

# Open-Domain, Content-based, Multi-modal Fact-checking of Out-of-Context Images via Online Resources

CVPR 2022

Sahar Adbelnabi, Rakibul Hasan, and Mario Fritz

CISPA Helmholtz Center for Information Security

2024. 01. 23

발제자: 윤  
예준



# 연구 배

## 경

- ‘가짜 뉴스’가 사회적, 개인적, 정치적으로 해로운 영향을 미칠 것이라는 우려가 커지고 있음
- 이미지를 이용한 가짜 뉴스 생성하는 기술로는 ‘Deep fake’, ‘image-repurposing’이 존재
- 특히, image-repurposing은 뛰어난 기술 없이 ‘가짜 뉴스’를 생성할 수 있어 실제 주로 사용되는 요소



Deep fake  
(CNN)

### Original Caption

Photographs taken on Thursday showed Air Force Intelligence s bombdamaged headquarters in Aleppo  
목요일에 촬영된 사진에는 알레포에 있는 공군 정보국 본부가 폭탄 피해를 입은 모습이 담겨 있습니다.



Original Image



Image-repurposing

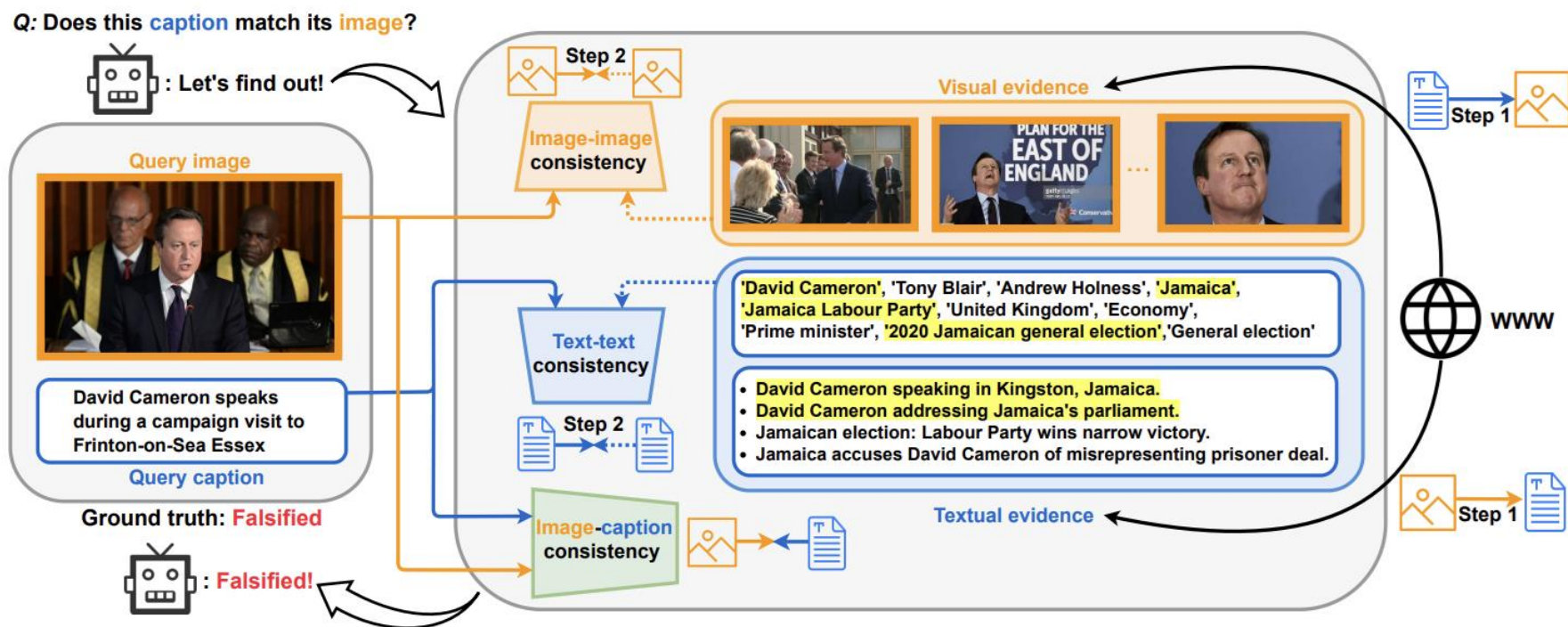
Image-repurposing (NewsCLIPpings)

