



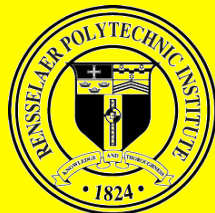
Visual Attention Model for Name Tagging in Multimodal Social Media

Di Lu¹, Leonardo Neves², Vitor Carvalho³, Ning Zhang², Heng Ji¹

¹Rensselaer Polytechnic Institute

²Snap Research

³Intuit



Snap Inc.

intuit.





Introduction



Trending of Social Media

TOP SOCIAL PLATFORMS

MONTHLY ACTIVE USERS (MAU) 2017

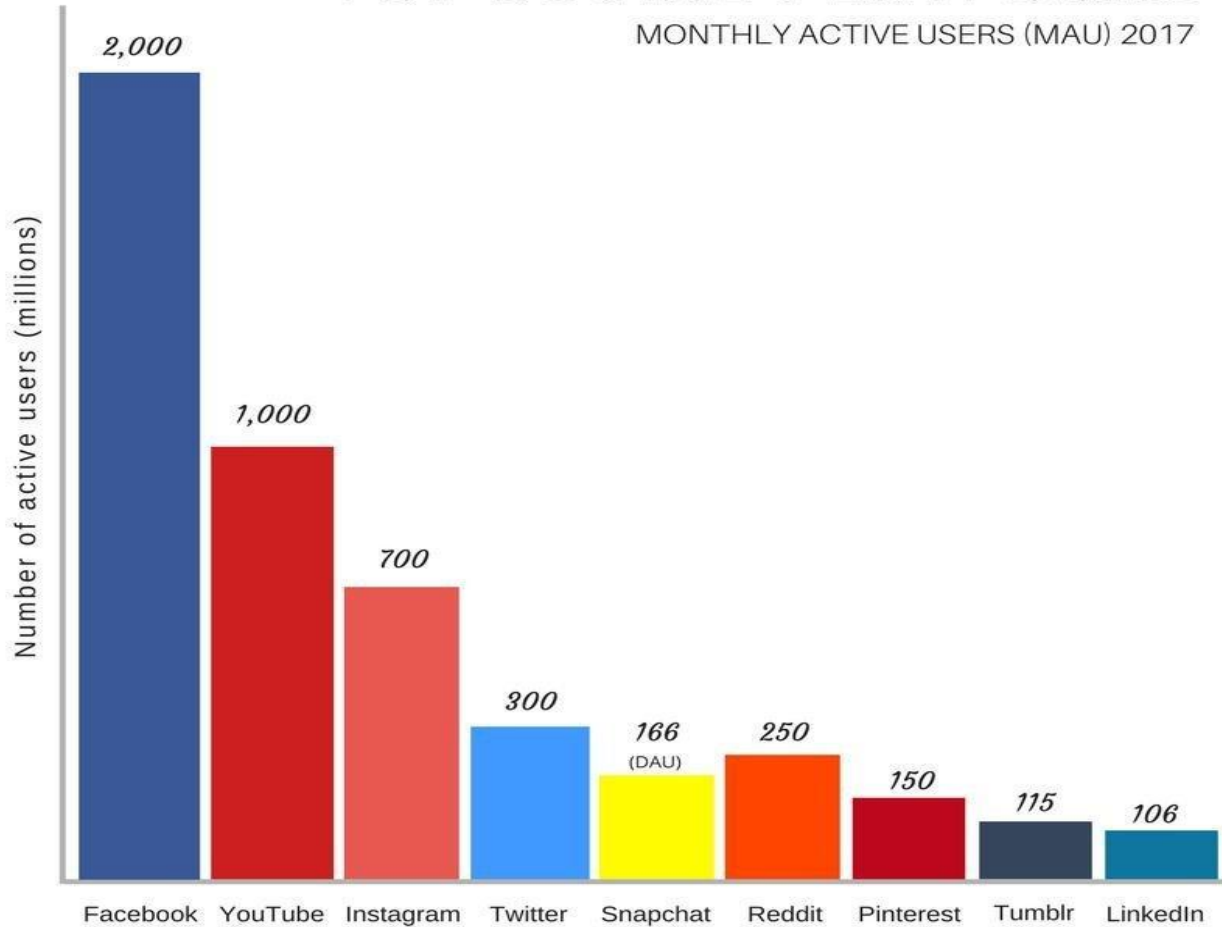


Image credit
(<https://blog.bufferapp.com/>)





Name Tagging

- What is Name Tagging?

