Web-Scale, Real-World Face Recognition in Movie Trailers

Enrique G. Ortiz, Alan Wright

(ortizeg@gmail.com, alanwright47@gmail.com)

University of Central Florida

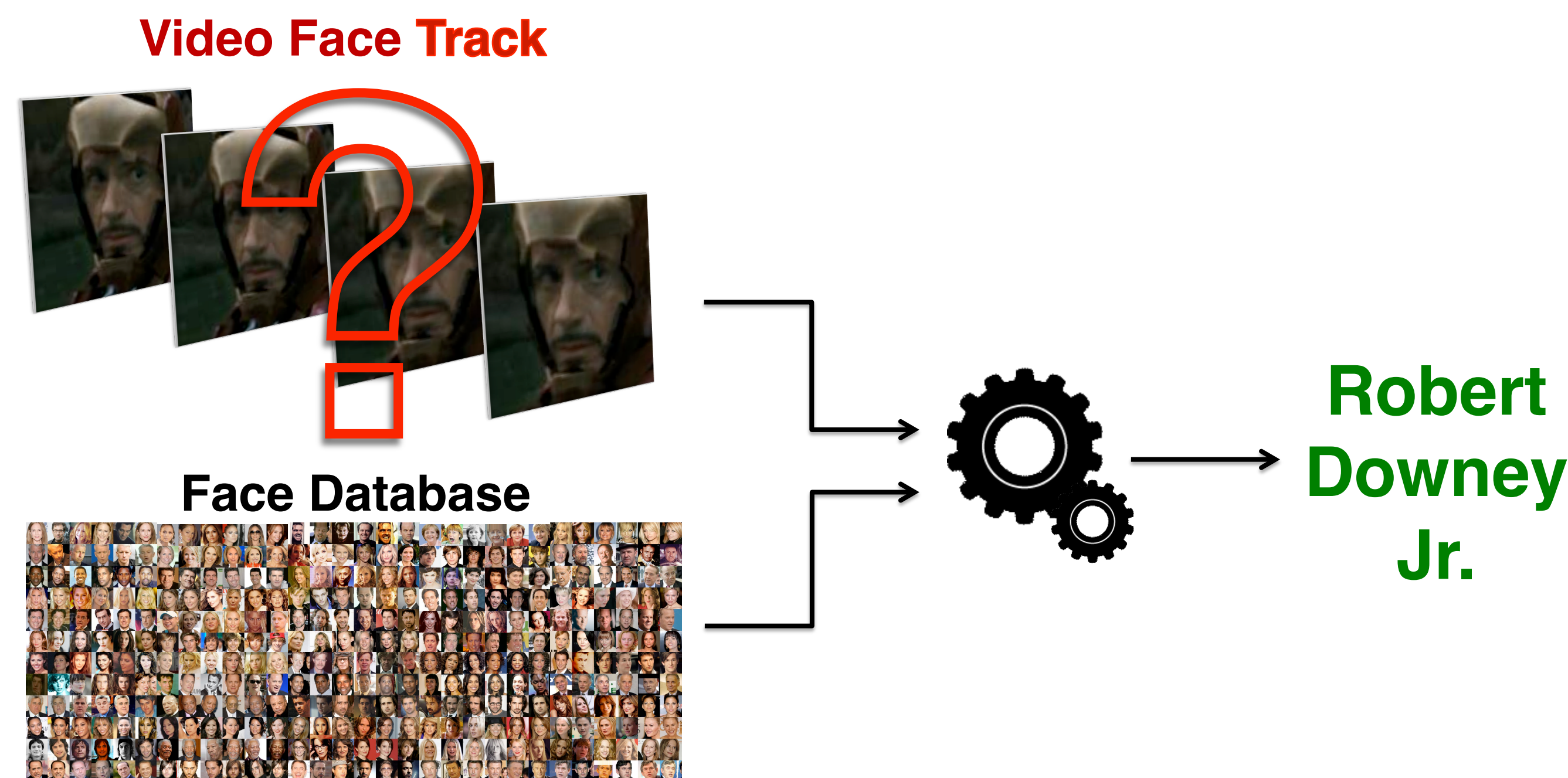University of Central Florida

Computer VISION Lab
