Multi-View Learning in the Presence of View Disagreement

C. Mario Christoudias, Raquel Urtasun, Trevor Darrell

UC Berkeley EECS & ICSI

MIT CSAIL

EECS

ICSI INTERNATIONAL COMPUTER SCIENCE INSTITUTE

MIT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY

The World is Multi-view

- Several datasets are comprised of multiple feature sets or views

Helsinki

From Wikipedia, the free encyclopedia

Coordinates: 60°10′15″N, 24°56′15″E

Helsinki (in Finnish; listen (help·info)), or **Helsingfors** (in Swedish; listen (help·info)) is the capital and largest city of Finland. It is in the southern part of Finland, on the shore of the Gulf of Finland, by the Baltic Sea. The population of the city of Helsinki is 569,892 (31 March 2008)[1], making it the most populous municipality in Finland by a wide margin. Foreign-born population stands at around 10%.

Helsinki, along with the neighbouring cities of Vantaa, Espoo, and Kauniainen, constitutes what is known as the capital region, with over 1,000,000 inhabitants. The Greater Helsinki area contains 24 municipalities and has a population of over 1,300,000.[1] The Greater Helsinki accounts for a quarter of population, 29% of jobs and a third of GDP.

Helsinki is Finland's capital for business, education, research, culture, and government. Greater Helsinki has eight universities, six technology parks and the largest technology campus in the Nordic countries.[2] Some 70% of foreign companies operating in Finland have settled in Helsinki region.[2] The immigration of rural residents has made it one of the fastest growing metropolitan areas in Europe.

Finland's main international airline hub, Helsinki-Vantaa airport is 40 minutes from the city center, with direct flights around the world. The busy Helsinki-Tallinn route takes 1.5 hours by sea and 18 minutes by helicopter. Two other big cities in Finland, Tampere and Turku, can be reached in 1.5 - 2 hours by train[3] and 1.5 - 2.5 hours by car.

Contents [hide]

City of Helsinki
Helsingin kaupunki - Helsingfors stad

Coat of arms

Nickname: Stadi, Hesa

Learning from Multiple Information Sources

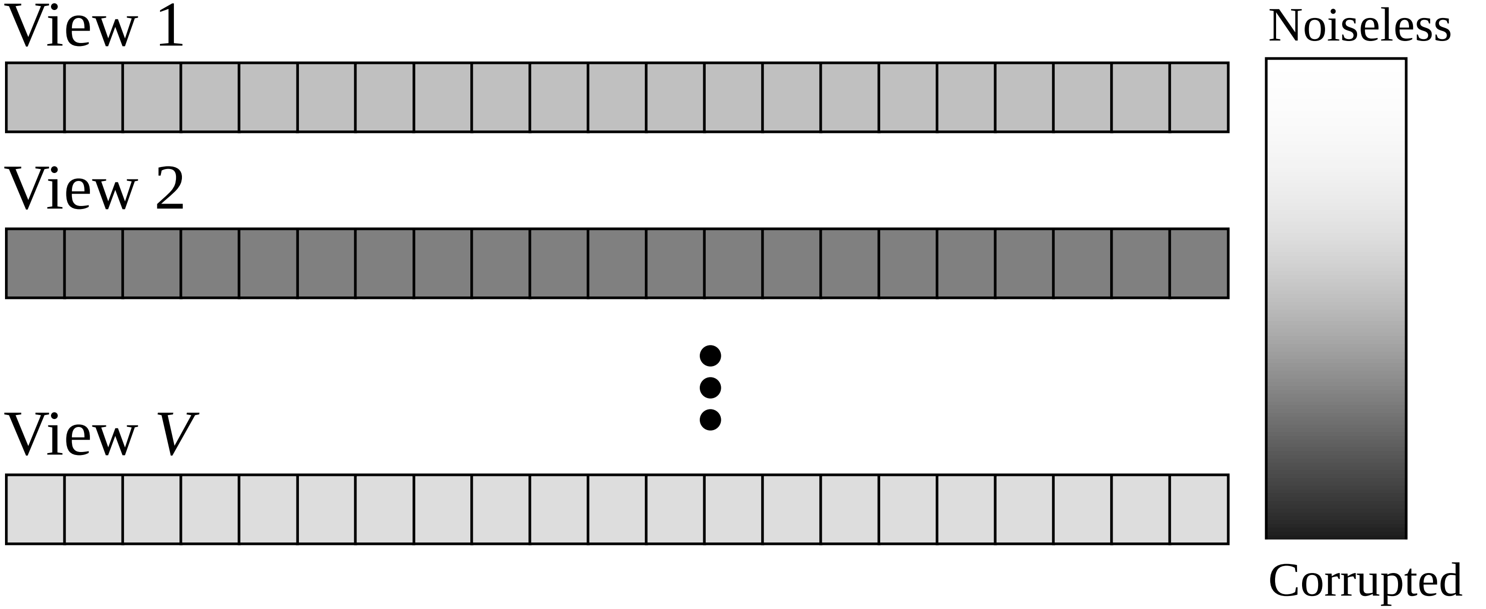
- Multi-view learning methods exploit view redundancy to learn from partially labeled data

- Can be advantageous to learning with only a single view [Blum et.al., '98], [Kakade et.al., '07]

  "Weaknesses of one view complement the strengths of the other"

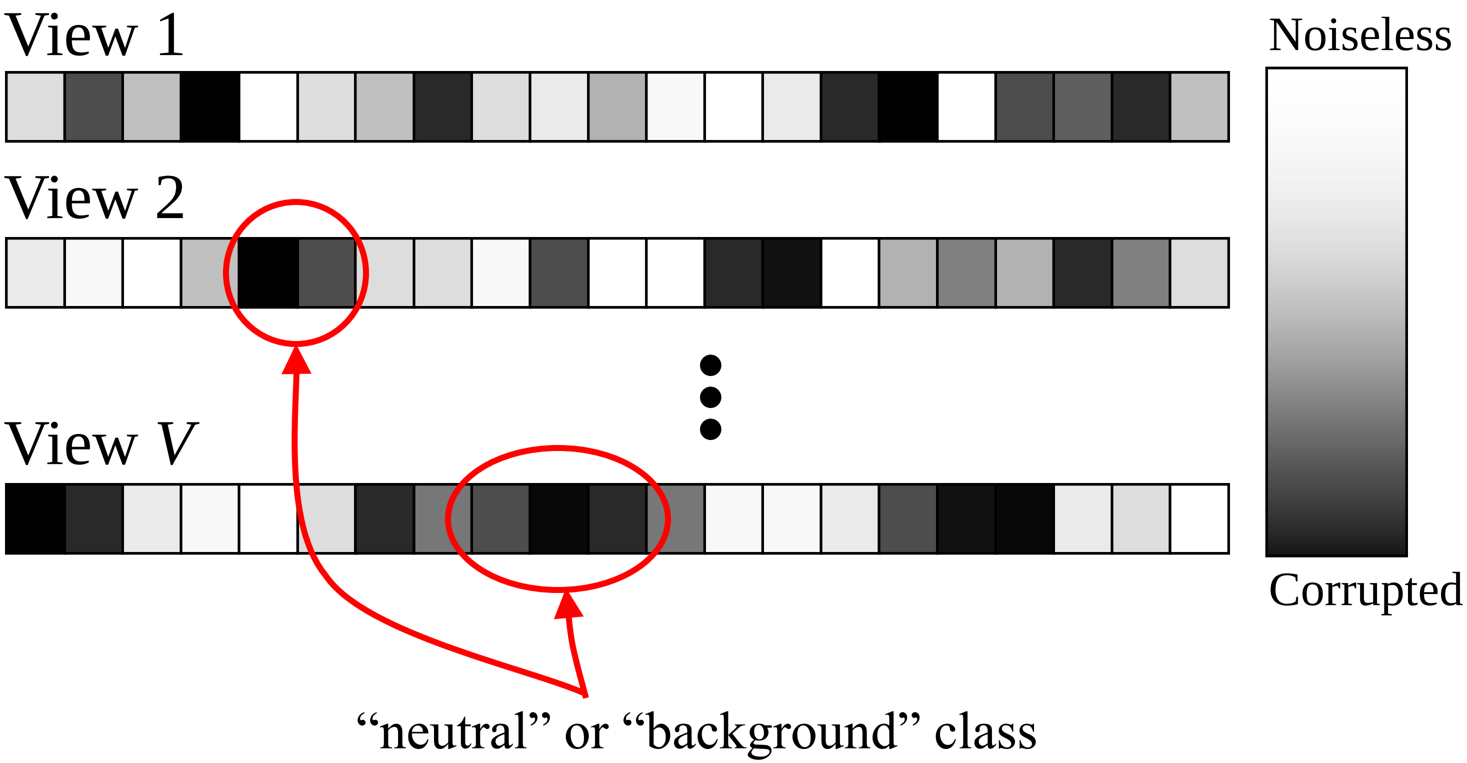
Dealing with Noise

- Multi-view learning approaches have difficulty dealing with noisy observations
- Methods proposed that model stream reliability [Yan et.al., '05], [Yu et.al., '07]



Dealing with Noise

- More generally view corruption is non-uniform:



View disagreement

- View disagreement can be caused by view corruption
  - Samples in each view belong to a different class

- Audio-Visual Examples:
  - *Uni-modal Expression*

    (person says `yes' without nodding)

  - *Temporary View Occlusions*

    (person temporarily covers mouth while speaking)

Our Approach

- Consider view disagreement caused by view corruption
- Detect and filter samples with view disagreement using an information theoretic measure based on conditional view entropy

Related Work

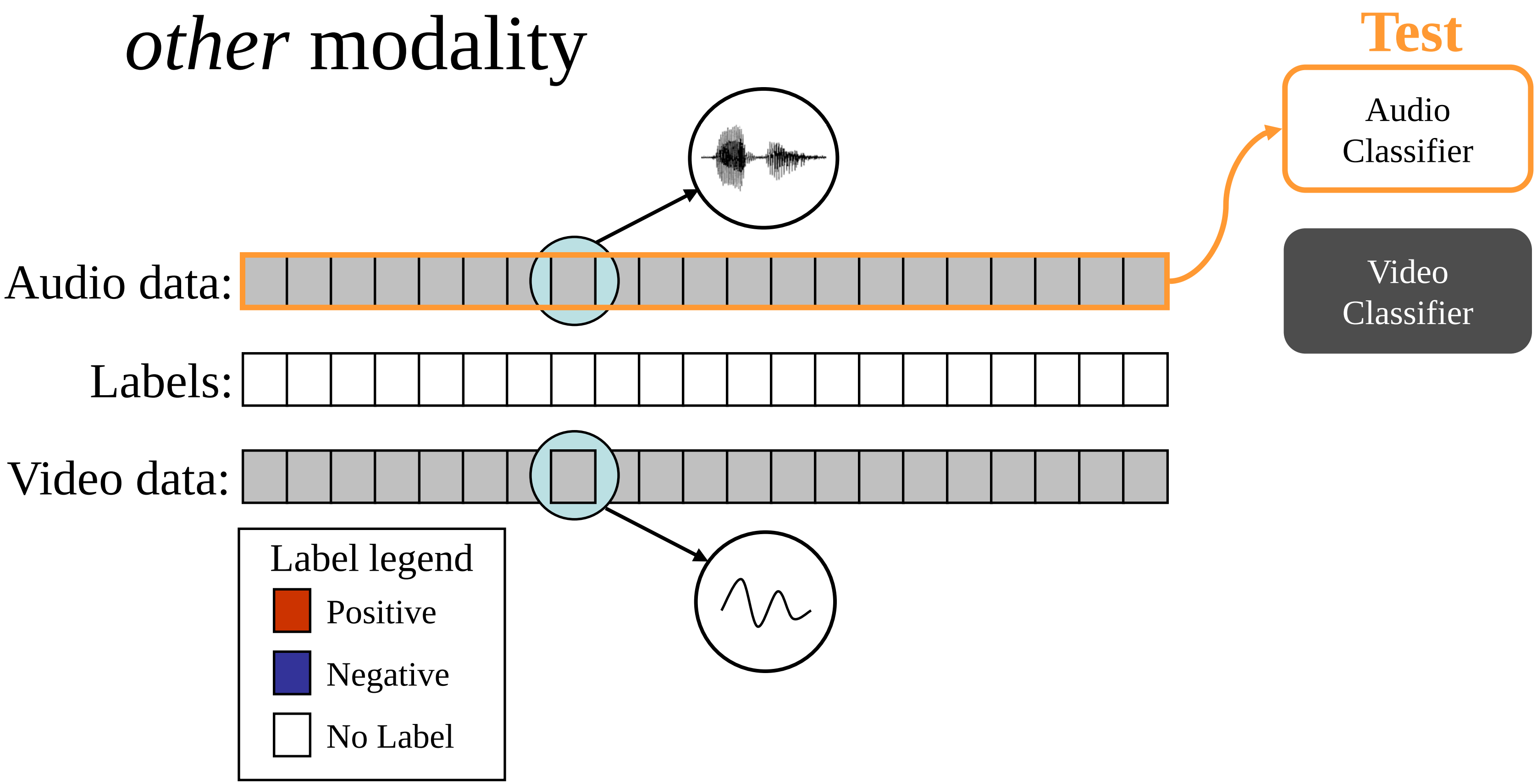
- View disagreement is a new type of *view in-sufficiency*
- Multi-view learning with insufficient views
  - Co-regularization
    [Collins et.al., ‘99], [Sindhwani et.al., ’05]
  - View validation
    [Muslea et.al., ’02], [Naphade et.al., ’05], [Yu et.al., ’07]
  - Multi-view manifold learning
    [Ando et.al., ‘07], [Kakade et.al., ‘07]
- Previous still rely on samples from all views belonging to the same class

Multi-View Bootstrapping

- Co-training [Blum & Mitchell, 98]
  – Mutually bootstrap a set of classifiers from partially labeled data
- Cross-view Bootstrapping
  – Learn a classifier in one modality from the labels provided by a classifier from another modality