- 이러한 이미지들은 실제 뉴스와 함께 사용되며 자신의 의견을 뒷받침하는 근거로 사용됨

# 연구 목

## 표

- Online Resources를 통해 Out-of-Context Images를 탐지하는 multi-modal fact-checking 제안
  - 웹에서 얻은 evidence와 주어진 Query Image-Text pair간의 consistency를 확인



# 데이





## 터

- NewsCLIPpings Datasets
  - 데이터 생성방법: text-text similarity, image-image similarity, etc.
  - Merged/Balanced train, valid, test: 71072, 7024, 7264
- Textual evidence
  - Google Vision API로 query 이미지에 관련된 텍스트 evidence 수집
  - API returns a list of entities that are associated with that image and the images' URLs and the containing pages' URLs  
=> entities 리스트와 URLs 사용해서 수집한 캡션, 제목 사용
- Visual evidence
  - Google custom search API로 query 텍스트에 관련된 이미지 evidence 수집

**Dataset.** Unless no search results were found, a single example in the dataset consists of the following:

- A query **image**  $I^q$ .
- A query **caption**  $C^q$ .
- **Visual evidence:**
  - A list of **images**:  $I^e = [I_1^e, \dots, I_K^e]$ .
- **Textual evidence:**
  - A list of **entities**:  $ENT = [E_1, \dots, E_M]$ .
  - A list of **captions/sentences**:  
 $S = [S_1, \dots, S_N]$ .

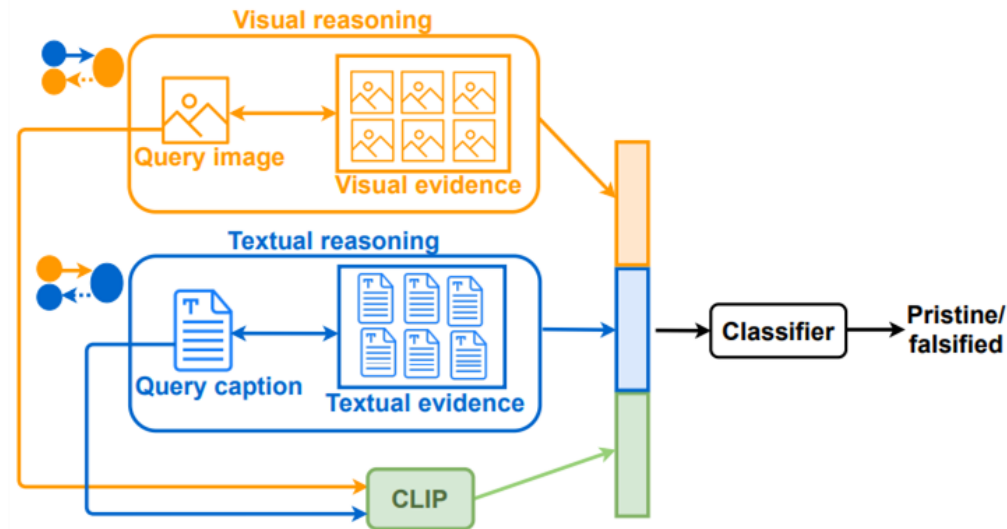
**Task.** Classify  $\{I^q, C^q\}$  to: *Pristine* or *Falsified*.

