Linking people in videos with “their” names using coreference resolution

Vignesh Ramanathan, Armand Joulin, Percy Liang, Li Fei-Fei




Computer Science Department

Stanford Vision Lab

Video Data

Video Data + Text Data

TV Script

Movie Script

Closed Caption

Video Data + Text Data

TV Script

Use this data to label people!

Movie Script

Closed Caption

Video Data + Text Data

TV Script

Names

Movie Script

Closed Caption

Video Data + Text Data

Everingham et al. 2006
Sivic et al. 2009
Cour et al. 2009
Tapaswi et al. 2012

TV Script

Names
Actions
Interactions
Events

Laptev et al. 2000
Marsazek et al. 2009
Bojanowski et al. 2013

Movie Script

Babaguchi et al. 2002
Xu et al. 2008

Closed Caption

Video Data + Text Data

Everingham et al. 2006
Sivic et al. 2009
Cour et al. 2009
Tapaswi et al. 2012

TV Script

Names

Actions

Interactions

Events

Laptev et al. 2000
Marsazek et al. 2009
Bojanowski et al. 2013

Movie Script

Babaguchi et al. 2002
Xu et al. 2008

Closed Caption

Text **to** Video

Roland arrives as Ian waits. He stands with Mary. She turns.

Challenging Setting

- No labelled instances
- Script is the only supervision

Unidirectional Models: Text **to** Video

Roland arrives as Ian waits. He stands with Mary. She turns.

Everingham et al. 2006
Sivic et al. 2009
Cour et al. 2009
Tapaswi et al. 2012
Bojanowski et al. '13

Unidirectional Models: Text **to** Video

Roland arrives as Ian waits. He stands with Mary. She turns.

Everingham et al. 2006
Sivic et al. 2009
Cour et al. 2009
Tapaswi et al. 2012
Bojanowski et al. '13

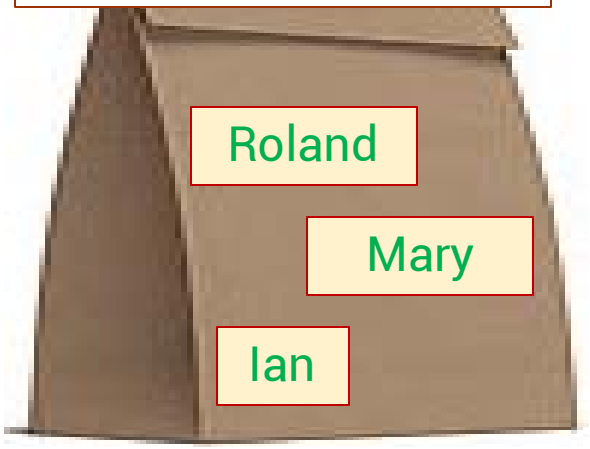
Use Limited Information

Only proper nouns

Unidirectional Models: Text **to** Video

Video = Bag of tracks

Script = Bag of names



Everingham et al. 2006
Sivic et al. 2009
Cour et al. 2009
Tapaswi et al. 2012
Bojanowski et al. '13

Use Limited Information

Only proper nouns

Treat Videos and Script as bag of tracks and names

Unidirectional Models: Text **to** Video

Unidirectional Models

Only proper nouns

Treat Videos and Script as bag of tracks and names

Temporal ordering of people

Time

Roland arrives as Ian waits. He stands with Mary. She turns

Time

Temporal ordering of people

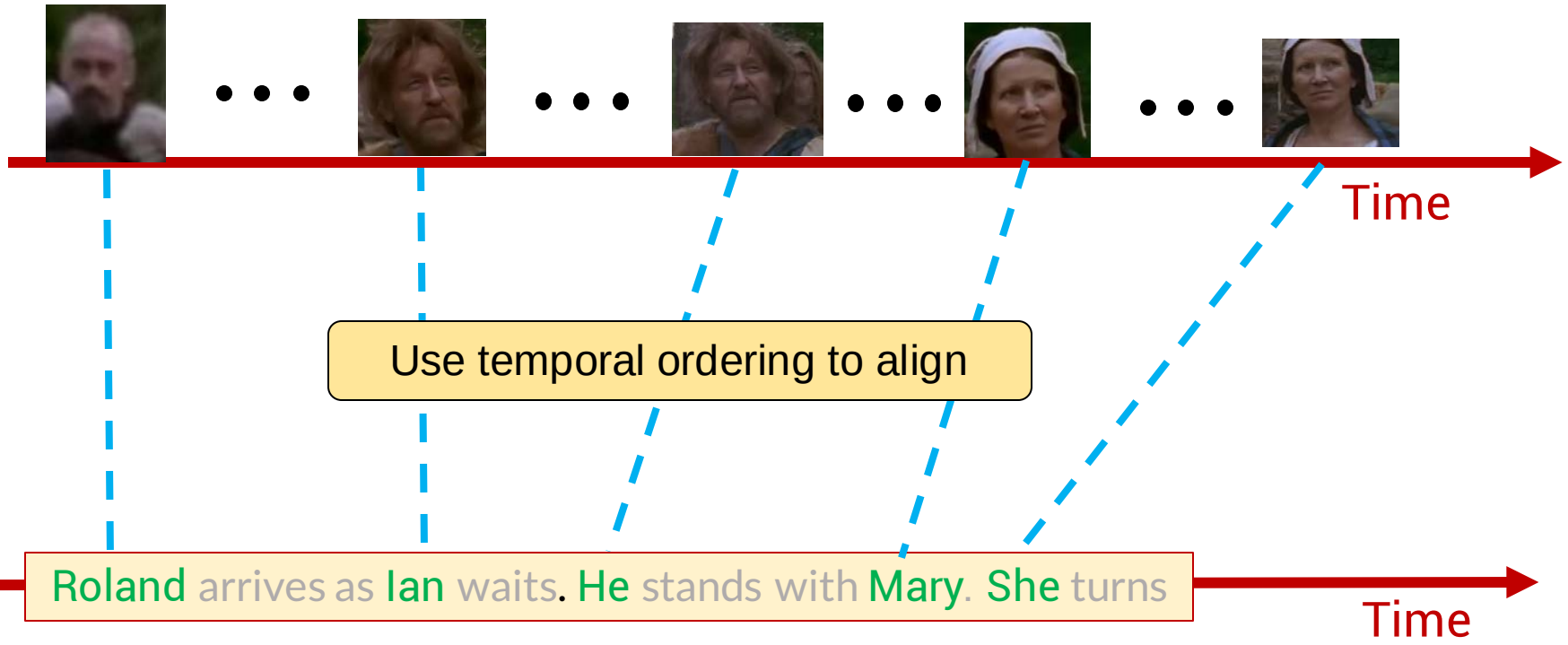
Time

People appear in the same order in both text and videos

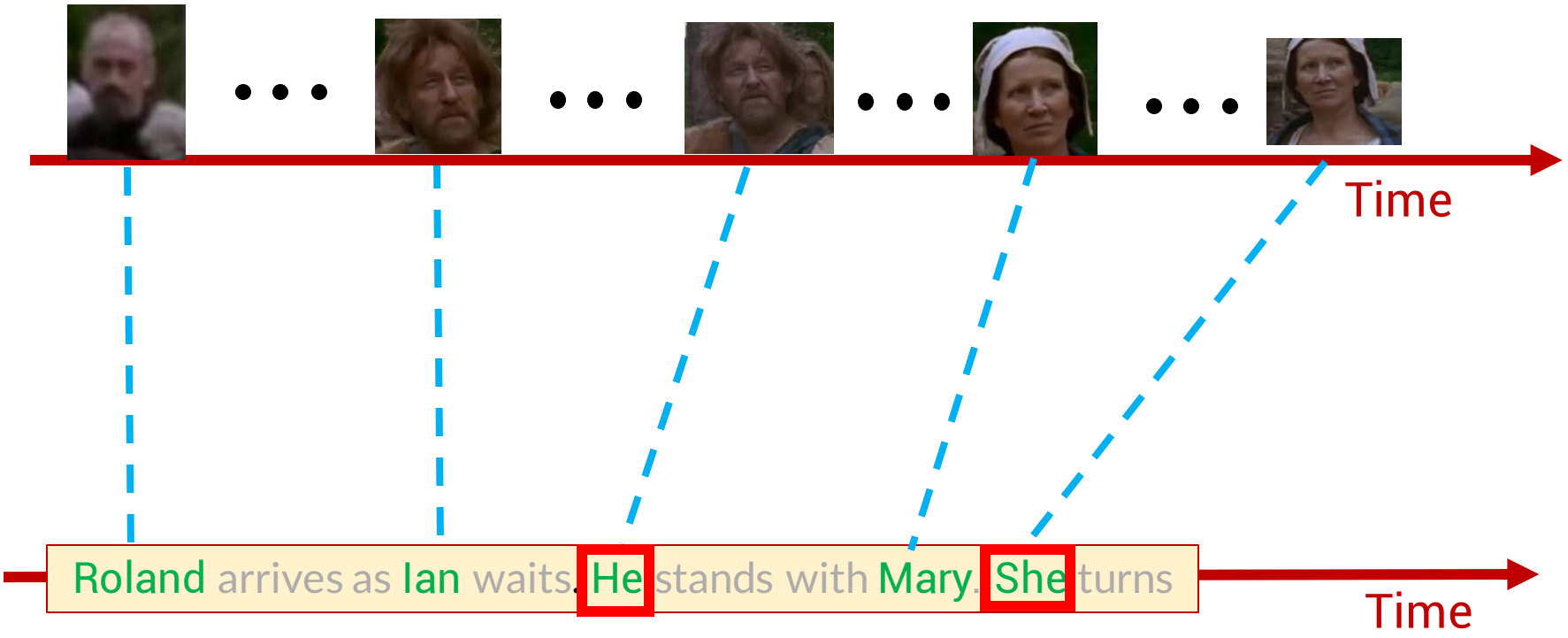
Roland arrives as Ian waits. He stands with Mary. She turns

Time

Temporal ordering of people

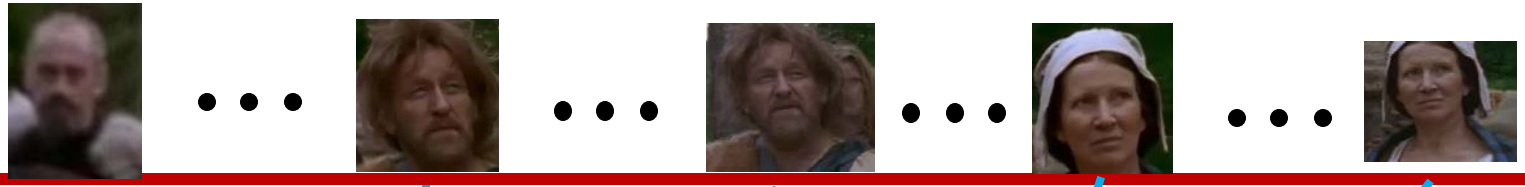


Coreference Resolution in NLP



To use the alignment we need to **identify the name of pronouns**

Coreference Resolution in NLP



Time

Roland arrives as Ian waits. He stands with Mary. She turns

Time

To use the alignment we need to **identify the name of pronouns**

→ Coreference Resolution

Coreference Resolution in NLP

Time

Roland arrives as Ian waits. He stands with Mary. She turns

Time

To use the alignment we need to **identify the name of pronouns**

Coreference Resolution

Coreference Resolution in NLP

Time

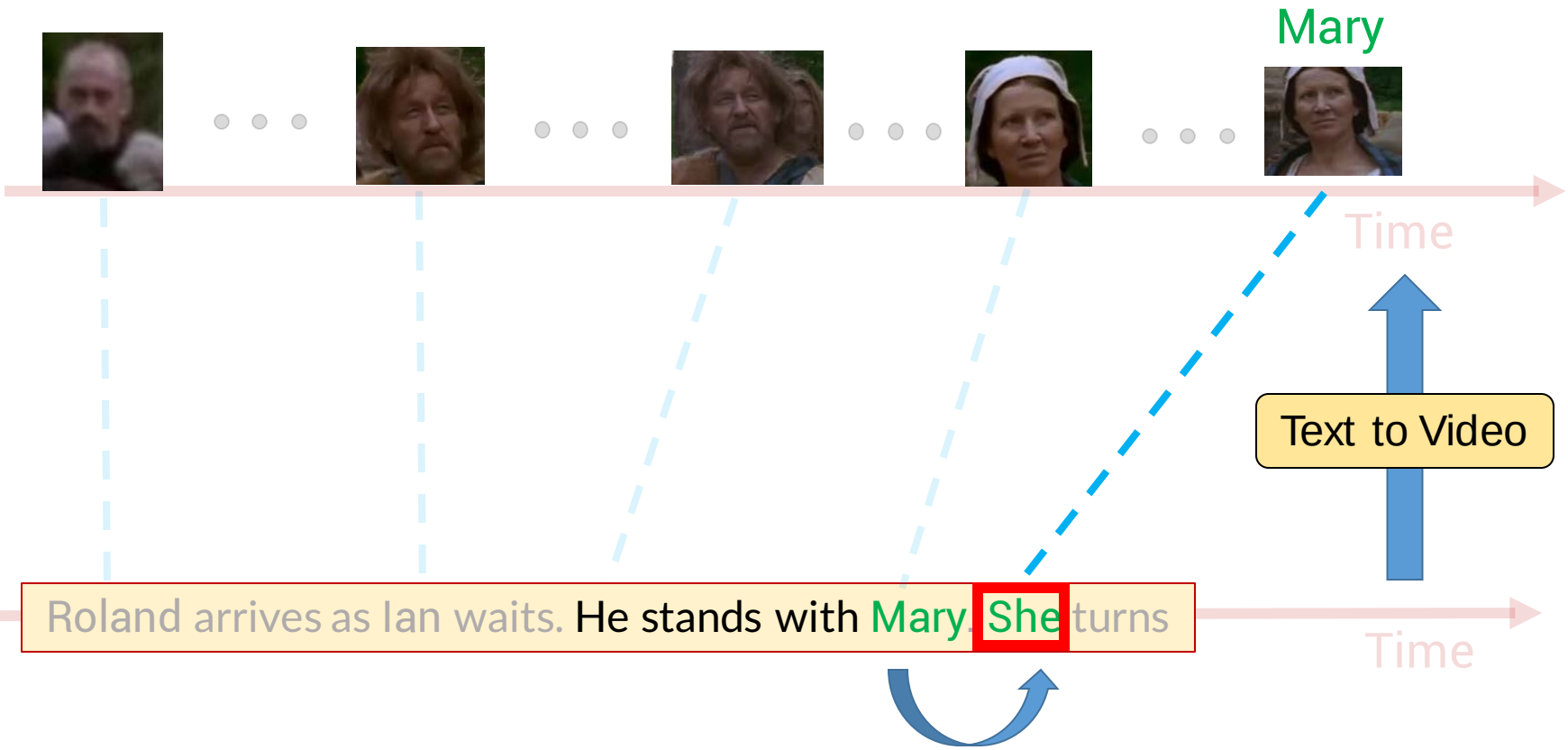
Roland arrives as Ian waits. He stands with Mary. She turns

Time

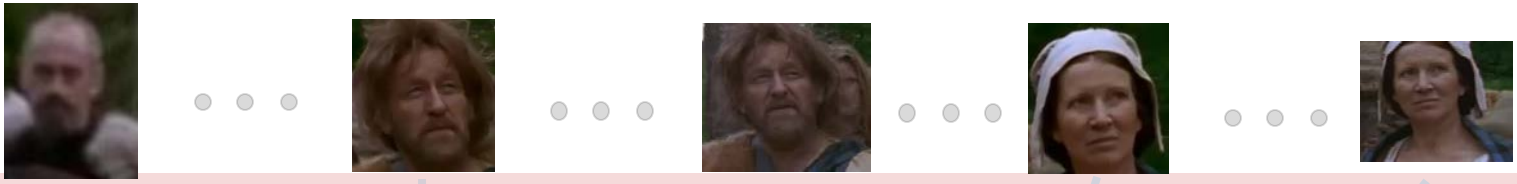
To use the alignment we need to **identify the name of pronouns**

Coreference Resolution

Coreference Resolution in NLP



Coreference Resolution in NLP



Time

Text to Video

Roland arrives as Ian waits. He stands with Mary. She turns

Time

A challenging NLP task

Coreference Resolution

Coreference Resolution in NLP

Time

Text to Video

?

