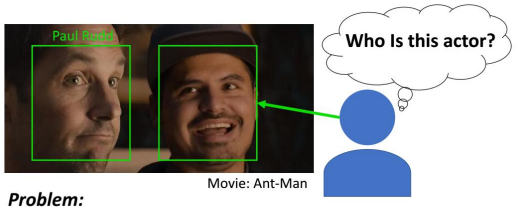
Actor Identification with Deep Learning

Alex Gabourie (gabourie@stanford.edu), Connor McClellan (cjmcc@stanford.edu)

Introduction & Motivation

- Prolific and lead actors/actresses are easily identifiable
- Supporting cast often feel familiar, but difficult to name



Problem:

- Netflix, Hulu, Google Play, etc. currently cannot ID actors

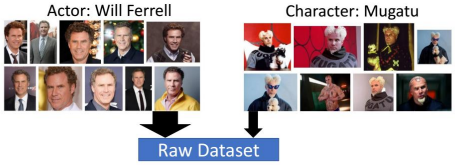
Solution:

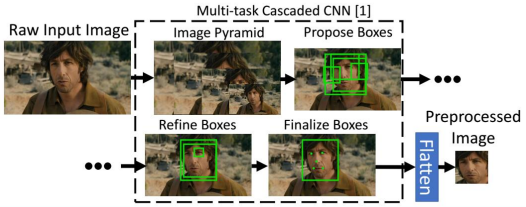
To develop an on-demand, actor identifier that performs:

1. Face Detection [Implemented already]
2. **Face Identification [CS 230 Project Objective]**
3. Object Tracking [Not implemented]

Data Collection & Preprocessing

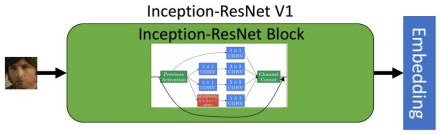
- Google image searches provide actor and character images





Facial Feature Embedding

- Facial features are extracted with an Inception-ResNet [2]
- Network is pre-trained and provided with FaceNet [3]



- Resulting embeddings used for face identification

Movie Specific Model

- Training set comes from actors and characters of one movie
- One small classification model trained for each title

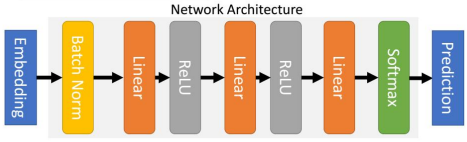
Advantages:

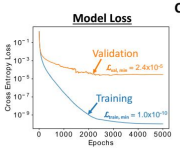
- Fewer classes to train for
- Training can be specific to character's costume

Disadvantages:

- Must train a model for each movie
- Each movie must have a dataset

Softmax Classifier





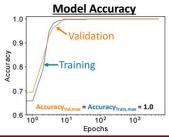
Case Study: Ridiculous 6

Training Information:

- Xavier initialization
- Minibatches
- 80/20 test/validation split (552 examples)
- Early stopping
- Decaying learning rate

Test Results:

- Test set taken from screenshots of movie
- 78/84 correct on test set (92.9% accuracy)



Incorrect Labeling

General Face Identification

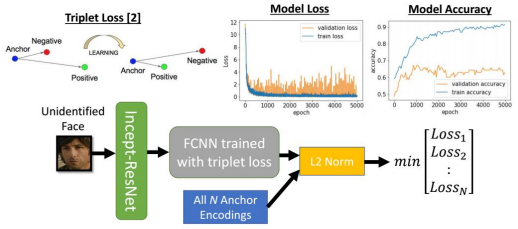
- A general face ID model is developed using triplet loss and transfer learning [2]
- Faces are ID'd by comparing face encodings with anchors and taking the min of the L2 norm

Advantages:

- Training only required once
- Needs one anchor face image

Disadvantages:

- Lower accuracy than a classifier for each movie



Model Comparisons

Model	# Actor Faces per New Movie	Test Accuracy	Retrain Each New Movie
SVM	50 to 100	95 %	yes
Softmax	50 to 100	93 %	yes
Triplet Loss	1*	88 %	no

* because network trained with triplet loss does not need to be retrained

Future Work

- Create better database creation algorithms
- Connect all models to complete video to actor ID solution
- Implement an object tracking algorithm
- Interface with a video player for real-time processing

References

[1] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks", *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499-1503, Oct 2016. [2] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In Proc. CVPR, 2015. [3] D. Sandberg, FaceNet, (2017), GitHub repository, https://github.com/davidsandberg/facenet