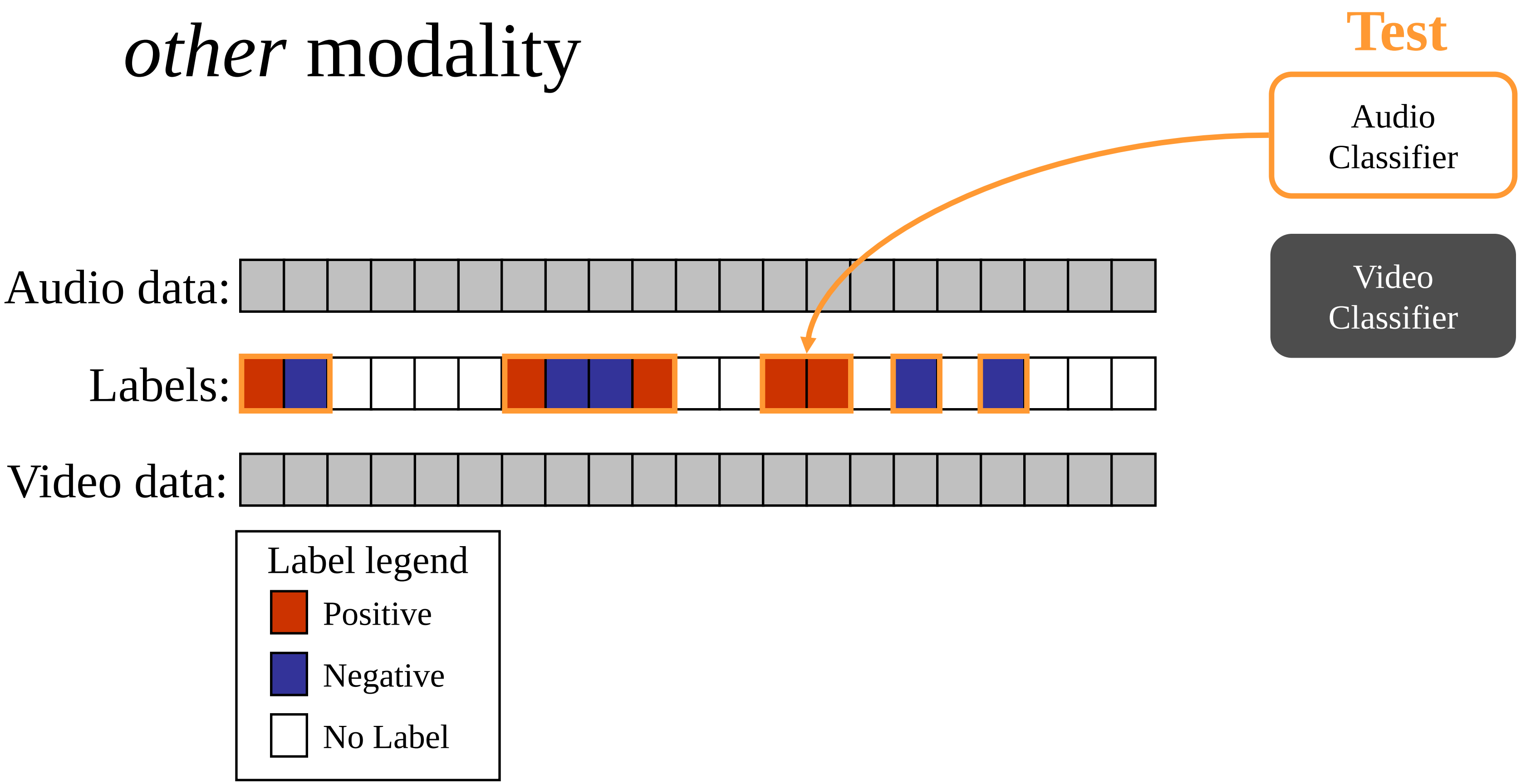
Bootstrapping One View from the Other

- Extrapolate from high-confidence labels in *other* modality



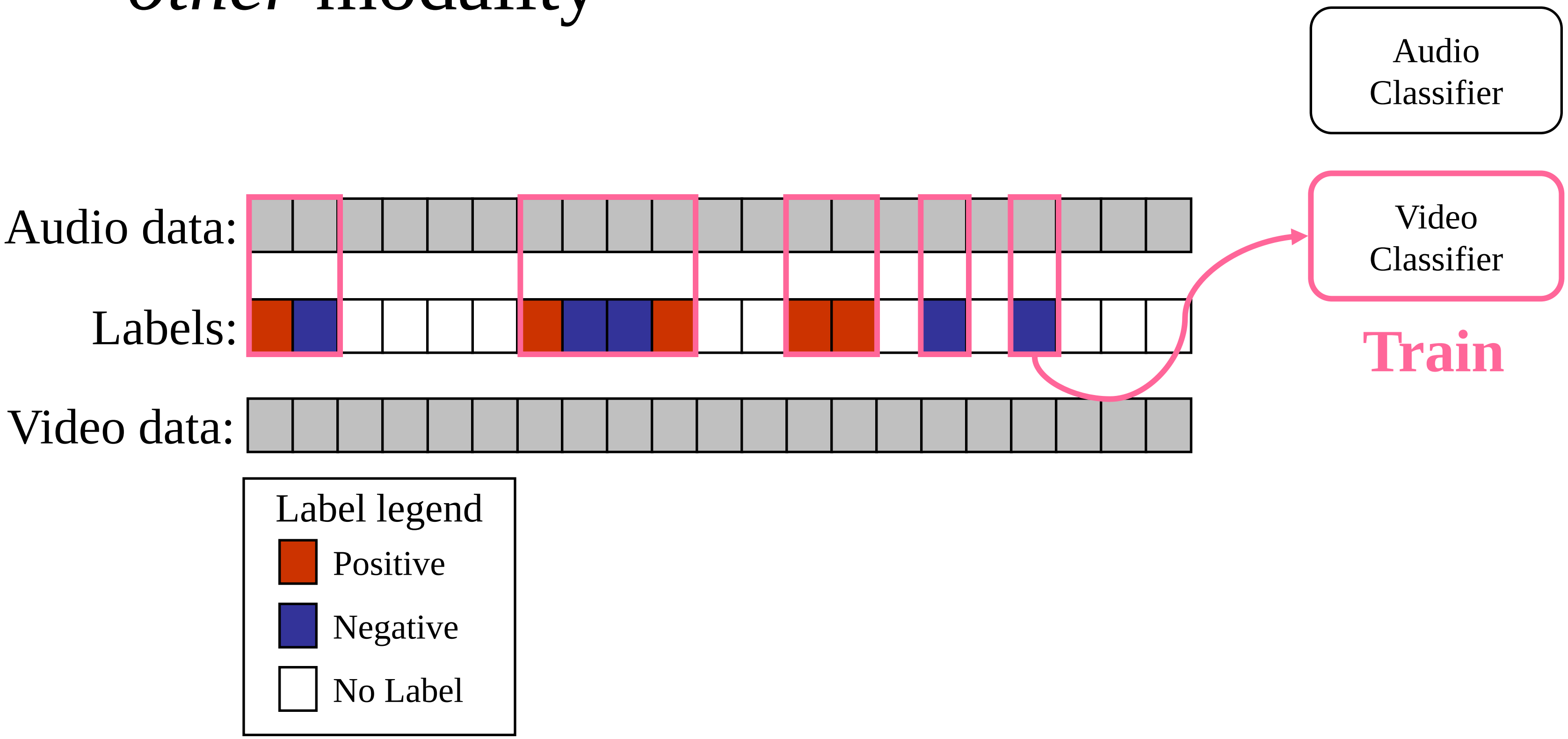
Bootstrapping One View from the Other

- Extrapolate from high-confidence labels in *other* modality



Bootstrapping One View from the Other

- Extrapolate from high-confidence labels in *other* modality



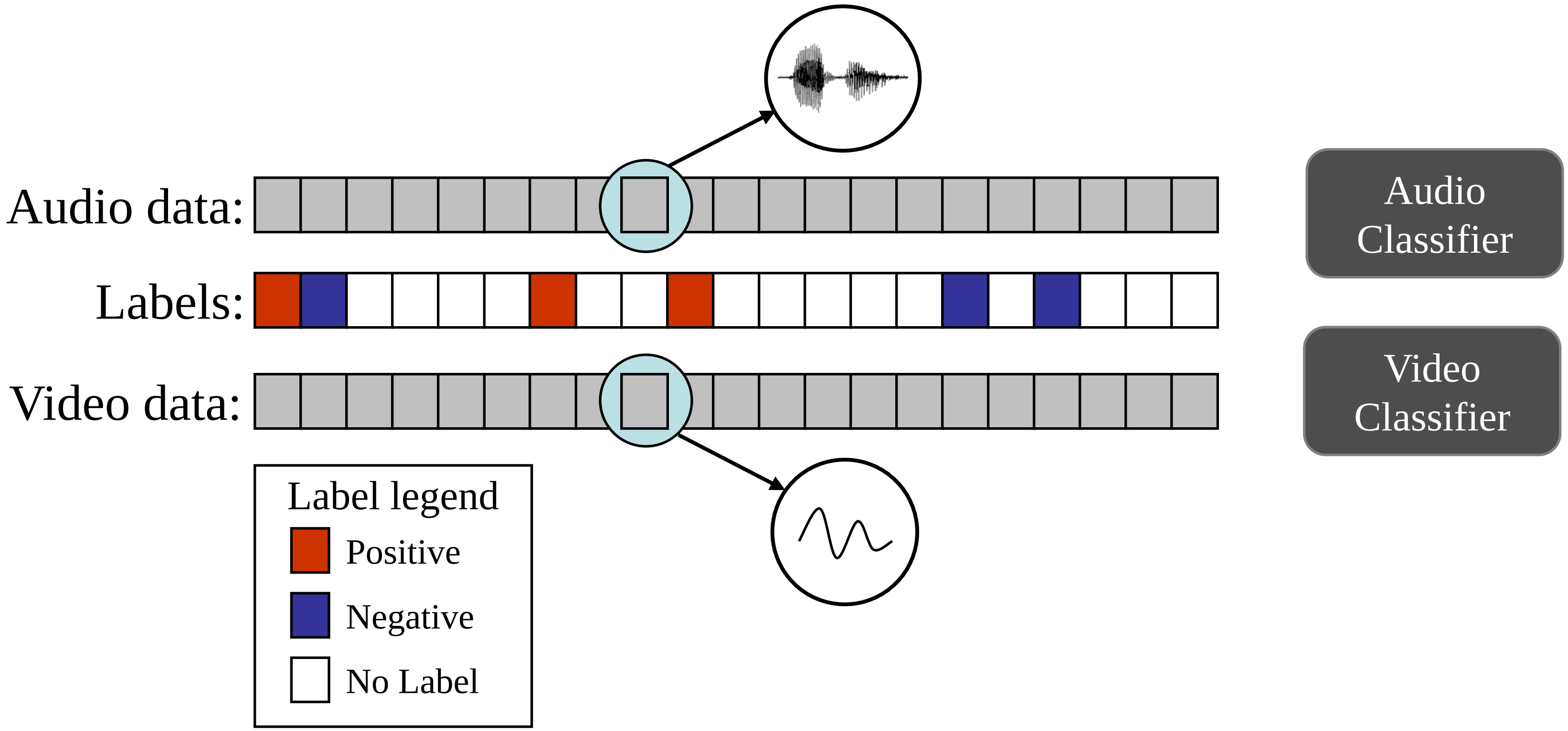
Co-training

[Blum and Mitchell, ’98]

- Learns from partially labeled data by mutually bootstrapping a set of classifiers on multi-view data
- Assumptions
  - Class conditional independence
  - Sufficiency
- Applied to:
  - Text classification (Collins and Singer, ‘99)
  - Visual object detection (Levin et al, ‘03)
  - Information retrieval (Yan and Naphade, ‘05)

