0): Conv2d(64, 128, kernel_size=(5, 5), stride=(1, 1))
    (1): Conv2d(128, 128, kernel_size=(3, 3), stride=(1, 1))
    (2): ReLU()
    (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (Lin1): Linear(in_features=15488, out_features=6000, bias=True)
  (Lin2): Linear(in_features=6000, out_features=600, bias=True)
  (Lin3): Linear(in_features=600, out_features=60, bias=True)
  (Lin4): Linear(in_features=60, out_features=6, bias=True)
  (Lin5): Linear(in_features=123008, out_features=6, bias=True)
)
```

Shot Detection Method

- **What's Shot Detection?**
  - Given a video, find the beginning and end of each shot.
- **Edge Change Ratio (ECR)**
  - Computes the outgoing and incoming edge pixels compared to existing edge pixels
  - A high dissimilarity is likely to be a new shot!
- **Global Threshold:**
  - Declares a cut when a global threshold is passed
- **Adaptive Threshold:**
  - Gets max in neighborhood

Global Classifier Method

- Global Descriptor: GIST
  - "Gist": the essence of the image
  - GIST feature extraction stage
  - CNN stage
- Train model to be able to detect indoor vs outdoor images
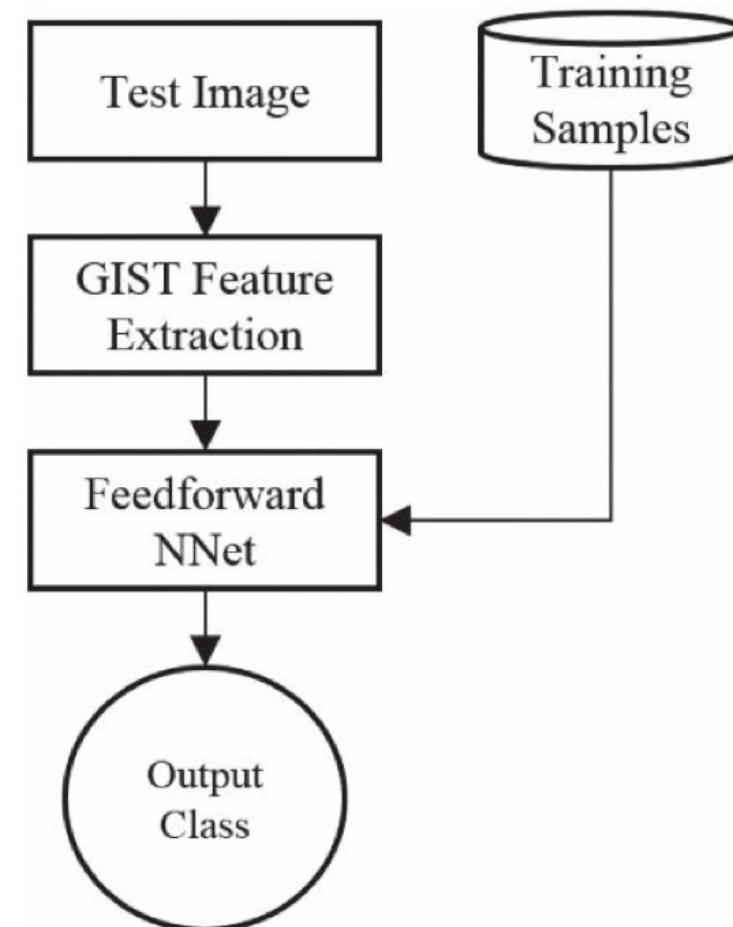
Source: (Tahir et al.)

