Two Heads Are Better Than One:
Improving Fake News Video Detection by Correlating with Neighbors

Peng Qi[1,2], Yuyang Zhao[3], Yufeng Shen[2], Wei Ji[3], Juan Cao[1,2*] and Tat-Seng Chua[3]
[1] Key Laboratory of Intelligent Information Processing,
Institute of Computing Technology, Chinese Academy of Sciences
[2] University of Chinese Academy of Sciences
[3] National University of Singapore
{qipeng,caojuan}@ict.ac.cn, yuyang.zhao@u.nus.edu,
shenyufeng22@mails.ucas.ac.cn, {jiwei,dcscts}@nus.edu.sg

ACL2023
code:https://github.com/ICTMCG/NEED

Reported by Xiaoke Li

Introduction

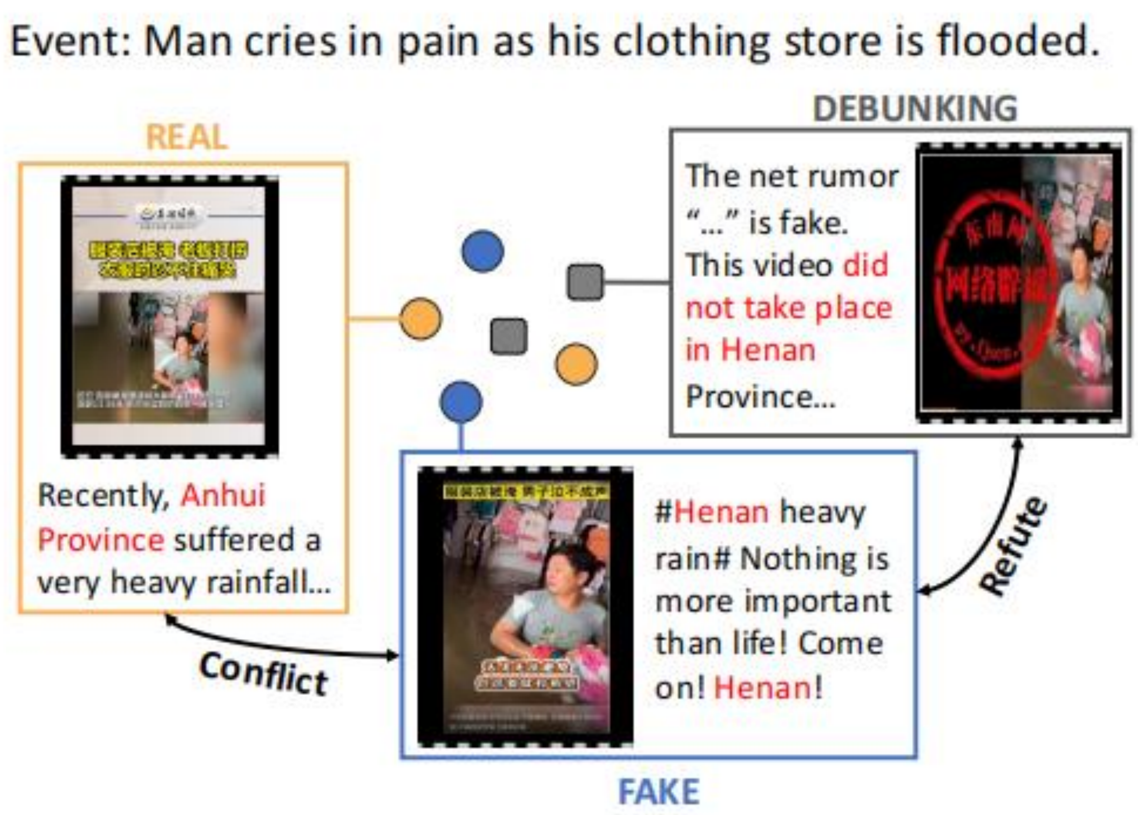
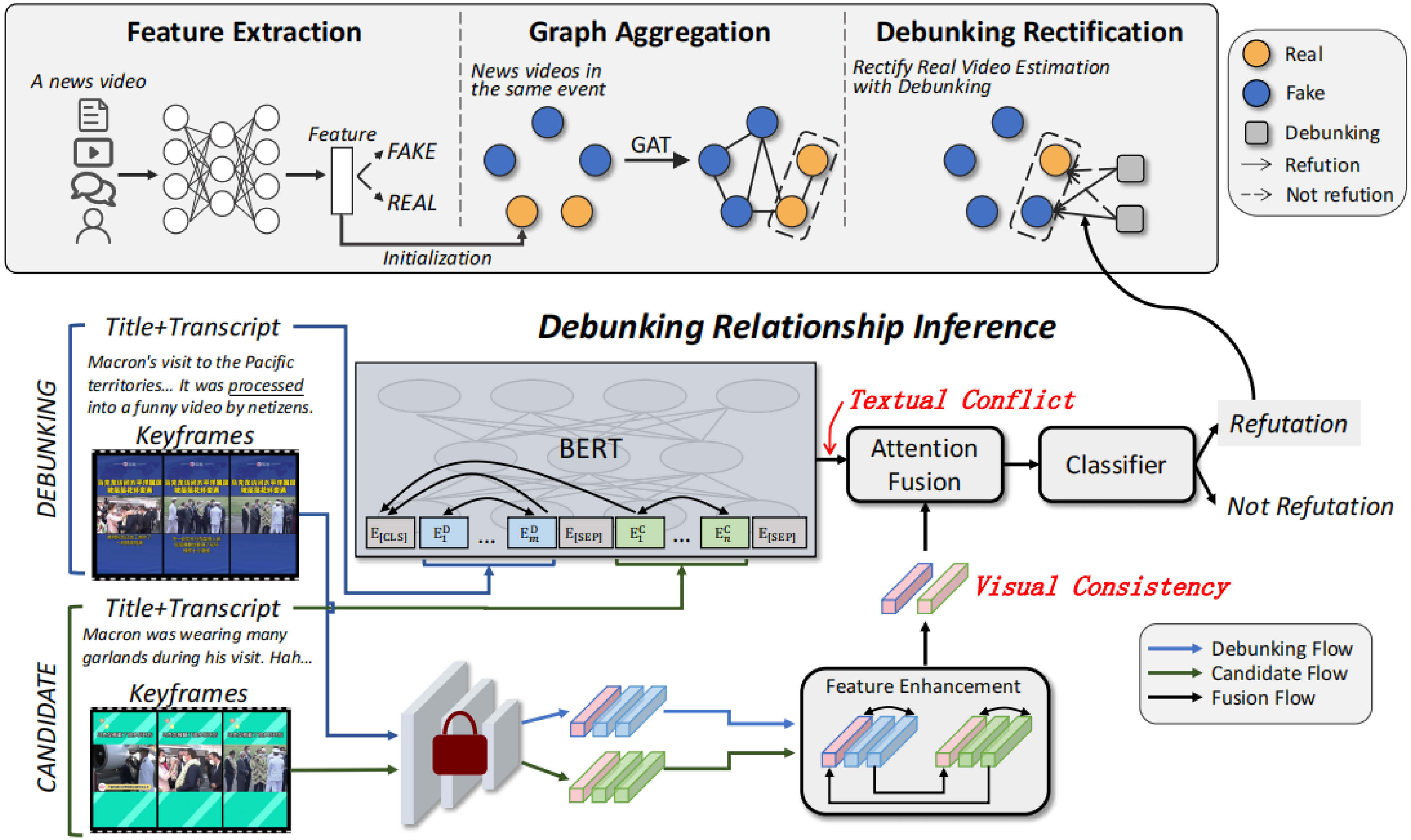
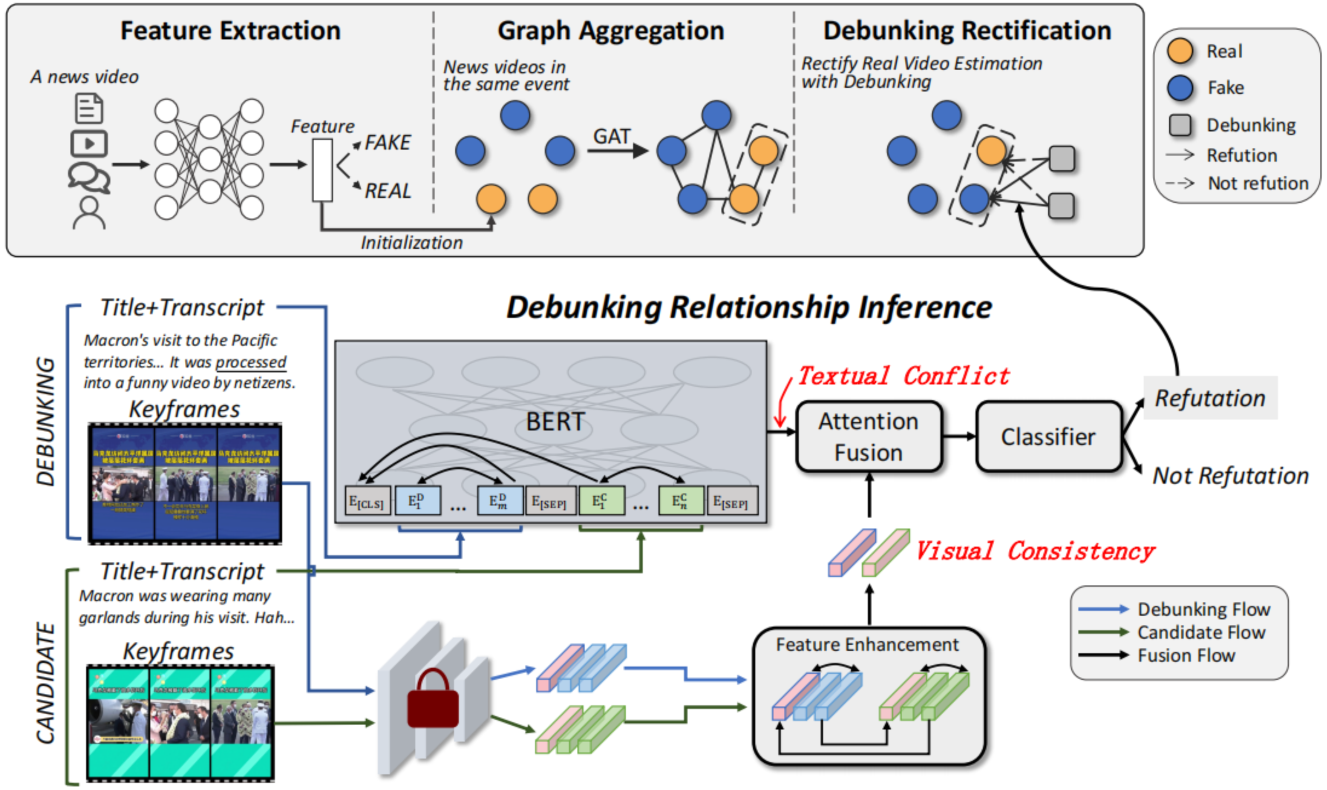