Roland arrives as Ian waits. He stands with Mary. She turns

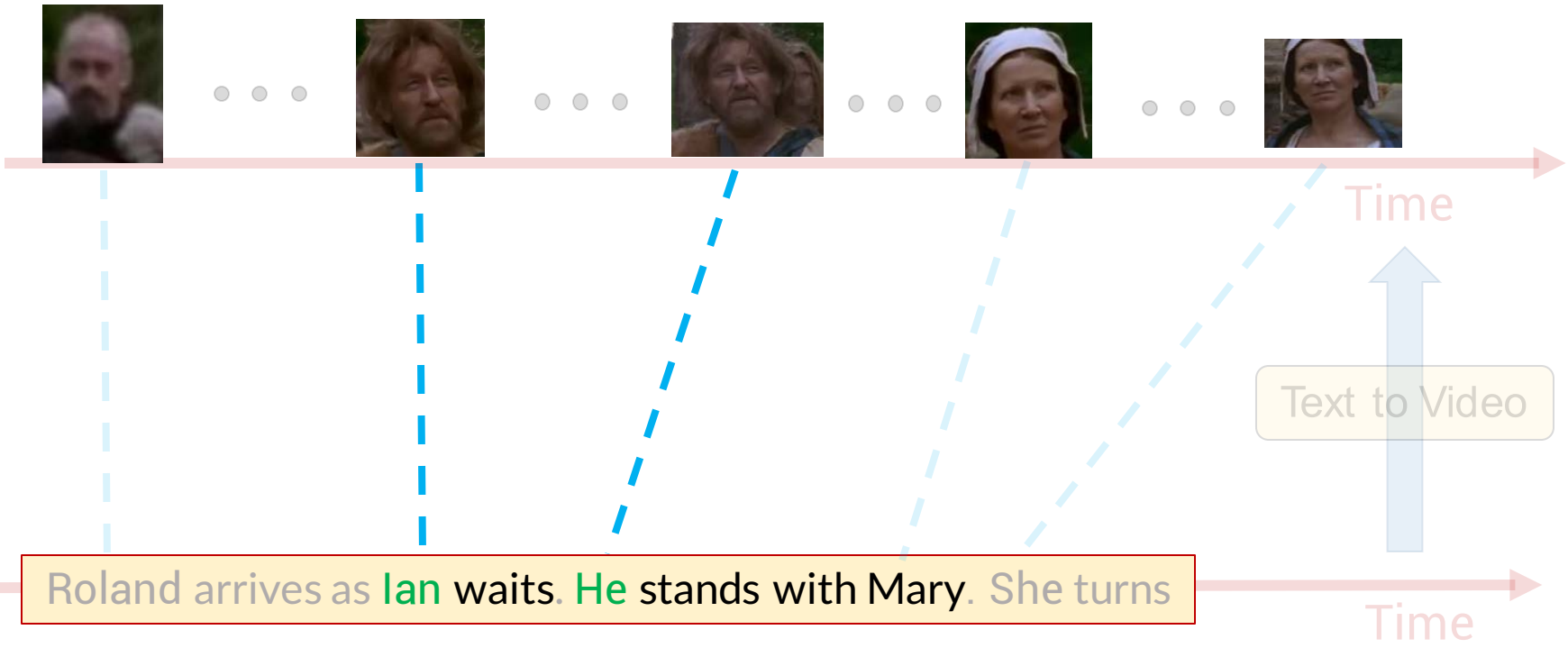
Time

?

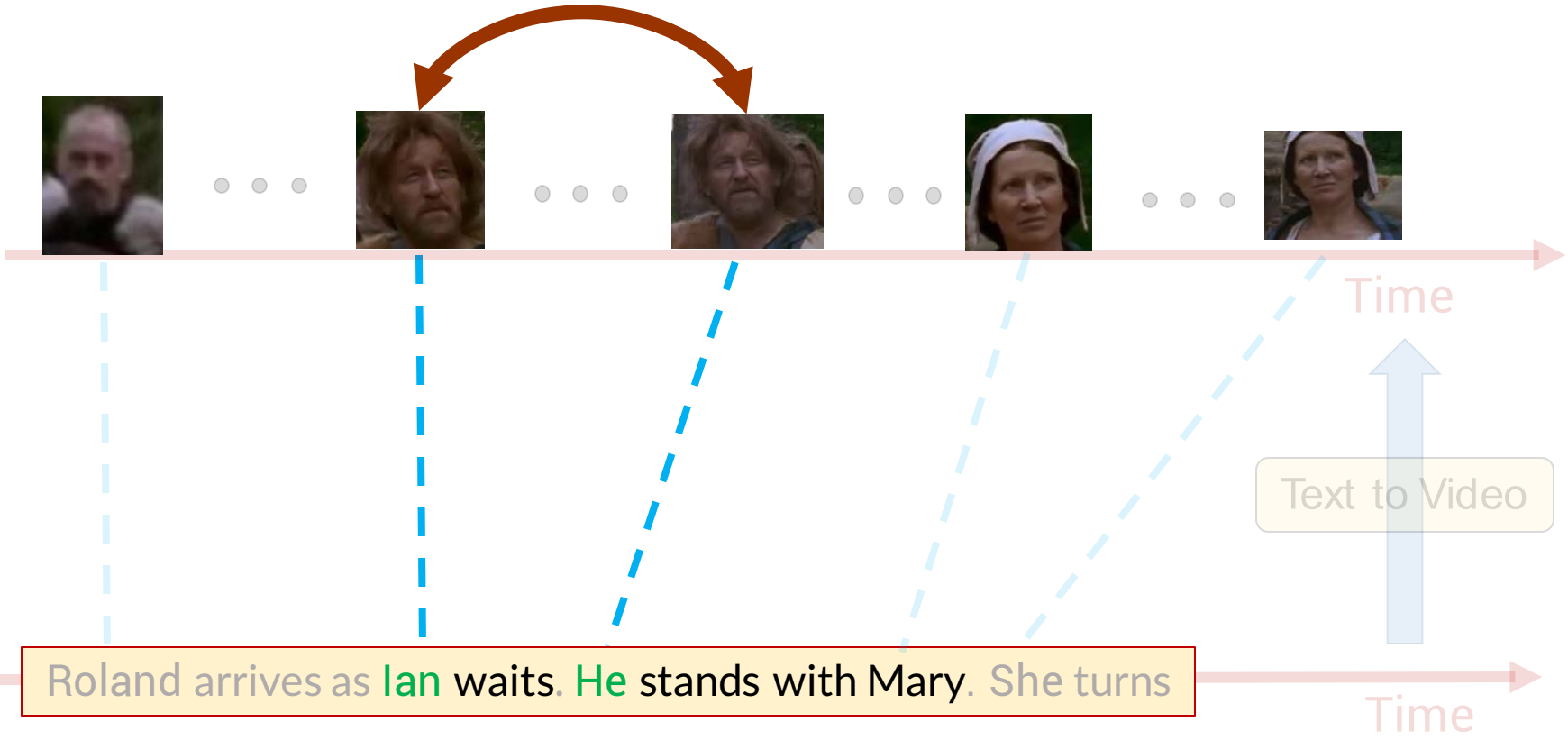
A challenging NLP task

Coreference Resolution

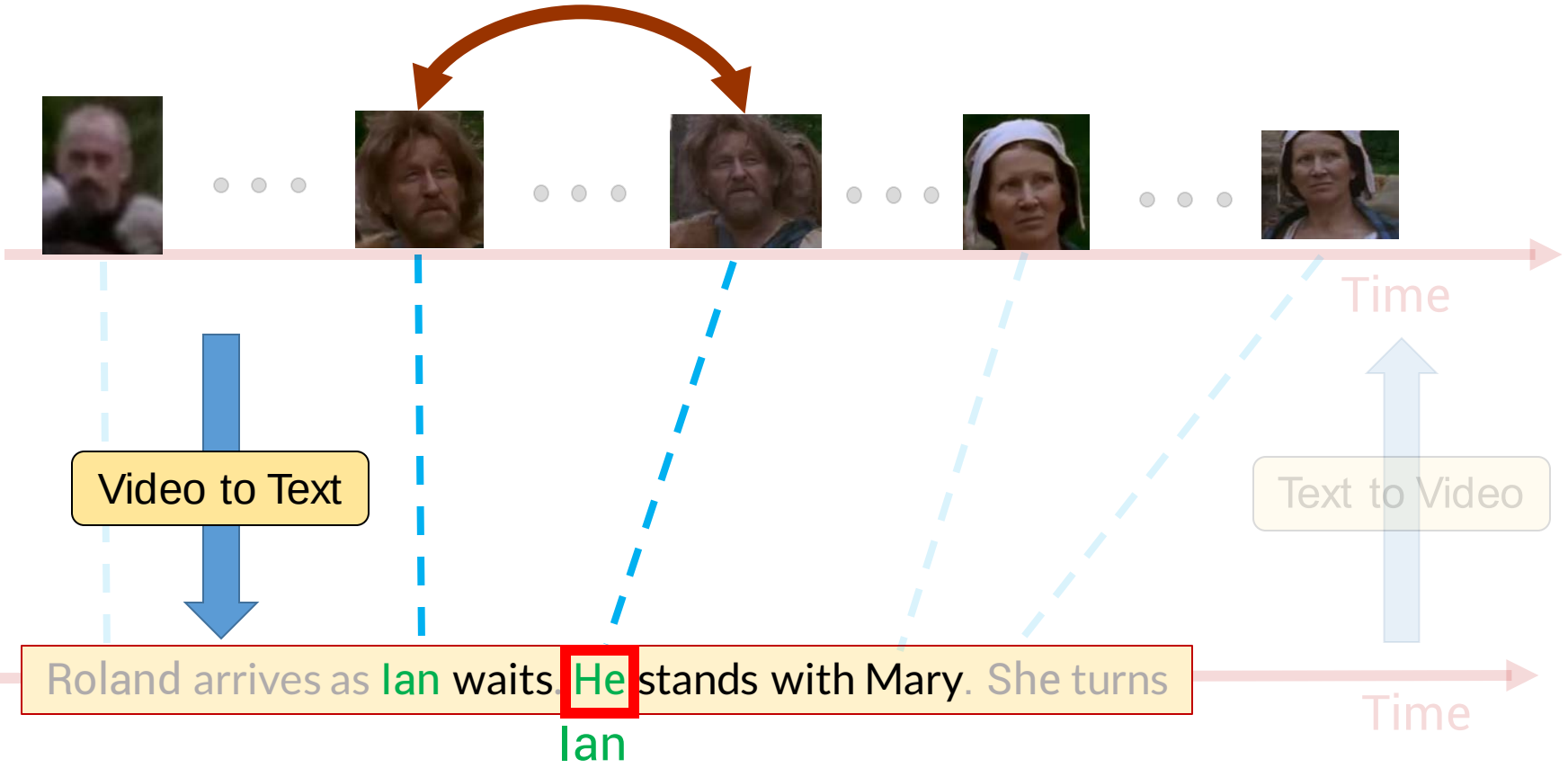
Coreference Resolution in NLP



Coreference Resolution in NLP



Coreference Resolution in NLP



Bidirectional model

Video to Text

Our Bidirectional Model

Text to Video

Roland arrives as Ian waits. He stands with Mary. She turns

Linking people with coreference resolution

Everingham et al. 2006
Sivic et al. 2009
Cour et al. 2009
Tapaswi et al. 2012
Bojanowski et al. '13

Our Bidirectional Model

Joint **coreference resolution** to use pronouns and nominals

Treat as **temporal sequences** to align video and text

Unidirectional Models

Only proper nouns

Treat Videos and Script as bag of tracks and names

Linking people with “their” names

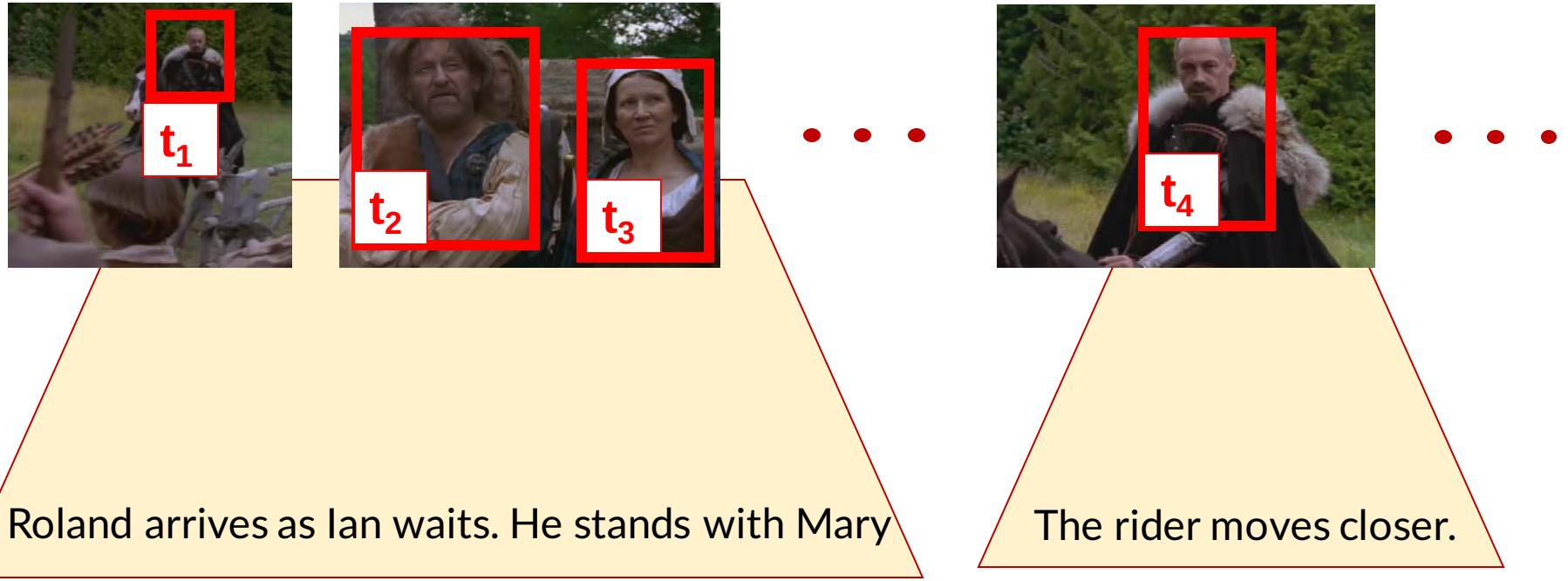
- Problem setup
- Our Bidirectional model
- Experiments
- Summary

Problem Setup (Input)

- Videos with detected human tracks

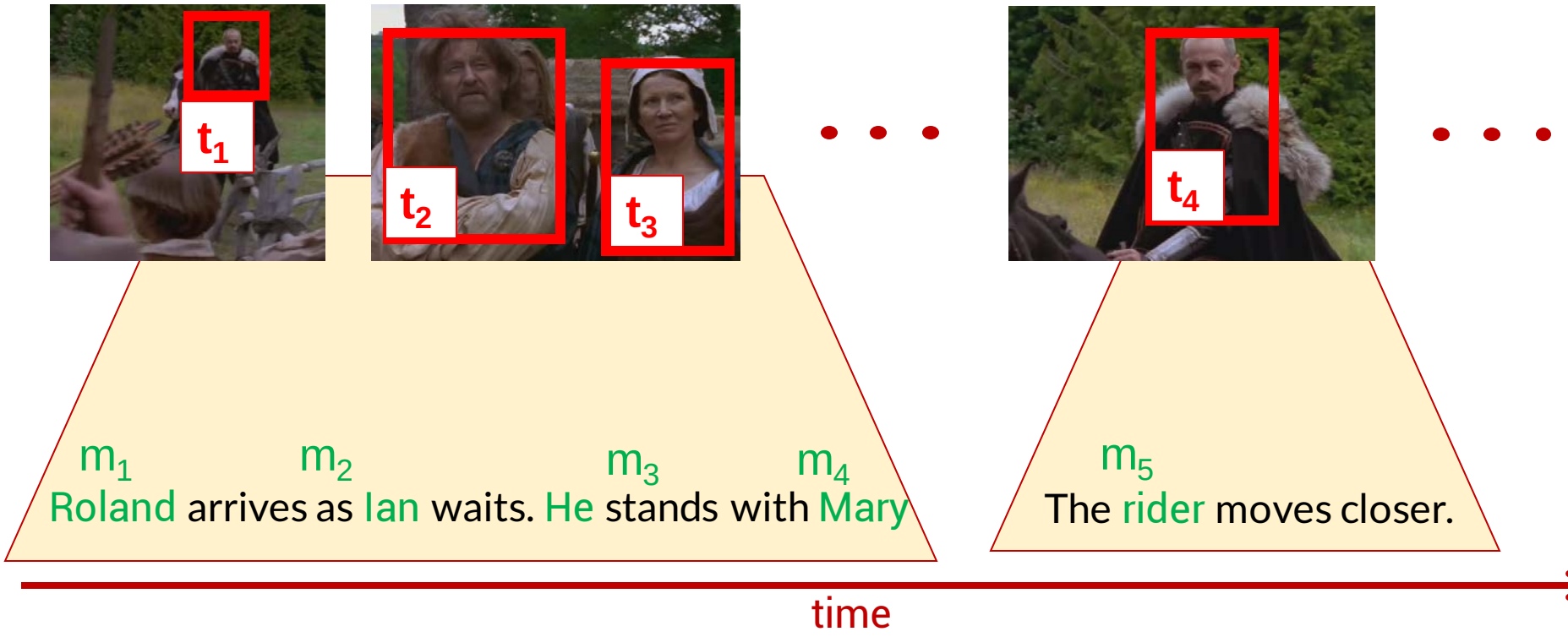
Problem Setup (Input)

- Videos with detected human tracks
- Descriptive script partially aligned with video segments



Problem Setup (Input)

- Videos with detected human tracks
- Descriptive script partially aligned with video segments
- Human mentions (nouns, pronouns) identified in the script.



