

Visual News: Benchmark and Challenges in News Image Captioning

Fuxiao Liu, Yinghan Wang, Tianlu Wang,
Vicente Ordonez



UNIVERSITY OF
MARYLAND



RICE UNIVERSITY



UNIVERSITY
of VIRGINIA

Descriptive or Interpretative ?



A bunch of people who are holding red umbrellas ¹



A baseball player hitting the ball during the game ¹

¹ Chen et al., 2015. Microsoft coco captions: Data collection and evaluation server. *arXiv preprint arXiv:1504.00325*.

Descriptive or Interpretative ?



President **Obama** and **Mitt Romney** debate in **Hempstead NY** on **Tuesday**.



Virginia Cavaliers fans celebrate on the court after the Cavaliers game against the **Duke Blue Devils** at **John Paul Jones Arena**.

Visual News Dataset



VATICAN CITY Pope Francis installed 19 new cardinals Saturday in a ceremony that unexpectedly included Pope Emeritus Benedict XVI marking the first time the two appeared together in public. This batch of new cardinals the first appointed by Francis is significant because the group includes prelates from developing countries such as Burkina Faso and Haiti in line with the pope's belief that the church should do more to help the world's poor Saturday's ceremony also helped move the spotlight away from more controversial topics ...

Pope Emeritus Benedict XVI left and Pope Francis greet each other in St Peter's Basilica

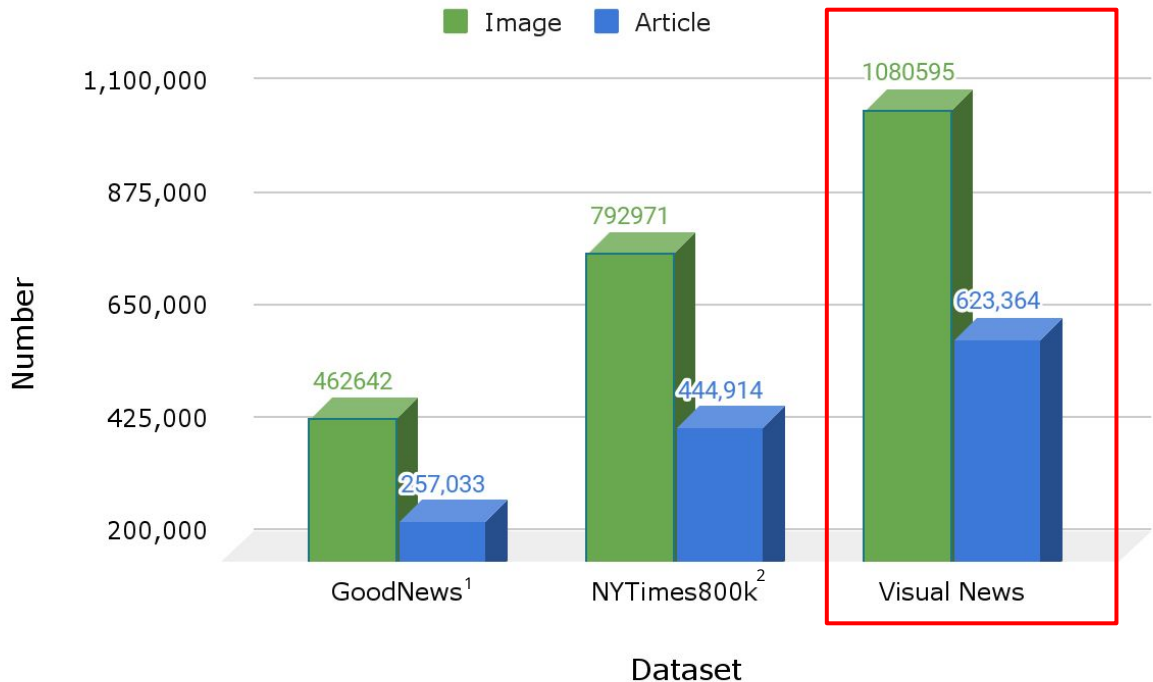


Hillary Clinton is the Democratic Party's presumptive presidential nominee according to the Associated Press securing enough support from superdelegates to push her over the top on the eve of the final round of state primaries. Both AP and NBC News reported Monday night that a sufficient number of superdelegates had indicated their support for Clinton to guarantee she will have the 2383 delegates needed at the party's July in convention in Philadelphia ...

Hillary Clinton arrives to the Los Angeles Get Out The Vote Rally at on June 6 2016 in Los Angeles



Visual News Dataset



A larger dataset: more images as well as articles !