[_{ORG} France] defeated [_{ORG} Croatia] in [_{MISC} World Cup] final at [_{LOC} Luzhniki Stadium].

- Why important?
 - Provide inputs to downstream applications
 - Searching
 - Recommendation
 - Knowledge graph construction





Challenges of Name Tagging in Social Media



Real Madrid midfielder Toni Kroos has revealed why he snubbed Cristiano Ronaldo's birthday party, following their humiliating derby defeat to Atletico Madrid.

Read: Khedira Doesn't Regret Attending CR7's Party

Ronaldo received a lot of criticism for hosting his birthday party just hours after his side lost 4-0 to Atletico, and although Kroos understands it was difficult to cancel the party, he feels the timing wasn't right.

"I was invited to Cristiano Ronaldo's party. I didn't go because I knew what could happen," he told German TV station *ZDF*.

"It wasn't the moment to have a party after losing 4-0 against Atletico. It's also true that many people had been invited and cancelling it wouldn't have been easy."

The 25-year-old, who won the World Cup with Germany in Brazil, also insisted that recent media reports of a Real Madrid 'crisis' were thrown out of proportion.

"We should take a step back and look at the whole picture in the face of what is being said. We have only lost the one game," Kroos added.

"I think that many teams would love to suffer a crisis like ours. Of course we should be criticised if we play a bad game, as we did that day, without doubt."

Read: Barca In Trouble For "Drunk Ronaldo" Chants?

Kroos joined Los Blancos for €30 million last summer and started all but two games for Real, assisting 12 goals and scoring one.

Do you think Real Madrid will return to their form from before Christmas? Have your say in the comments section below.

News VS Tweet

- Limited Textual Context
- Performs much worse on social media data



CR7 or TK8





Challenges of Name Tagging in Social Media

- Language Variations

I luv juuustin 

- Bad segmentation

Alison wonderlandxDiploxjuaz B2B ayee'

- Within word white spaces

LETS GO L A K E R S





Utilization of Vision



[Modern Baseball](#) played an intimate surprise set at Shea



[Karl-Anthony Towns](#) named unanimous 2015-2016 NBA Rookie of the Year

Difficult cases based on text only



Task Definition



Colts Have 4th Best QB Situation in NFL with Andrew Luck #ColtStrong

Multimedia Input:
image-sentence pair



[_{ORG} **Colts**] Have 4th Best QB Situation in [_{ORG} **NFL**] with [_{PER} **Andrew Luck**] #ColtStrong

