

# Bad Guy Identifier

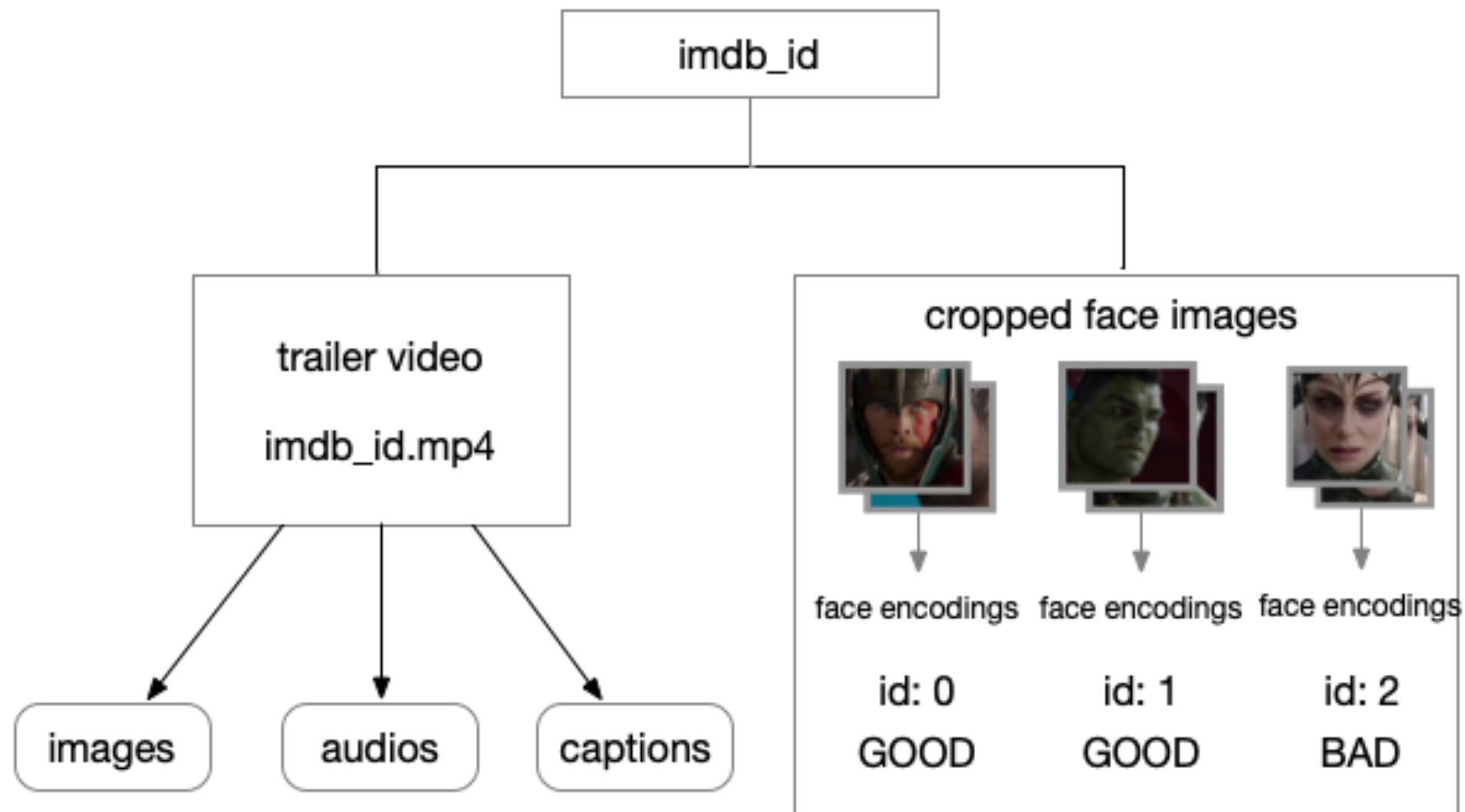
Actor role recognition based on movie trailers

Yue Cao (yuec1)