University of Central Florida

1. Problem

- **Objective:**
  - Recognize Identities of Faces in Video
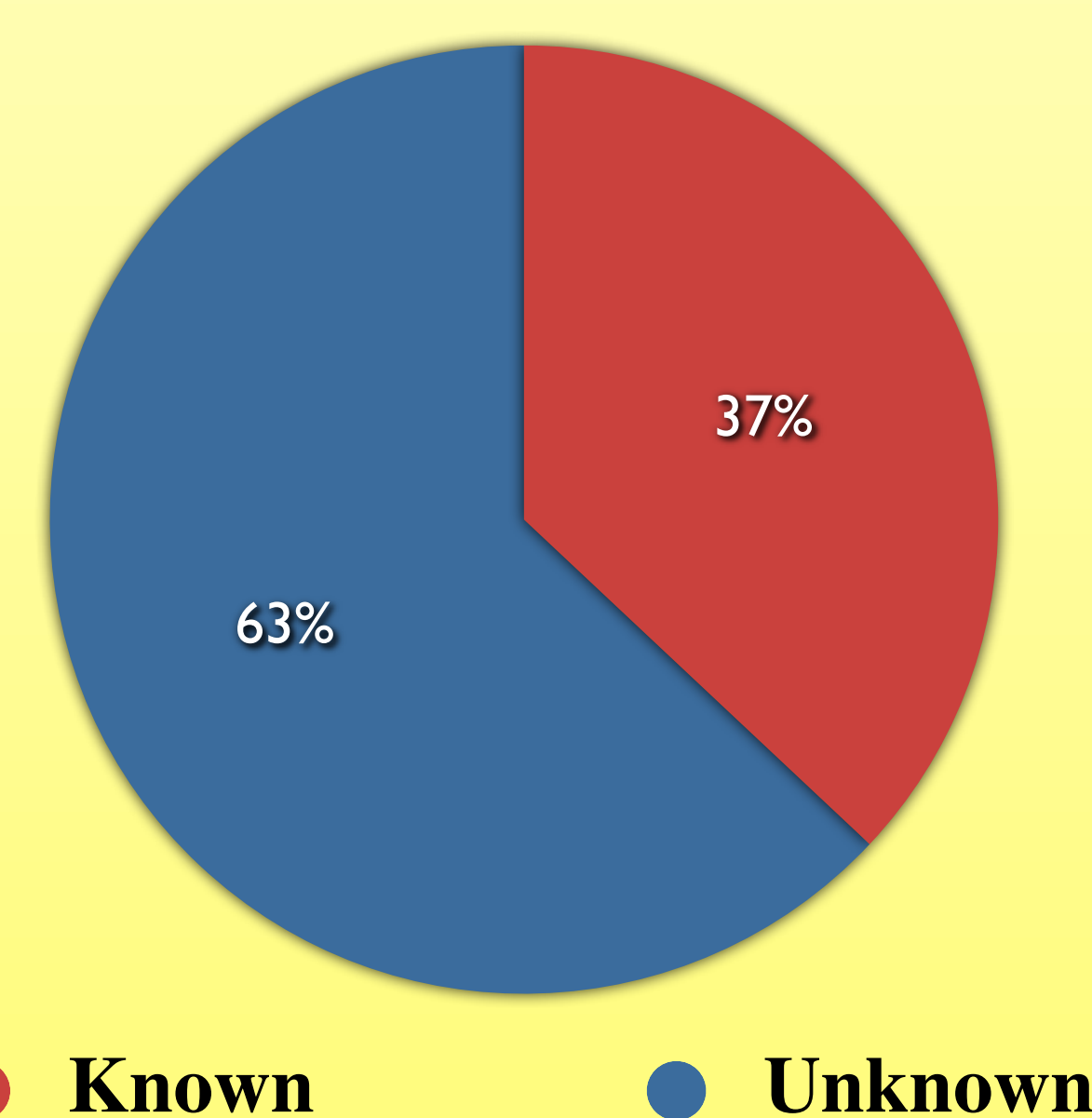- **Applications:**
  - Video Retrieval
  - Video Surveillance

Video Face Track

Face Database

Robert Downey Jr.