Output: tagging results
on sentence





Overview



Our work

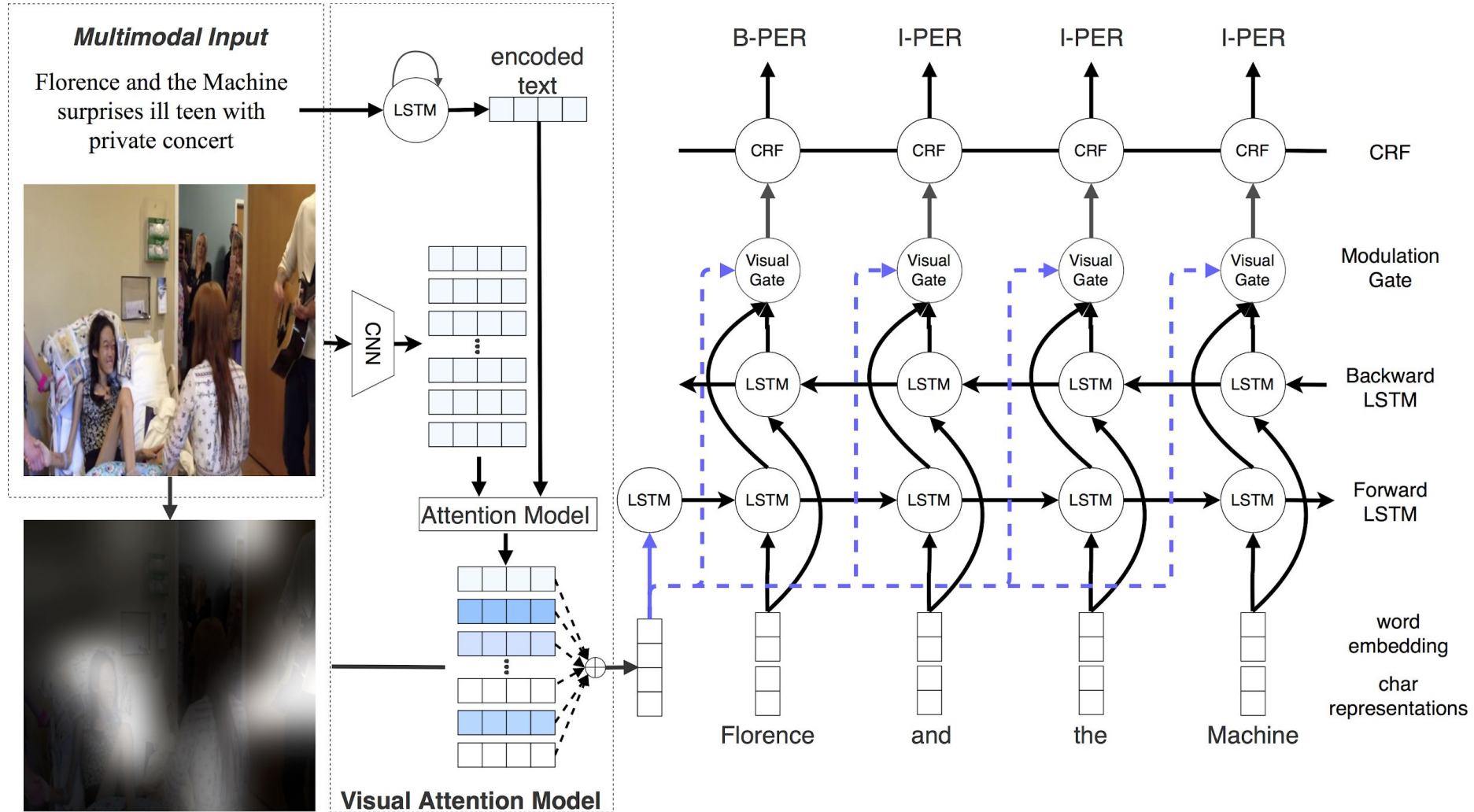
- Sequence Labeling Model
 - Bidirectional Long-short-term-memory-networks (BLSTM)
 - Word representations Generations
 - Conditional-random-fields (CRF)
 - Joint tags prediction
 - State-of-the-art for news articles (Lample et al. 2016)
- Visual attention model (Bahdanau et al., 2014)
 - Extract visual features from image regions that are most related to accompanying sentence
- Modulation Gate before CRFs
 - Combine word representation with visual features based on their relatedness