Supervisor: Prof. Alex Hauptmann

*MCDS Analytics Capstone 2018*

# Dataset



- #Movies: 198 → 271
- #Labels: 61/271 movies are labeled
  - GOOD, BAD, N, NA

# Hypothesis

## Definition of labels

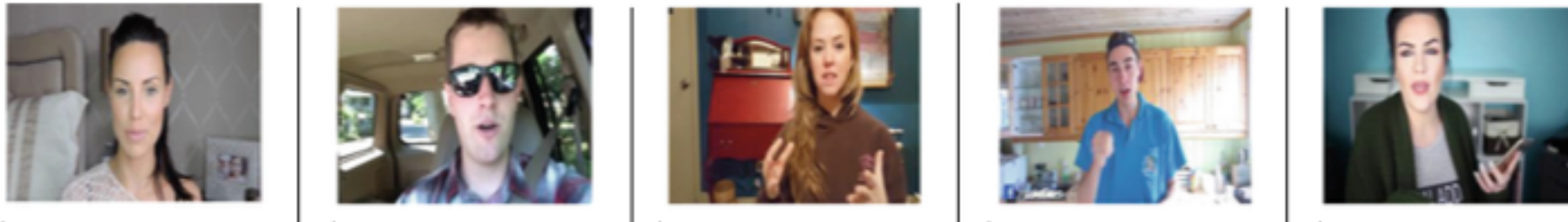
- NA: not a main character
- BAD: unlawful actions OR shows a lack of empathy
- GOOD: positive human traits OR (lawful actions AND looks like a normal person we meet everyday)
- Neutral: mixed traits OR appears to be a main character but lacks any obvious traits

## Takeaways

- unavoidable fuzziness due to the task nature
- coarse but agreeable, applicable standards are better than vague / hard-to-use ones

# Related Works

## ChaLearn First Impressions Challenge 2016



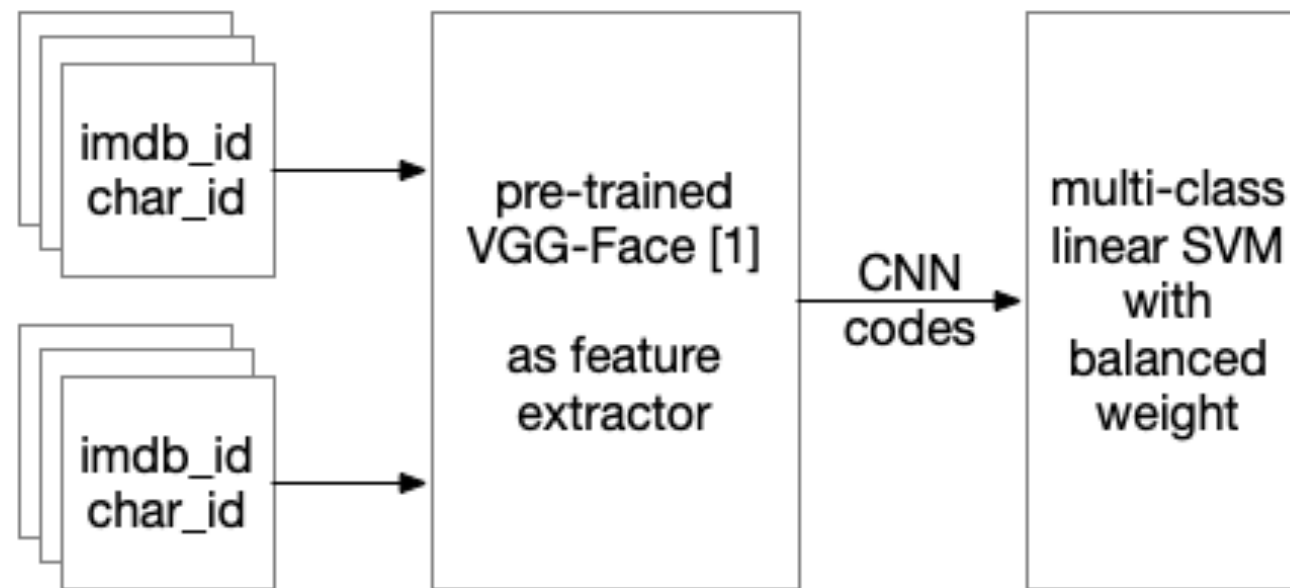
- short YouTube videos → big-5 personality traits [1]
- top runners' model: visual & audio modality with early/late fusion
  - Deep Bimodal Regression [2]
    - face features: VGG-Face features fine tuned on FER-2013
    - scene features: VGG-VD-19 trained on ILSVRC 2012
    - audio features: low level descriptors (LLD) from openSMILE
  - Multimodal LSTM [3]
    - visual features: Recurrent CNN
    - audio features: MFCCs, Energy, Zero Crossing Rate, etc

[1] Escalante, Hugo Jair, et al. "ChaLearn Joint Contest on Multimedia Challenges Beyond Visual Analysis: An overview." *ICPR*. 2016.