Problem Setup (Input)

- Videos with detected human tracks
- Descriptive script partially aligned with video segments
- Human mentions (nouns, pronouns) identified in the script.
- A list of cast names

Cast List: Roland, Ian, Mary, …

m_1 m_2 m_3 m_4
Roland arrives as Ian waits. He stands with Mary

m_5
The rider moves closer.

time

Problem Setup (Output)

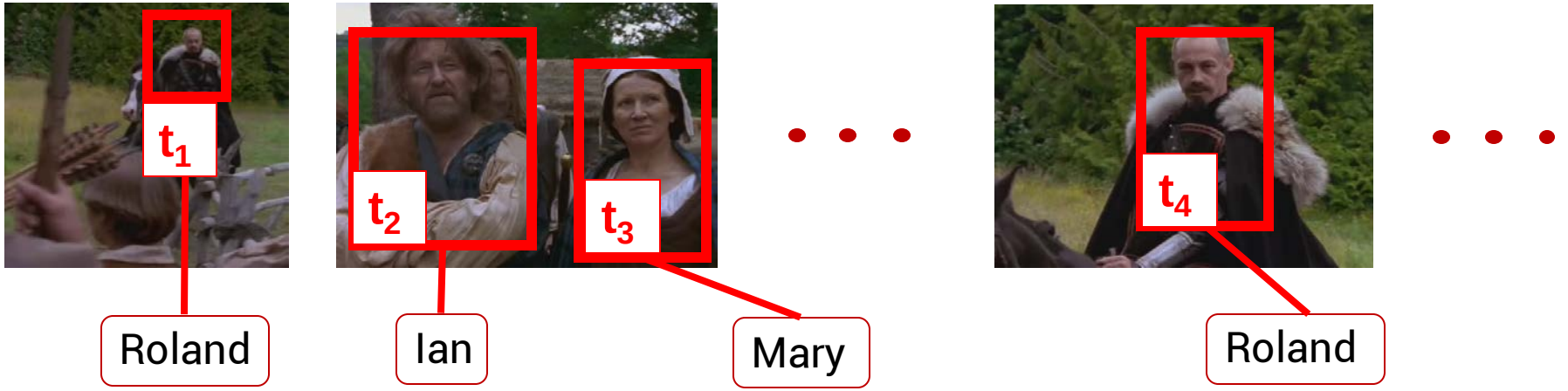
m_1 m_2 m_3 m_4 m_5

Roland arrives as Ian waits. He stands with Mary ...

The rider moves closer.

Problem Setup (Output)

Name assignment to human tracks



m_1 m_2 m_3 m_4 m_5

Roland arrives as Ian waits. He stands with Mary ... The rider moves closer.

Problem Setup (Output)

Name assignment to human tracks

m_1 m_2 m_3 m_4

Roland arrives as Ian waits. He stands with Mary ...

m_4

The rider moves closer.

Problem Setup (Output)

Name assignment to human tracks

Name assignment to human mentions

Roland — m_1

Ian — m_2

Ian — m_3

Mary — m_4

Roland — m_4

Roland arrives as Ian waits. He stands with Mary ...

The rider moves closer.

Linking people with “their” names

- Problem setup
- Our Bidirectional model
- Experiments
- Summary

Our Bidirectional Model

Minimize joint cost of name assignment to tracks and mentions with alignment

Our Bidirectional Model

Minimize joint cost of name assignment to tracks and mentions with alignment

Roland

Ian

Mary

Tracks

Track name assignment

Names

Our Bidirectional Model

Minimize joint cost of name assignment to tracks and mentions with alignment

Roland Ian Mary

Tracks

Track name assignment

Names

Roland Ian Ian Mary

m_1 m_2 m_3 m_4

Roland arrives as Ian waits. He stands with Mary

Names

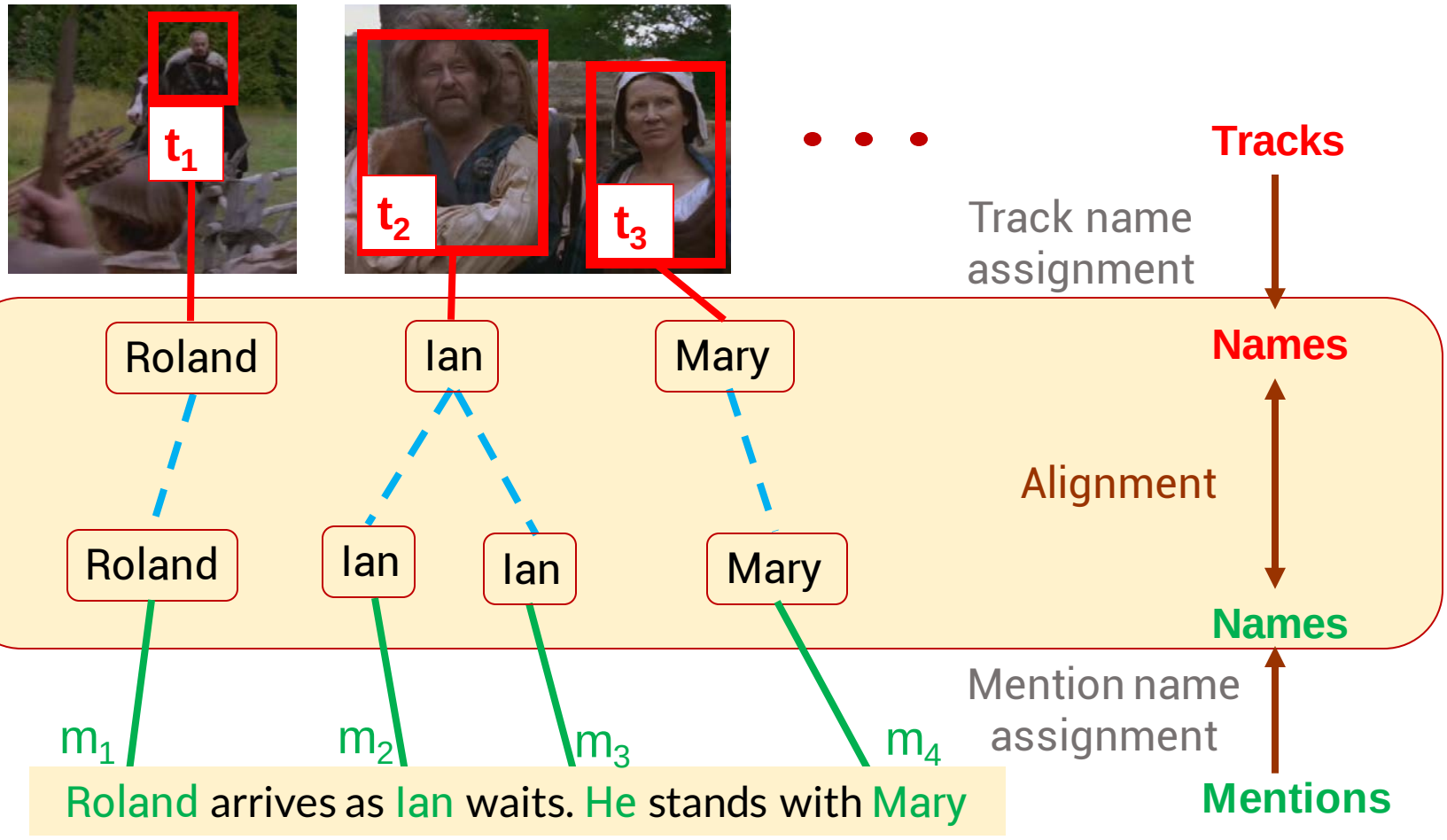
Mention name assignment

Mentions

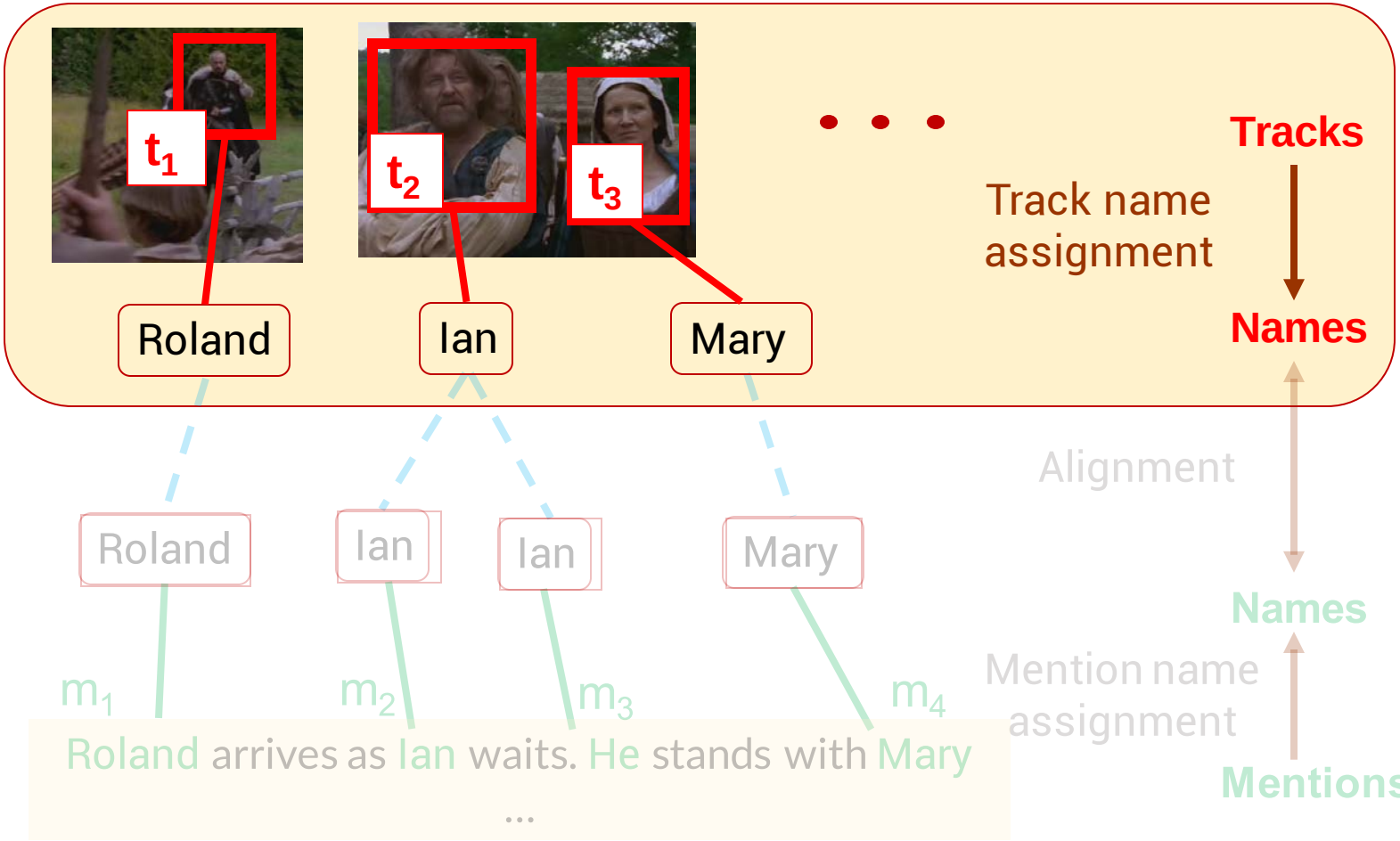
...

Our Bidirectional Model

Minimize joint cost of name assignment to tracks and mentions with alignment



Our Bidirectional Model: **Track name assignment**



Our Bidirectional Model: **Track name assignment**

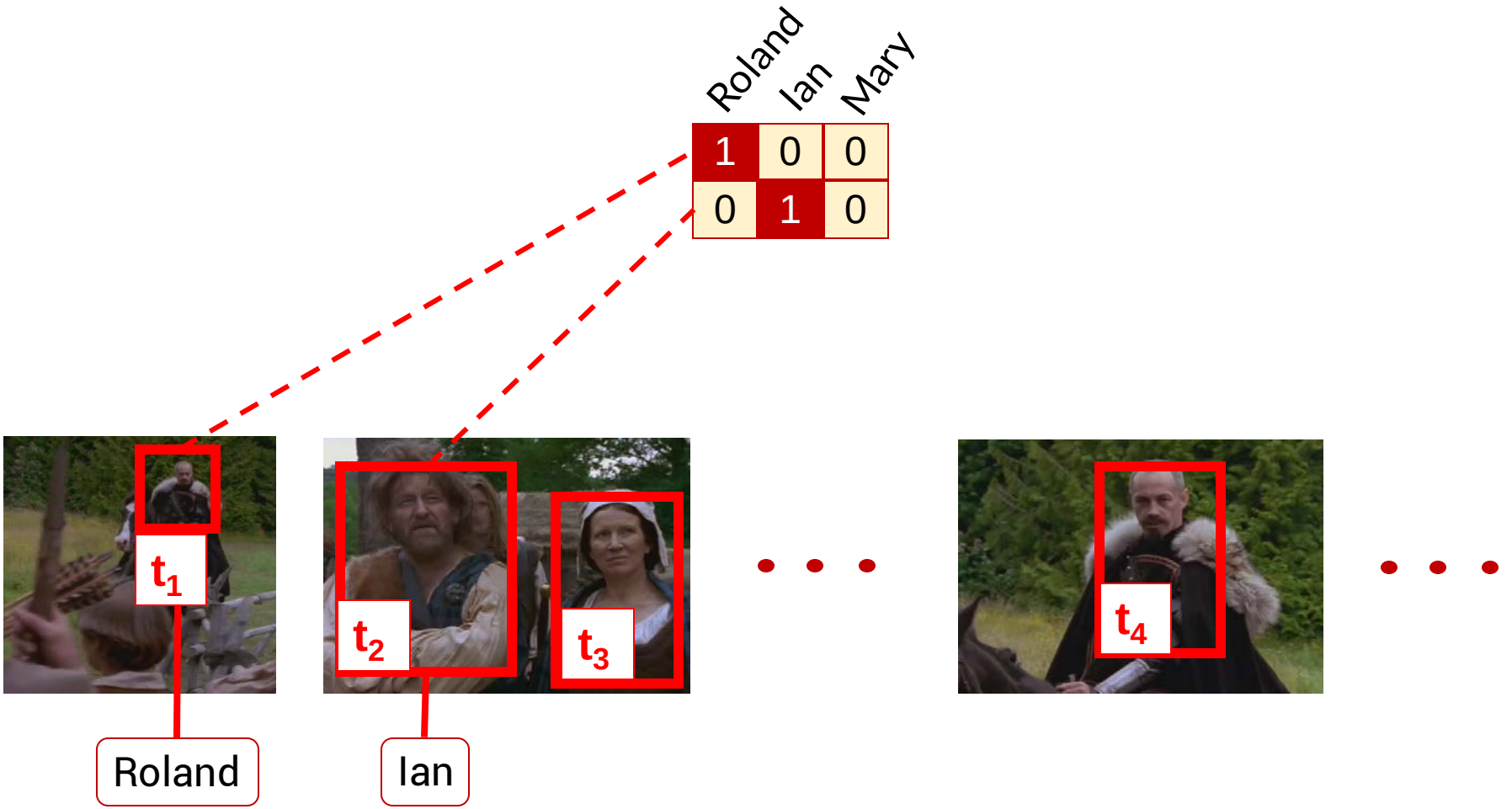
$\mathbf{Y} \in \{0,1\}^{T \times P}$ Name assignment for tracks