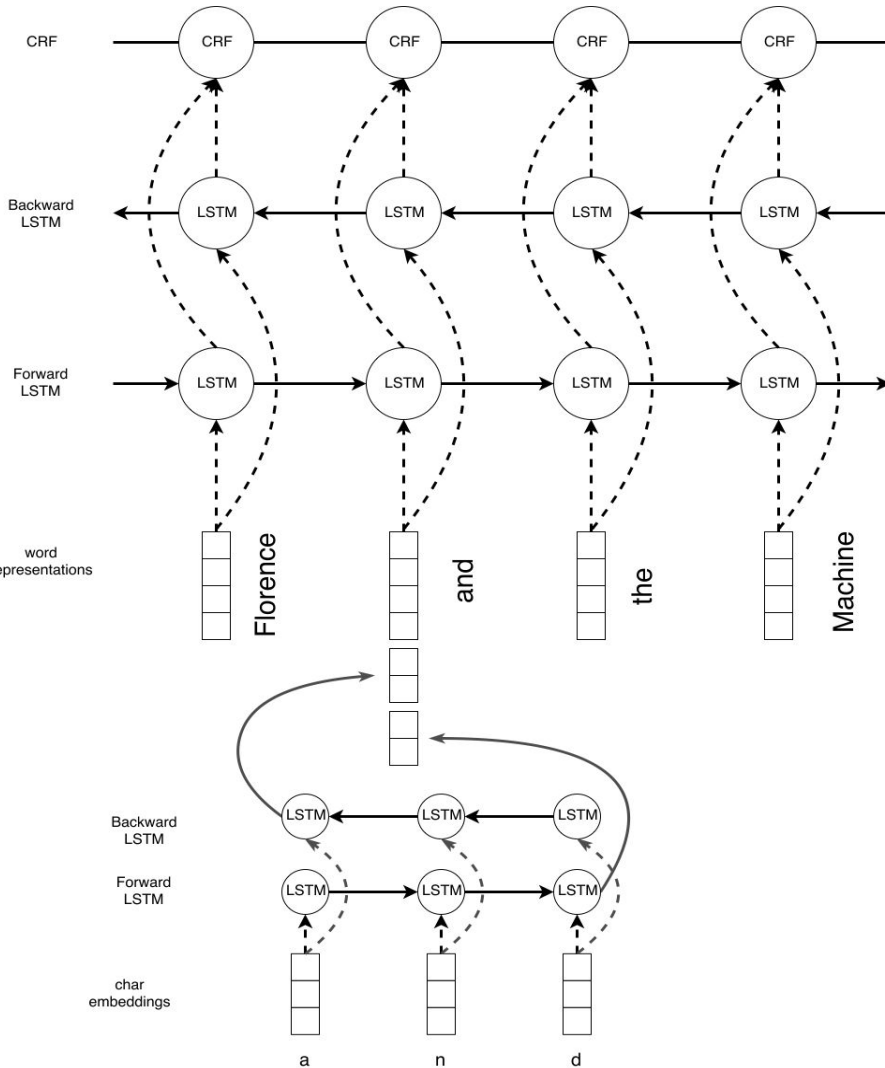
Model

Overall Framework





Sequence Labeling—BLSTM-CRF (Lample et al. 2016)



$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o)$$

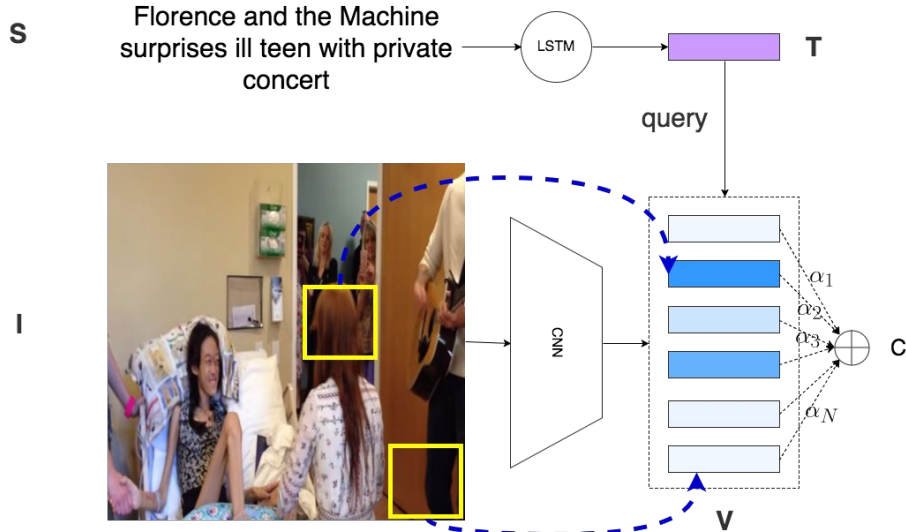
$$h_t = o_t \odot \tanh(c_t)$$

x_t , c_t and h_t are the input, memory and hidden state at time t respectively. W_{xi} , W_{hi} , W_{xf} , W_{hf} , W_{xc} , W_{hc} , W_{xo} and W_{ho} are weight matrices. \odot is the element-wise product functions and σ is the element-wise sigmoid function