[2] Zhang, Chen-Lin, et al. "Deep bimodal regression for apparent personality analysis." *European Conference on Computer Vision*. Springer, Cham, 2016.

[3] Subramaniam, Arulkumar, et al. "Bi-modal first impressions recognition using temporally ordered deep audio and stochastic visual features." *European Conference on Computer Vision*. Springer, Cham, 2016.

# Baseline

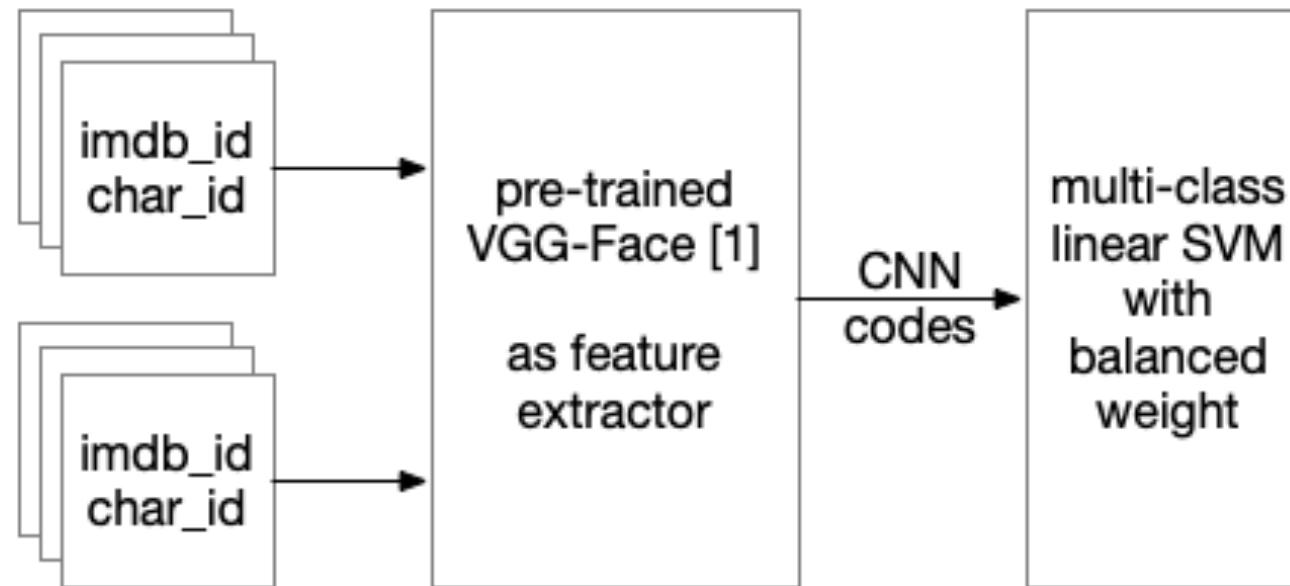


- **Evaluation:** weighted F1

$$F_1 = \left( \frac{\text{recall}^{-1} + \text{precision}^{-1}}{2} \right)^{-1}$$

- averaged F1 of each class weighted by support
- Smaller iterations (<200) works best, easy to overfit

# Baseline



	weighted F1
random	0.25165
all GOOD	0.25256
weighted random	0.29715
baseline	<b>0.36515</b>

# Timeline

Task	Steps	status
Dataset generation	prepare base dataset	DONE
	character identification	DONE
	manual labeling	TBD Nov 26
Baseline	preprocessing	DONE
	pretrained vgg16 + SVM	DONE
Model training	audio features extraction using pyAudioAnalysis	TBD Nov 18
	audiovisual features early/late fusion	TBD Nov 28