Image-caption pair	Textual evidence	Visual evidence
 The Futenma marine corps airbase on the southern Japanese island of Okinawa	<div>'United States', 'Ginowan', 'Governor', 'Military base', 'Politics', 'Japan', 'Takeshi Onaga', 'Governor of Okinawa Prefecture', 'Hirokazu Nakaima', 'Shinzo Abe', 'Okinawa', 'airport'</div> <div>1- Hercules aircraft parked on the tarmac at Marine Corps Air Station Futenma in Ginowan on Okinawa. 2- Japan Decides to Stop Works on US Airbase Relocation in Okinawa. 3- Japan Decides to Restart Relocation of US Base in Okinawa Despite Protests.</div>	  

# 방

# 버

- Consistency-Checking Network (CCN)
  - Visual Reasoning
  - Textual Reasoning
  - CLIP & Classifier
- Challenges
  - query로 evidence를 수집할 때 대부분의 검색 결과는 관련이 없어 잡음으로 작용할 수 있음
  - query와 evidence를 비교하려면 깊은 이해와 추론이 필요



# 방

# 변

- Consistency-Checking Network (CCN)
  - Visual Reasoning

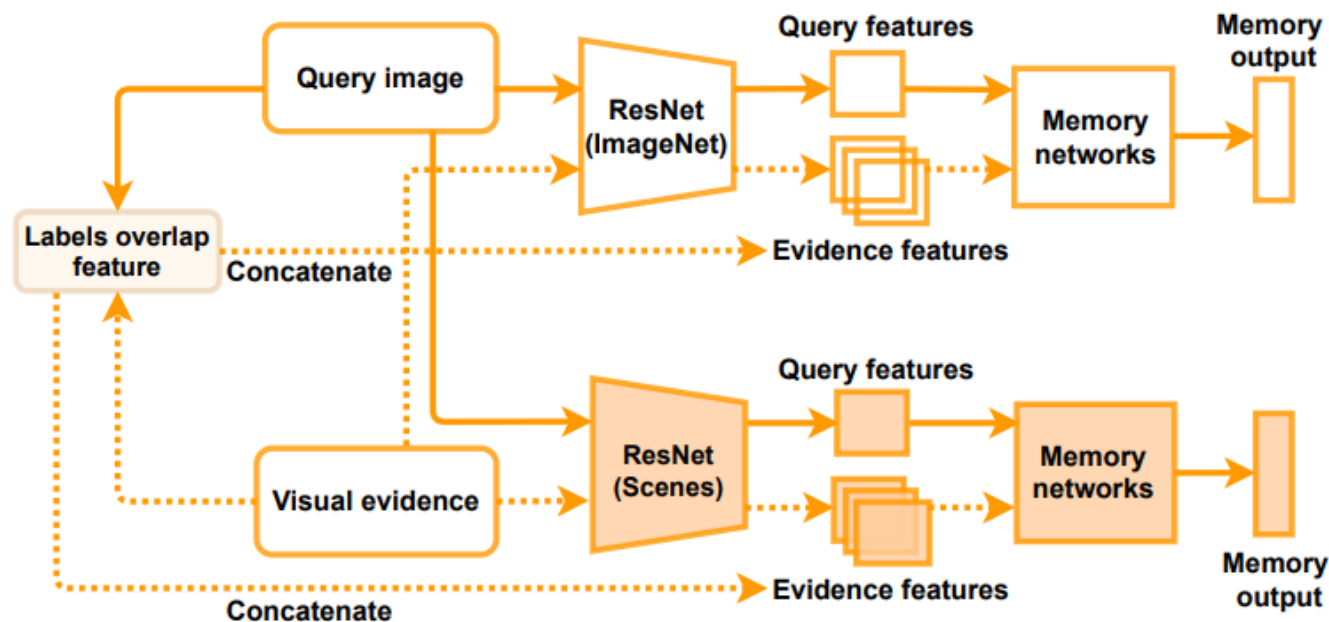
- 1) ResNet152로 query와 evidence 표현 추출
- 2) Google API 이용하여 query와 evidence간 겹치는 label count 후 evidence feature에 concatenate
- 3) Memory networks