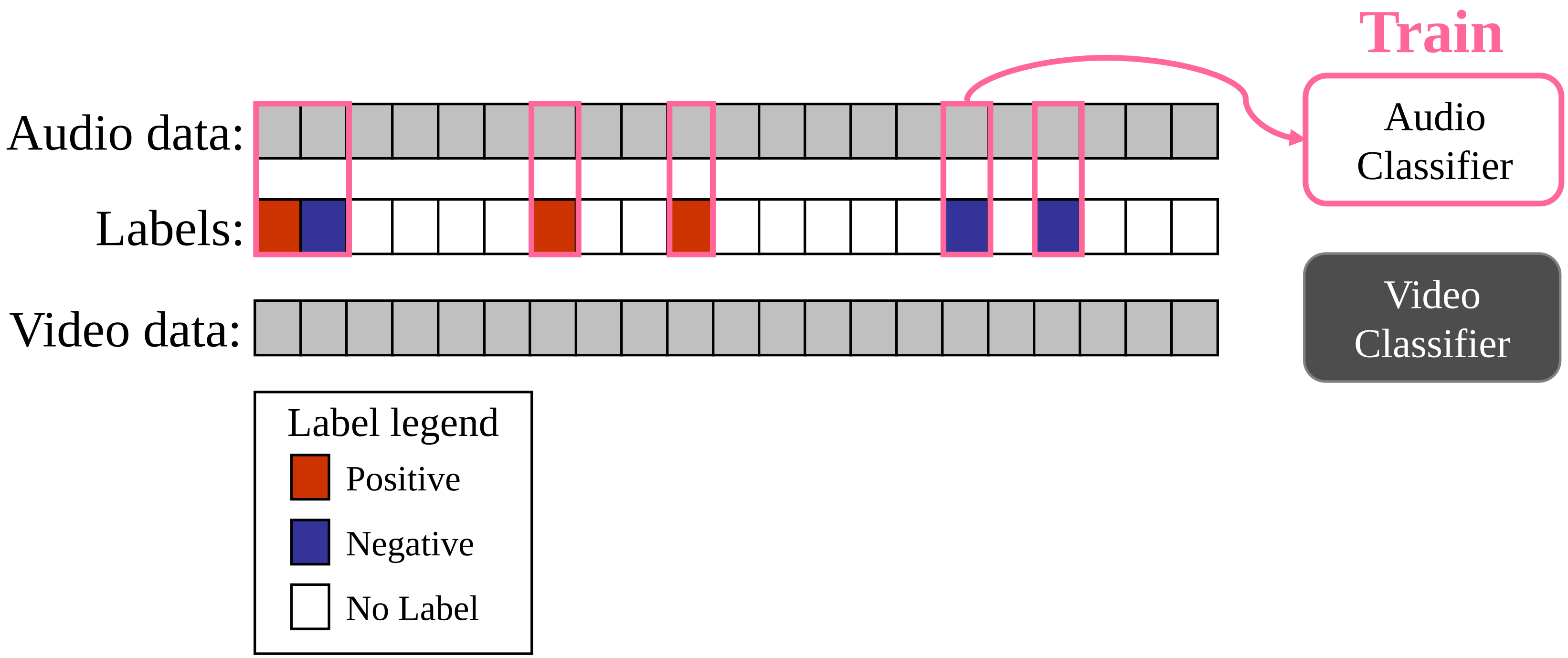
Co-training Algorithm

- Start with seed set of labeled examples



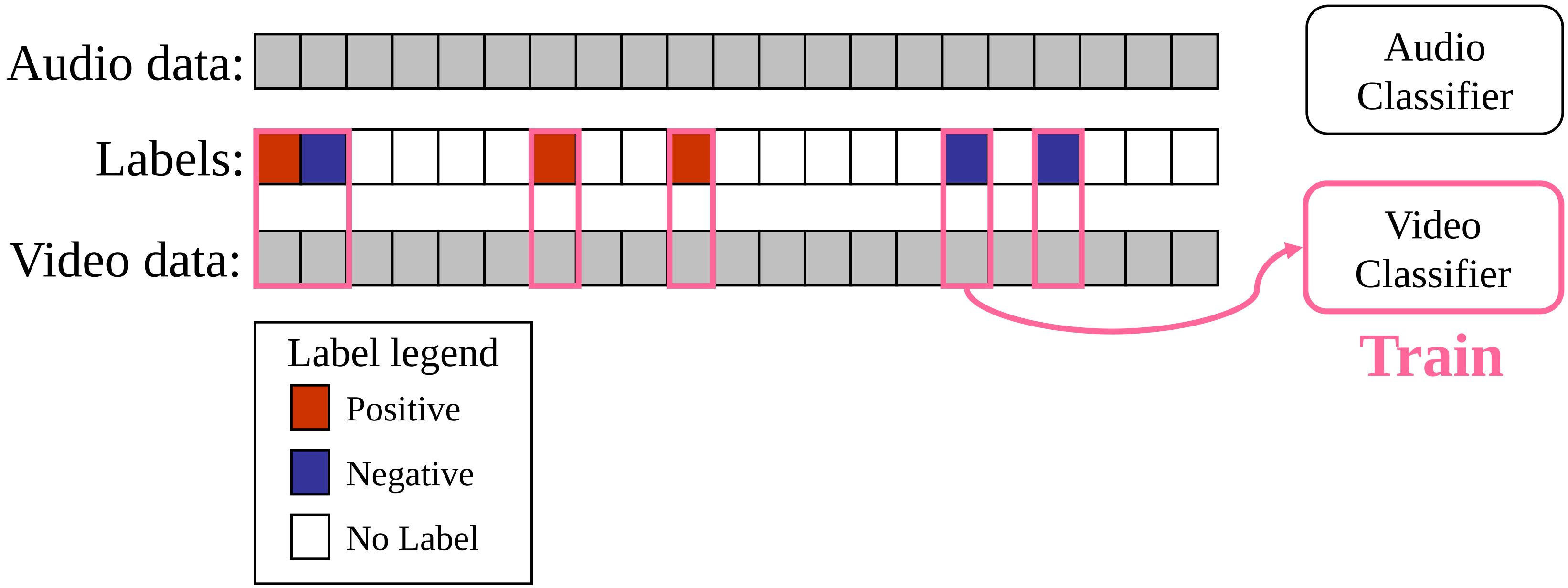
Co-training Algorithm

- Step 1: Train classifiers on seed set



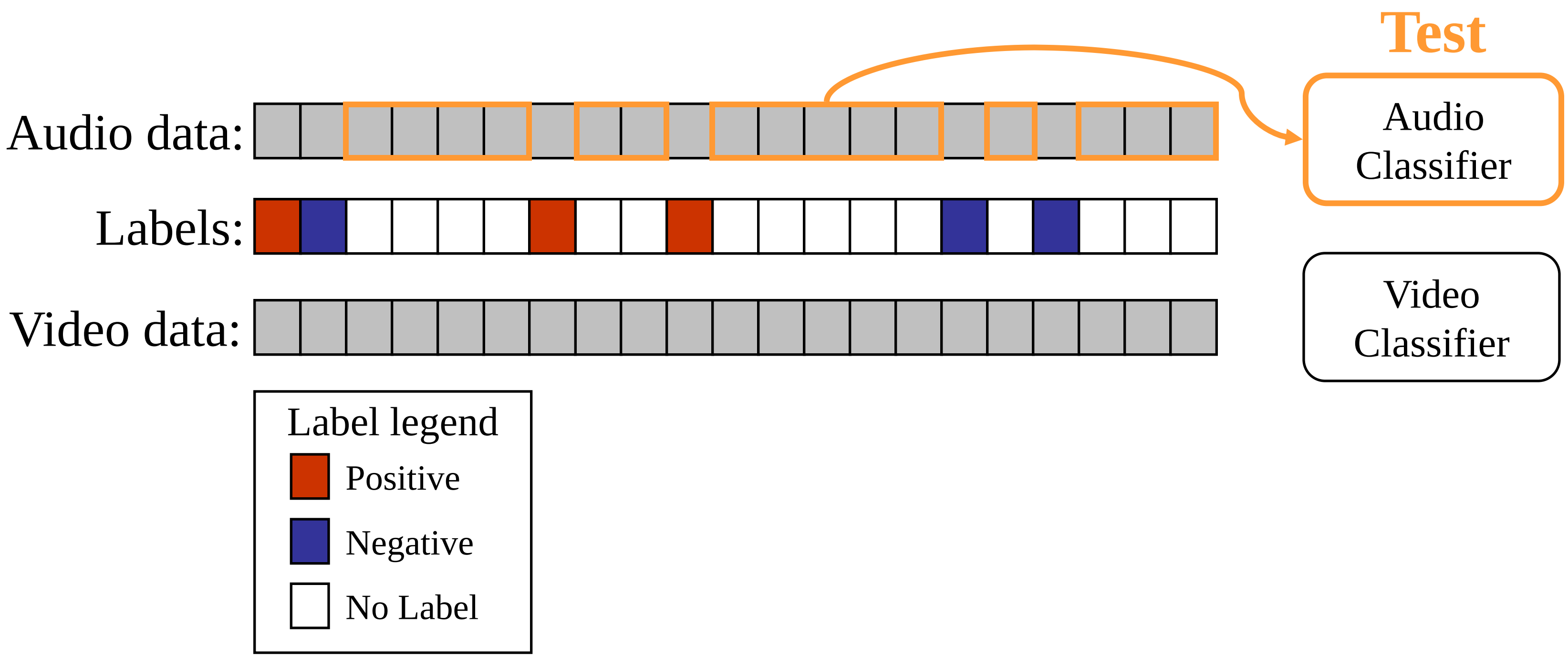
Co-training Algorithm

- Step 1: Train classifiers on seed set



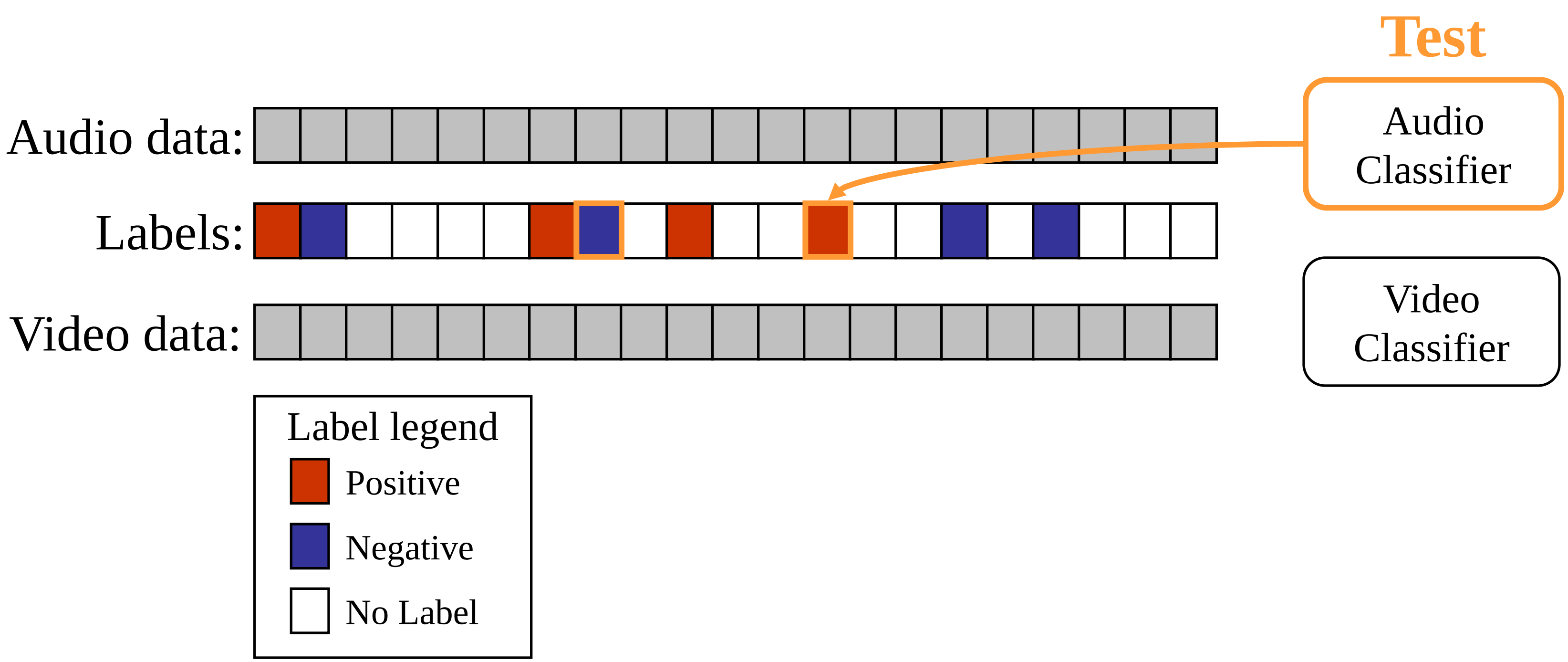
Co-training Algorithm

- Step 2: Evaluate on unlabeled data, add *N* most confident examples from each view



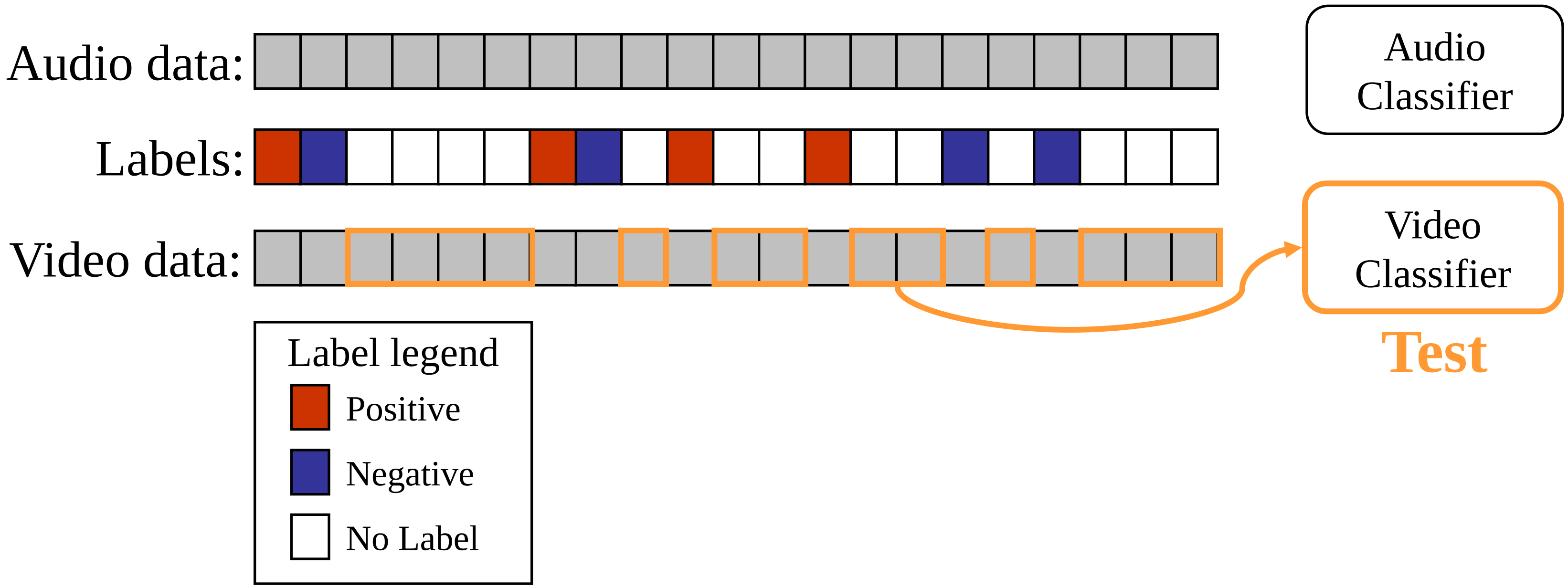
Co-training Algorithm

- Step 2: Evaluate on unlabeled data, add *N* most confident examples from each view