Face Classifier Results

robert_downey_jr: 0.17%
chris_evans: 99.79%
scarlett_johansson: 0.0%
mark_ruffalo: 0.0%
gwyneth_paltrow: 0.0%
other: 0.05%
Prediction: chris_evans
True Label: chris_evans

robert_downey_jr: 0.0%
chris_evans: 0.0%
scarlett_johansson: 64.09%
mark_ruffalo: 1.03%
gwyneth_paltrow: 30.27%
other: 4.61%
Prediction: scarlett_johansson
True Label: scarlett_johansson

robert_downey_jr: 0.0%
chris_evans: 0.01%
scarlett_johansson: 0.0%
mark_ruffalo: 0.0%
gwyneth_paltrow: 99.98%
other: 0.01%
Prediction: gwyneth_paltrow
True Label: gwyneth_paltrow

- Achieved 84% accuracy on our test data using base model
- Achieved 60% accuracy on our augmented model
  - Data augmentation lowered our accuracy
- High confidence on some images and lower confidence on others
  - Likely due to training data

Shot Detection Results

- ECR was accurately able to detect hard cuts
- Soft cuts were harder to detect
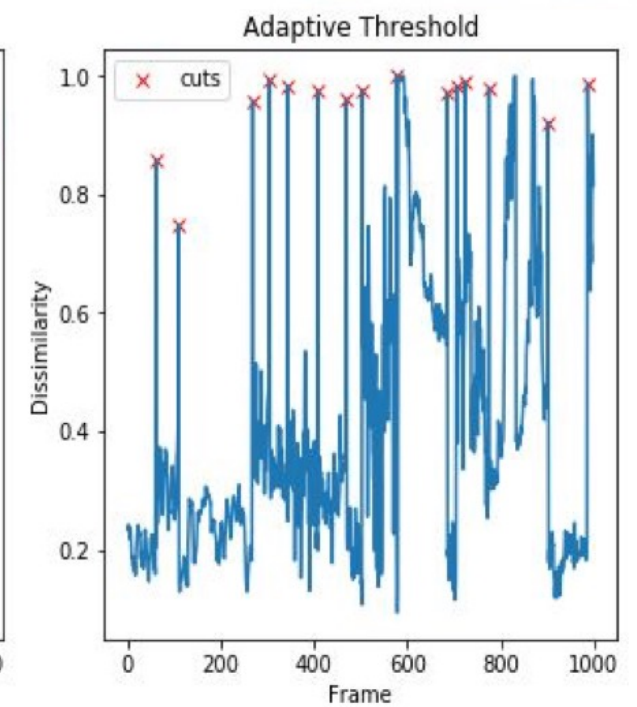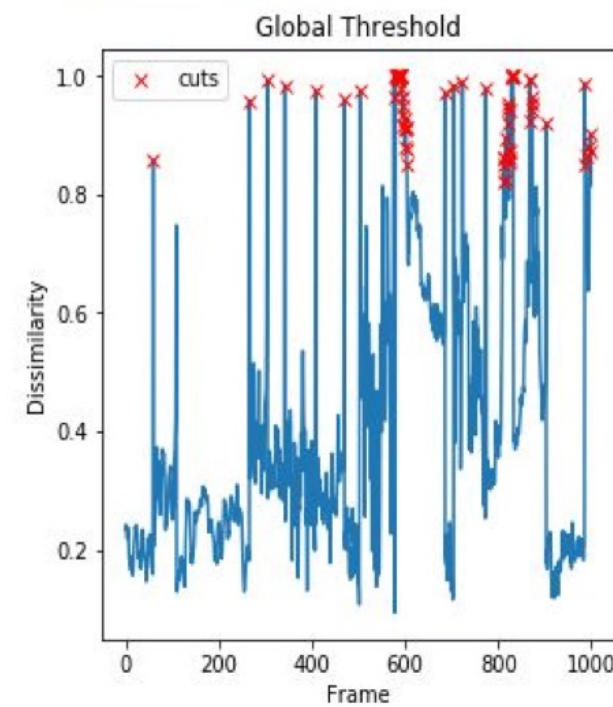
Missing: [832]
Overgenerated: []

Global Classifier Results

- Achieved 50% accuracy on our test set
  - Low amount of data for our classifier to train on
  - Long training time was a constraint (6 minutes per epoch)

```
[ ]  model, best_model = train_gist(indoor_outdoor_traindataloader, indoor_outdoor_te

       0%|            | 0/3 [00:00<?, ?it/s]
      33%|███         | 1/3 [05:30<11:00, 330.24s/it]Running loss: 352.9105854034424
    Accuracy of the network on the test images: 50.0

      67%|██████      | 2/3 [10:59<05:30, 330.03s/it]Running loss: 348.8928301334381
    Accuracy of the network on the test images: 50.0

     100%|██████████  | 3/3 [16:29<00:00, 330.04s/it]
    Running loss: 347.38148605823517
    Accuracy of the network on the test images: 50.0
    Finished training
```

Comparison

- Face classifier
  - Model trained on augmented data performed worse
  - Model trained on non-augmented data performed better
- Shot detection
  - Adaptive threshold selects most of the true cuts
  - Global threshold over detects cuts
- Global classifier
  - Same accuracy as a random coin flip

Topics Covered in the Course

- Filtering / Image Manipulation
  - Applying different image manipulations such as zooms + upscaling, changing the brightness, applying a gaussian filter
- Deep Learning
  - Developing and training a CNN for face classification
- Global Classification
  - Developing and training a model for global classification of an image

Team Contributions

- Frank Karunaratna
  - Creating the face classifier
  - Finding the datasets and setting up the training
  - Finding test video
- Michael Ly
  - Creating the shot detection
  - Researching how to do shot detection
  - Creating the global classifier
  - Setting up the training and finding the datasets for the global classifier

robert_downey_jr

Project Citations

"Avengers Assemble Scene - The Avengers (2012) Movie Clip HD." Youtube, ShortClips, https://www.youtube.com/watch?v=oBqqI6NMeaM&fbclid=IwAR0I_RNvYfXTt3WRM5W6VZViQhbKOi_Mw6nDiaWV3F9Yy43pUvSRu6hdN1A. Accessed 6 Dec. 2019.