$$m_i^a = \text{ReLU}(W_i^a I^e + b_i^a), \quad (1)$$

$$m_i^c = \text{ReLU}(W_i^c I^e + b_i^c) \quad (2)$$

$$p_{ij} = \text{Softmax}(\hat{I}^q T m_{ij}^a), \quad (3)$$

$$o_i = \sum_j p_{ij} m_{ij}^c + \hat{I}^q \quad (4)$$





방

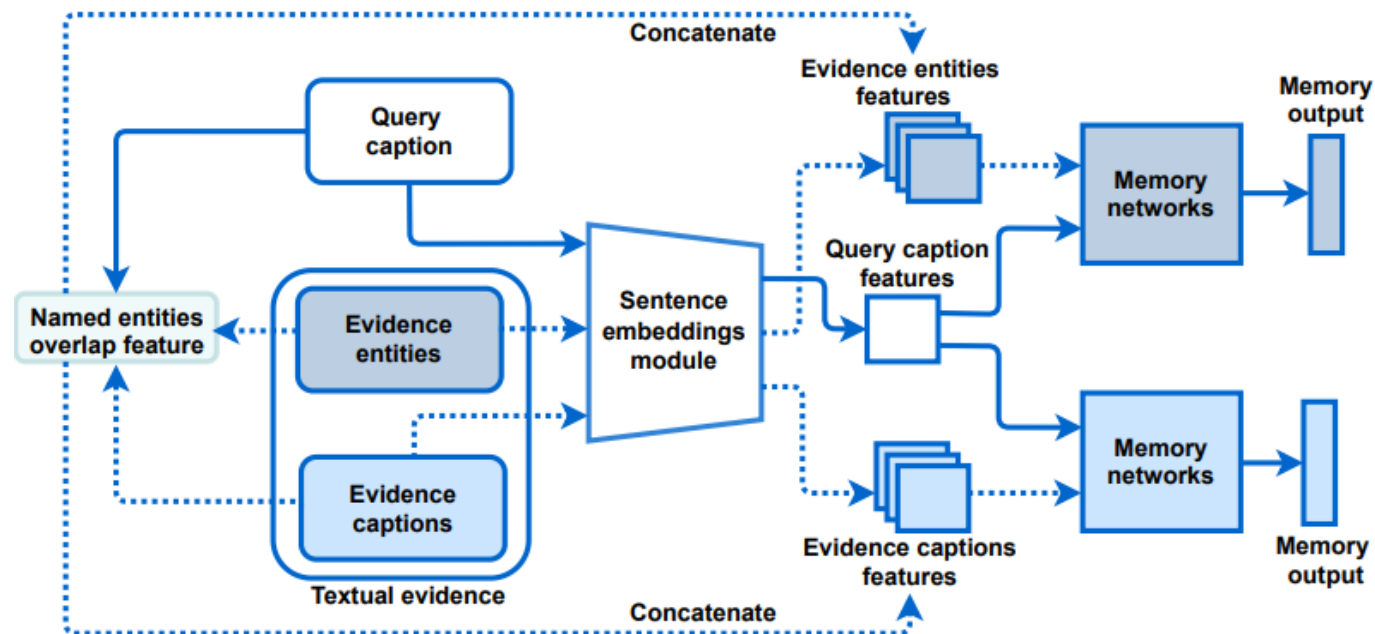
면

- Consistency-Checking Network (CCN)
  - Textual Reasoning

- Sentence embeddings module
  - sentence transformer
  - bert\_lstm
- Named entities overlap feature  
overlap 되면 1 아니면 0
- Memory networks