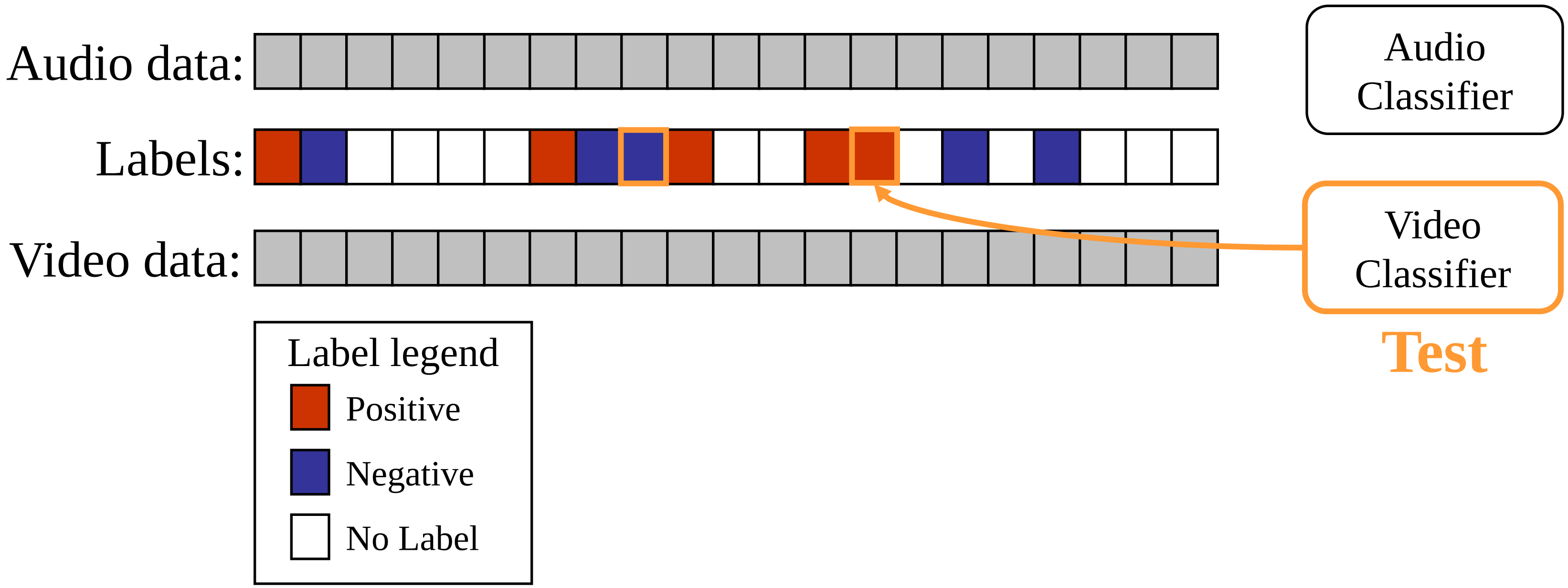
Co-training Algorithm

- Step 2: Evaluate on unlabeled data, add *N* most confident examples from each view



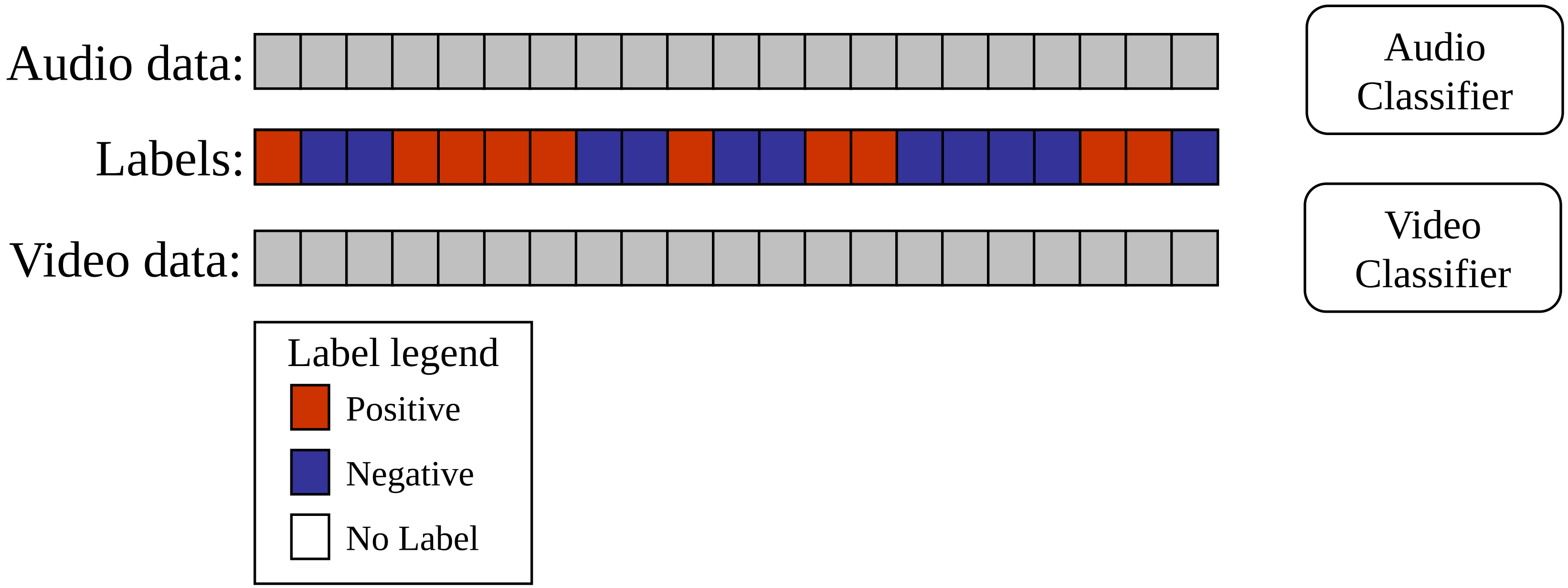
Co-training Algorithm

- Step 2: Evaluate on unlabeled data, add *N* most confident examples from each view

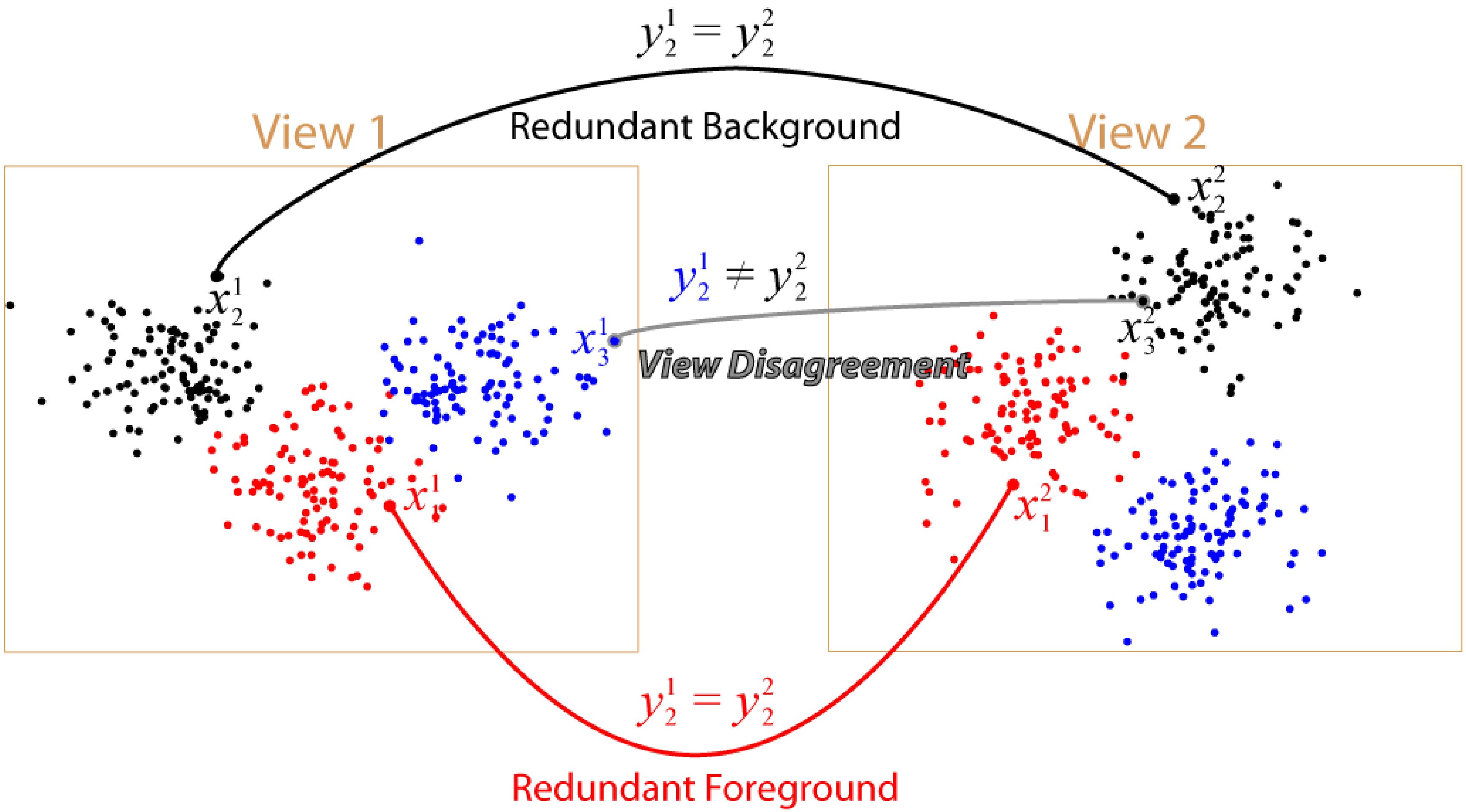


Co-training Algorithm

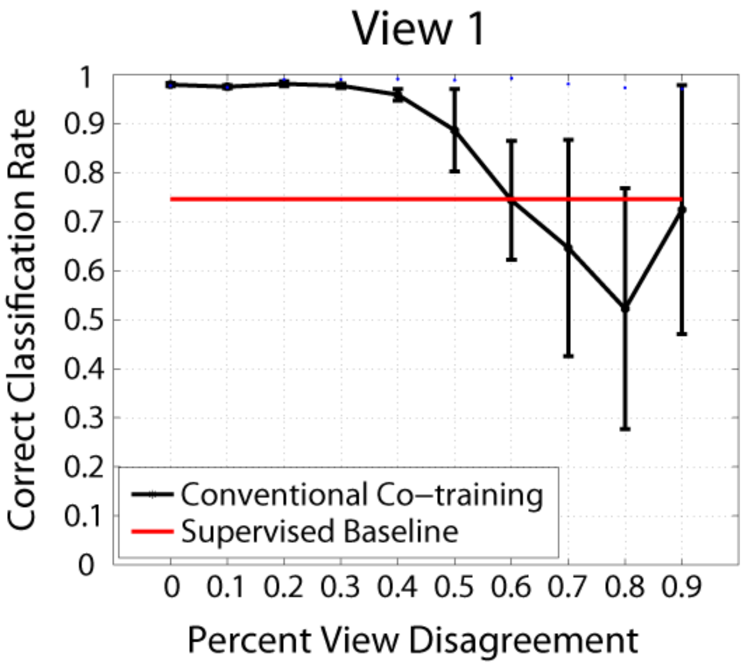
- Iterate steps 1 and 2 until done

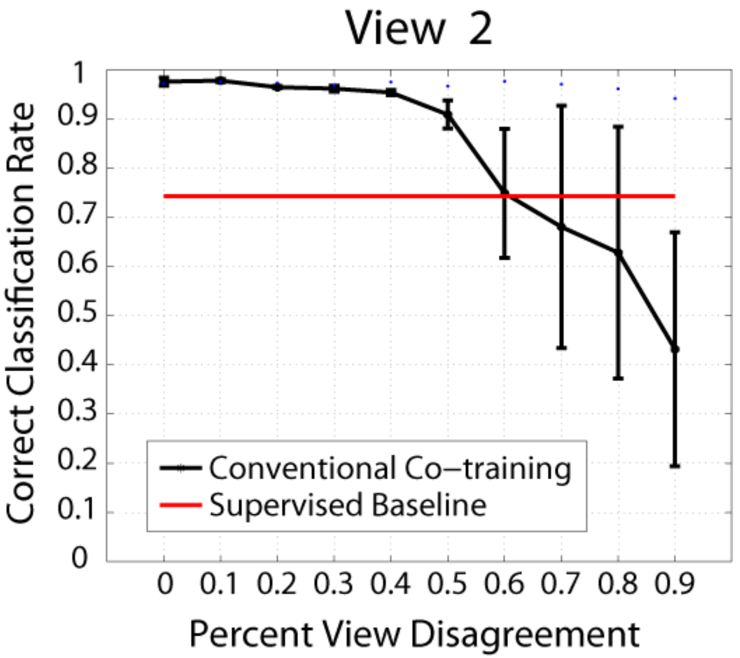


View Disagreement Example: Normally Distributed Classes



Conventional Co-training under View Disagreement

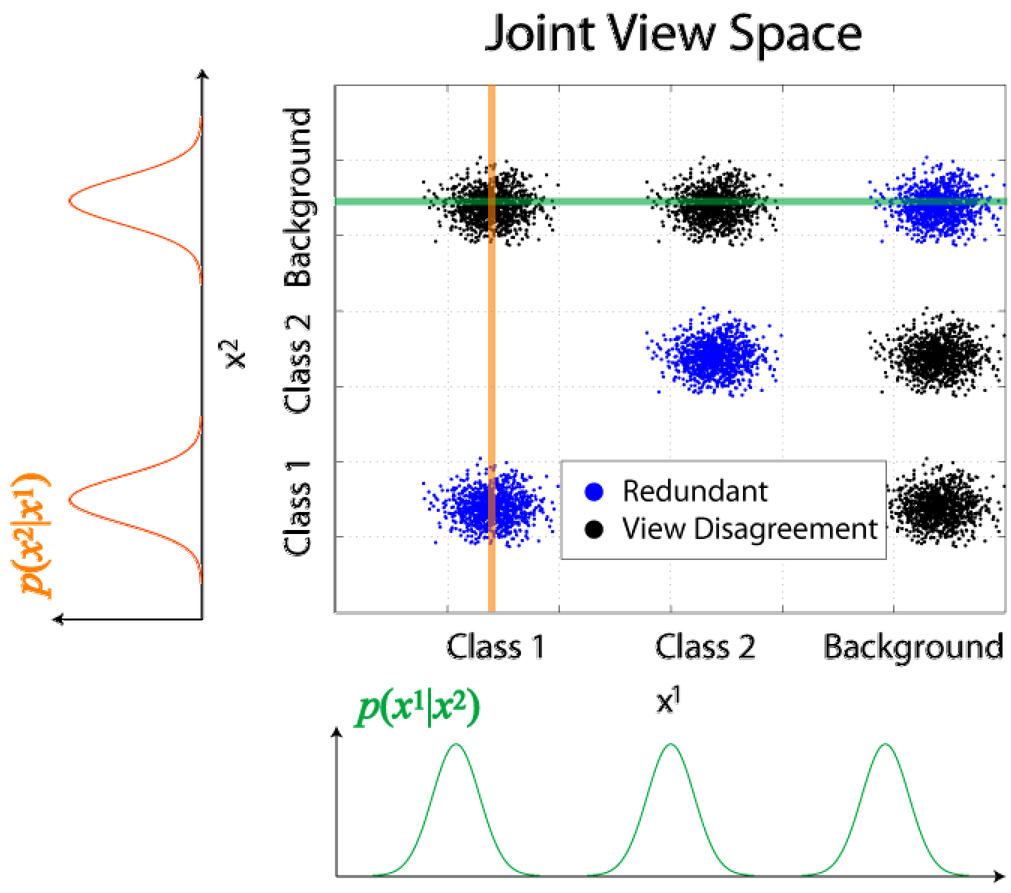




Our Approach: Key Assumption

- Given *n* foreground classes and background
  - Foreground classes can only co-occur with the same class or background
  - Background class can co-occur with either of the *n*+1 classes
- Reasonable assumption for audio-visual problems

Our Approach: Notional Example



Conditioning on a **foreground sample** gives distribution with `*low*' *entropy*.