¹Biten et al. Good News, Everyone! Context driven entity-aware captioning for news images. CVPR 2019

²Tran et al. Transform and Tell: Entity-Aware News Image Captioning. CVPR 2020.



LIVE
Coronavirus cases near 1.5 million worldwide

With countries still struggling to curb the outbreak Italy warns the EU could fall over its response.

World

- A visual guide to lockdown
- Why it is so hard to buy flour?



More diverse: various new agencies !



DATA COLLECTION

Visual News Dataset

<i>Visual News</i>	<i>Guardian</i>	<i>BBC</i>	<i>USA</i>	<i>Wash.</i>
Avg. Article Length	787	630	700	978
Avg. Caption Length	22.5	14.2	21.5	17.1

<i>GoodNews¹</i>	<i>NYTimes800k²</i>
451	974
18	18

More diverse: various text length !

¹Biten et al. Good News, Everyone! Context driven entity-aware captioning for news images. CVPR 2019

²Tran et al. Transform and Tell: Entity-Aware News Image Captioning. CVPR 2020.

Visual News Dataset

<i>Visual News</i>	<i>Guardian</i>	<i>BBC</i>	<i>USA</i>	<i>Wash.</i>
% of Words are NE	0.18	0.17	0.22	0.33
% of Captions w/NE	0.89	0.85	0.95	0.92
% of Captions w/People's Name	0.72	0.46	0.84	0.69

<i>GoodNews¹</i>	<i>NYTimes800k²</i>
0.27	0.26
0.97	0.96
0.68	0.68

* w/NE means with named entities.

* More statistic in the appendix

More diverse: various named entity distributions!

¹Biten et al. Good News, Everyone! Context driven entity-aware captioning for news images. CVPR 2019

²Tran et al. Transform and Tell: Entity-Aware News Image Captioning. CVPR 2020.

Visual News Dataset

PERSON:
*Donald Trump, Hillary Clinton,
Barack Obama, Lebron James, Pope
Francis, Michelle Obama, Novak
Djokovic...*

GPE:
*New York, Florida, Washington, US,
Los Angeles, Chicago...*

ORG:
*NBA, Apple, The White House,
Capitol Hill, Boeing,
United States Senate...*



PERSON:
*Donald Trump, Bill Clinton,
Michelle Obama, Barack Obama,
Hillary Clinton, Jeb Bush, Pope
Francis...*

GPE:
*Washington, US, New York, Maryland,
Virginia...*

ORG:
*The White House, United States
Senate, GOP, Apple, Capitol Hill...*



PERSON:
*David Cameron, Nicola Sturgeon,
Ed Miliband, Angela Merkel,
Carwyn Jones, Nigel Farage, Nick
Clegg, Prince Charles...*

GPE:
*Scotland, China, UK, London, US,
India, Syria...*

ORG:
BBC, EU, UN, the White House...



PERSON:
*David Cameron, Nicola Sturgeon,
Ed Miliband, Angela Merkel, Nigel
Farage, Nick Clegg, vladimir putin,
Luis Suarez, George Osborne...*

GPE:
*London, Australia, US, England,
Paris, UK, China, India...*

ORG:
*UN, EU, BBC, Apple, Chelsea,
Tesco...*



Visual News Dataset

More diverse: various named entities from different agencies !



PERSON:
Donald Trump, Hillary Clinton, Barack Obama, LeBron James, Pope Francis, Michelle Obama, Novak Djokovic...

GPE:
New York, Florida, Washington, US, Los Angeles, Chicago...

ORG:
NBA, Apple, The White House, Capitol Hill, Boeing, United States Senate...

PERSON:
Donald Trump, Bill Clinton, Michelle Obama, Barack Obama, Hillary Clinton, Jeb Bush, Pope Francis...

GPE:
Washington, US, New York, Maryland, Virginia...

ORG:
The White House, United States Senate, GOP, Apple, Capitol Hill...



PERSON:
David Cameron, Nicola Sturgeon, Ed Miliband, Angela Merkel, Carwyn Jones, Nigel Farage, Nick Clegg, Prince Charles...

GPE:
Scotland, China, UK, London, US, India, Syria...

ORG:
BBC, EU, UN, the White House...

PERSON:
David Cameron, Nicola Sturgeon, Ed Miliband, Angela Merkel, Nigel Farage, Nick Clegg, Vladimir Putin, Luis Suarez, George Osborne...

