

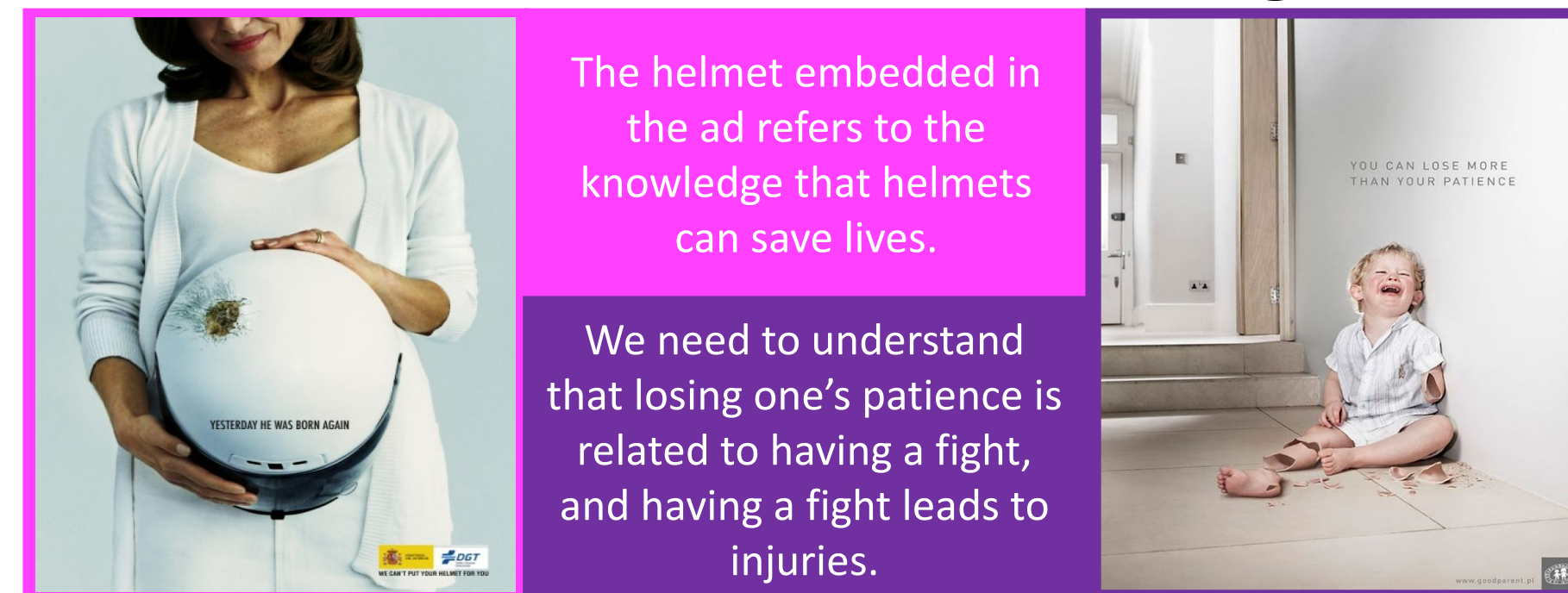
ADVISE: Symbolism and External Knowledge for Decoding Advertisements