Conditioning on a **background sample** gives distribution with `*high*' *entropy*.

Conditional Entropy Measure

- Let $\mathbf{x}_k = (x_k^1, ..., x_k^V)$
- Indicator function $m(\cdot)$ over view pairs (x^i, x^j)

$$m(x^i, x_k^j) = \begin{cases} 1, & H(x^i | x_k^j) < \bar{H}_{ij} \\ 0, & \text{otherwise} \end{cases}$$

with

$$H(x^i | x_k^j) = -\sum_{x^i \in U^i} p(x^i | x_k^j) \log p(x^i | x_k^j)$$

$m()$ detects foreground samples x_k^j

- $\bar{H}_{ij}$ is the *mean conditional entropy*

$$\bar{H}_{ij} = \frac{1}{M} \sum_{\mathbf{x}_k \in U} H(x^i | x_k^j)$$

- $p(\mathbf{x})$ is a kernel density estimate [Silverman, 70]

Redundant Sample Detection

- A sample $\mathbf{x}_k$ is a *redundant foreground* sample if it satisfies

$$\prod_{i=1}^{V} \prod_{j \neq i} m(x^i, x_k^j) = 1$$

- A sample $\mathbf{x}_k$ is a *redundant background* sample if it satisfies

$$\sum_{i=1}^{V} \sum_{j \neq i} m(x^i, x_k^j) = 0$$

View Disagreement Detection

- Two views (x_k^i, x_k^j) of a multi-view sample $\mathbf{x}_k$ are in view disagreement if

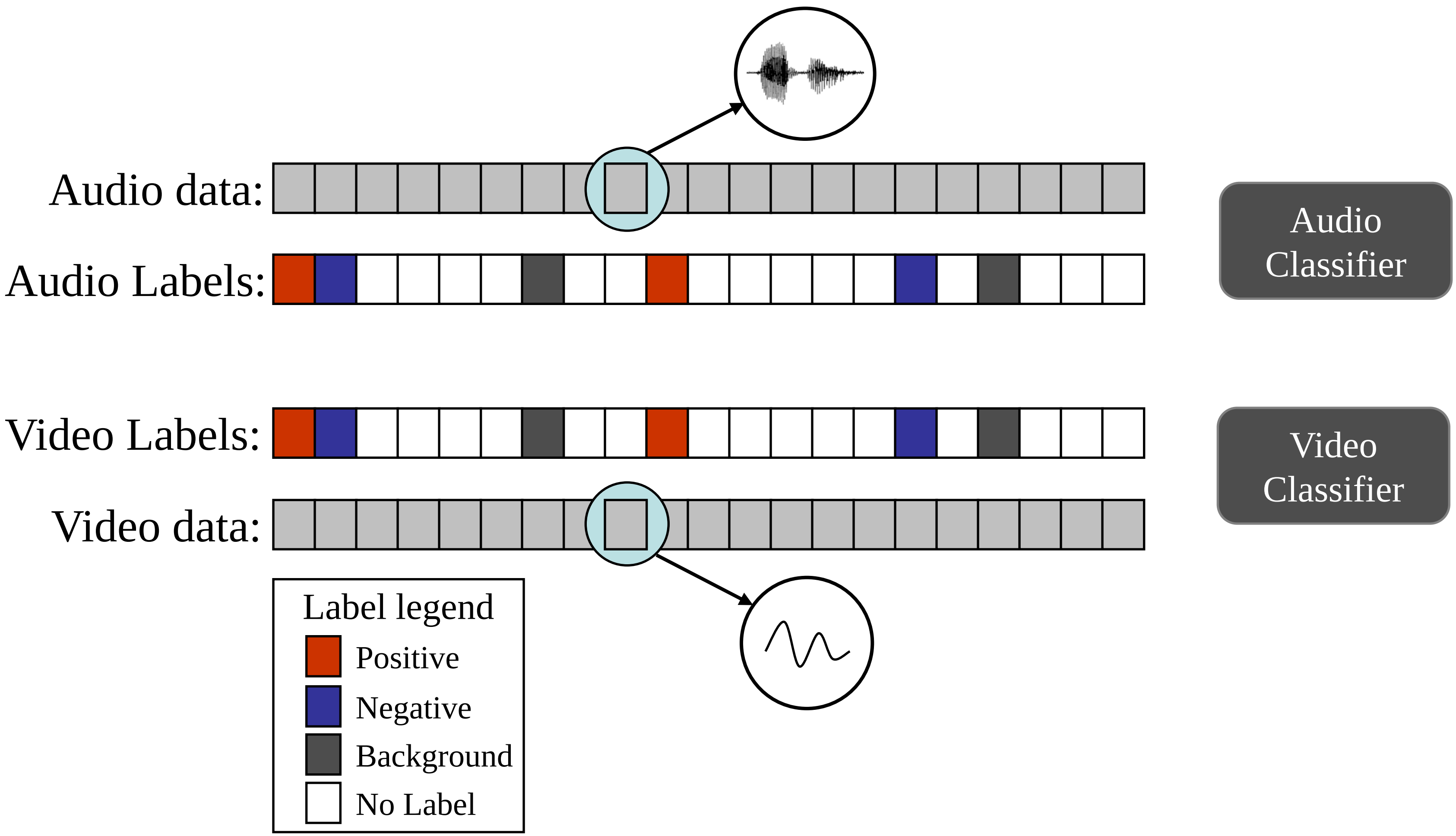
$$m(x^i, x_k^j) \oplus m(x^j, x_k^i) = 1$$

  where $\oplus$ is the logical xor operator.

- Define modified co-training algorithm

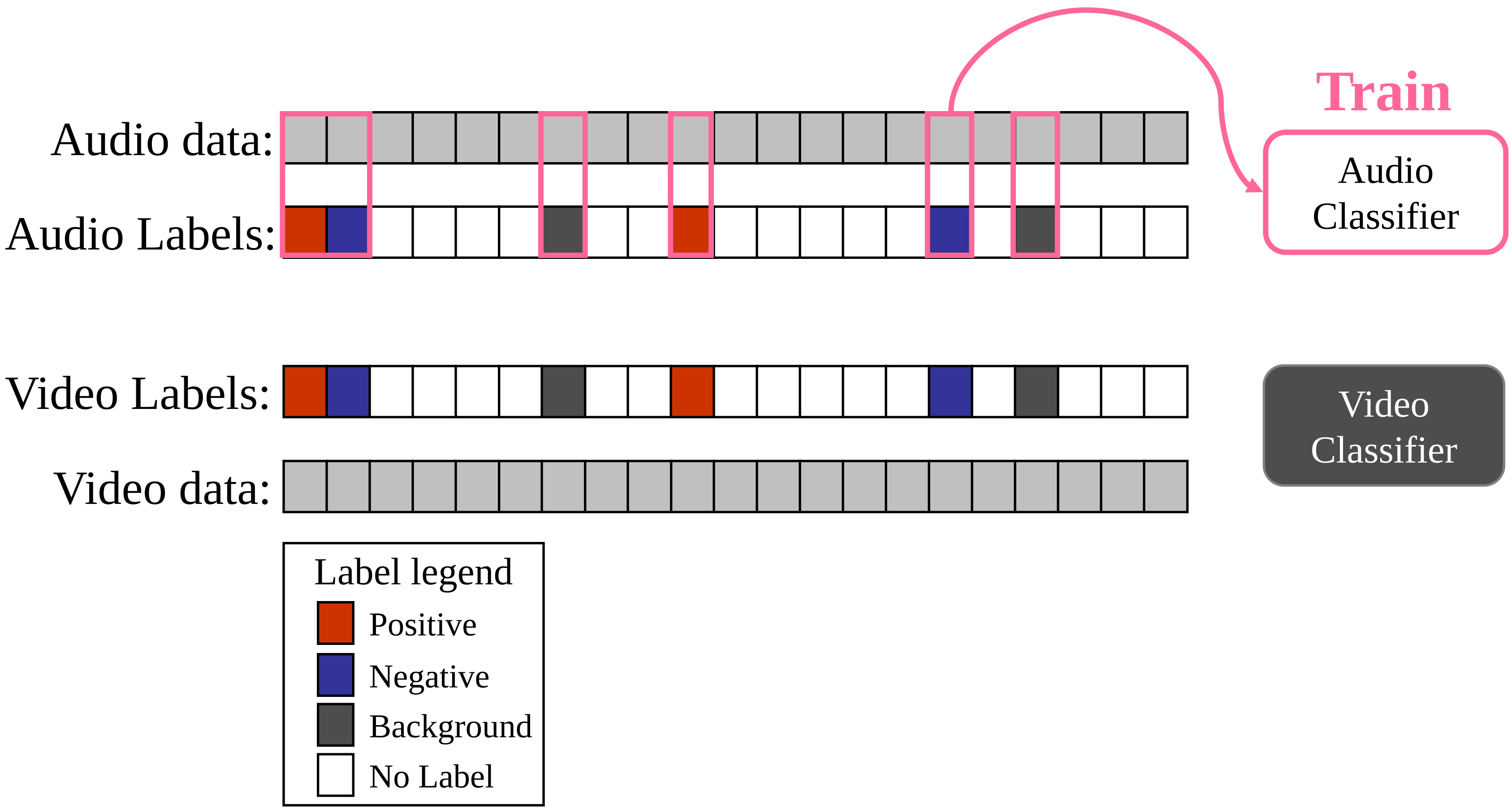
Co-training in the Presence of View Disagreement

- Start with seed set of labeled examples



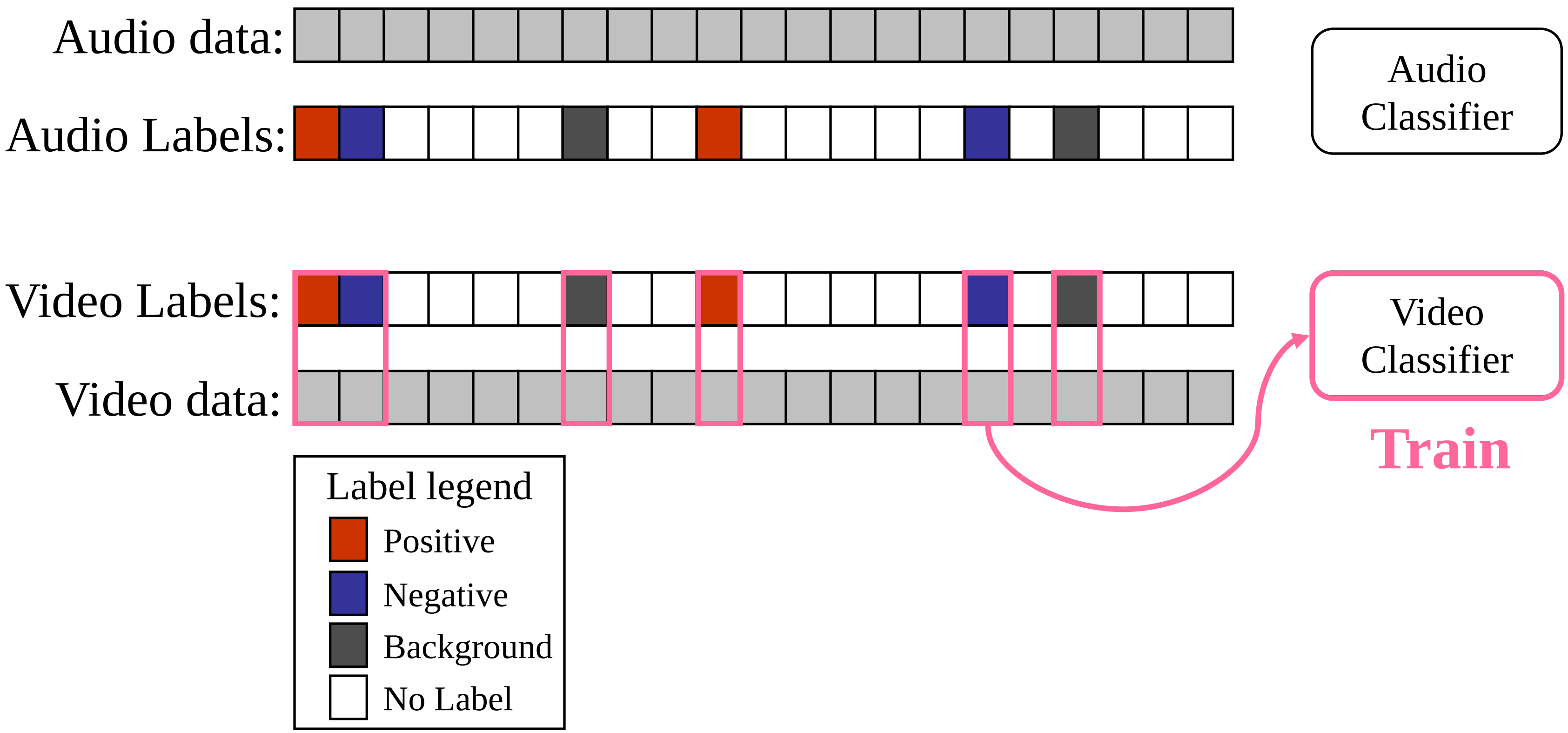
Co-training in the Presence of View Disagreement