Attention Model for Text Related Visual Features Localization



S Input sentence

I Input image

$$T = LSTM(S)$$

$$V = CNN(I) \quad \text{Outputs from convolutional layer}$$

$$V = \{V_1, V_2, V_3, \dots, V_N\}$$

$$a_i = \tanh(W_{T,A}T + W_{V,A}V_i + b_A)$$

$$e_i = W_P a_i + b_p$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^N \exp(e_j)}$$

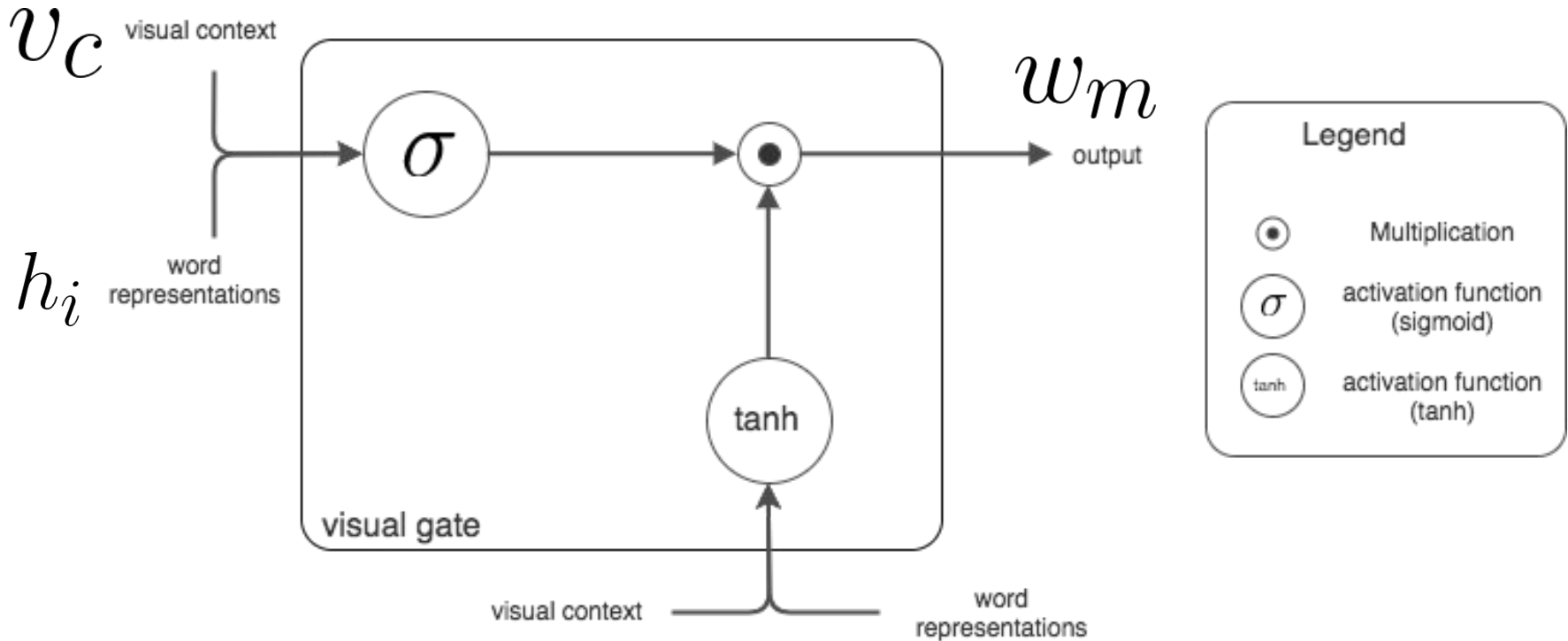
Attention calculate

$$C = \sum_{i=1}^N \alpha_i V_i \quad \text{Context Vector}$$





Modulation Gate



$$\beta_v = \sigma(W_v h_i + U_v v_c + b_v)$$

$$\beta_w = \sigma(W_w h_i + U_w v_c + b_w)$$

$$m = \tanh(W_m h_i + U_m v_c + b_m)$$

$$w_m = \beta_w \cdot h_i + \beta_v \cdot m$$

v_c Visual context

h_i Word representation

w_m Visually tuned word representation





Experiments



Dataset

- Snap Caption Dataset and Twitter DataSet (image+text)
- Topics: Sports, concerts and other social events
- Named Entity Types: Person, Organization, Location and MISC