2. YouTube Trailers Dataset

- **Collected Movie Trailers From YouTube**
- **Extended Public Figures – 210 classes**
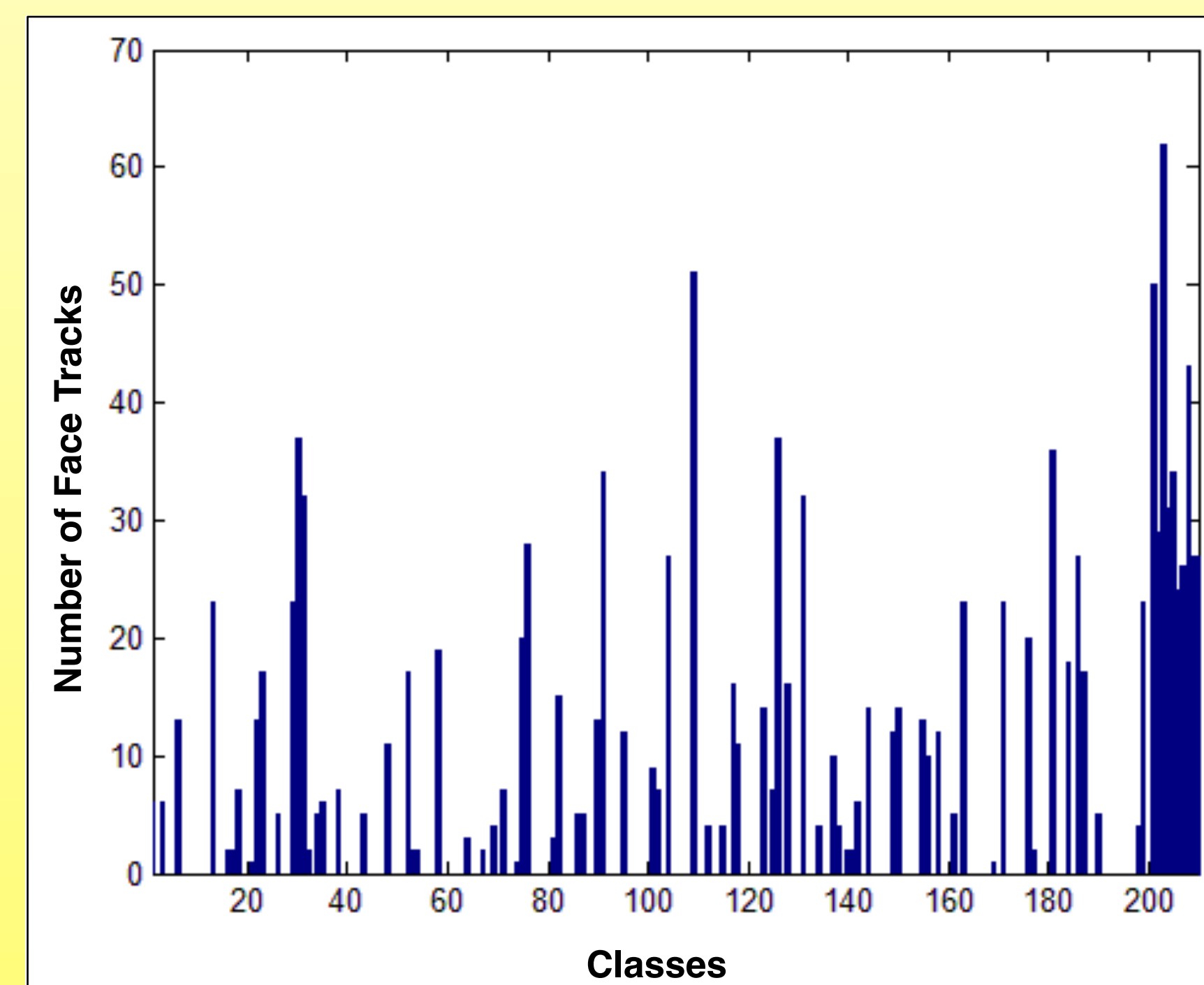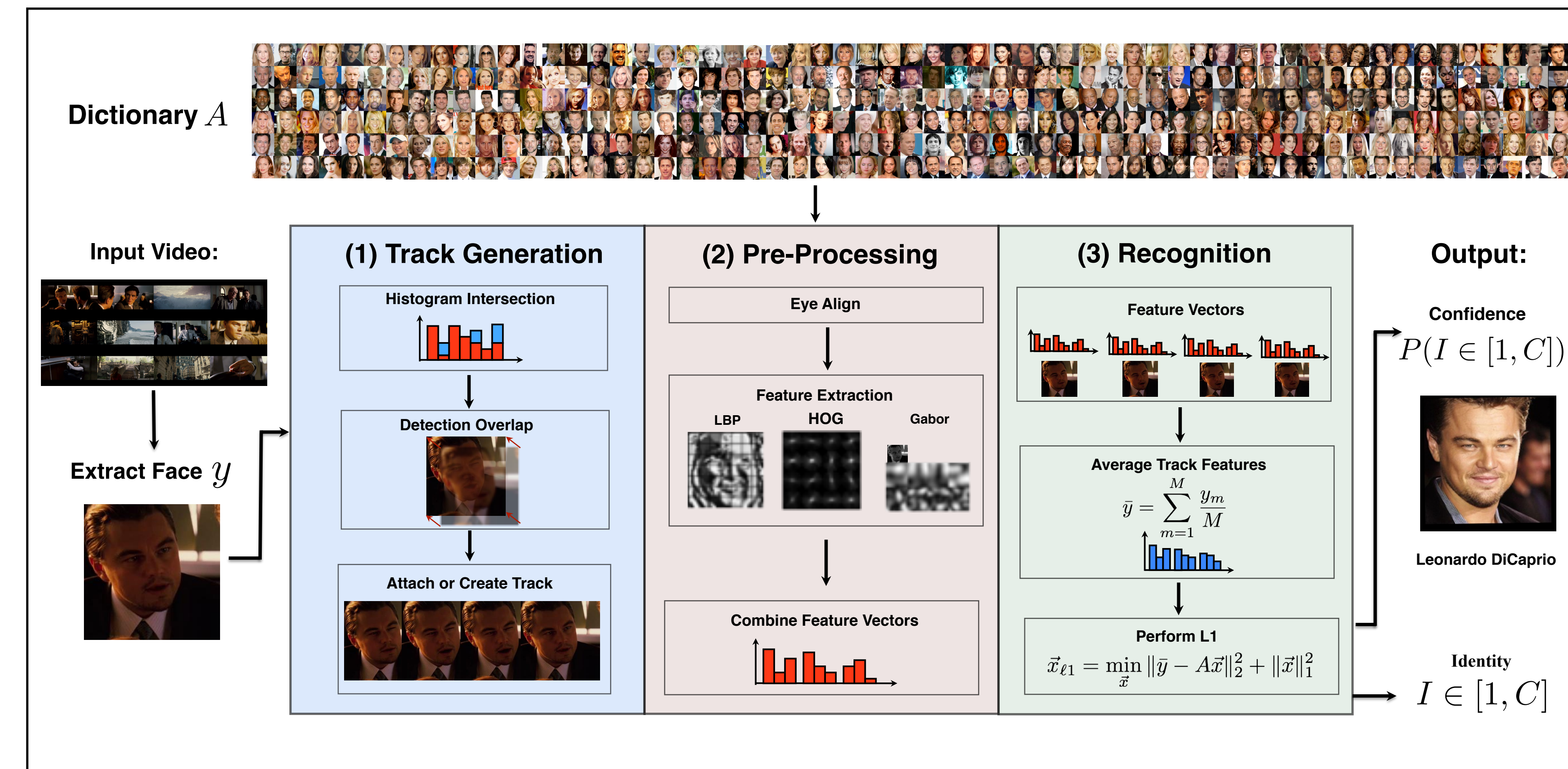- **108 videos**
- **3585 face tracks**