Our Bidirectional Model: **Track name assignment**

$\mathbf{Y} \in \{0,1\}^{T \times P}$ Name assignment for tracks

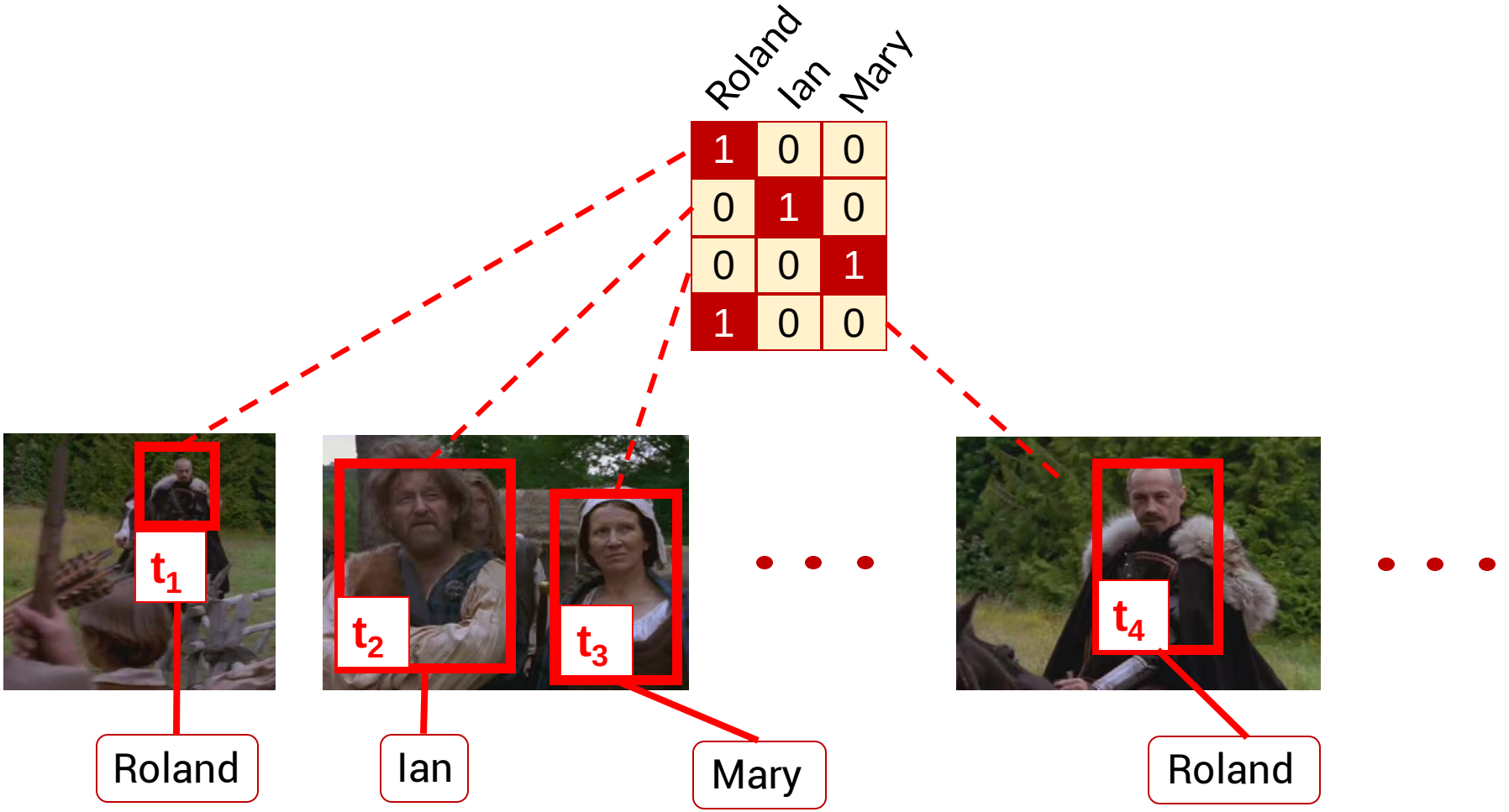
Our Bidirectional Model: **Track name assignment**

$\mathbf{Y} \in \{0,1\}^{T \times P}$ Name assignment for tracks



Our Bidirectional Model: **Track name assignment**

$\mathbf{Y} \in \{0,1\}^{T \times P}$ Name assignment for tracks



Our Bidirectional Model: **Track name assignment**

$\mathbf{Y} \in \{0,1\}^{T \times P}$ Name assignment for tracks

• • •

• • •

Our Bidirectional Model: **Track name assignment**

$\mathbf{Y} \in \{0,1\}^{T \times P}$ Name assignment for tracks

Clustering cost*

$$tr\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right)$$

Convex quadratic in $\mathbf{Y}$

*Bojanowski et al. ICCV'13

- Discriminative clustering.
- Tracks with **similar features** should have **same name**.

Our Bidirectional Model: **Track name assignment**

$\mathbf{Y} \in \{0,1\}^{T \times P}$ Name assignment for tracks

Clustering cost*

$$tr\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right)$$

Convex quadratic in $\mathbf{Y}$

*Bojanowski et al. ICCV'13

Our Bidirectional Model: **Track name assignment**

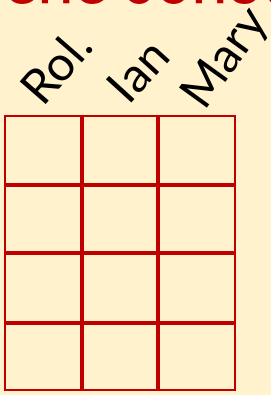
$\mathbf{Y} \in \{0,1\}^{T \times P}$ Name assignment for tracks

Clustering cost

$$tr\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right)$$

Convex quadratic in $\mathbf{Y}$

Scene constraint



Scene (Ian, Mary)

Our Bidirectional Model: **Track name assignment**

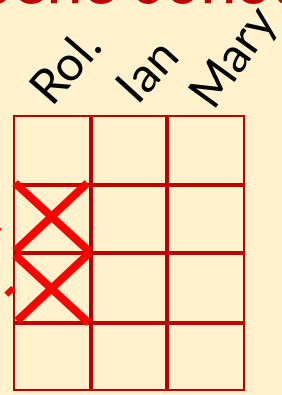
$\mathbf{Y} \in \{0,1\}^{T \times P}$ Name assignment for tracks

Clustering cost

$$tr\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right)$$

Convex quadratic in $\mathbf{Y}$

Scene constraint



The tracks are not “Roland”

Scene (Ian, Mary)

Our Bidirectional Model: **Track name assignment**

$\mathbf{Y} \in \{0,1\}^{T \times P}$ Name assignment for tracks

Clustering cost

$$tr\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right)$$

Convex quadratic in $\mathbf{Y}$

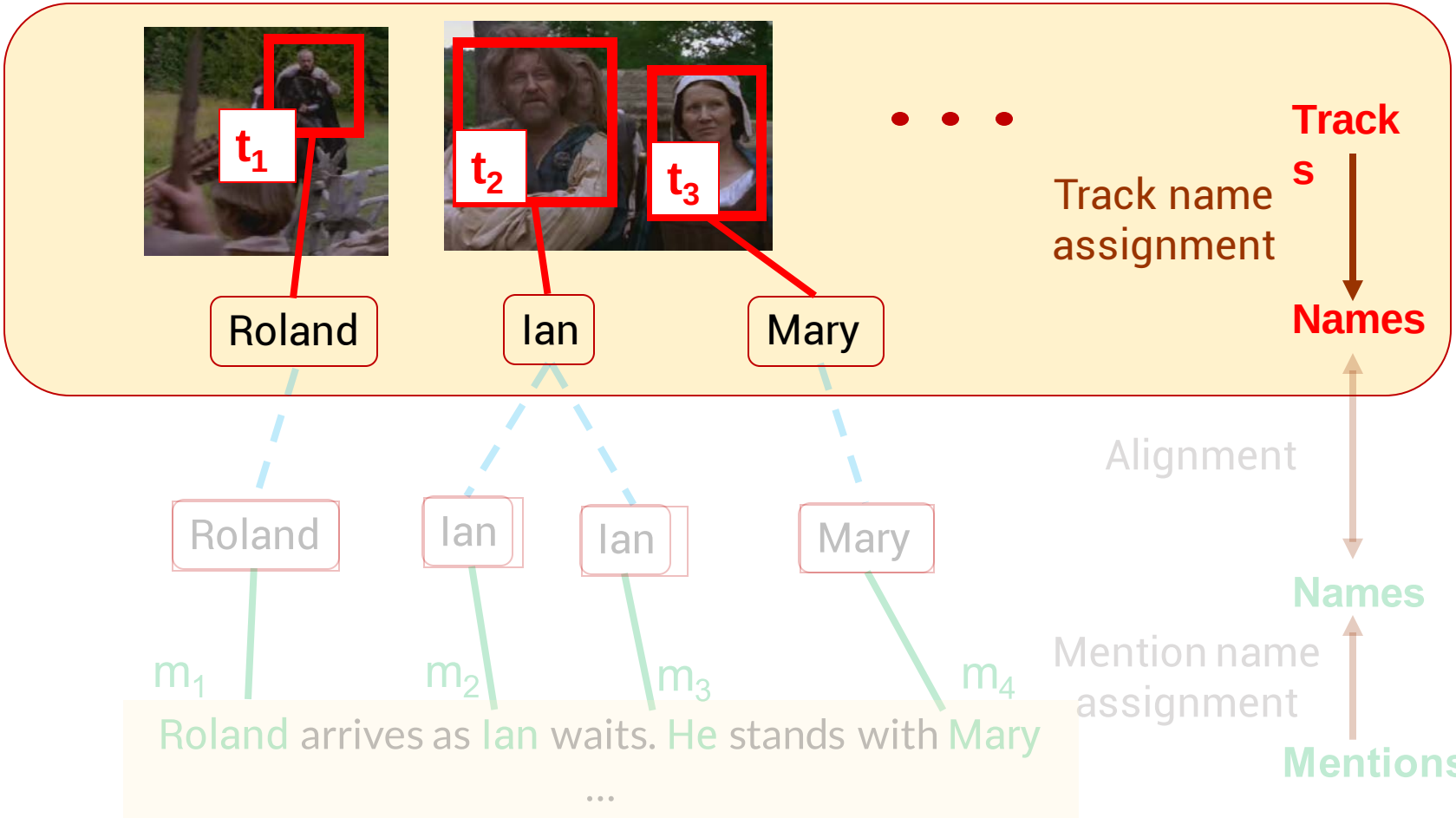
Scene constraint

$$\sum_{\substack{t \in \text{scene} \\ p \notin \text{scene}}} \mathbf{Y}_{t,p} = 0$$

Linear constraint

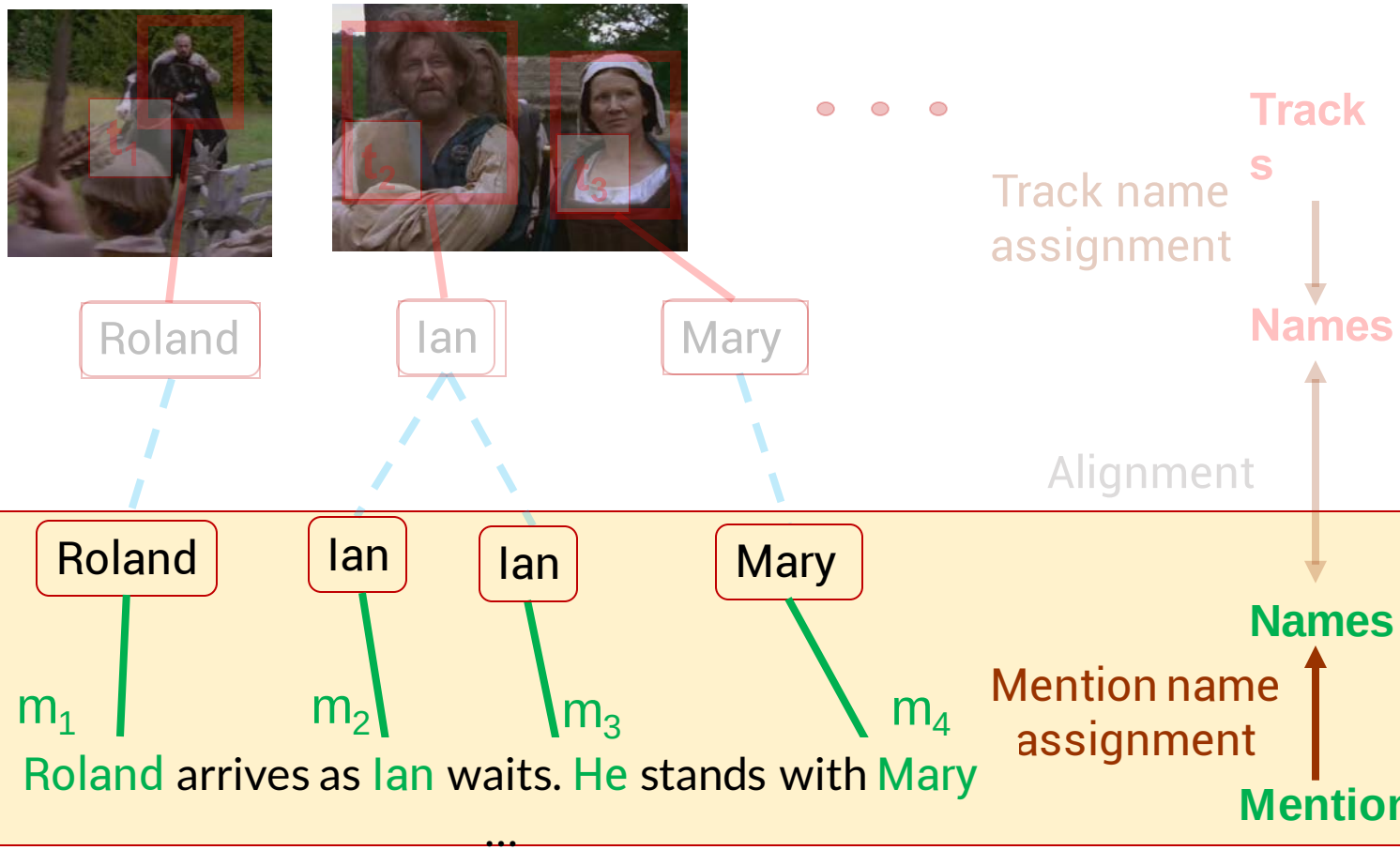
Our Bidirectional Model

$$\min \quad \mathrm{tr}\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right)$$
$$\text{s.t} \quad \mathbf{Y} \in C_Y$$



Our Bidirectional Model

$$\min \quad \mathrm{tr}\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right)$$
$$\text{s.t} \quad \mathbf{Y} \in C_Y$$



Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

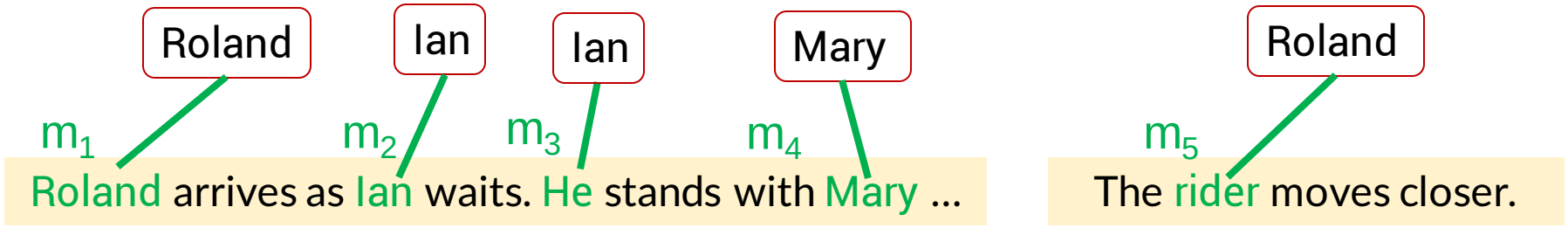
m_1 m_2 m_3 m_4 m_5

Roland arrives as Ian waits. He stands with Mary ... The rider moves closer.

Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

	Roland	Ian	Mary
m_1	1	0	0
m_2	0	1	0
m_3	0	1	0
m_4	0	0	1
m_5	1	0	0



Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

Mentions independently are not informative!

m_3

Roland arrives as Ian waits. He stands with Mary ...

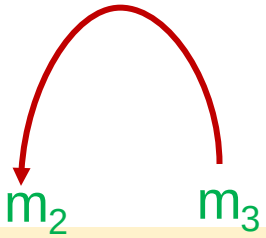
The rider moves closer.

Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

We rely on context to identify the mention-name.

Name of an ambiguous mention depends on an antecedent mention



Roland arrives as Ian waits. He stands with Mary ...

The rider moves closer.

Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

$\mathbf{R} \in \{0,1\}^{M^2}$ Antecedence variable

m_1 m_2 m_3 m_4 m_5

Roland arrives as Ian waits. He stands with Mary ...

The rider moves closer.

Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

$\mathbf{R} \in \{0,1\}^{M^2}$ Antecedence variable

Every pair of mentions has a **text feature vector**, which determines antecedence.

m_1 m_2 m_3 m_4 m_5

Roland arrives as Ian waits. He stands with Mary ...

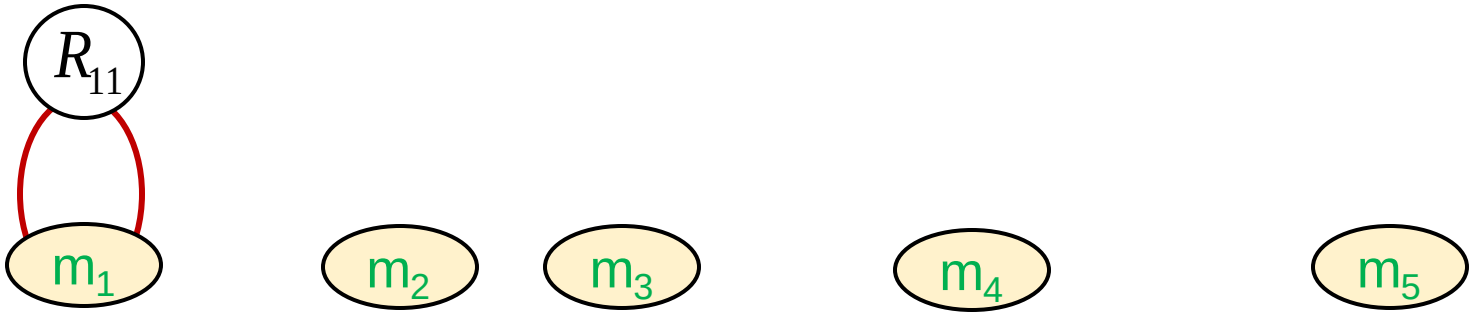
The rider moves closer.

Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

$\mathbf{R} \in \{0,1\}^{M^2}$ Antecedence variable

Every pair of mentions has a **text feature vector**, which determines antecedence.

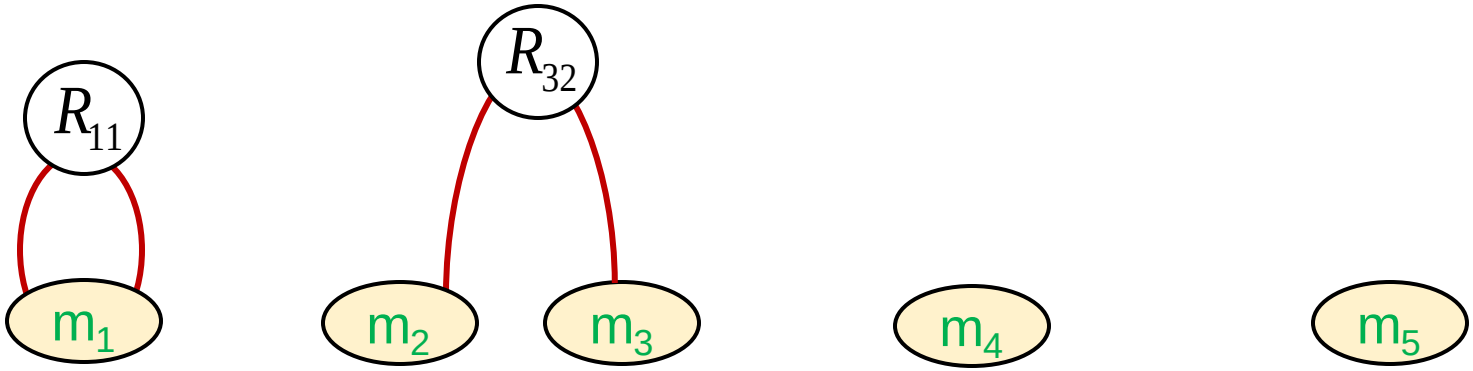


Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

$\mathbf{R} \in \{0,1\}^{M^2}$ Antecedence variable

Every pair of mentions has a **text feature vector**, which determines antecedence.



Roland arrives as Ian waits. He stands with Mary ...

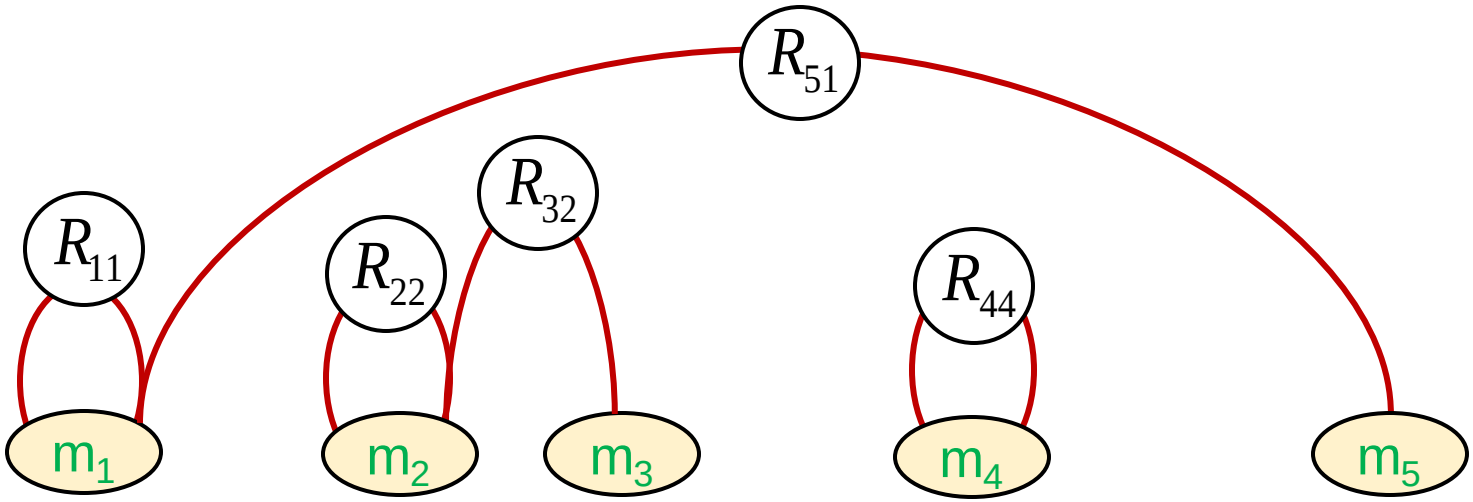
The rider moves closer.

Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

$\mathbf{R} \in \{0,1\}^{M^2}$ Antecedence variable

Every pair of mentions has a **text feature vector**, which determines antecedence.



Roland arrives as Ian waits. He stands with Mary ...

The rider moves closer.

Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

$\mathbf{R} \in \{0,1\}^{M^2}$ Antecedence variable

m_1 m_2 m_3 m_4 m_5

Roland arrives as Ian waits. He stands with Mary ...

The rider moves closer.

Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

$\mathbf{R} \in \{0,1\}^{M^2}$ Antecedence variable

Antecedent clustering cost

$$\mathbf{R}^T \Pi_{men} \mathbf{R}$$

Convex quadratic in $\mathbf{R}$

- Discriminative clustering
- Mention-pairs with **similar features** have **similar antecedence**

m_1 m_2 m_3 m_4 m_5

Roland arrives as Ian waits. He stands with Mary ...

The rider moves closer.

Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

$\mathbf{R} \in \{0,1\}^{M^2}$ Antecedence variable

Antecedent clustering cost

$$\mathbf{R}^T \Pi_{men} \mathbf{R}$$

Convex quadratic in $\mathbf{R}$

m_2 m_3

Roland arrives as Ian waits. He stands with Mary ...

The rider moves closer.

Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

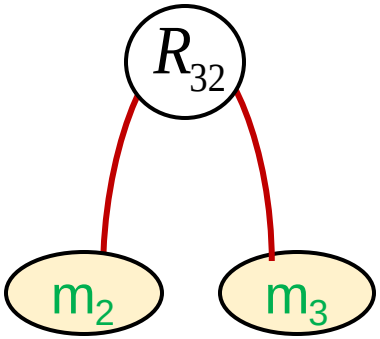
$\mathbf{R} \in \{0,1\}^{M^2}$ Antecedence variable

Antecedent clustering cost

$$\mathbf{R}^T \Pi_{men} \mathbf{R}$$

Convex quadratic in $\mathbf{R}$

R constrains **Z**



Antecedent mention shares the same name.

Roland arrives as Ian waits. He stands with Mary ...

The rider moves closer.

Our Bidirectional Model: **Mention name assignment**

$\mathbf{Z} \in \{0,1\}^{M \times P}$ Name assignment for mentions

$\mathbf{R} \in \{0,1\}^{M^2}$ Antecedence variable

Antecedent clustering cost

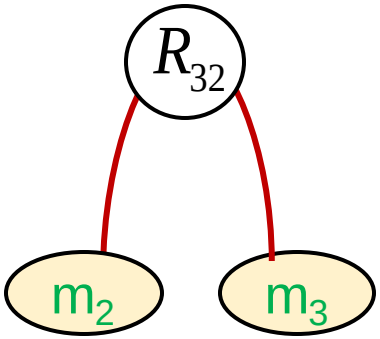
$$\mathbf{R}^T \Pi_{men} \mathbf{R}$$

Convex quadratic in $\mathbf{R}$

R constrains **Z**

$$\left\| \mathbf{Z}_i - \mathbf{Z}_j \right\|_\infty \leq \left(1 - \mathbf{R}_{ij}\right)$$

Convex constraint



Antecedent mention shares the same name.

Roland arrives as Ian waits. He stands with Mary ...

The rider moves closer.

Our Bidirectional Model

$$\min \quad \mathrm{tr}\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right) + \mathbf{R}^T \Pi_{men} \mathbf{R}$$
$$\text{s.t} \quad \mathbf{Y} \in C_Y \ , \qquad \mathbf{Z}, \mathbf{R} \in C_{Z,R}$$

Roland Ian Mary

Track s

Track name assignment

Names

Alignment

Roland Ian Ian Mary

m_1 m_2 m_3 m_4

Roland arrives as Ian waits. He stands with Mary ...

Names

Mention name assignment

Mentions

Our Bidirectional Model

$$\min \quad \mathrm{tr}\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right) + \mathbf{R}^T \Pi_{men} \mathbf{R}$$

$$\text{s.t} \quad \mathbf{Y} \in C_Y \ , \qquad \mathbf{Z}, \mathbf{R} \in C_{Z,R}$$

Roland

Ian

Mary

Roland

Ian

Ian

Mary

Tracks

Track name assignment

Names

Alignment

Names

Mention name assignment

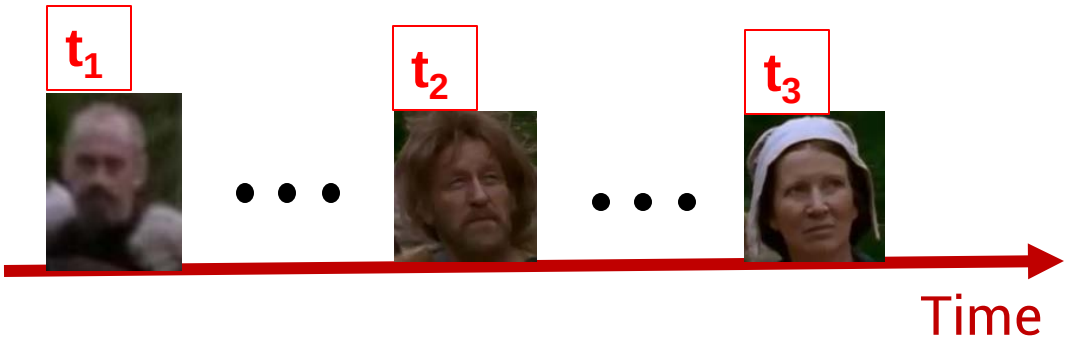
Mentions

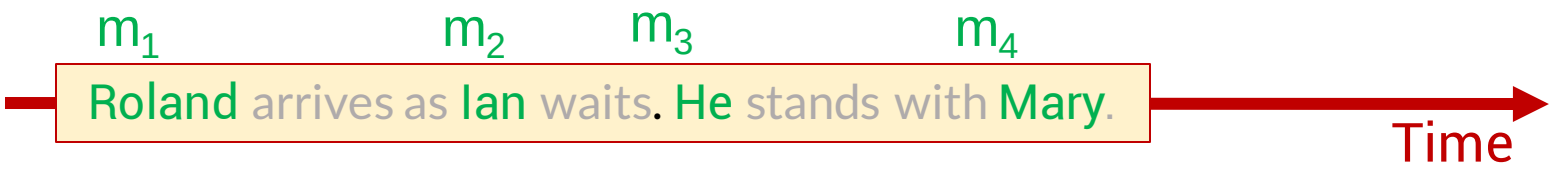
m_1 m_2 m_3 m_4

Roland arrives as Ian waits. He stands with Mary ...

Our Bidirectional Model: **Alignment**

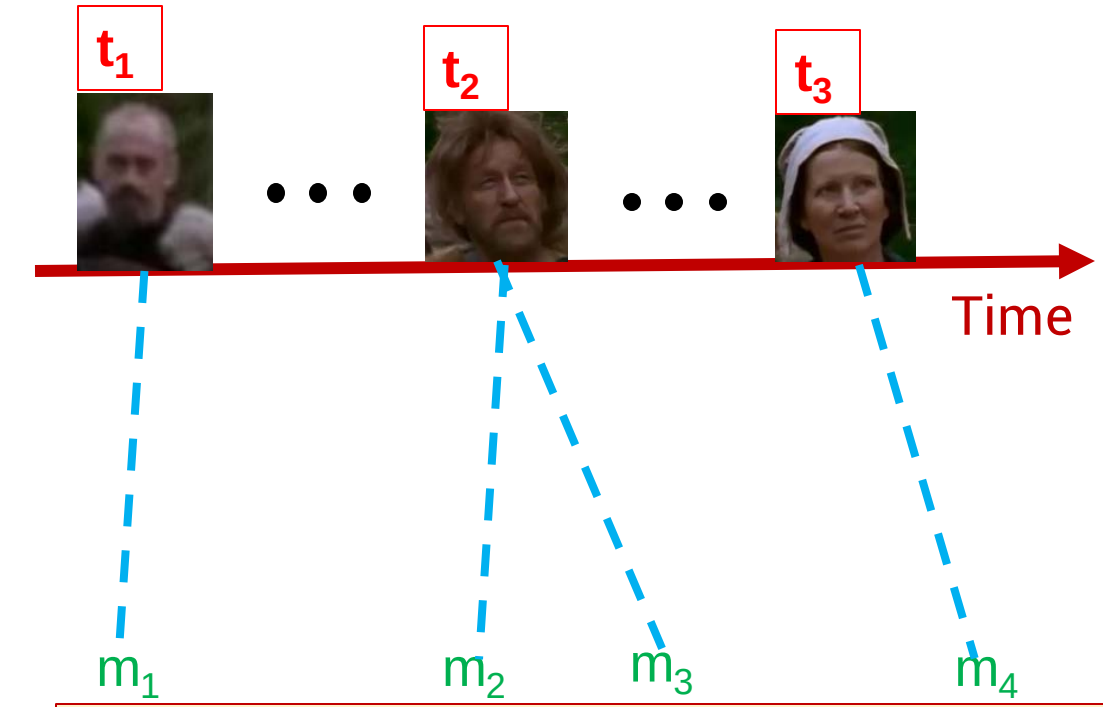
$\mathbf{A} \in \{0,1\}^{T \times M}$ Alignment matrix between tracks and mentions