GPE:
London, Australia, US, England, Paris, UK, China, India...

ORG:
UN, EU, BBC, Apple, Chelsea, Tesco...



The Washington Post



Image Caption Model

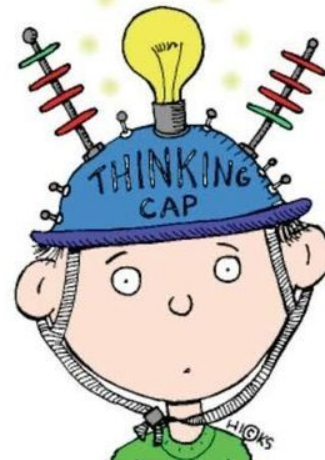
Seven days into free agency miami heat president pat riley made his first roster moves to show lebron james why he should stick around this much is clear riley is confident james chris bosh and dwyane wade will return according to people who have had phone conversations with riley in the last week the

People spoke to sports because of the sensitive nature of the conversations of course riley's confidence doesn't guarantee

...



Model



Lebron James hugs
Pat Riley after
winning in **Miami**

Previous Work

1. Template based methods ¹



Miss some contextual clues

the exterior of the **ORG_** in **PLACE_** → the exterior of the **Brooklyn Academy of Music** in **New York**

¹Biten et al. Good News, Everyone! Context driven entity-aware captioning for news images. CVPR 2019

Previous Work

1. Template based methods ¹



Miss some contextual clues

the exterior of the **ORG_** in **PLACE_** → the exterior of the **Brooklyn Academy of Music** in **New York**

2. Transformer + BPE ² based methods ³



Tremendous training parameters

Sun rise ' s executive director . → **Sunrise's** executive director.

¹Biten et al. Good News, Everyone! Context driven entity-aware captioning for news images. CVPR 2019

²Sennrich et al . Neural machine translation of rare words with subword units. ACL 2016.

³Tran et al. Transform and Tell: Entity-Aware News Image Captioning. CVPR 2020.

Visual News Captioner

Article

Seven days into free agency Miami Heat President Pat Riley made his first roster moves to show LeBron James why ...

Image



Visual News Captioner

Article

Seven days into
free agency Miami
Heat President Pat
Riley made his first
roster moves to
show LeBron
James why ...