Sample Face Tracks

Face Track Distribution

37%
63%

- Known
- Unknown

Class Distribution

(Number of Face Tracks vs Classes)

3. System Pipeline

Dictionary A

Input Video:

Extract Face y

(1) Track Generation
- Histogram Intersection
- Detection Overlap
- Attach or Create Track

(2) Pre-Processing
- Eye Align
- Feature Extraction: LBP, HOG, Gabor
- Combine Feature Vectors

(3) Recognition
- Feature Vectors
- Average Track Features
$$\bar{y} = \sum_{m=1}^{M} \frac{y_m}{M}$$
- Perform L1
$$\vec{x}_{\ell 1} = \min_{\vec{x}} \|\vec{y} - A\vec{x}\|_2^2 + \|\vec{x}\|_1^2$$

Output:
- Confidence $P(I \in [1, C])$
- Leonardo DiCaprio
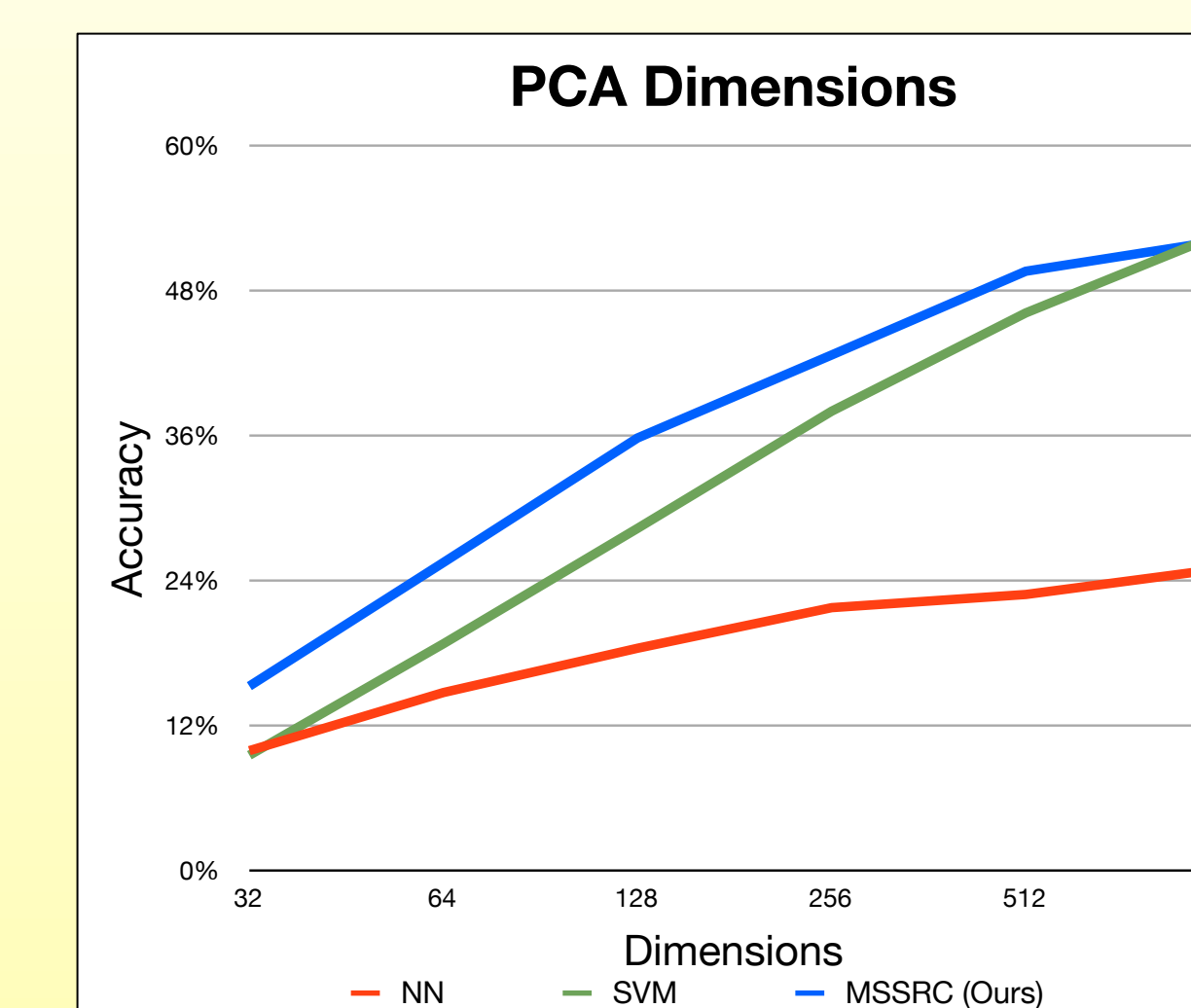- Identity $I \in [1, C]$