Our Bidirectional Model: **Alignment**

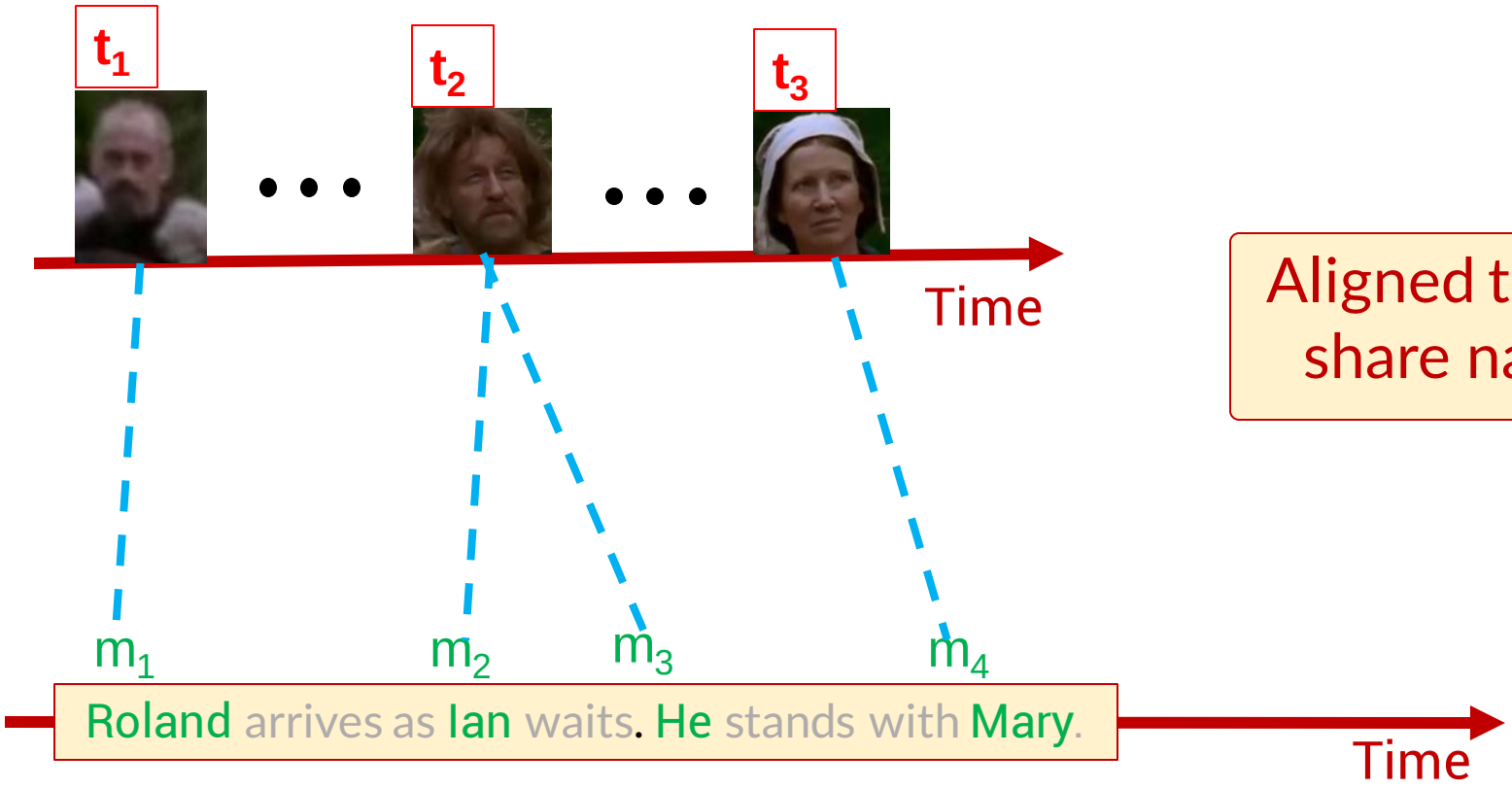
$\mathbf{A} \in \{0,1\}^{T \times M}$ Alignment matrix between tracks and mentions



	t_1	t_2	t_3
m_1	1	0	0
m_2	0	1	0
m_3	0	1	0
m_4	0	0	1

Our Bidirectional Model: **Alignment**

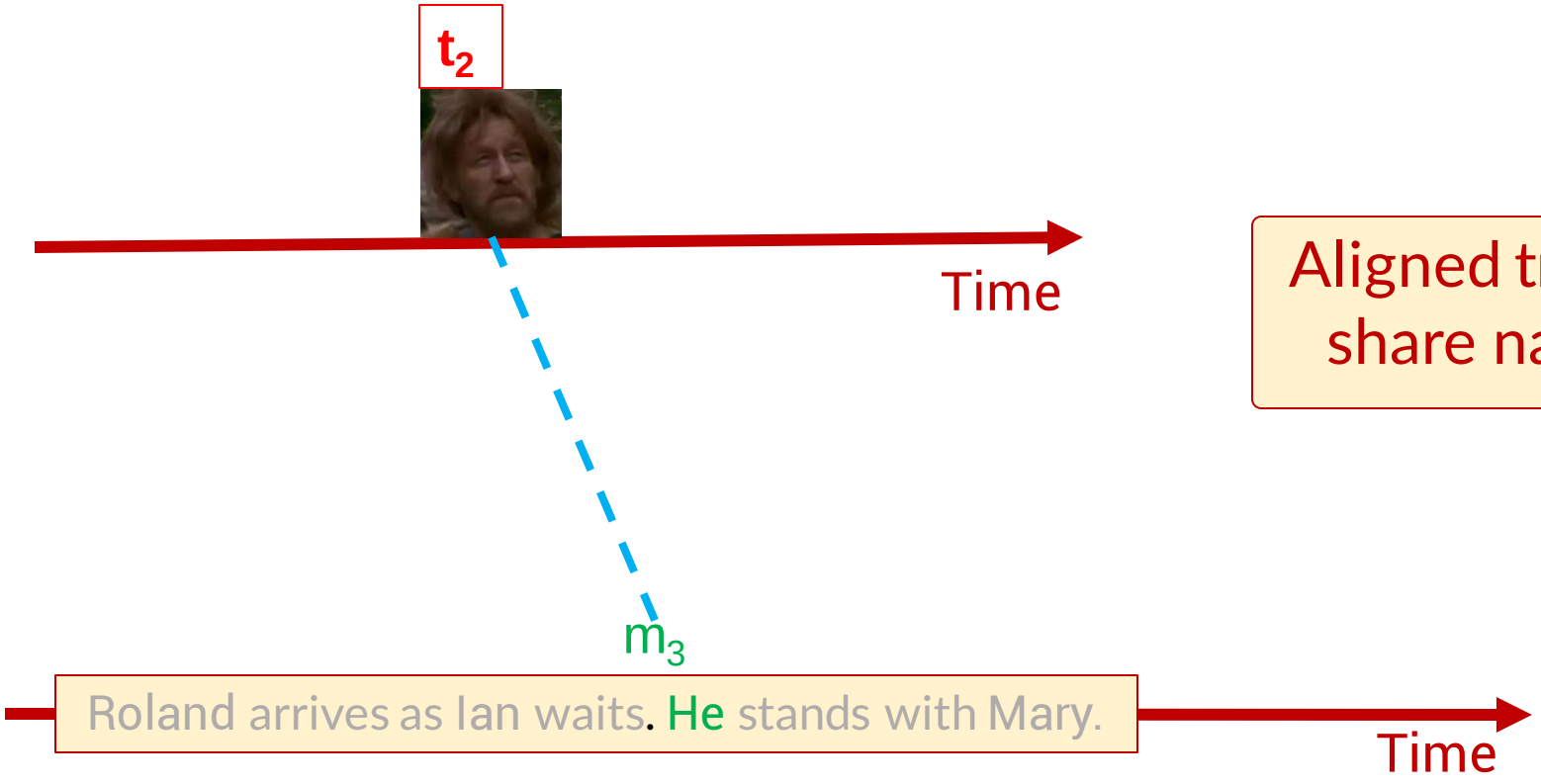
$\mathbf{A} \in \{0,1\}^{T \times M}$ Alignment matrix between tracks and mentions



Aligned tracks share name

Our Bidirectional Model: **Alignment**

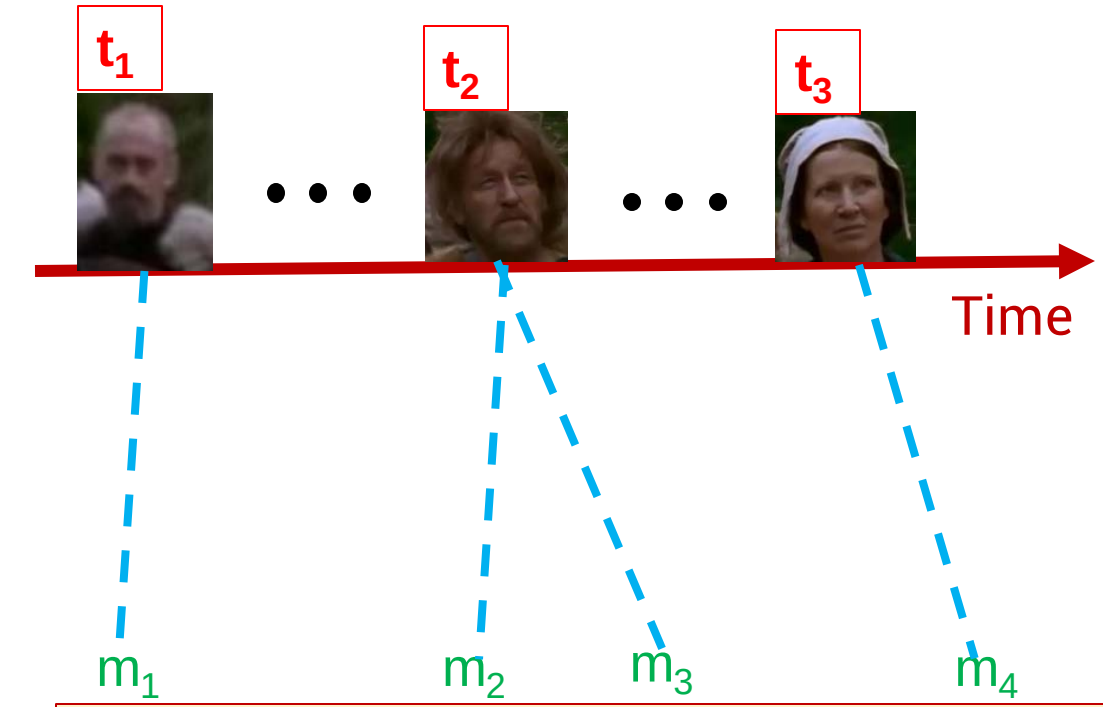
$\mathbf{A} \in \{0,1\}^{T \times M}$ Alignment matrix between tracks and mentions



Aligned tracks share name

Our Bidirectional Model: **Alignment**

$\mathbf{A} \in \{0,1\}^{T \times M}$ Alignment matrix between tracks and mentions



Alignment cost

$$\|\mathbf{Y} - \mathbf{AZ}\|_F^2$$

Our Bidirectional Model: **Alignment**

$\mathbf{A} \in \{0,1\}^{T \times M}$ Alignment matrix between tracks and mentions

Alignment cost

$$\|\mathbf{Y} - \mathbf{AZ}\|_F^2$$

Dynamic Time Warping

Dynamic program to optimize in $\mathbf{A}$ (enforces monotonic alignment)

Our Bidirectional Model

$$\min \quad \mathrm{tr}\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right) + \mathbf{R}^T \Pi_{men} \mathbf{R} + \left\|\mathbf{Y} - \mathbf{AZ}\right\|_F^2$$

$$\text{s.t} \quad \mathbf{Y} \in C_Y \ , \qquad \mathbf{Z}, \mathbf{R} \in C_{Z,R} \ , \qquad \mathbf{A} \in C_A$$

Roland | Ian | Mary

Roland | Ian | Ian | Mary

Alignment

Names

Names

Track name assignment

Tracks

Mention name assignment

Mentions

m_1 m_2 m_3 m_4

Roland arrives as Ian waits. He stands with Mary ...

Optimization

$$\min \quad \mathrm{tr}\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right) + \mathbf{R}^T \Pi_{men} \mathbf{R} + \left\|\mathbf{Y} - \mathbf{AZ}\right\|_F^2$$

$$\text{s.t} \quad \mathbf{Y} \in C_Y \ , \qquad \mathbf{Z}, \mathbf{R} \in C_{Z,R} \ , \quad \mathbf{A} \in C_A$$

- Relax $\mathbf{Y}, \mathbf{Z}, \mathbf{R}$ to be in $[0,1]$
- Use Block Coordinate Descent

Optimization

$$\min \quad \mathrm{tr}\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right) + \mathbf{R}^T \Pi_{men} \mathbf{R} + \left\|\mathbf{Y} - \mathbf{A}\mathbf{Z}\right\|_F^2$$

$$\text{s.t} \quad \mathbf{Y} \in C_Y \;, \quad \mathbf{Z}, \mathbf{R} \in C_{Z,R} \;, \quad \mathbf{A} \in C_A$$

- Relax $\mathbf{Y}, \mathbf{Z}, \mathbf{R}$ to be in $[0,1]$
- Use Block Coordinate Descent
- Quadratic Program to optimize $\mathbf{Y}$

Optimization

$$\min \quad \mathrm{tr}\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right) + \mathbf{R}^T \Pi_{men} \mathbf{R} + \left\|\mathbf{Y} - \mathbf{A}\mathbf{Z}\right\|_F^2$$

$$\text{s.t} \quad \mathbf{Y} \in C_Y \;, \quad \mathbf{Z}, \mathbf{R} \in C_{Z,R} \;, \quad \mathbf{A} \in C_A$$

- Relax $\mathbf{Y}, \mathbf{Z}, \mathbf{R}$ to be in $[0,1]$
- Use Block Coordinate Descent
- Quadratic Program to optimize $\mathbf{Y}$
- Quadratic Program to optimize $\mathbf{Z}, \mathbf{R}$

Optimization

$$\min \quad \mathrm{tr}\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right) + \mathbf{R}^T \Pi_{men} \mathbf{R} + \left\|\mathbf{Y} - \mathbf{AZ}\right\|_F^2$$

$$\text{s.t} \quad \mathbf{Y} \in C_Y \;, \quad \mathbf{Z}, \mathbf{R} \in C_{Z,R} \;, \quad \mathbf{A} \in C_A$$

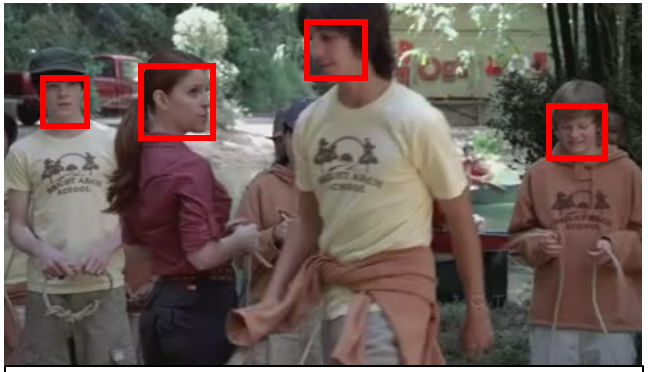
- Relax $\mathbf{Y}, \mathbf{Z}, \mathbf{R}$ to be in $[0,1]$
- Use Block Coordinate Descent
- Quadratic Program to optimize $\mathbf{Y}$
- Quadratic Program to optimize $\mathbf{Z}, \mathbf{R}$
- Dynamic Program to optimize in terms of $\mathbf{A}$

Linking people with “their” names

- Problem setup
- Our Bidirectional model
- Experiments
- Summary

Dataset

We reveal Lynette holding Porter by his feet, while he clings to Preston's desk.



Missy points to the larger kid. The big kid walks off. Other kids jeer.

Cary eyes the siblings, as Alicia looks across the bullpen

Dataset

We reveal Lynette holding Porter by his feet, while he clings to Preston's desk.

Missy points to the larger kid. The big kid walks off. Other kids jeer.

Cary eyes the siblings, as Alicia looks across the bullpen

pronoun/nominal

Dev. Set (14 episodes)	**3329** tracks (3 eps.)	**811** mentions

Dataset

We reveal Lynette holding Porter by his feet, while he clings to Preston's desk.

Missy points to the larger kid. The big kid walks off. Other kids jeer.

Cary eyes the siblings, as Alicia looks across the bullpen

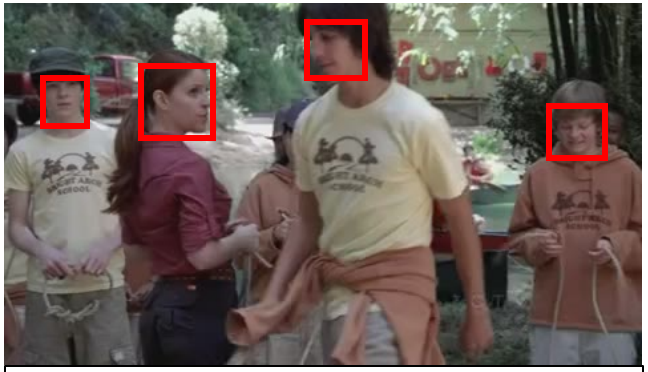
pronoun/nominal

Dev. Set (14 episodes)	**3329** tracks (3 eps.)	**811** mentions

- **No training set** since we have no labelled examples.
- The model **parameters** such as regularization constant are tuned on the **dev. set**.