Figure 1: A set of videos belonging to the same event. Fake news videos contain conflicting information with the real ones, and the debunking videos can refute the mismatched information in the fake news videos.

Method


Feature Extraction
A news video
Feature
FAKE
REAL
Initialization
Graph Aggregation
News videos in the same event
GAT
Debunking Rectification
Rectify Real Video Estimation with Debunking
Real
Fake
Debunking
Refution
Not refution
Debunking Relationship Inference
DEBUNKING
Title+Transcript
Macron's visit to the Pacific territories... It was processed into a funny video by netizens.
Keyframes
CANDIDATE
Title+Transcript
Macron was wearing many garlands during his visit. Hah...
Keyframes
BERT
Textual Conflict
Attention Fusion
Classifier
Refutation
Not Refutation
Visual Consistency
Feature Enhancement
Debunking Flow
Candidate Flow
Fusion Flow

Method



$$e_{ij} = \mathrm{LeakyReLU}(\boldsymbol{a}^\top[\boldsymbol{W}\boldsymbol{v}_i, \boldsymbol{W}\boldsymbol{v}_j]),$$

$$\alpha_{ij} = \mathrm{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k\in\mathcal{N}_i}\exp(e_{ik})}, \quad (1)$$

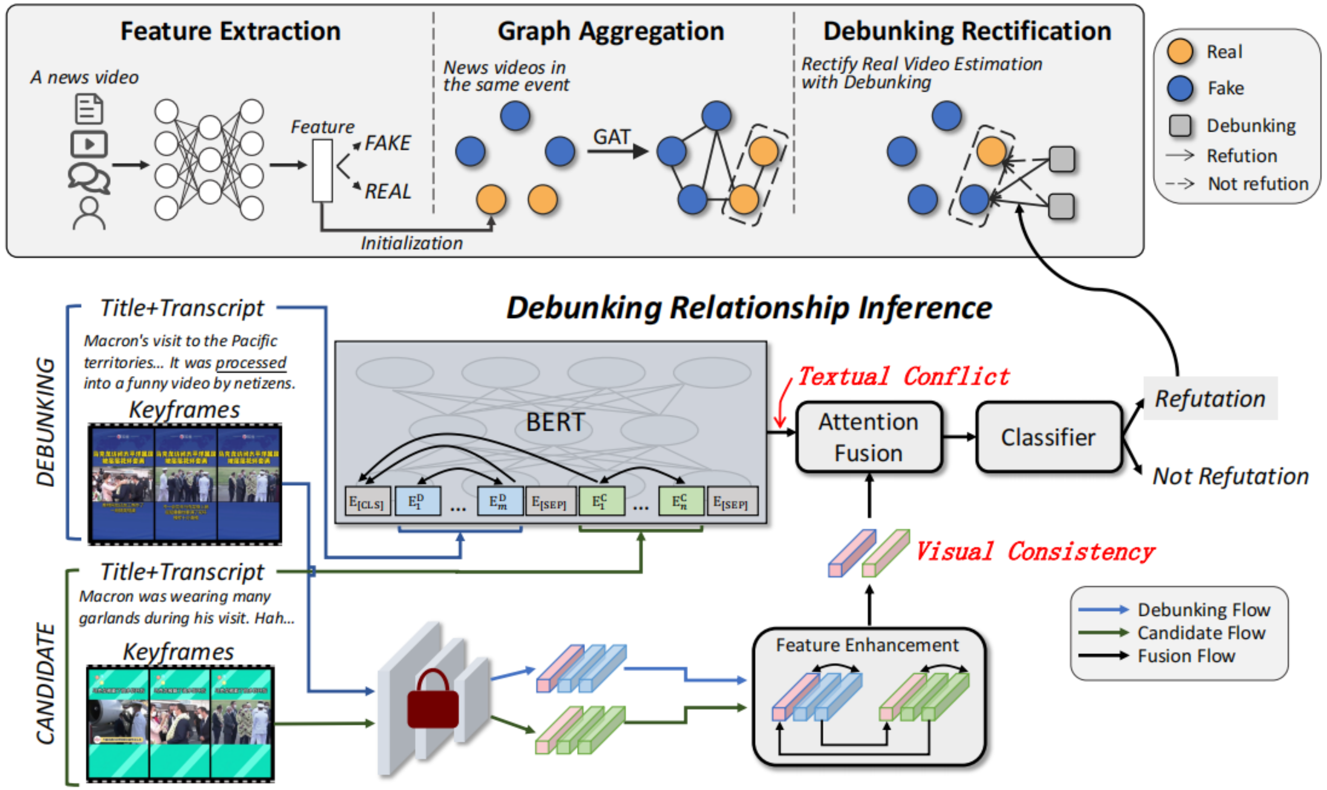