$$m_e^{a/c} = \text{ReLU}(W_e^{a/c} E + b_e^{a/c}), \quad (6)$$

$$m_s^{a/c} = \text{ReLU}(W_s^{a/c} S + b_s^{a/c}), \quad (7)$$



- Evidence's domain:  
3번 이상 나타난 domain에 대해 embedding 추가

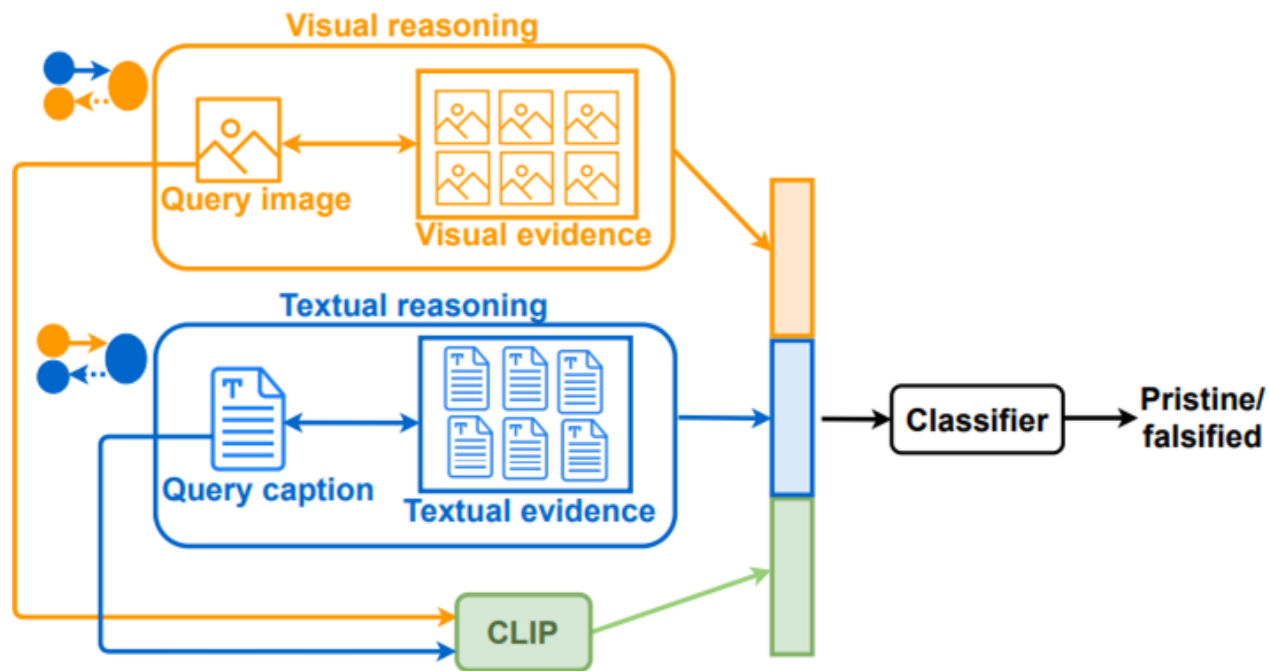
방

방

- Consistency-Checking Network (CCN)
  - CLIP & Classifier

$$o_t = \text{BN}(o_i) \oplus \text{BN}(o_p) \oplus \text{BN}(o_e) \oplus \text{BN}(o_s) \oplus \text{BN}(J_{\text{clip}}), \quad (8)$$

$$L = -y_{\text{true}} \log(p_f) - (1 - y_{\text{true}}) \log(1 - p_f) \quad (9)$$





# 결과

- CCN에서 사용한 방법들의 중요성과 CCN의 우수성을 보여줌

#	Evidence type	Separate mem.	BN	Dataset filter	CLIP	ResNet (ImageNet)	ResNet (Scenes)	Labels	Sent. transformer	BERT+LSTM	NER	Accuracy
1	all	✓	✗	✗	✗	✓	✗	✗	✓	✗	✗	73.5%
2	all w/o Images	✓	✗	✗	✗	-	-	-	✓	✗	✗	62.5%
3	all w/o Captions	✓	✗	✗	✗	✓	✗	✗	✓	✗	✗	57.4%
4	all w/o Entities	✓	✗	✗	✗	✓	✗	✗	✓	✗	✗	71.8%
5	all	✓	✓	✗	✗	✓	✗	✗	✓	✗	✗	84.2%
6	all	✗	✓	✗	✗	✓	✗	✗	✓	✗	✗	81.7%
7	all	✓	✓	✓	✗	✓	✗	✗	✓	✗	✗	80.3%
8	all	✓	✓	✓	✗	✓	✗	✗	✓	✗	✓	81.2%
9	all	✓	✓	✓	✓	✓	✗	✗	✓	✗	✓	82.6%
10	all	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	83.4%
11	all	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	83.9%
12	all	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	<b>84.7%</b>
13	all w/o domains	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	83.9%