Keren Ye
Department of Computer Science, University of Pittsburgh



Introduction

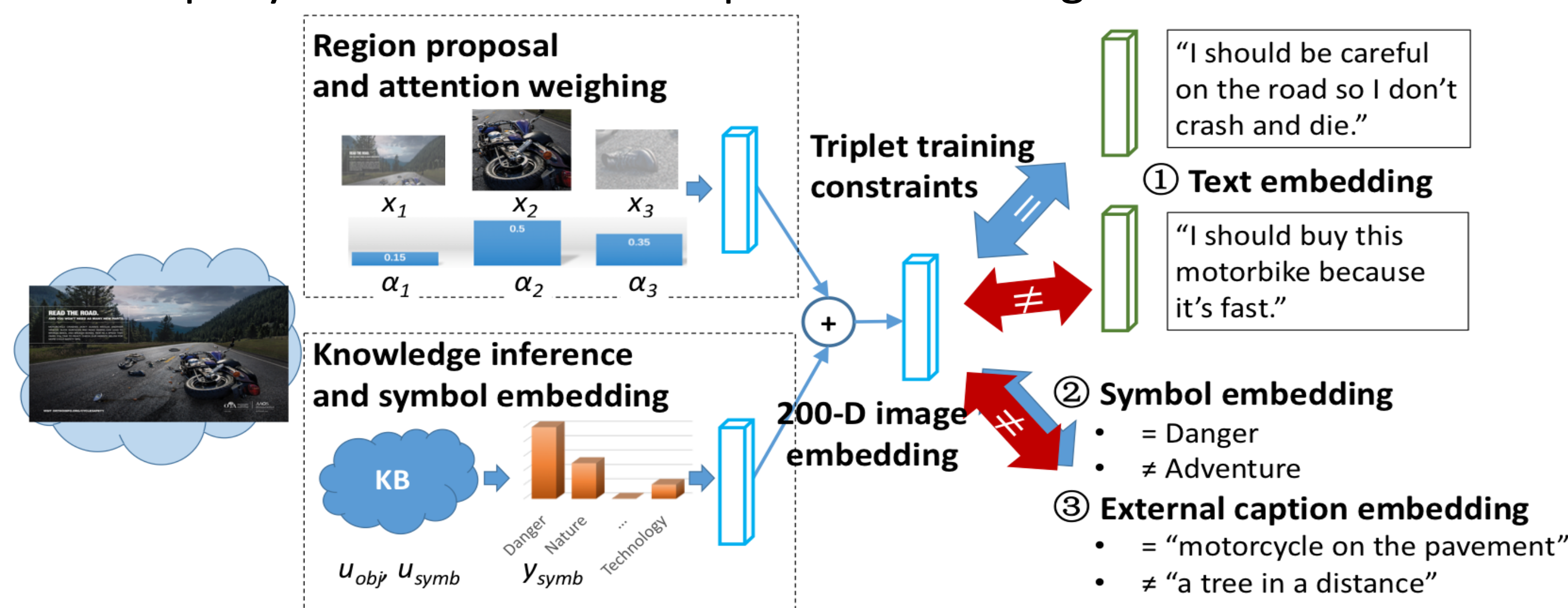
- Advertisements embed references to outside knowledge, and inspire us to ask:
 - How to utilize **symbolic references and knowledge** to understand the meaning of an ad?
- We formulate the ad understanding task as matching an ad image to human-written statements about the ad's message.
- We interpret an ad using symbolic region proposals and apply bottom-up attention to aggregate information.
- We use external knowledge as a constraint to regularize the model, and incorporate discovered object-symbol mappings.



- Constrains via symbols and external captions
 - We use these external resources as pivots to enforce similar examples to be closer.
 - Symbols are abstract words such as “danger” and “strength”



- External captions are descriptions of the image regions extracted using the DenseCap model (Johnson et al., 2016)
- Additive external knowledge (knowledge branch)
 - KB Symbols** – uses an external classifier to link certain visuals to symbolic concepts, then embeds them into the same feature space
 - KB Objects** – infers symbols from real-world objects first, then maps symbols to the same space as the images and statements



Dataset

- We use the PITT image ads dataset (Hussain et al., CVPR 2017)

statement	202,090	symbol	64,131	topic	204,340
sentiment	102,340	slogan	11,130	strategy	20,000

- We use the **action-reason statements**, and require the model to rank the 3 statements paired w/ the image higher than 47 statements for other images



Ranking result:
... I should report domestic abuse because ignoring the problem will not make anything better
... I should buy this makeup because it causes love

Method

- Basic image-text triplet embedding
 - The distance between an image and its corresponding statement should be smaller than the distance between that image and any other statement, or between other images and that statement.



I should stop smoking because it can save my life.

$$L(v, t; \theta) = \sum_{i=1}^K \left[\sum_{j \in N_{vt}(i)} [\|v_i - t_i\|_2^2 - \|v_i - t_j\|_2^2 + \beta]_+ \right] + \sum_{j \in N_{tv}(i)} [\|t_i - v_i\|_2^2 - \|t_i - v_j\|_2^2 + \beta]_+$$

image as anchor, rank statements
statement as anchor, rank images

- Image embedding using symbol regions
 - We use Huang et al., 2017 to train a region proposal network and fine-tune on symbol box annotations of Hussain et al., 2017
 - We use the bottom-up attention mechanism (Anderson et al., 2017) to aggregate features from different proposals.

Experiments

- Evaluate on the main ranking task
 - Rank of the highest-ranked true matching statement
 - Recall@3: number of correct statements ranked in the Top-3

Method	Rank		Recall@3	
	PSA	Product	PSA	Product
2-WAY NETS (Eisenschstat et al., 2017)	4.836	4.170	0.923	1.212
VSE (Kiros et al., 2015)	4.155	3.202	1.146	1.447
VSE++ (Faghri, et al., 2017)	4.139	3.110	1.197	1.510
HUSSAIN (Hussain et al., 2017)	3.854	3.093	1.258	1.515
ADVISE (Ours)	3.013	2.469	1.509	1.725

- We show the top-5 ranked statements from the 50 candidates
- Statements in **bold** are the ones written for the image



[0.350] I should buy Revlon makeup because they are pretty and natural
[0.355] I should use Revlon lip balms and mascara because it will enhance the look of my lips and lashes
[0.392] I should buy Revlon makeup because it will enhance my features
[0.444] I should use Heinz because it does not have unnatural things in it
[0.614] I should drink this bacardi because it makes the world seem different



[0.630] I should wear a helmet because it will prevent brain damage
[0.741] I should put a helmet on my child because its preventative for head injuries
[0.791] I should put a helmet on my child because I don't want my child's head to end up like that melon
[0.869] I should but always because it will hold up to leaks
[0.898] I should eat Munch Nuts because I will go crazy over them