$$\hat{\boldsymbol{v}}_i = \sigma\Big(\sum_{j\in\mathcal{N}_i}\alpha_{ij}\boldsymbol{W}\boldsymbol{v}_j\Big), \quad (2)$$

$$\mathcal{L} = -[(1-y)\log(1-p_{\mathrm{GA}}) + y\log p_{\mathrm{GA}}], \quad (3)$$

where p_{GA} is the predicted probability and $y \in \{0,1\}$ denotes the ground-truth label.

$$p^i = \max\{p^i_{\mathrm{GA}}, p^i_{\mathrm{DR}}\},$$
$$p^i_{\mathrm{DR}} = \max_{\eta^j_{\mathrm{D}}\in I_{\mathrm{D}}} \mathtt{DRI}(\eta^i_{\mathrm{C}}, \eta^j_{\mathrm{D}}). \quad (4)$$

Method



$$\boldsymbol{x}_{\mathrm{t}} = \mathrm{BERT}([\mathrm{CLS}]S_{\mathrm{D}}[\mathrm{SEP}]S_{\mathrm{C}}[\mathrm{SEP}]). \quad (5)$$

$$\hat{\boldsymbol{F}}_{\mathrm{D}} = [\boldsymbol{f}_{\mathrm{D}}^{[\mathrm{CLS}]}, \boldsymbol{f}_{\mathrm{D}}^{1}, \ldots, \boldsymbol{f}_{\mathrm{D}}^{l}] + \boldsymbol{f}_{\mathrm{tem}},$$
$$\hat{\boldsymbol{F}}_{\mathrm{C}} = [\boldsymbol{f}_{\mathrm{C}}^{[\mathrm{CLS}]}, \boldsymbol{f}_{\mathrm{C}}^{1}, \ldots, \boldsymbol{f}_{\mathrm{C}}^{k}] + \boldsymbol{f}_{\mathrm{tem}}. \quad (6)$$

$$\boldsymbol{x}_{\mathrm{v}} = [\boldsymbol{f}_{\mathrm{D}}^{[\mathrm{CLS}]}, \boldsymbol{f}_{\mathrm{C}}^{[\mathrm{CLS}]}]. \quad (7)$$

Experiments

Table 1: Statistics on the number of news videos in each event.