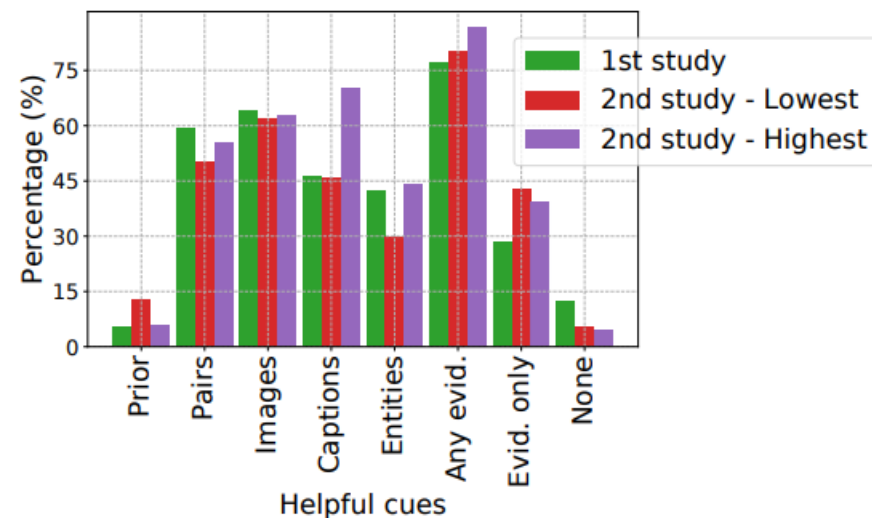
Method	Evidence	Pair	All	Falsified	Pristine
CLIP	✗	✓	66.1%	68.1%	64.2%
Averaged	✓	✗	70.6%	72.4%	68.9%
CCN	✓	✓	<b>84.7%</b>	<b>84.8%</b>	<b>84.5%</b>

# 결과

## Human Performance Baseline

- study 1:  
제시된 evidence를 가지고 캡션과 이미지가 일치하는지, 어떤 정보 출처가 도움되는지 레이블링
- study 2:  
인간이 참여하여 제공하는 evidence에서 가장 관련성이 높고 유용한 evidence를 검색 가능하다고 가정  
Highest: 모델 attention 높은 evidence, Lowest: 모델 attention 낮은 evidence

	Study	All	Falsified	Pristine
Average	1 <sup>st</sup>	81.0%±4.71	79.5%±8.31	82.3%±9.31
	2 <sup>nd</sup> , Highest	86.2%±4.9	84.5%±9.3	88.0%±7.2
	2 <sup>nd</sup> , Lowest	77.7%±6.0	76.0%±9.0	79.5%±7.5
Best worker	1 <sup>st</sup>	89.0%	92.0%	93.7%
	2 <sup>nd</sup> , Highest	94.0%	98.0%	98.0%
	2 <sup>nd</sup> , Lowest	88.0%	90.0%	86.0%