Dataset

We reveal Lynette holding Porter by his feet, while he clings to Preston's desk.



Missy points to the larger kid. The big kid walks off. Other kids jeer.

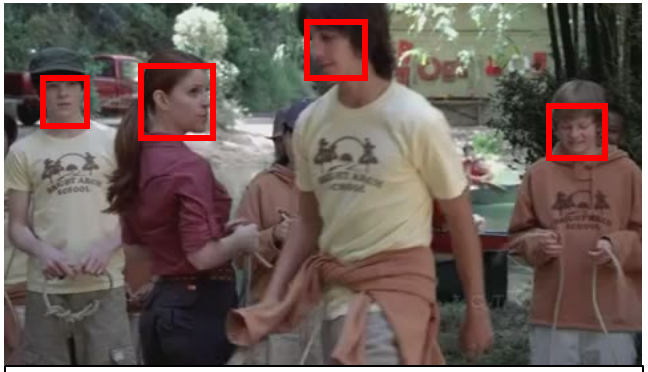
Cary eyes the siblings, as Alicia looks across the bullpen

pronoun/nominal

Dev. Set (14 episodes)	**3329** tracks (3 eps.)	**811** mentions
Test Set (5 episodes)	**4757** tracks	**300** mentions

Dataset

We reveal Lynette holding Porter by his feet, while he clings to Preston's desk.



Missy points to the larger kid. The big kid walks off. Other kids jeer.

Cary eyes the siblings, as Alicia looks across the bullpen

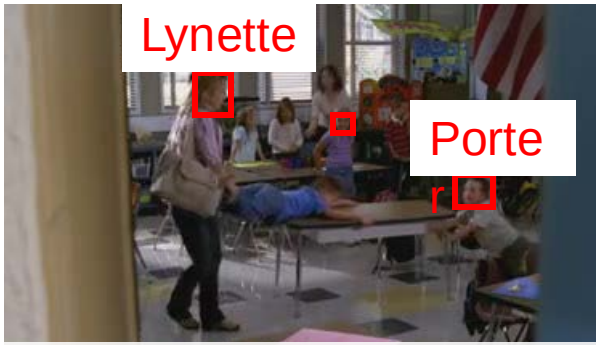
pronoun/nominal

Dev. Set (14 episodes)	**3329** tracks (3 eps.)	**811** mentions
Test Set (5 episodes)	**4757** tracks	**300** mentions

Tasks:

- Name assignment to tracks;
- Name assignment to mentions;

Dataset



We reveal Lynette holding Porter by his feet, while he clings to Preston's desk.

Missy points to the larger kid. The big kid walks off. Other kids jeer.

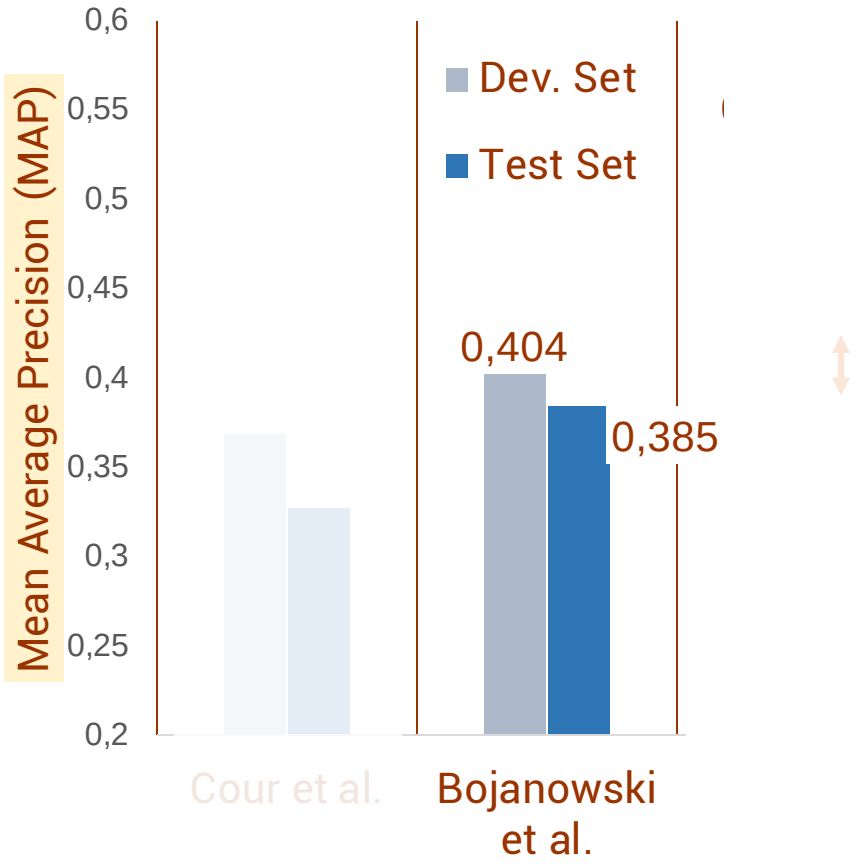
Cary eyes the siblings, as Alicia looks across the bullpen

		pronoun/nominal
Dev. Set (14 episodes)	**3329** tracks (3 eps.)	**811** mentions
Test Set (5 episodes)	**4757** tracks	**300** mentions

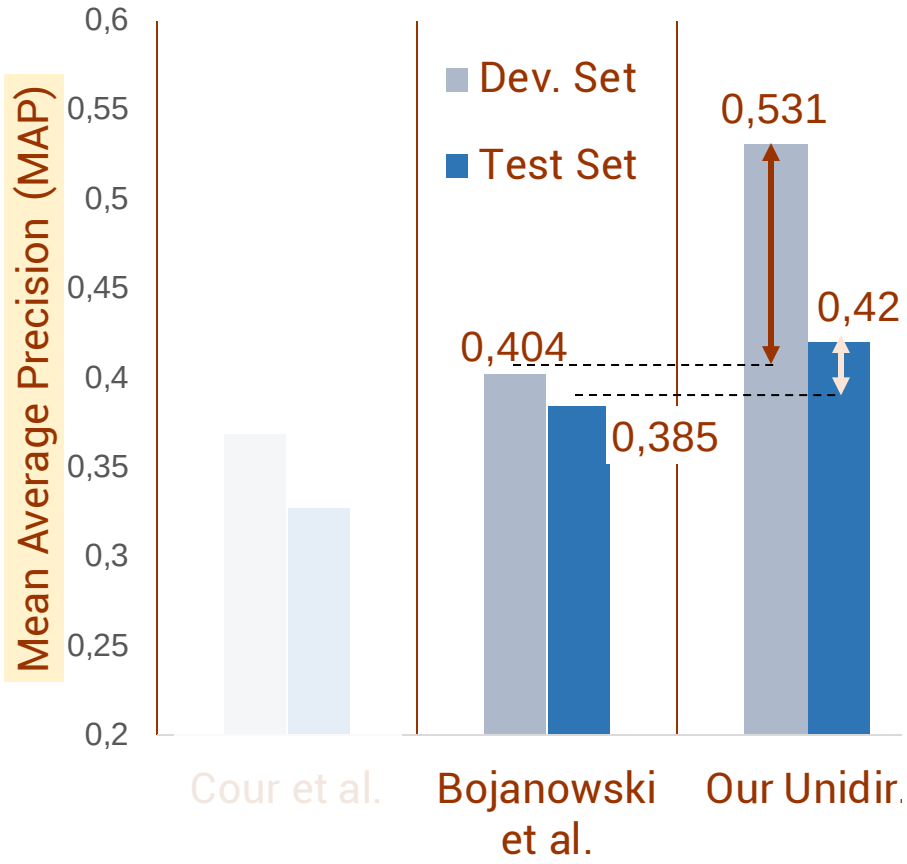
Tasks:

- **Name assignment to tracks;**
- Name assignment to mentions;

Experiments: Name Assignment to Tracks



Experiments: Name Assignment to Tracks



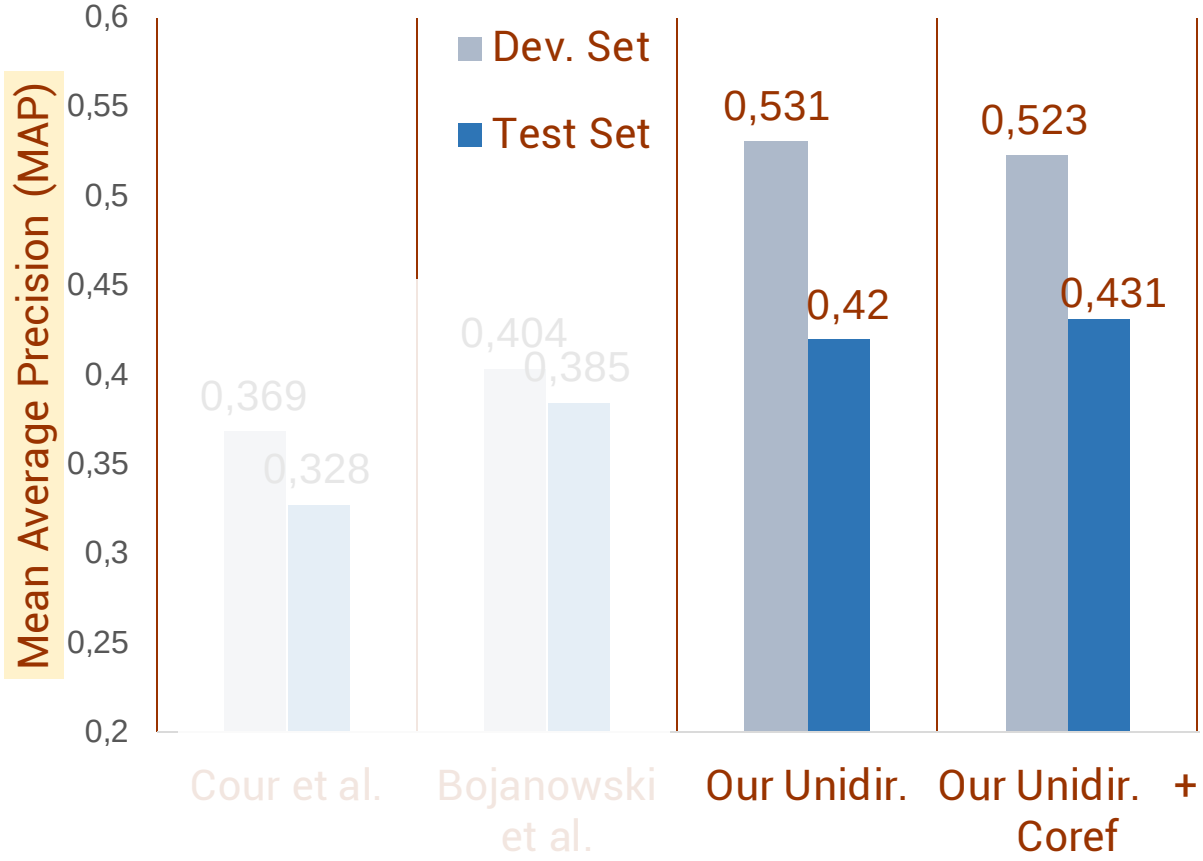
Our Unidirectional Model

Bojanowski et al (ICCV'13) with **Additional Constraints**

$$\min \quad \mathrm{tr}\left(\mathbf{Y}^T \Pi_{track} \mathbf{Y}\right)$$

$$\text{s.t} \quad \mathbf{Y} \in C_Y$$

Experiments: Name Assignment to Tracks

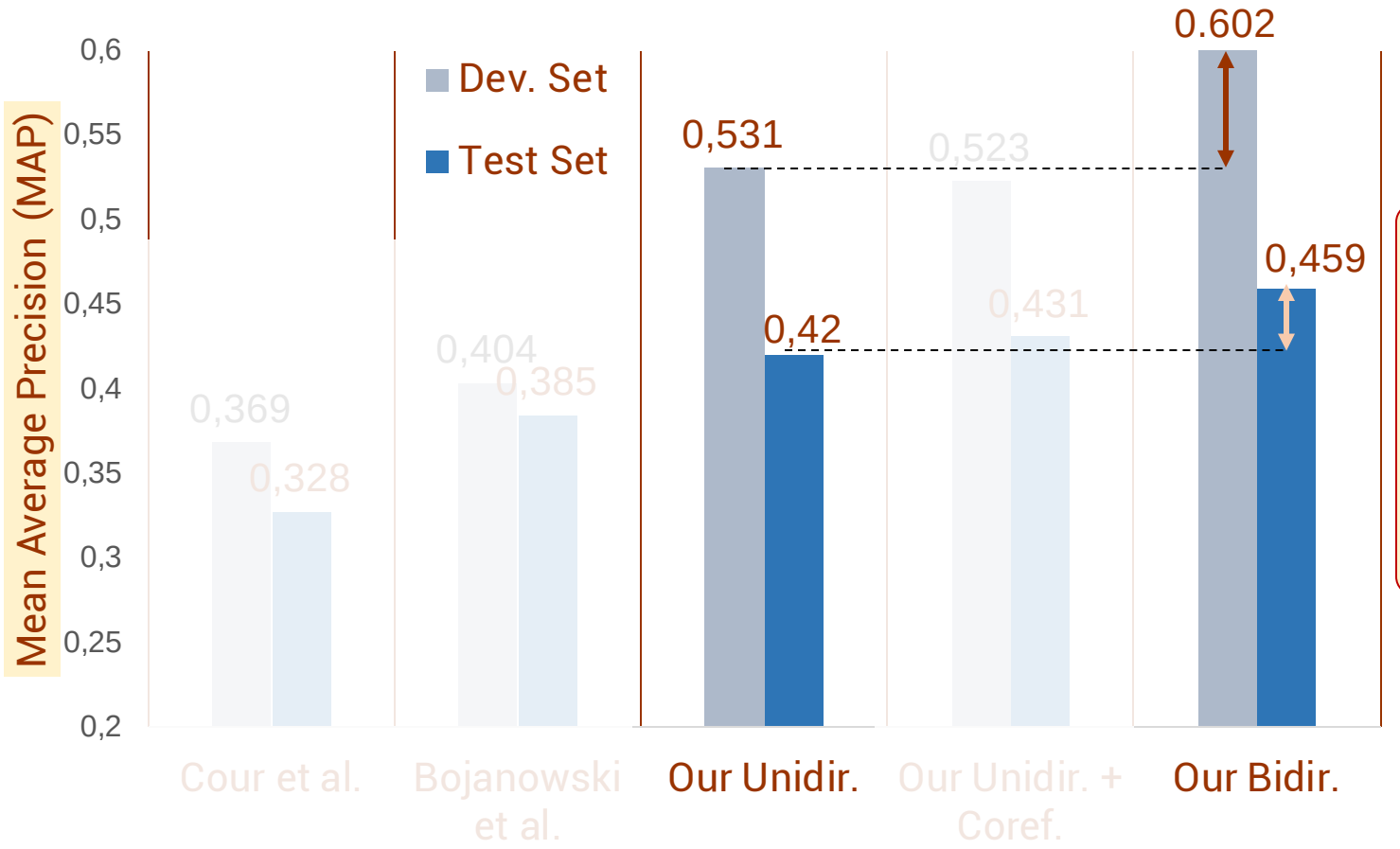


No Alignment

Unidirectional Track Name
+
Unidirectional Mention Name

without Alignment

Experiments: Name Assignment to Tracks



Bidirectional Model

Full Model with alignment

Dataset

We reveal Lynette **Porter** Porter by his feet, while **he** clings to Preston's desk.

Bryan points to the larger kid. The **big kid** walks off. Other kids jeer.