	#Fake	#Real	#Debunking	All
Avg.	3	3	3	8
Min.	0	0	0	1
Max.	24	21	20	25

Experiments

Table 2: Performance (%) comparison of base models with and without NEED. The better result in each group using the same base model are in **boldface**, and the absolute gain is calculated. We report the mean and standard deviation of the five-fold cross-validation.

	Method	Acc.		F1		Prec.		Recall	
single-modal	BERT	$77.05_{\pm 3.24}$	–	$77.02_{\pm 3.27}$	–	$77.21_{\pm 3.12}$	–	$77.07_{\pm 3.20}$	–
	+ NEED	**82.99**$_{\pm 3.86}$	5.94↑	**82.96**$_{\pm 3.87}$	5.94↑	**83.19**$_{\pm 3.87}$	5.98↑	**82.99**$_{\pm 3.88}$	5.92↑
	Faster R-CNN +Att	$70.19_{\pm 2.70}$	–	$70.00_{\pm 2.68}$	–	$70.68_{\pm 2.89}$	–	$70.15_{\pm 2.69}$	–
	+ NEED	**78.48**$_{\pm 3.30}$	8.29↑	**78.45**$_{\pm 3.28}$	8.45↑	**78.71**$_{\pm 3.45}$	8.03↑	**78.50**$_{\pm 3.28}$	8.35↑
	VGGish	$66.91_{\pm 1.33}$	–	$66.82_{\pm 1.30}$	–	$67.07_{\pm 1.41}$	–	$66.89_{\pm 1.32}$	–
	+ NEED	**75.25**$_{\pm 1.61}$	8.34↑	**75.12**$_{\pm 1.63}$	8.30↑	**75.73**$_{\pm 1.67}$	8.66↑	**75.22**$_{\pm 1.61}$	8.33↑
	Wu et al. (2022)	$77.10_{\pm 2.04}$	–	$74.71_{\pm 2.13}$	–	$76.43_{\pm 2.16}$	–	$73.98_{\pm 2.05}$	–
	+ NEED	**82.96**$_{\pm 3.42}$	5.86↑	**82.93**$_{\pm 3.44}$	8.22↑	**83.14**$_{\pm 3.44}$	6.71↑	**82.95**$_{\pm 3.46}$	8.97↑
Multimodal	FANVM	$76.00_{\pm 2.29}$	–	$75.98_{\pm 2.30}$	–	$76.07_{\pm 2.28}$	–	$76.01_{\pm 2.30}$	–
	+ NEED	**80.97**$_{\pm 4.05}$	4.97↑	**80.90**$_{\pm 4.10}$	4.92↑	**81.36**$_{\pm 3.96}$	5.29↑	**80.96**$_{\pm 4.04}$	4.95↑
	SV-FEND	$79.95_{\pm 1.97}$	–	$79.89_{\pm 2.01}$	–	$80.23_{\pm 1.78}$	–	$79.94_{\pm 1.98}$	–
	+ NEED	**84.62**$_{\pm 2.13}$	4.67↑	**84.61**$_{\pm 2.12}$	4.72↑	**84.81**$_{\pm 2.24}$	4.58↑	**84.64**$_{\pm 2.14}$	4.70↑

Experiments

Table 3: Ablation studies on each component in NEED. GA: graph aggregation, DR: debunking rectification. The standard deviation values are ignored for simplicity.

Method	Acc.	F1	Prec.	Recall
SV-FEND	79.95	79.89	80.23	79.94
+ DR	80.94	80.90	81.15	80.93
+ GA	83.43	83.41	83.61	83.45
+ NEED (DR&GA)	**84.62**	**84.61**	**84.81**	**84.64**
VGGish	66.91	66.82	67.07	66.89
+ DR	72.84	72.70	73.30	72.84
+ GA	74.83	74.64	75.54	74.80
+ NEED (DR&GA)	**75.25**	**75.12**	**75.73**	**75.22**
DR	82.95	81.05	81.36	81.04