(1) Track Generation

- **Detect Face**
- **Distance Metrics:**
  - **I. Detection Overlap**
  - **II. Appearance (Histogram Intersection)**
- **Based on Distance Metrics, Create New Track or Attach to Previous Track**

(2) Pre-Processing

- **Eye Align the Face**
- **Extract Features:**
  - **I. LBP**
  - **II. HOG**
  - **III. Gabor**
- **Reduce Dimensionality with PCA**
- **Combine Feature Vectors**

PCA Dimensions

(Accuracy vs Dimensions: NN, SVM, MSSRC (Ours))

(3) Recognition

Standard Sparse Representation-based Classification

- **Assumption: Test image y is a linear combination of a few training images in dictionary A corresponding to the matching identity.**

$$y = Ax$$

$$\vec{x}_{\ell 1} = \min_{\vec{x}} \|\vec{y} - A\vec{x}\|_2^2 + \|\vec{x}\|_1^2$$

Mean Sequence SRC

- **Known: Faces in face track $Y = [y_1, y_2, \ldots, y_M]$ have the same identity.**
- **Assumption: There exists a coefficient vector agreement over all images.**

$$\hat{\vec{x}}_{\ell 1} = \min_{\vec{x}} \sum_{m=1}^{M} \|\vec{y}_m - A\vec{x}\|_2^2 + \|\vec{x}\|_1^2$$

$$\hat{\vec{x}}_{\ell 2} = \min_{\vec{x}} \sum_{m=1}^{M} \|\vec{y}_m - A\vec{x}\|_2^2 \Rightarrow \sum_{m=1}^{M} \frac{\vec{y}_m}{M} = A\vec{x}$$

$$\vec{x}_{\ell 1} = \min_{\vec{x}} \|\sum_{m=1}^{M} \frac{\vec{y}_m}{M} - A\vec{x}\|_2^2 + \|\vec{x}\|_1^2$$