		Training	Develement	Testing
Snap	Sentence	4,817	1,032	1,033
	Tokens	39,035	8,334	8,110
Twitter	Sentence	4,290	1,432	1,459
	Tokens	68,655	22,872	23,051

Size of the dataset in numbers of sentences and tokens





Results

Model	Snap Captions			Tweets		
	Precision	Recall	F1	Precision	Recall	F1
BLSTM-CRF	57.71	58.65	58.18	78.88	77.47	78.17
+Global Image Vector	61.49	57.84	59.61	79.75	77.32	78.51
+Visual Attention	65.53	57.03	60.98	80.81	77.36	79.05
Gate controlled visual attention	66.67	57.84	61.94	81.62	79.90	80.75





Attention Visualization



(a). [PER Klay Thompson] [ORG Warriors] overwhelm [ORG TrailBlazers] 110-99 go up 2-0 in series



(b). [PER Radiohead] offers old and new at first concert in four years...



(c). [MISC Cannes] just became the [PER Blake Lively] show #Cannes2016 @[PER blakelively]



(d). #iPhoneAt10: How [PER Steve Jobs] and [ORG Apple] changed modern society



(e). [PER Florence and the Machine] surprises ill teen with private concert



(f). [ORG Warriorette] Basketball Campers ready for Day 2



(g). Is defending champ [PER Sandeul] able to win for the third time on '[MISC Duet Song Festival]'?



(h). Shirts at the ready for our hometown game today #[ORG Leicester] #pgautomotive #[ORG premierleague]



(i). ARMY put up a huge ad in [LOC Times Square] for [PER BTS] '4th anniversary!





Error analysis



Nice image of [PER Kevin Love] and [PER Kyle Korver] during 1st half #NBAFinals #CavsIn9 # [LOC Cleveland]



Very drunk in a #magnum concert



Looking forward to editing some SBU baseball shots from Saturday.



Poor Alignment between Image and Sentence

Blur Images

Wrong Visual Attention

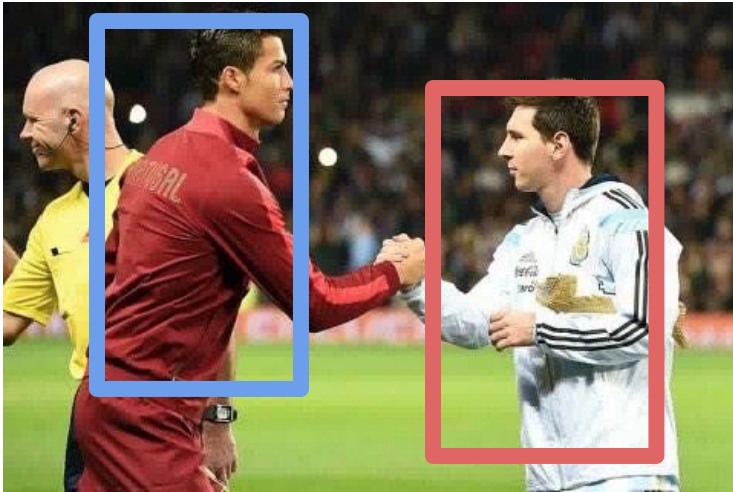




Future Work



- Joint Multimodal Grounding and Name Tagging



[PER **CR7**] & [PER **Messi**]
shake hands

- Fine Grained Name Tagging



San Francisco Giants



New York Giants



Belfast Giants

Giants won the game





Thank You.



Questions?