- Synonyms learnt by the extra constraints

Symbol	Statement	DenseCap
comfort	couch, sofa, soft	pillow, bed, blanket
speed, excitement, adventure	cool	sunglasses, sleeve, jacket
safety, danger, injury	driving	car, windshield, van
delicious, hot, food	ketchup	beer, pepper, sauce
environment, nature, adventure	wilderness, outdoors, terrain	rock
food, healthy, hunger	salads, food, salad	tomato

- Ablation study (% improvement over basic embedding)

Method	PSA		Product	
	Rank	Recall@3	Rank	Recall@3
GENERIC REGION	17%	15%	15%	11%
SYMBOL REGION	8%	5%	4%	2%
+ATTENTION	-3%	-1%	2%	2%
+SYMBOL/OBJECT	3%	3%	1%	<1%
+KB OBJECTS	1%	1%	<1%	<1%
+KB SYMBOLS	4%	3%	<1%	<1%

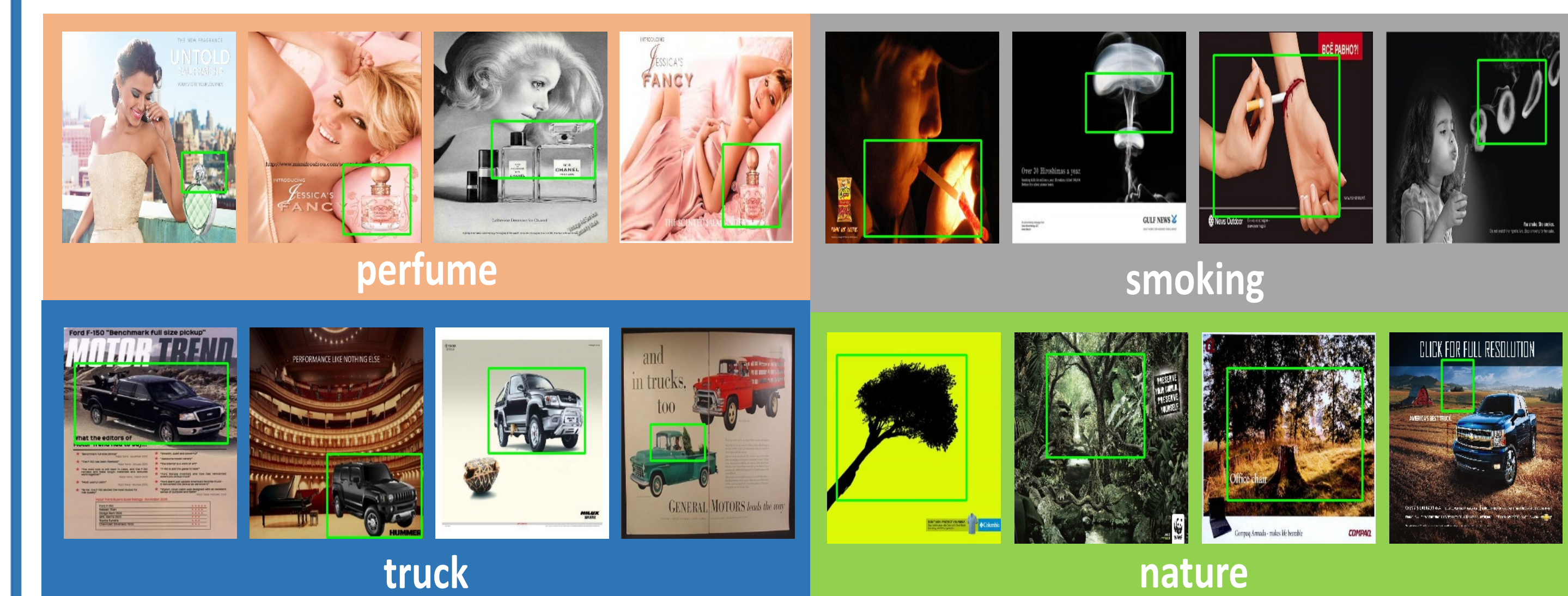
- Results on hard-statements, slogan ranking, clustering

- Hard-statements: negatives are chosen from the same ad topic
- Slogan: rank the creative captions from the PITT ads dataset
- Topic clustering: how well the models clusters ad images, wrt ground-truth clustering defined by the topics of the ads

Method	Hard statements (Rank)	Slogans (Rank)	Clustering (Homogeneity)
HUSSAIN (Hussain et al., 2017)	5.595	4.082	0.291
VSE++ (Faghri et al., 2017)	5.635	4.102	0.292
ADVISE (Ours)	4.827	3.331	0.355

- Association of image regions and words

- Given the query words, we use k-NN to retrieve the most related image regions from the test images



Acknowledgement



NSF
Grant Nr 1566270



Google
Faculty Research Award



NVIDIA
hardware grant