Baidya, Arindam. "How to Increase Your Small Image Dataset Using Keras ImageDataGenerator." Medium, 30 Sept. 2018, https://medium.com/@arindambaidya168/https-medium-com-arindambaidya168-using-keras-imagedatagenerator-b94a87cdefad.

Centeno, Iván de Paz. Ipazc/Mtcnn. 2018. 2019. GitHub, https://github.com/ipazc/mtcnn.

Dahal, Paras. "Classification and Loss Evaluation - Softmax and Cross Entropy Loss." DeepNotes, 28 May 2017, http://deepnotes.io/softmax-crossentropy.

Saglani, Vatsal. "Multi-Class Image Classification Using CNN over PyTorch, and the Basics of CNN." Medium, 28 June 2019, https://medium.com/@thevatsalsaglani/multi-class-image-classification-using-cnn-over-pytorch-and-the-basics-of-cnn-fdf425a11dc0.

Skvarenina, Peter. "Detecting Facial Features Using Deep Learning." Medium, 4 Aug. 2017, https://towardsdatascience.com/detecting-facial-features-using-deep-learning-2e23c8660a7a.

Tahir, Waleed, et al. "Indoor/Outdoor Image Classification Using GIST Image Features and Neural Network Classifiers." 2015 12th International Conference on High-Capacity Optical Networks and Enabling/Emerging Technologies (HONET), IEEE, 2015, pp. 1–5. DOI.org (Crossref), doi:10.1109/HONET.2015.7395428.

Zhang, Kaipeng, et al. "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks." IEEE Signal Processing Letters, vol. 23, no. 10, Oct. 2016, pp. 1499–503. DOI.org (Crossref), doi:10.1109/LSP.2016.2603342.

Project Citations cont.

Brownlee, Jason. "How to Perform Face Detection with Deep Learning." *Machine Learning Mastery*, 2 June 2019, https://machinelearningmastery.com/how-to-perform-face-detection-with-classical-and-deep-learning-methods-in-python-with-keras/.

Brownlee, Jason. "How to Develop a Face Recognition System Using FaceNet in Keras." *Machine Learning Mastery*, 6 June 2019, https://machinelearningmastery.com/how-to-develop-a-face-recognition-system-using-facenet-in-keras-and-an-svm-classifier/.

Wang, Chi-Feng. "How Do You Train a Face Detection Model?" *Medium*, 8 Oct. 2018, https://towardsdatascience.com/how-do-you-train-a-face-detection-model-a60330f15fd5.

Yang, S., Luo, P., Loy, C., & Tang, X. (2016). WIDER FACE: A Face Detection Benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

"Yonatankatz/Edge-Change-Ratio-Example." *GitHub*, https://github.com/yonatankatz/edge-change-ratio-example. Accessed 3 Dec. 2019.

Lin, Heshan. "Video Shot Detection." Microsoft PowerPoint file. https://cis.temple.edu/~latecki/Courses/CIS750-03/Lectures/Heshan_ShotDet.ppt

Nasreen, Azra, and Shobha G. "Key Frame Extraction Using Edge Change Ratio for Shot Segmentation ." *Ijarcce*, vol. 2, no. 11, Nov. 2013, doi:10.17148/ijarcce.2013.24.

Big Buck Bunny » Download. https://peach.blender.org/download/. Accessed 3 Dec. 2019.

Project Citations cont.

Tong, Zhehang, et al. "A Review of Indoor-Outdoor Scene Classification." *Proceedings of the 2017 2nd International Conference on Control, Automation and Artificial Intelligence (CAAI 2017)*, Atlantis Press, 2017. *DOI.org (Crossref)*, doi:10.2991/caai-17.2017.106.

Oliva, Aude, and Antonio Torralba. *Spatial Envelope*. 2001, http://people.csail.mit.edu/torralba/code/spatialenvelope/.

Xiao, Jianxiong, et al. "SUN Database: Exploring a Large Collection of Scene Categories." *International Journal of Computer Vision*, vol. 119, no. 1, Aug. 2016, pp. 3–22. *DOI.org (Crossref)*, doi:10.1007/s11263-014-0748-y.

Murthy, Krishna. "Gabor Filters : A Practical Overview." *Computer Vision Tutorials*, 27 Apr. 2014, https://cvtuts.wordpress.com/2014/04/27/gabor-filters-a-practical-overview/.

"Python - How to Perform Max/Mean Pooling on a 2d Array Using Numpy." *Stack Overflow*, https://stackoverflow.com/questions/42463172/how-to-perform-max-mean-pooling-on-a-2d-array-using-numpy. Accessed 3 Dec. 2019.

https://www.kaggle.com/frules11/pins-face-recognition. Accessed 24 Oct. 2019.

Face Detection Using Haar Cascades — OpenCV-Python Tutorials 1 Documentation. https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html. Accessed 3 Dec. 2019.