# 결

# 과

## • Human Performance Baseline











Image-caption pair	Textual evidence	Visual evidence
 <p>The Futenma marine corps airbase on the southern Japanese island of Okinawa</p>	<p>'United States', 'Ginowan', 'Governor', 'Military base', 'Politics', 'Japan', 'Takeshi Onaga', 'Governor of Okinawa Prefecture', 'Hirokazu Nakaima', 'Shinzo Abe', 'Okinawa', 'airport'</p> <p>1- Hercules aircraft parked on the tarmac at Marine Corps Air Station Futenma in Ginowan on Okinawa. 2- Japan Decides to Stop Works on US Airbase Relocation in Okinawa. 3- Japan Decides to Restart Relocation of US Base in Okinawa Despite Protests.</p>	
	Prediction: Pristine	
 <p>The soaring number of Syrian refugees has sparked increasing resentment in Lebanon</p>	<p>'Syria', 'Lebanon', 'United Kingdom', 'Tent', 'Syrians', 'Language', 'Refugee', 'Recreation', 'Tourism', 'Camping', 'Language barrier', 'rural area'</p> <p>1- Syrian refugees at a camp in eastern Lebanon, December 2014. 2- Syrians entering Lebanon face new restrictions 3- Among those displaced, 1.6 million children have fled Syria. 4- Syrian refugees in the UK: 'We will be good people. We will build this country'</p>	
	Prediction: Pristine	
 <p>Healthcare activists say the ruling against Novartis ensures poor people will be able to access cheap versions of cancer medicines</p>	<p>'United States Capitol', 'Affordable Care Act', 'Supreme Court of the United States', 'Presidency of Donald Trump', 'President of the United States', 'United States', 'us capitol grounds'</p> <p>1- Demonstrators from Doctors for America in support of Obamacare march in front of the Supreme Court on March 4, 2015. 2- The Affordable Care Act Is Back In Court, 5 Facts You Need To Know. 3- As Court Hears Arguments in Lawsuit To Eliminate Obamacare, Conn. Senators Plead Their Case.</p>	
	Prediction: Falsified	
 <p>Smoke rises following an Israeli air strike in Gaza City</p>	<p>'Kobane', 'Kurdistan Region', 'United States', 'Peshmerga', 'Turkey', 'Kurds', 'Syria', 'Iraq', 'kobani war'</p> <p>1- Smoke rises after a U.S.-led airstrike in the Syrian town of Kobani 2- The border town of Kobani is under threat after the Islamists drove 180,000 Kurds into Turkey. 3- Former Kurdish Sniper Claims To Have Killed Around 250 ISIS Fighters.</p>	
	Prediction: Falsified	
 <p>How can our young readers persuade their parents to get them a Playstation 3</p>	<p>'Grand Theft Auto V', 'Gamer', 'Grand Theft Auto IV', 'Wii', 'Grand Theft Auto VI', 'PlayStation 3', 'Rockstar Games', 'Rockstar Leeds', 'Terry seeborne marshall', 'Gordon Hall', 'Rockstar Games'</p> <p>1- A court order banning Sony from importing PS3s into the Netherlands has been lifted. 2- Rockstar Games, creators of the Grand Theft Auto franchise, said it was "very saddened" to hear of Mr Hall's death 3- Oakland Athletics to Begin Accepting Bitcoin for Private Suites</p>	
	Prediction: Falsified	

Figure 6. Qualitative examples of news pairs along with the collected evidence. Examples with green background are pristine, red background are falsified. Highlighted items are the ones with the highest attention. Only a subset of the evidence is shown for display purposes.

## 결론

---

- 웹을 통해 수집한 multi-modal evidence와 주어진 query 이미지-캡션 쌍간의 일관성을 확인하여 자동으로 복잡한 fact-checking process를 자동적으로 해주는 프레임워크 CCN 제안
- 이전 baseline보다 훨씬 능가하며 multi-modal fact-checking의 새로운 task를 formalize 함

# 한계

---

## 점

- 자동화된 도구에 모든 것을 의존하면 위험한 결과를 초래할 수 있음
- 제안한 접근 방식은 검색 엔진의 검색 결과에 의존
- 웹사이트에 존재하는 bias로 일부 evidence는 모순적일 수 있음

# Open

---

## Questions

- 웹사이트에 대한 bias를 줄이는 multi-modal fact-checking 방법은?
- 사람이 참여하지 않는 완전한 자동 fact-checking을 위해선 어떻게 발전해야하는가?

감사합니  
다.