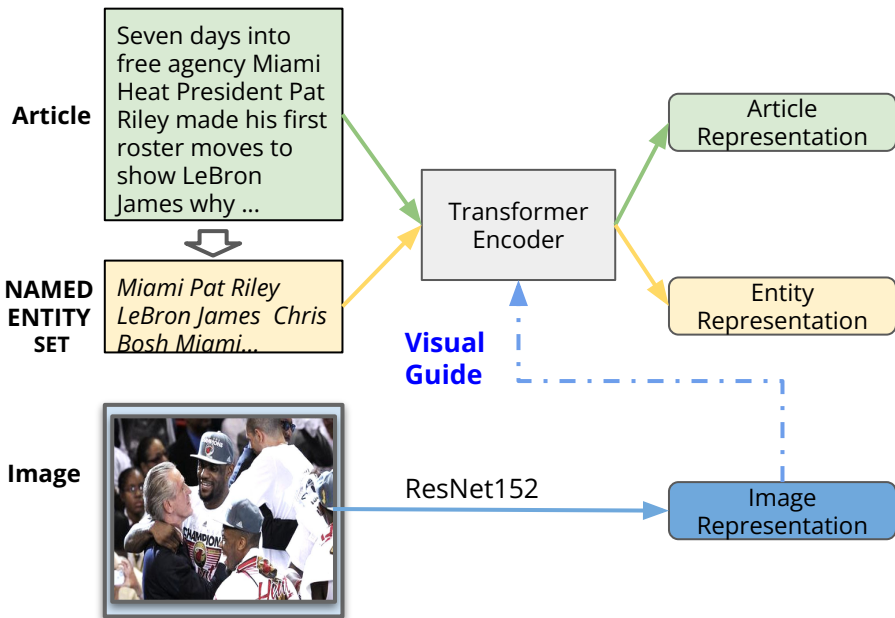
NAMED
ENTITY
SET

*Miami Pat Riley
LeBron James Chris
Bosh Miami...*

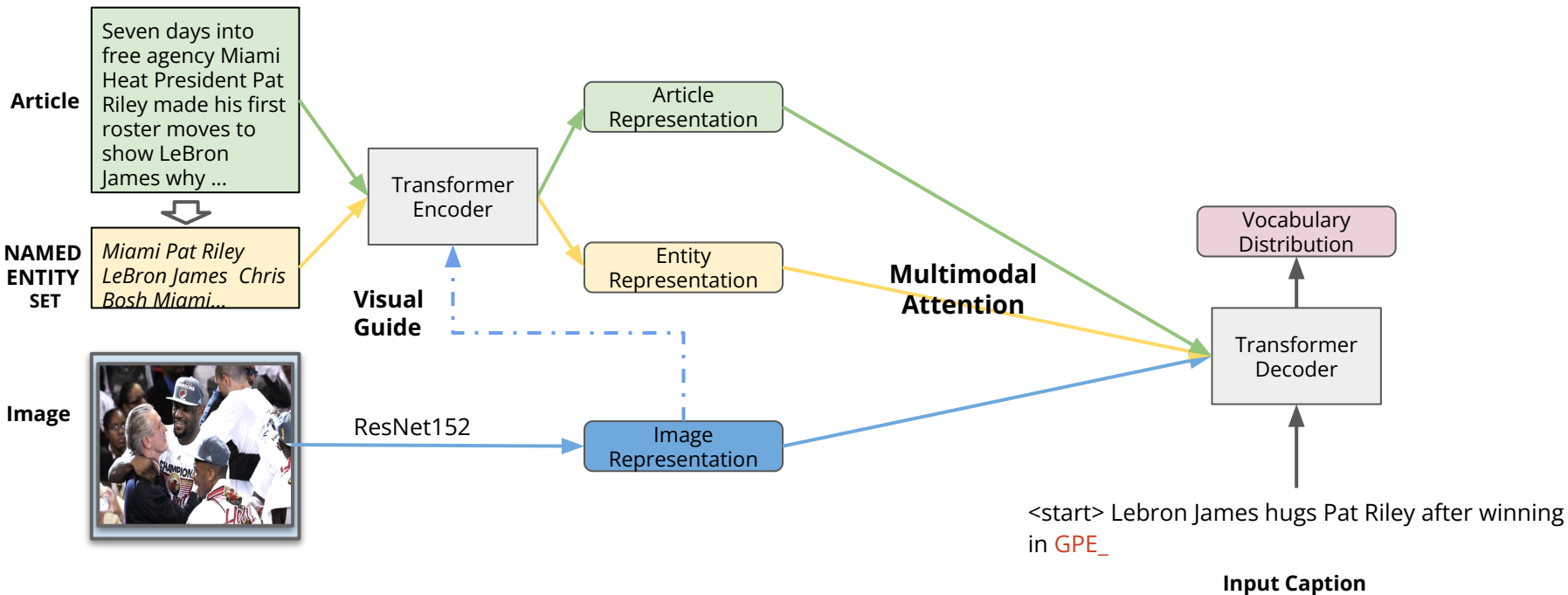
Image



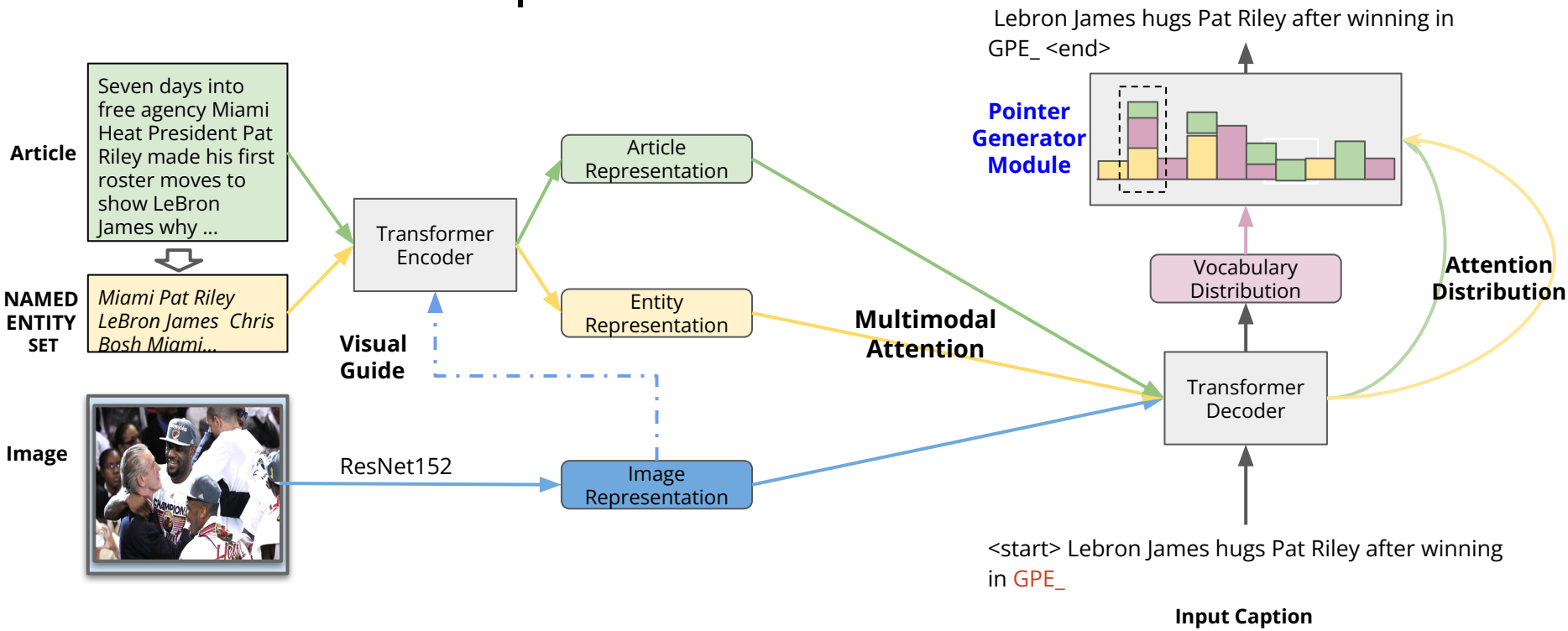
Visual News Captioner



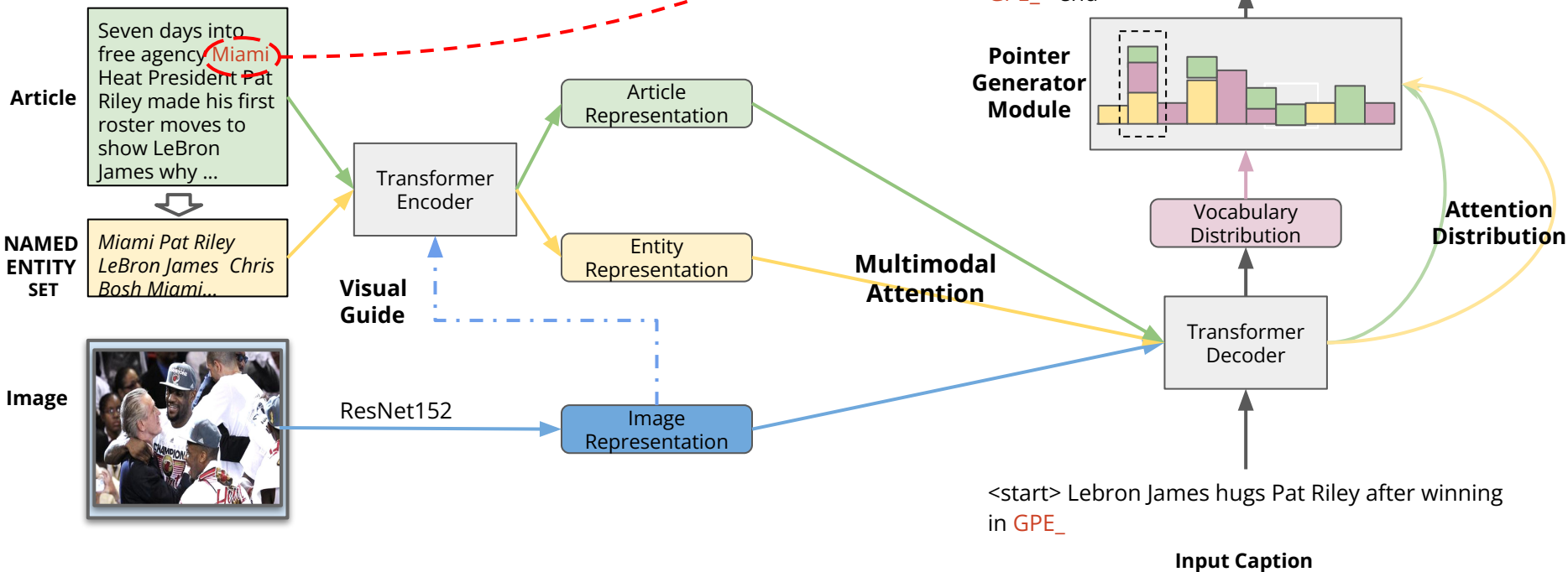
Visual News Captioner



Visual News Captioner



Visual News Captioner



Evaluation on GoodNews¹

Model	BLEU-4	METEOR	ROUGE	CIDer	Precision	Recall
TextRank ³	1.7	7.5	11.6	9.5	1.7	5.1
Show and Tell ⁴	0.7	4.1	11.9	12.2	---	---
GoodNews ¹	0.8	4.3	12.1	12.7	8.2	7.2
Transform and Tell ²	6.0	---	21.4	53.8	22.2	18.7
Visual News Captioner	6.1	8.3	21.6	55.4	22.9	19.3

* We use precision and recall evaluate the model ability to predict named entities.

¹Biten et al. Good News, Everyone! Context driven entity-aware captioning for news images. CVPR 2019

²Tran et al. Transform and Tell: Entity-Aware News Image Captioning. CVPR 2020.

³Barrios et al. Variations of the similarity function of textrank for automated summarization. ASAI 2015.

⁴Xu et al. Show, attend and tell: Neural image caption generation with visual attention. ICML 2015.

Evaluation on NYTimes800k²

Model	BLEU-4	METEOR	ROUGE	CIDer	Precision	Recall
TextRank ³	1.9	7.3	11.4	9.8	3.6	4.9
GoodNews ¹	0.8	4.1	11.3	12.2	8.6	7.3