- Step 1: Train classifiers on seed set



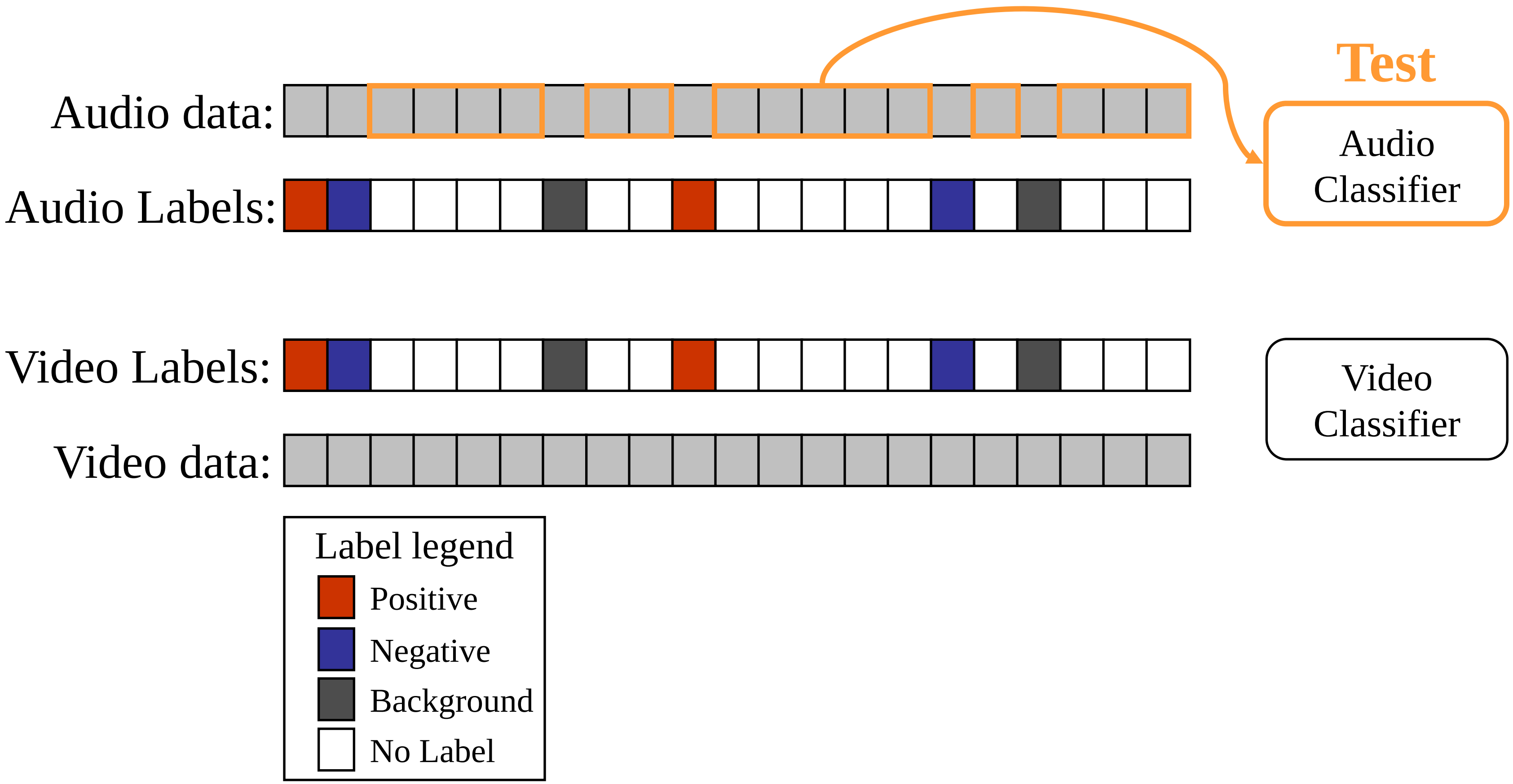
Co-training in the Presence of View Disagreement

- Step 1: Train classifiers on seed set



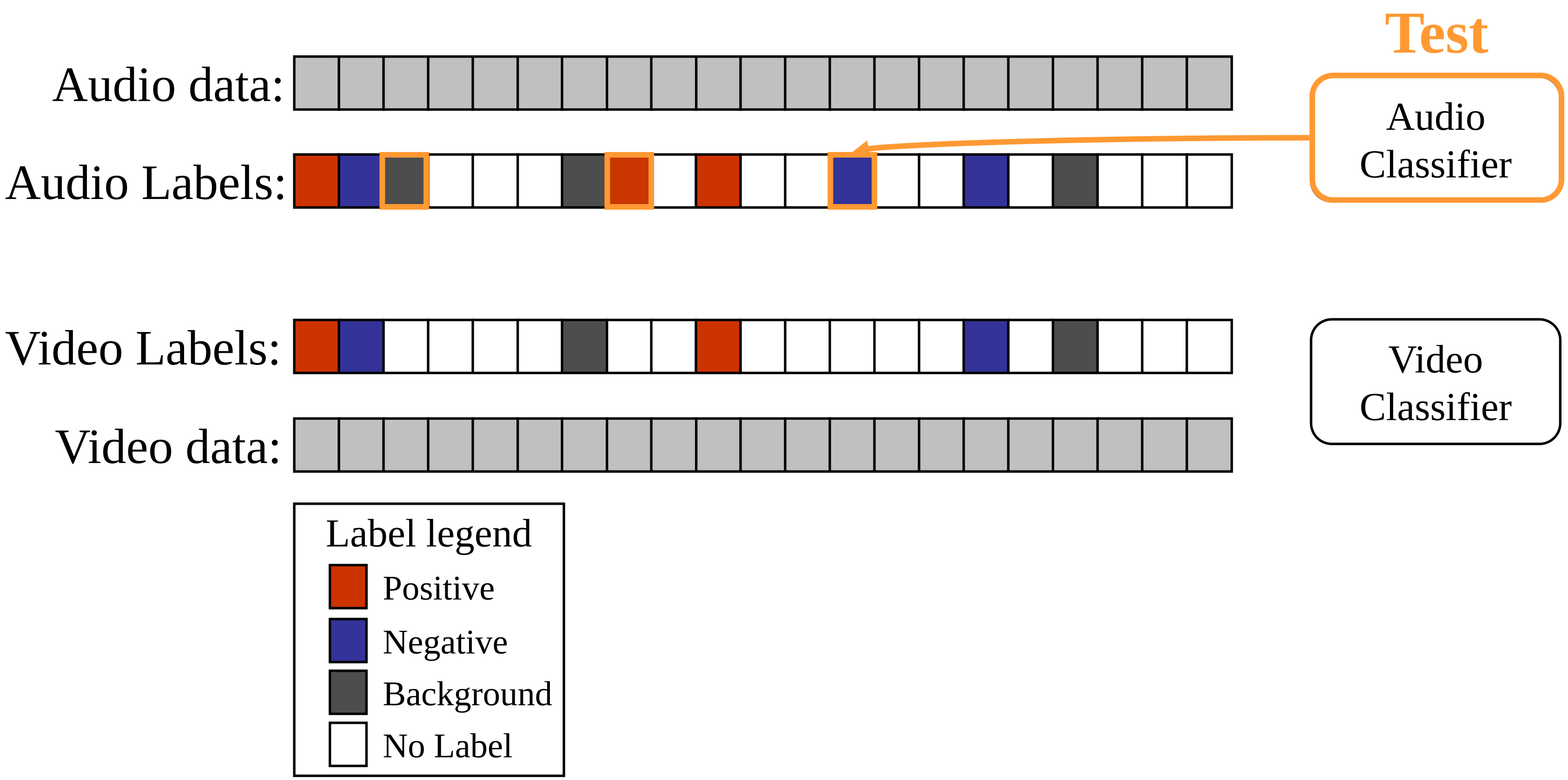
Co-training in the Presence of View Disagreement

- Step 2: Evaluate on unlabeled data, add *N* most confident examples from each view



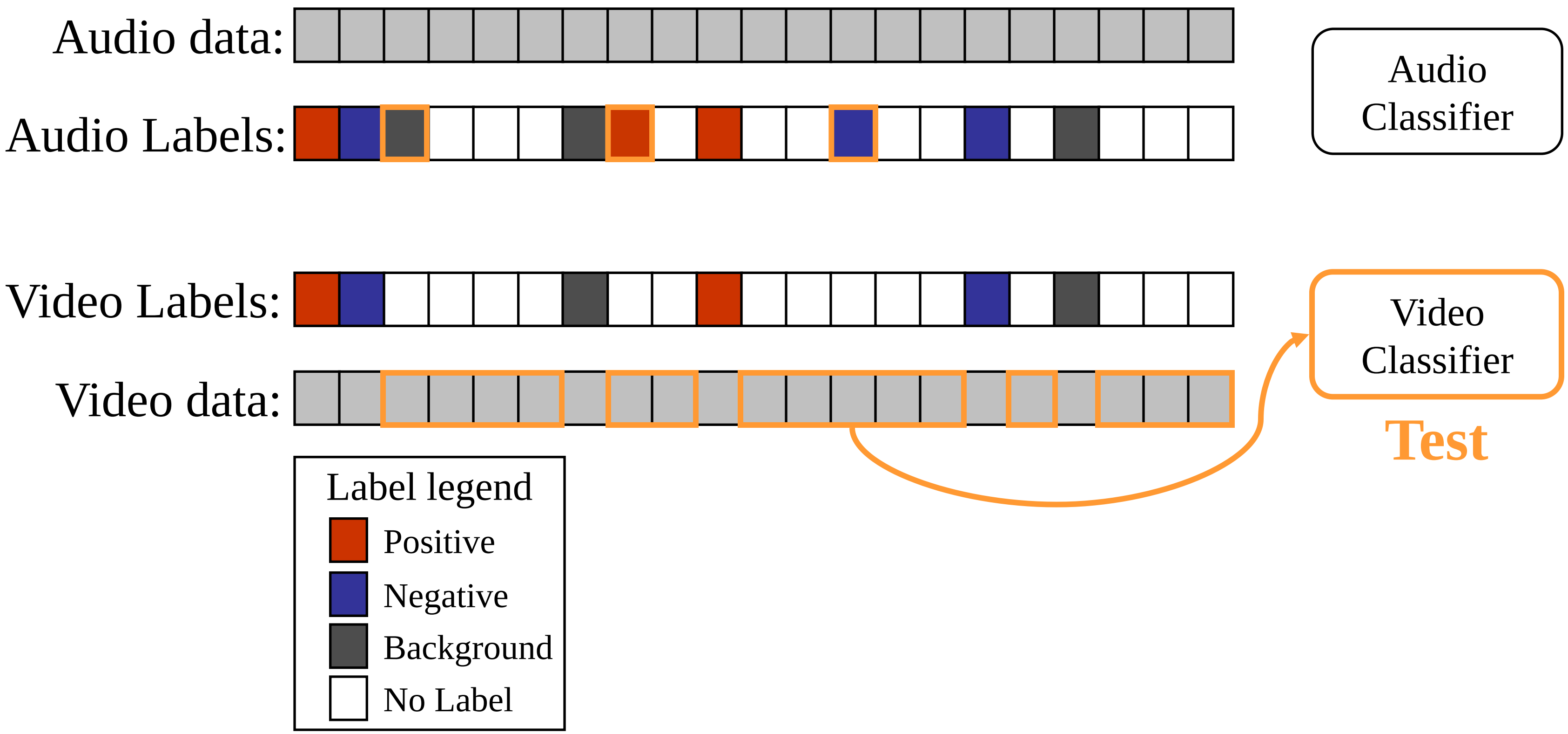
Co-training in the Presence of View Disagreement

- Step 2: Evaluate on unlabeled data, add *N* most confident examples from each view



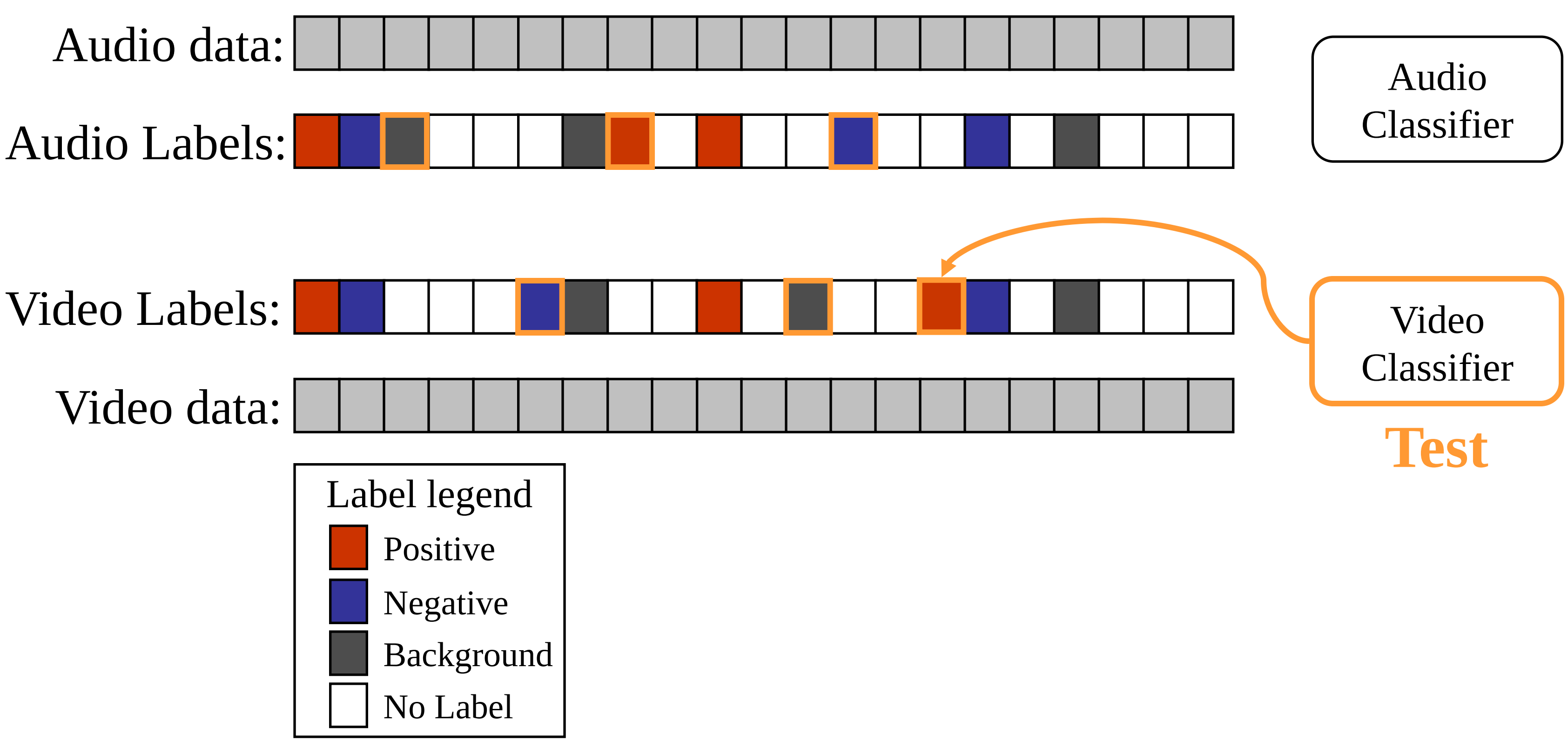
Co-training in the Presence of View Disagreement