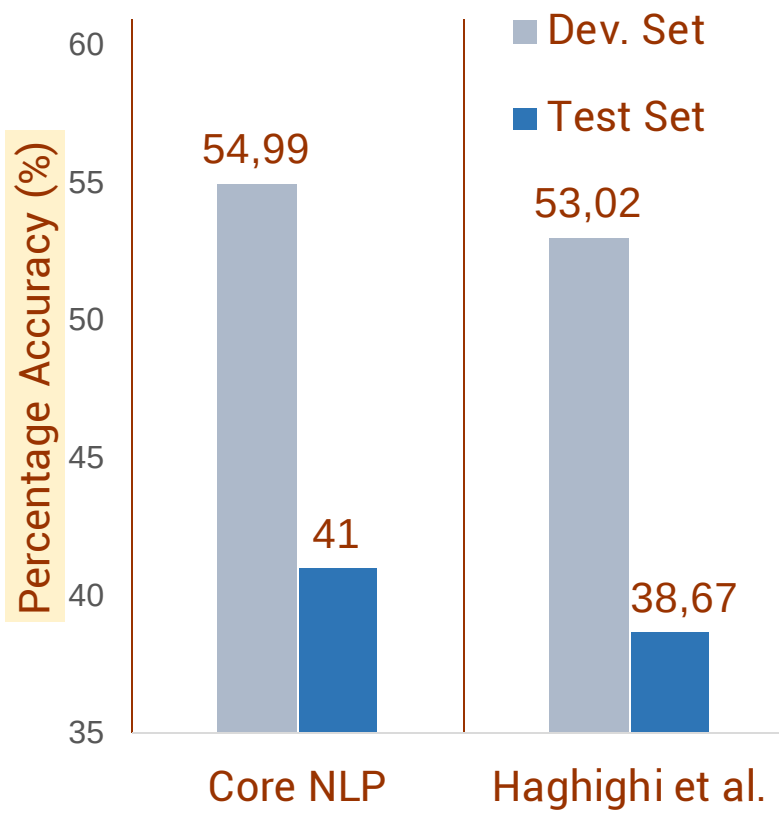
Cary eyes the siblings, as Alicia looks across the bullpen

		pronoun/nominal
Dev. Set (14 episodes)	**3329** tracks (3 eps.)	**811** mentions
Test Set (5 episodes)	**4757** tracks	**300** mentions

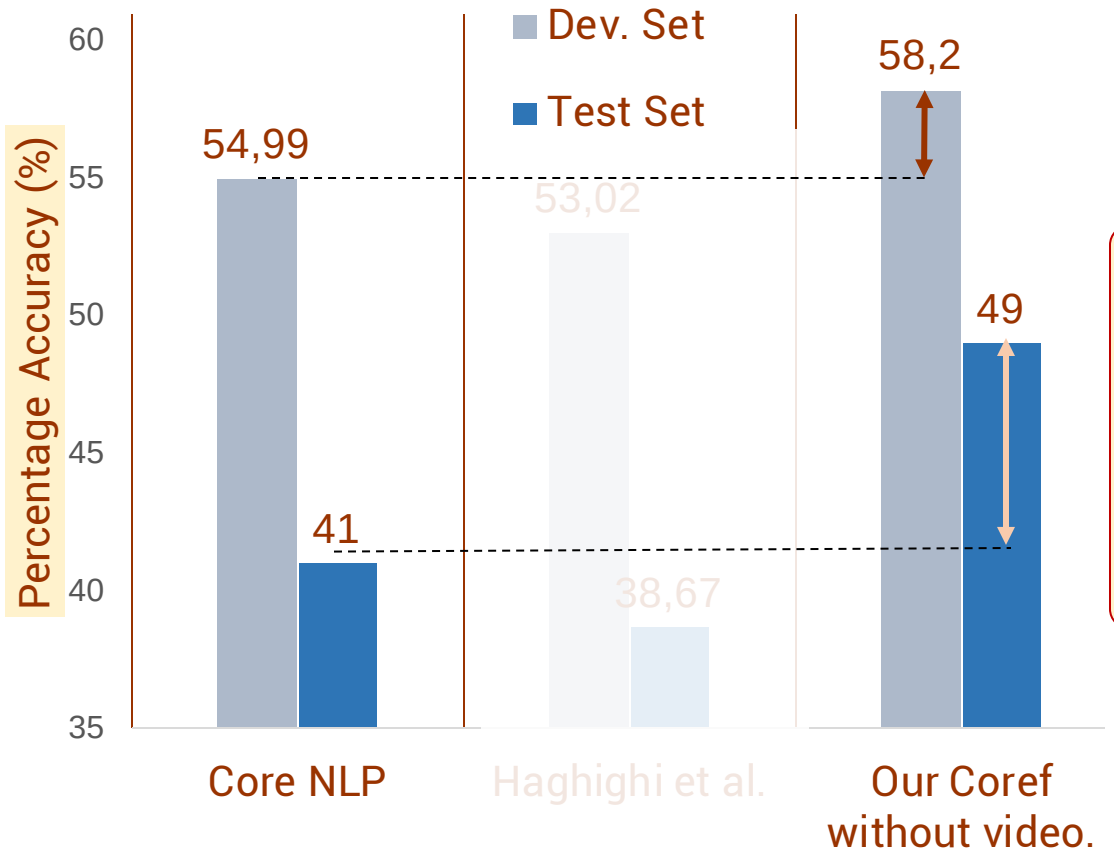
Tasks:

- Name assignment to tracks;
- **Name assignment to mentions;**

Experiments: Name Assignment to Mentions



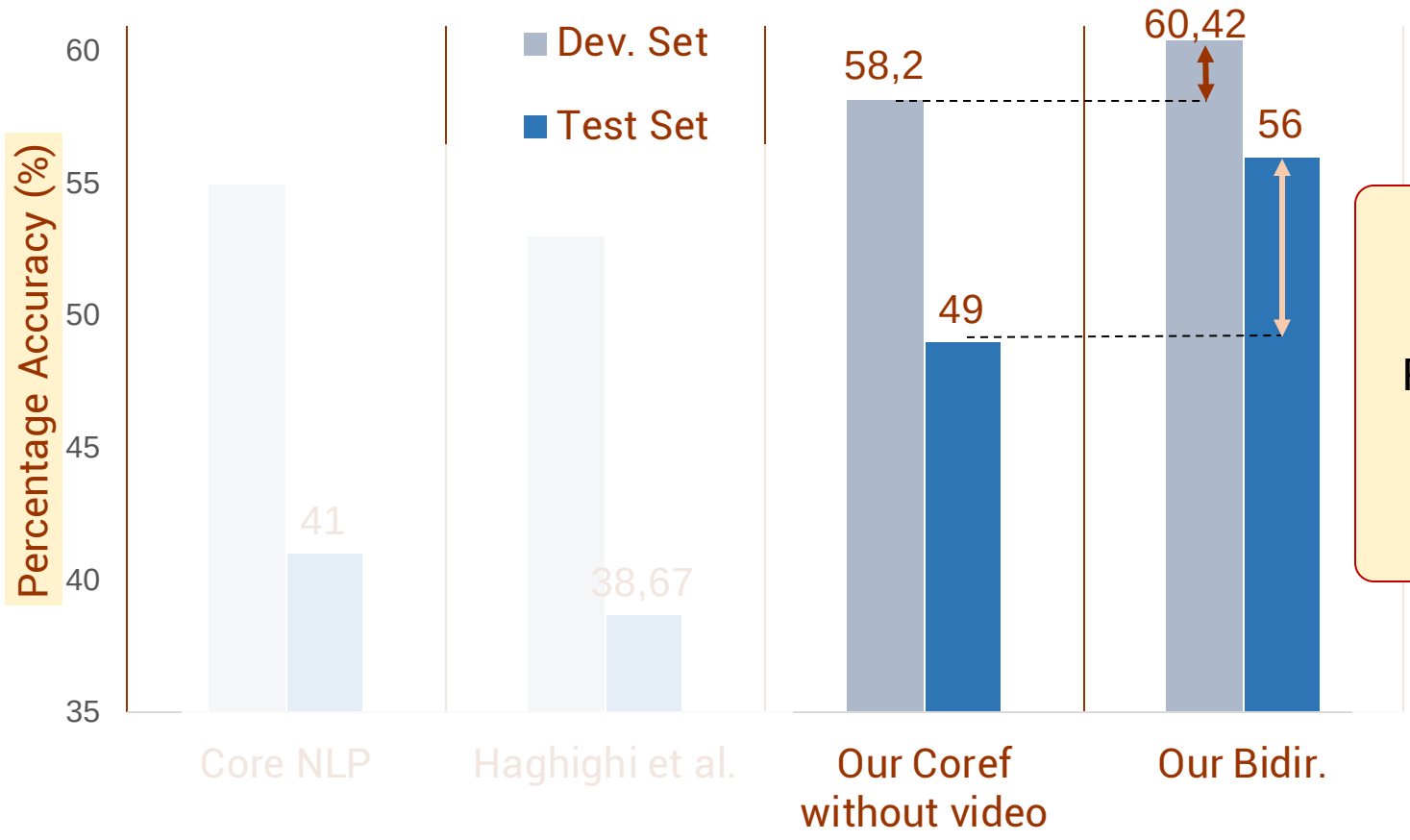
Experiments: Name Assignment to Mentions



No Video information used

Additional constraints specific to script data

Experiments: Name Assignment to Mentions



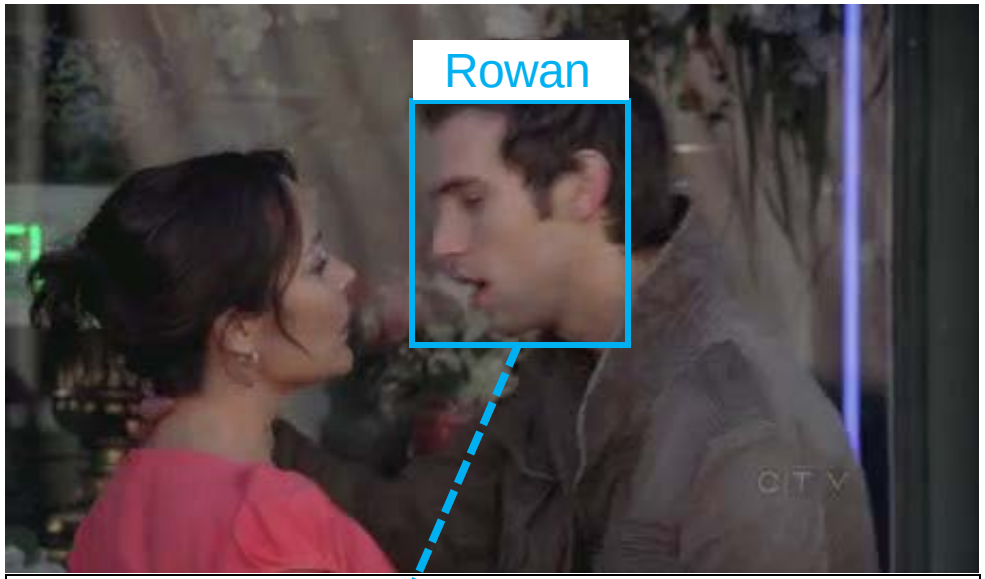
Bidirectional Model

Full Model with alignment
+
track name

Experiments: Benefit of bidirectional model

Experiments: Benefit of bidirectional model

Experiments: Benefit of bidirectional model



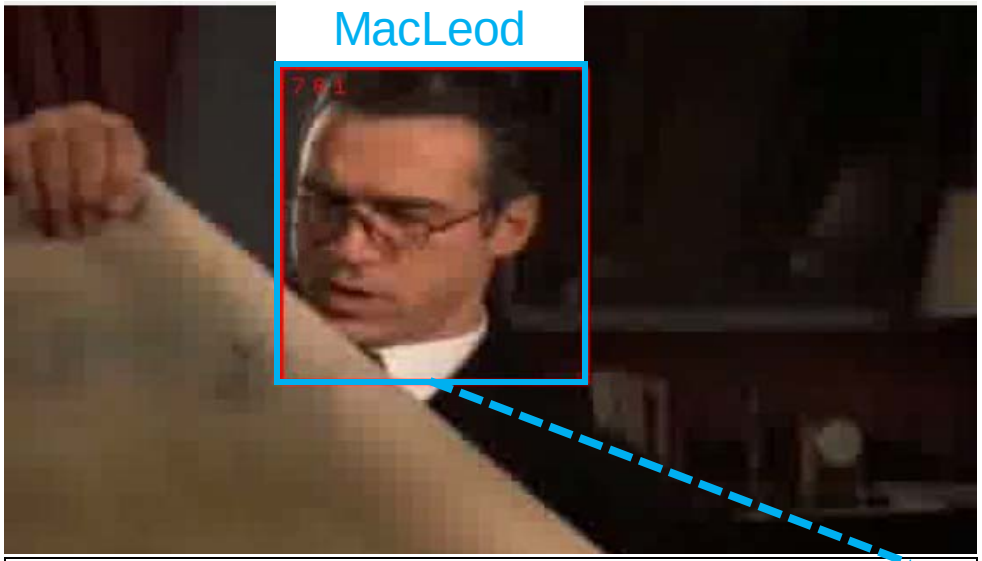
Gabriel cues the entry of Rowan. Rose doesn't notice him. **He** takes her in his arms.

Gabriel(unidir), Rowan(bidir)

Experiments: Benefit of bidirectional model

Edouard & MacLeod unfurl the canvas, searching for the name. He then peers at it. She turns

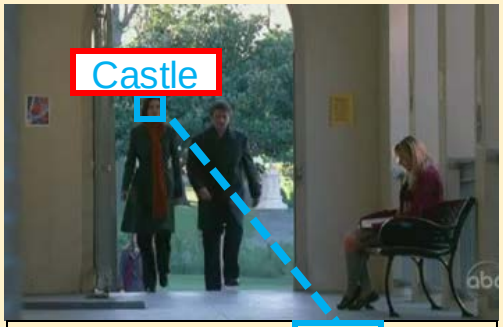
Experiments: Benefit of bidirectional model



Edouard & MacLeod unfurl the canvas. **He** then peers at it. She turns.

Edouard(unidir), MacLeod(bidir)

Error Analysis: Name Assignment to Mentions

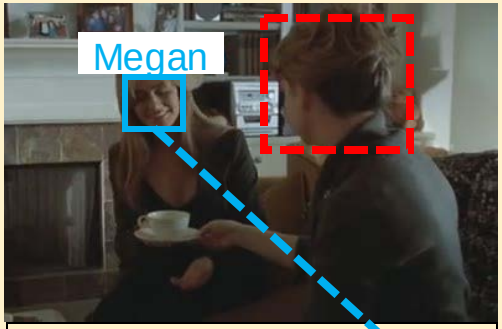


Low resolution faces lead to bad name assignment!

Error Analysis: Name Assignment to Mentions

Beckett turns… **She** bites her lips and shakes her head

Beckett(unidir), Castle(bidir)



Elaine Tillman, fragile but with inner strength. **She** looks to Megan.

Elaine(unidir), Megan(bidir)

Faces not detected or face not visible!

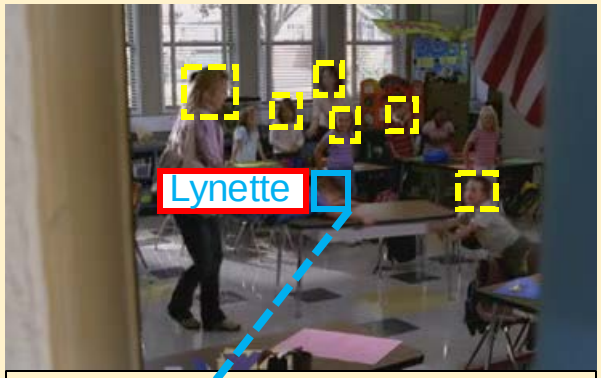
Error Analysis: Name Assignment to Mentions

Beckett turns… **She** bites her lips and shakes her head

Beckett(unidir), Castle(bidir)

Elaine Tillman, fragile but with inner strength. **She** looks to Megan.

Elaine(unidir), Megan(bidir)



Lynette holds Porter by feet, while **he** clings to the desk.

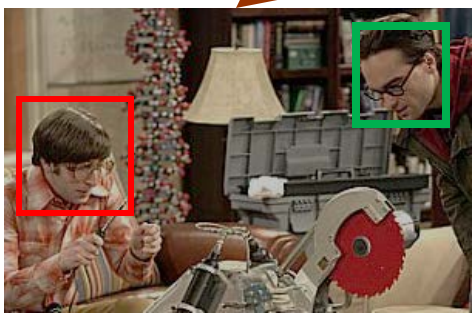
Long clip with too many people, leads to wrong alignment

Linking people with “their” names

- Problem setup
- Our Bidirectional model
- Experiments
- Summary

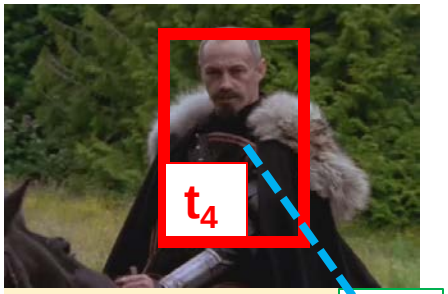
Summary

Bidirectional model to handle challenging problems in Vision and NLP



Leonard looks at the robot, while the engineer

Temporal ordering based alignment



The foreigner m_4

Model evaluated on new TV episode Dataset

Cary eyes the siblings, as Alicia looks across the bullpen

Thank You

Microsoft® Research

NIST National Institute of Standards and Technology

Google Research

intel®

Mind's Eye

DARPA