Transform and Tell ²	6.3	---	21.7	54.4	24.6	22.2
Visual News Captioner	6.1	8.1	21.9	56.1	24.8	22.3

* We use precision and recall evaluate the model ability to predict named entities.

¹Biten et al. Good News, Everyone! Context driven entity-aware captioning for news images. CVPR 2019

²Tran et al. Transform and Tell: Entity-Aware News Image Captioning. CVPR 2020.

³Barrios et al. Variations of the similarity function of textrank for automated summarization. ASAI 2015.

Evaluation on Visual News

Model	BLEU-4	METEOR	ROUGE	CIDer	Precision	Recall
TextRank ³	2.1	8.0	12.0	8.4	4.1	6.1
Show Attend Tell ⁴	1.5	4.6	12.6	11.3	---	---
GoodNews ¹	2.1	5.2	13.5	13.2	5.3	5.3
Visual News Captioner	5.3	8.2	17.9	50.5	19.7	17.6

* We use precision and recall evaluate the model ability to predict named entities.

¹Biten et al. Good News, Everyone! Context driven entity-aware captioning for news images. CVPR 2019

²Tran et al. Transform and Tell: Entity-Aware News Image Captioning. CVPR 2020.

³Barrios et al. Variations of the similarity function of textrank for automated summarization. ASAI 2015.

⁴Xu et al. Show, attend and tell: Neural image caption generation with visual attention. ICML 2015.

Lightweight !

Model	Num of Parameters
Transform and Tell ²	200 M