Experiments

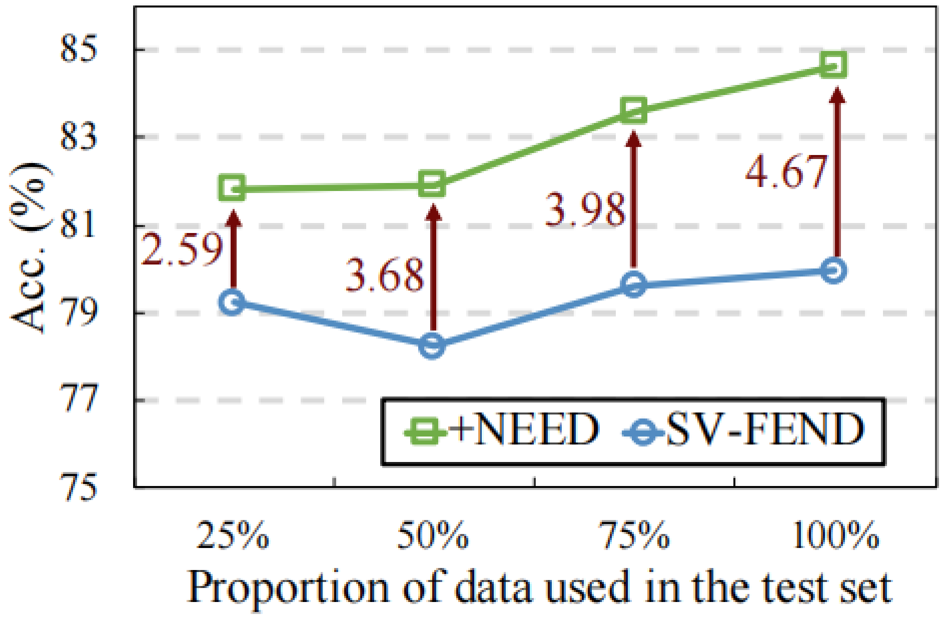


Figure 3: Performance of NEED in early detection.

Table 4: Peformance of NEED under the temporal split.

Method	Acc.	F1	Prec.	Recall
SV-FEND	82.20	81.47	82.89	80.99
+NEED	**89.67**	**89.37**	**90.16**	**88.97**

Experiments

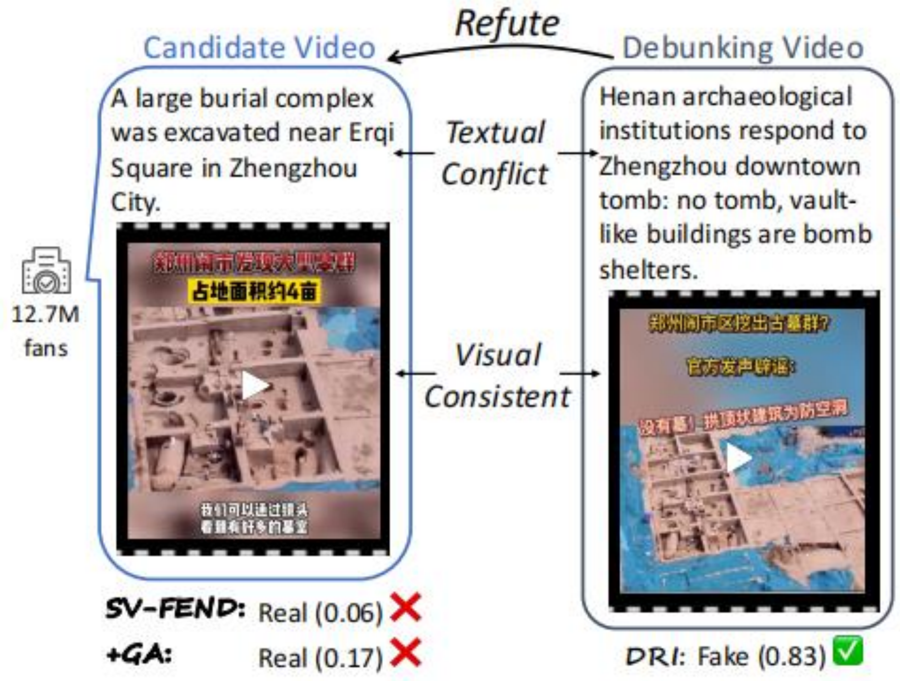


Figure 5: An example where the debunking video helps spot the "hard" fake news video missed by previous modules. The number denotes the predicted probability of labels being 0 (real) and 1 (fake), respectively.

Thanks