- Step 2: Evaluate on unlabeled data, add *N* most confident examples from each view



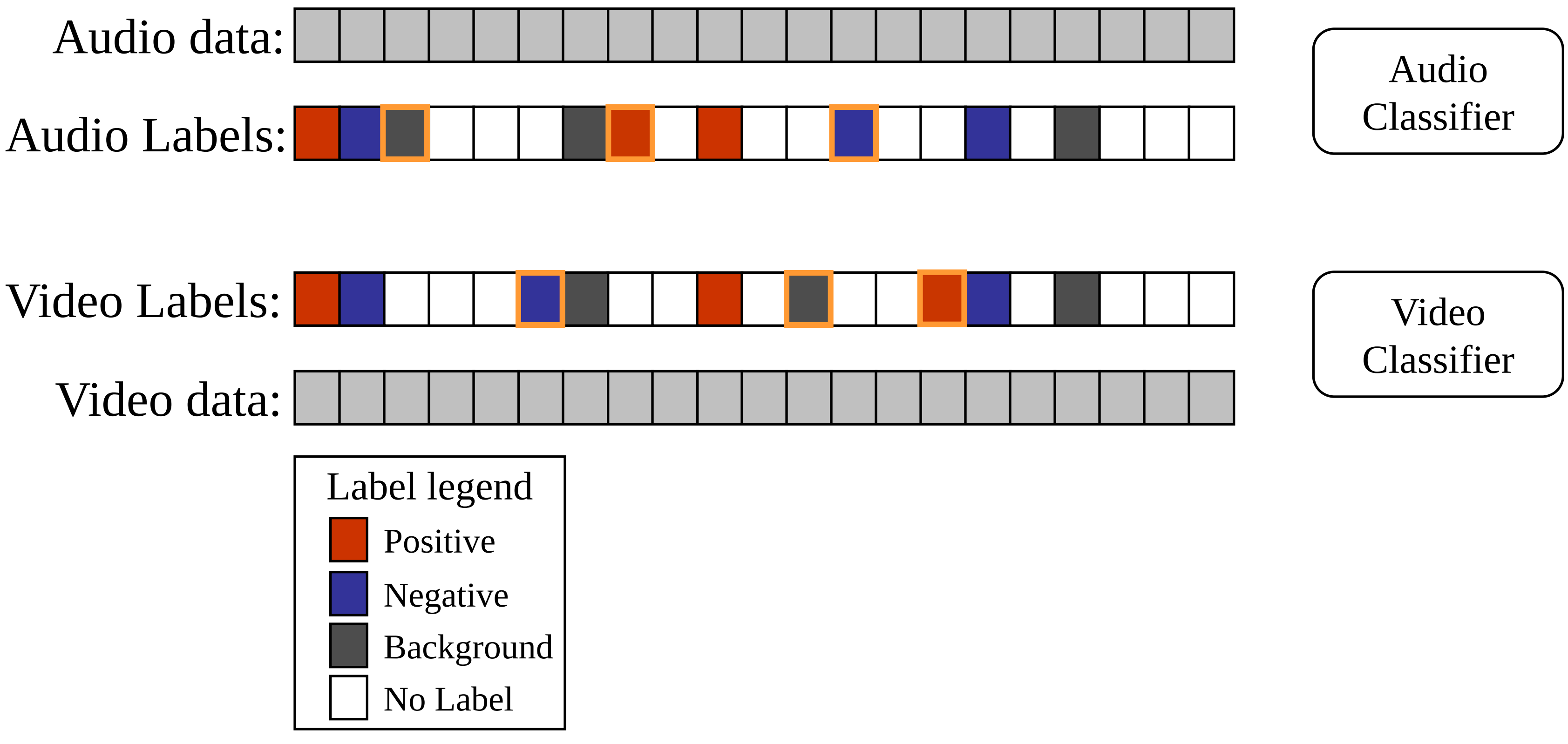
Co-training in the Presence of View Disagreement

- Step 2: Evaluate on unlabeled data, add *N* most confident examples from each view



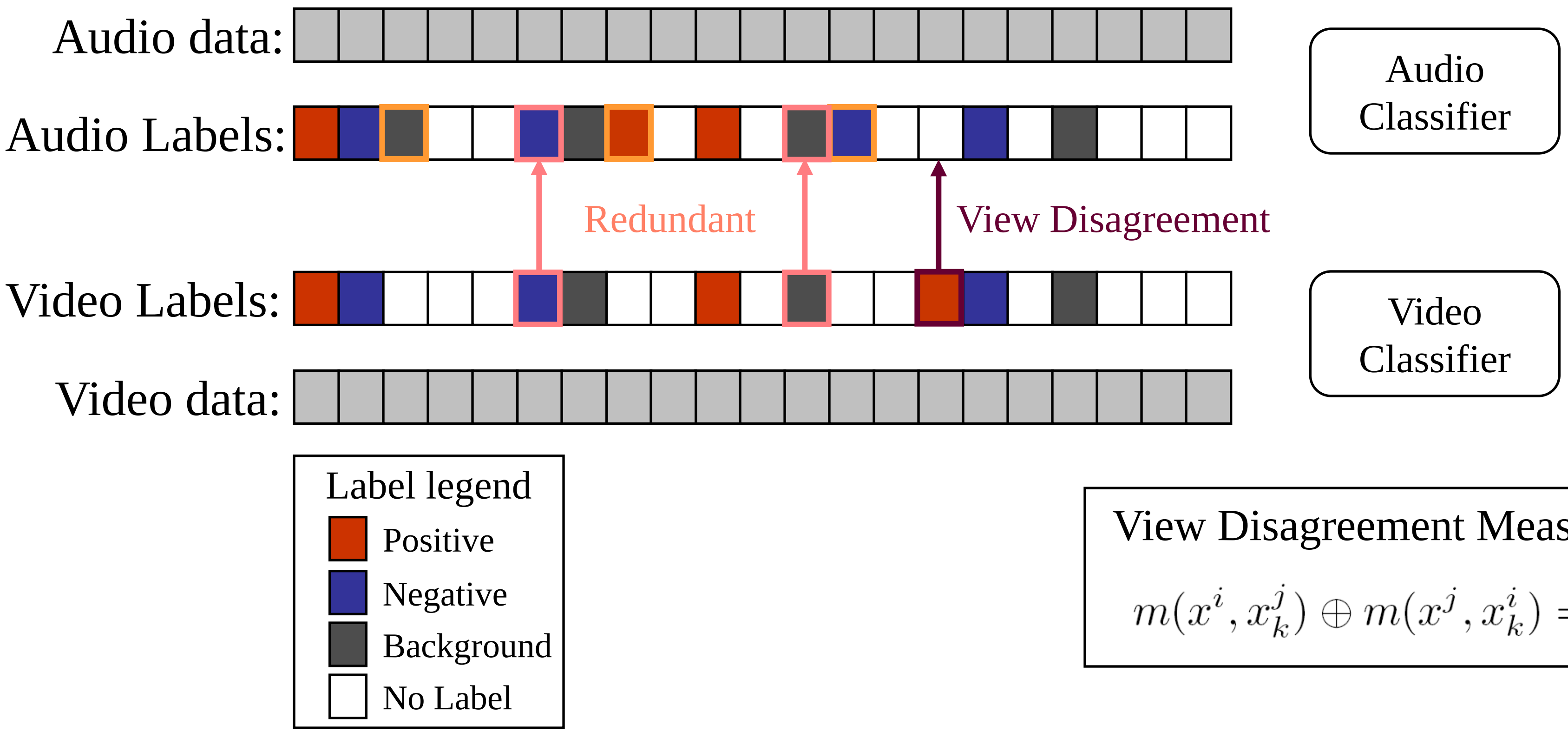
Co-training in the Presence of View Disagreement