GoodNews ¹	157 M
Visual News Captioner	93 M

Visual News Captioner is more lightweight!

¹Biten et al. Good News, Everyone! Context driven entity-aware captioning for news images. CVPR 2019

²Tran et al. Transform and Tell: Entity-Aware News Image Captioning. CVPR 2020..

Conclusion

1. We construct Visual News, the largest and most diverse news image captioning dataset.
2. We propose Visual News Captioner, an entity-aware captioning method.
3. We benchmarked both template-based and end-to-end captioning methods on large-scale news image datasets, revealing the challenges in the task of news image captioning.

Dataset and Code available at:

<https://github.com/FuxiaoLiu/VisualNews-Repository>

Appendix

<i>Visual News</i>	<i>Guardian</i>	<i>BBC</i>	<i>USA</i>	<i>Wash.</i>
% of Captions w/"GPE_"	0.37	0.29	0.52	0.50
% of Captions w/"ORG_"	0.37	0.29	0.57	0.43
% of Captions w/"DATE_"	0.23	0.22	0.47	0.29
% of Captions w/"FAC_"	0.06	0.44	0.15	0.11

***More diverse: various named
entity distributions!***

* w/NE means with named entities.

¹Ali Furkan Biten et al. Good News, Everyone! Context driven entity-aware captioning for news images. CVPR 2019

²Alasdair Tran et al. Transform and Tell: Entity-Aware News Image Captioning. CVPR 2020.