Classification and Confidence

- **Select identity based on minimum per class residual error.**

$$I(\vec{y}) = \min_j r_j(\vec{y}) = \min \|\vec{y} - A_j \vec{x}_j\|_2$$

- **Compute identity confidence as a percentage of maximum coefficient coverage.**
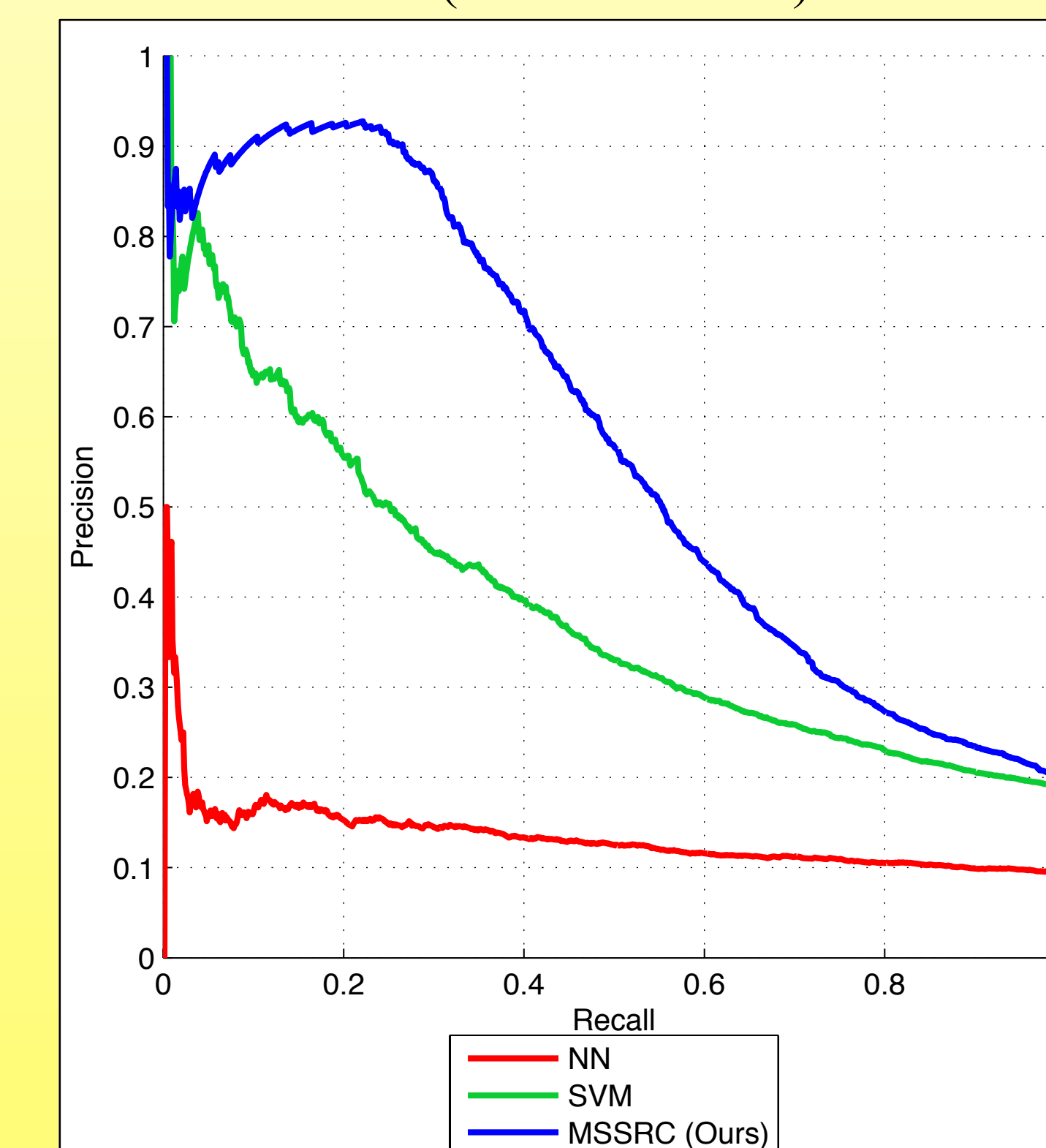
$$SCI = \frac{C \cdot \max_j \|\vec{x}_j\|_1 / \|\hat{\vec{x}}_{\ell_1}\|_1 - 1}{C - 1} \in [0, 1]$$

5. Results

- **Training: PubFig + 10**
  - **Max 200 Images Per Class**
  - **34522 Total Images**
- **Testing: YouTube Trailers Dataset**
  - **3585 Face Tracks**

Method	Accuracy (%)	Recall at 90% Precision (%)
NN	22.85	N/A
SVM	46.19	0.80
MSSRC (Ours)	49.60	26.45

Precision-Recall Curve (with unknowns)

(Precision vs Recall: NN, SVM, MSSRC (Ours))

Precision-Recall Curve (without unknowns)

(Precision vs Recall: NN, SVM, MSSRC (Ours))