- Step 3: Map labels using conditional-entropy measure



Co-training in the Presence of View Disagreement

- Step 3: Map labels using conditional-entropy measure

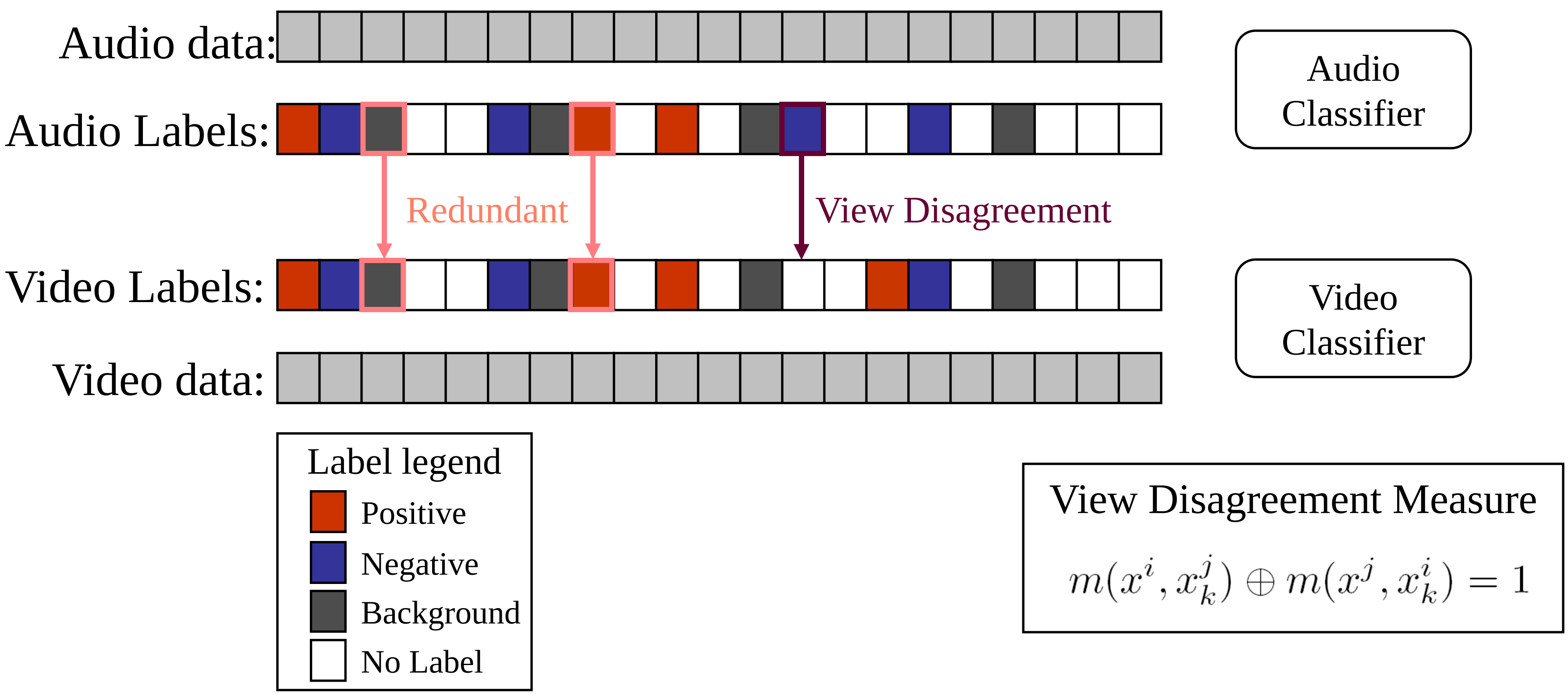


View Disagreement Measure

$$m(x^i, x^j_k) \oplus m(x^j, x^i_k) = 1$$

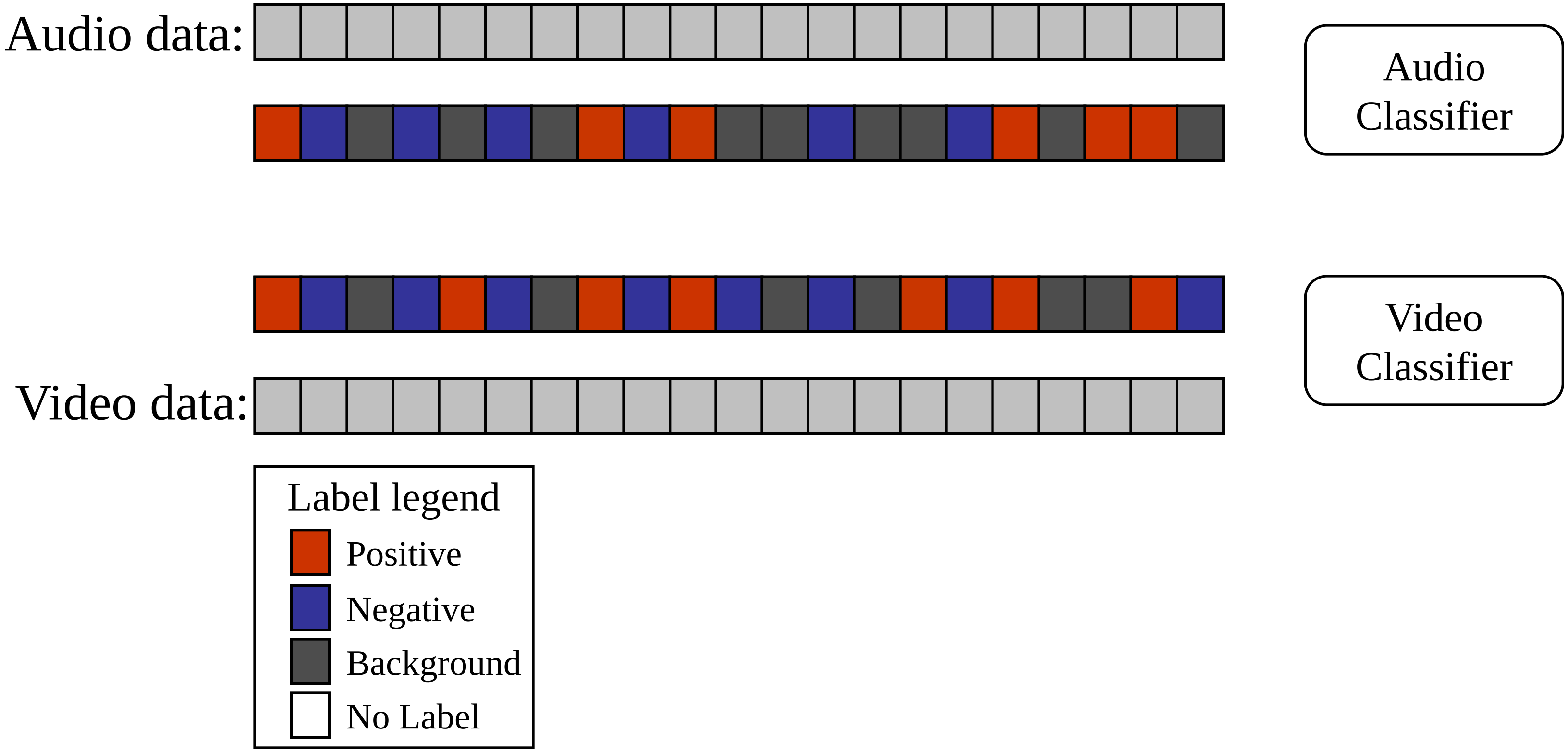
Co-training in the Presence of View Disagreement

- Step 3: Map labels using conditional-entropy measure

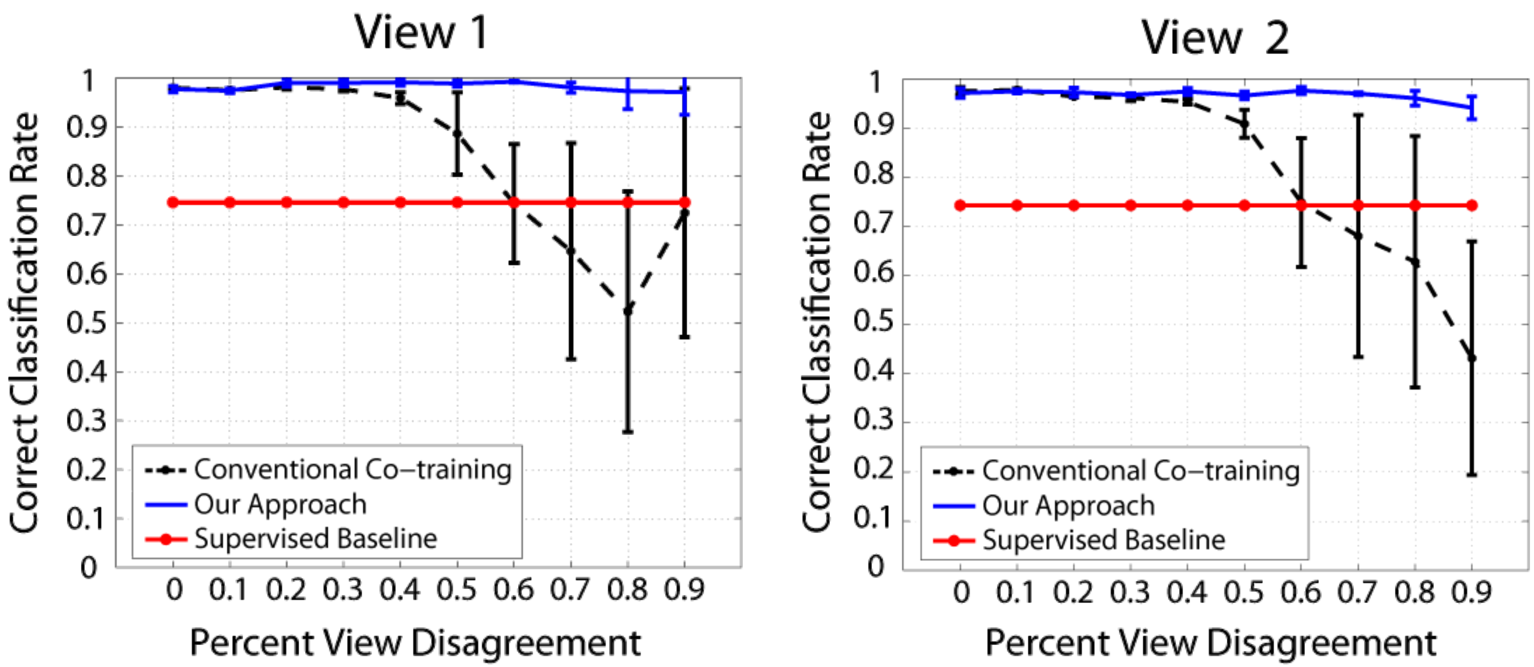


Co-training in the Presence of View Disagreement

- Iterate steps 1 through 3 until done



Normally Distributed Classes: Results



Real Data

- Agreement from head gesture and speech
  - Head gesture: nod/shake
  - Speech: ‘yes’ or ‘no’
  - 15 subjects, 103 questions

- Simulated view disagreement
  - Background segments in visual domain
  - Babble noise in audio

Experimental Setup

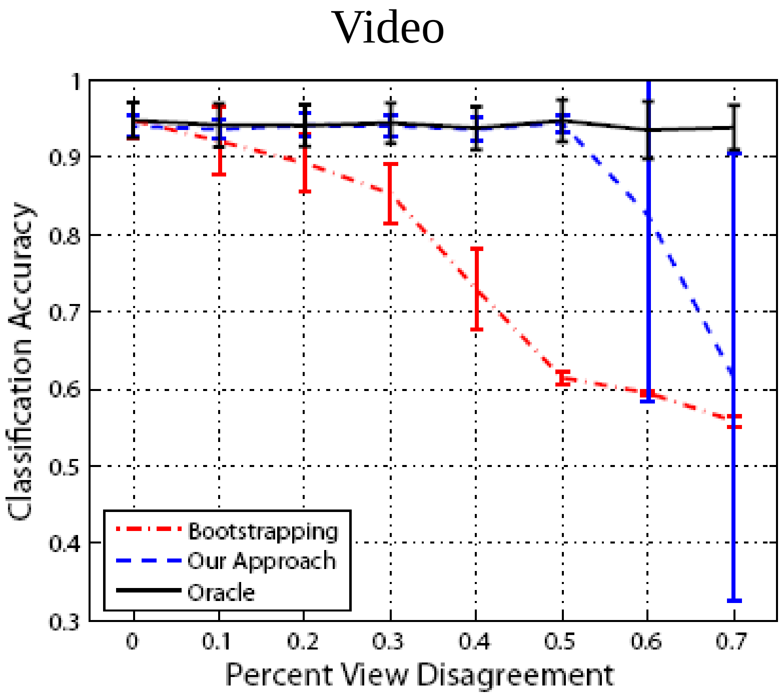
- Single frame audio and video observations
- Bayes classifier for audio and visual gesture recognition,

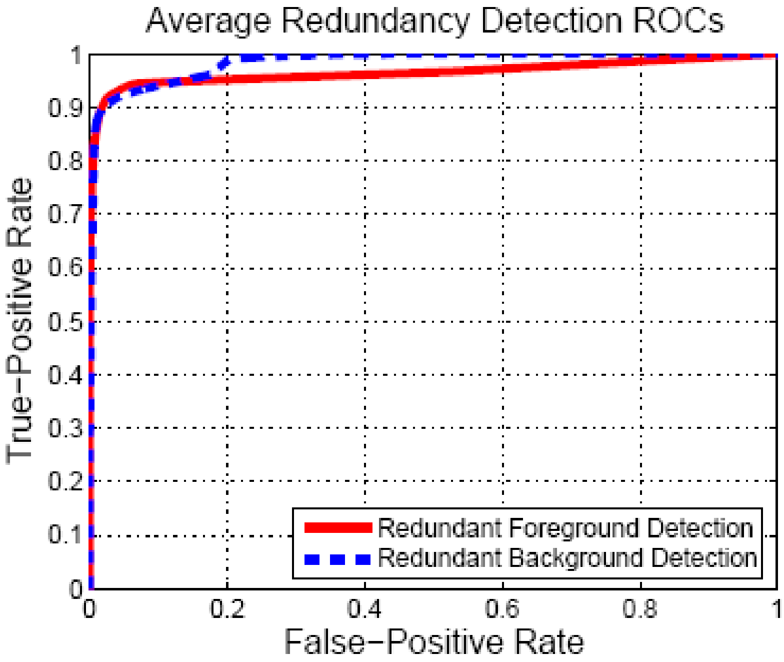
$$\underbrace{p(y}_{\text{label}} \mid \underbrace{x}_{\text{audio or video observation}}) = \frac{p(x \mid y)}{\sum_y p(x \mid y)}$$

  $p(x|y)$ is Gaussian.
- Randomly separated subjects into 10 train and 5 test subjects
- Show results averaged over 5 splits

Cross-View Bootstrapping Experiment

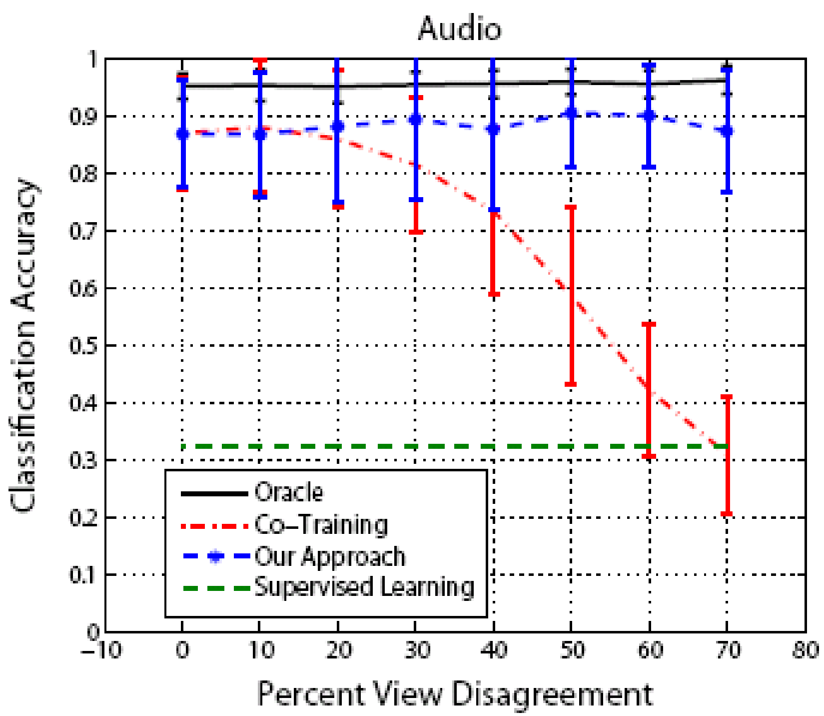
- Bootstrap visual classifier from audio labels

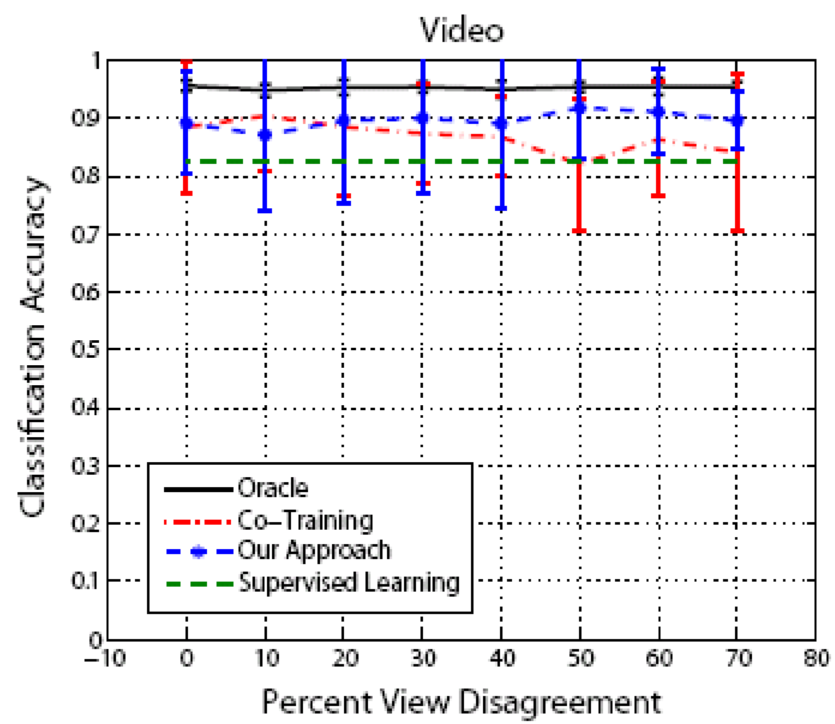




Co-training Experiment

- Learn both audio and video classifiers





Conclusions and Future Work

- Investigated the problem of view disagreement in multi-view learning
- Information theoretic measure to detect view disagreement due to view corruption
- On audio-visual user agreement task our method was robust to gross amounts of view disagreement (50%-70%)
- Future Work
  - More general view disagreement distributions
  - Integrate view disagreement uncertainty into co-training