

# Analysis of diversity-accuracy tradeoff in image captioning

Ruotian Luo

Greg Shakhnarovich



# AIISPICE

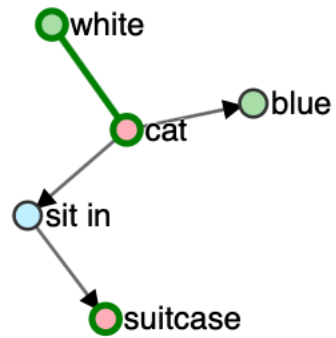
- A new metric for evaluating a set of captions, on both accuracy and diversity.



# SPICE

## Generated caption

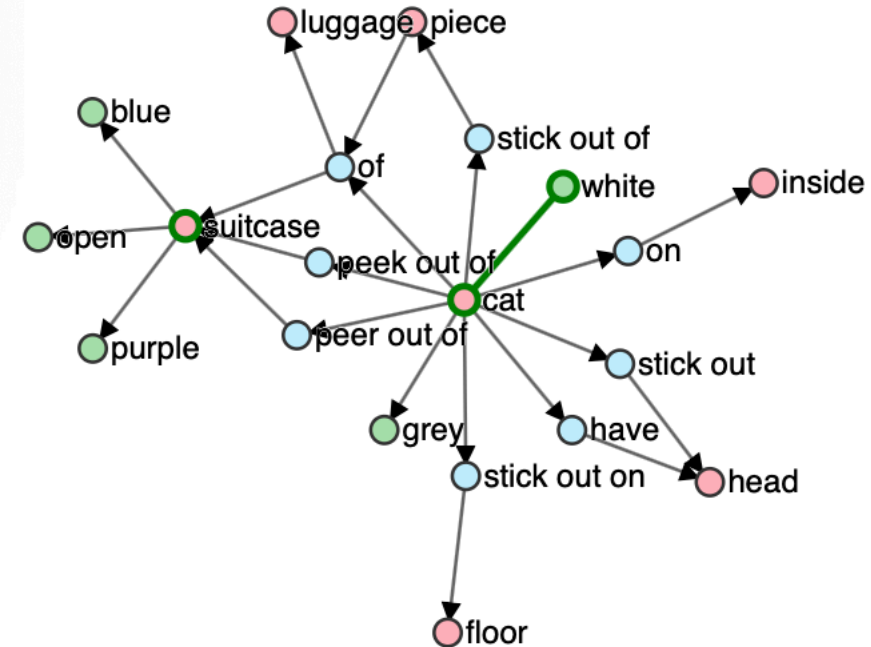
"A



suitcase"

## Reference captions

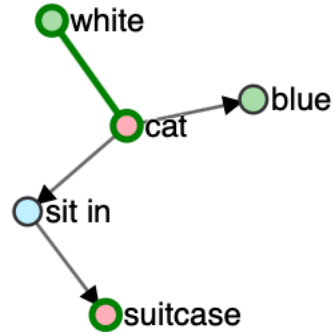
"A  
"A  
"A  
"A  
"A



on the floor."  
suitcase."

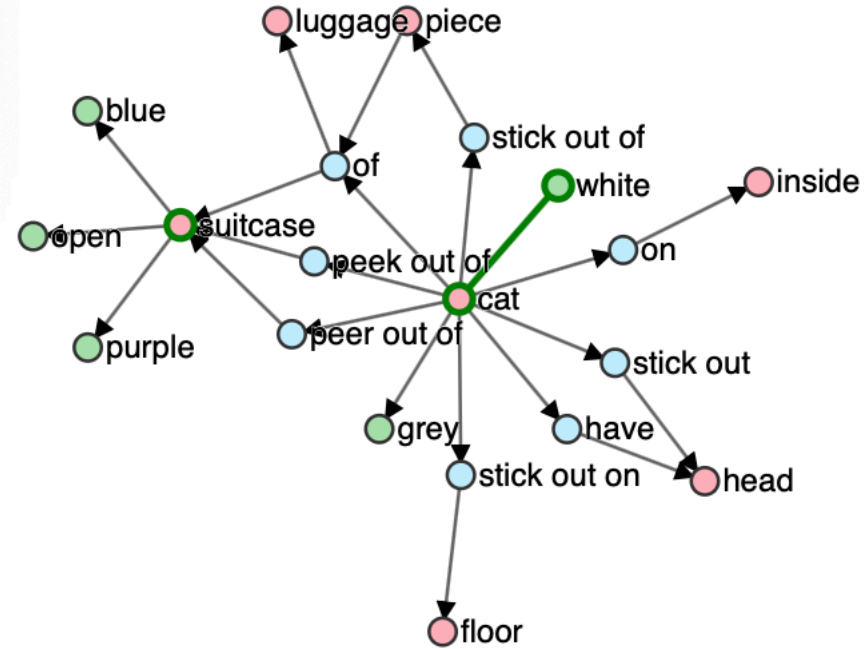


## Generated caption



"a blue and white cat sitting in a suitcase"

## Reference captions



"A cat peers out of an open suitcase."

"A cat sticking its head out of a piece of luggage on the floor."

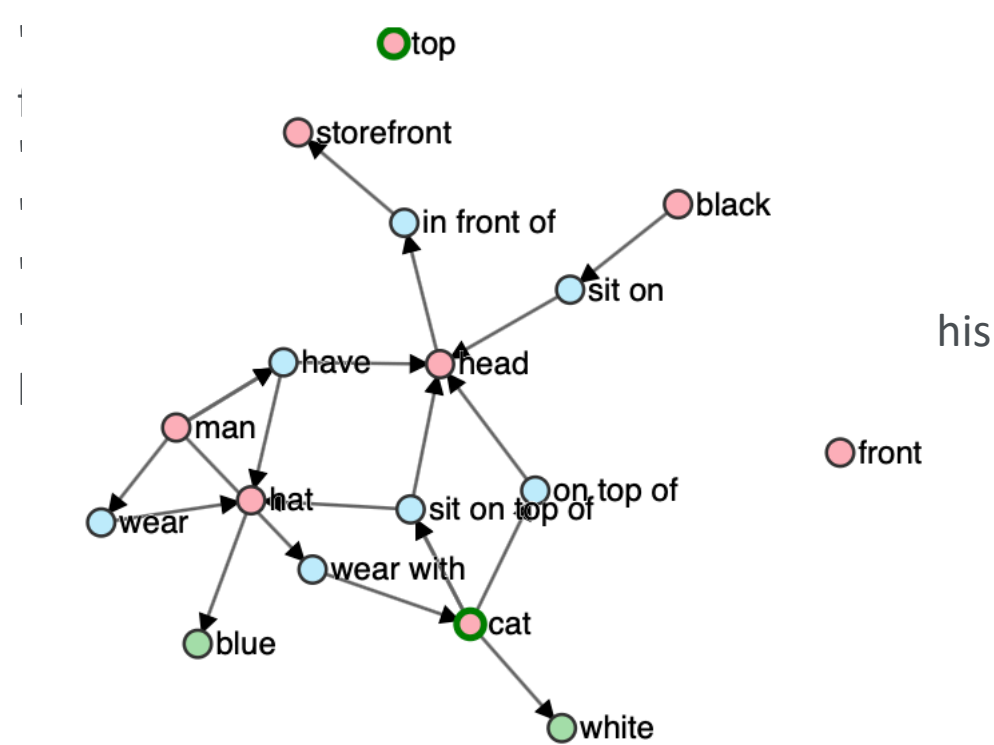
"A grey and white cat on the inside of a purple suitcase."

"A cat peeking out of a partially open suitcase."

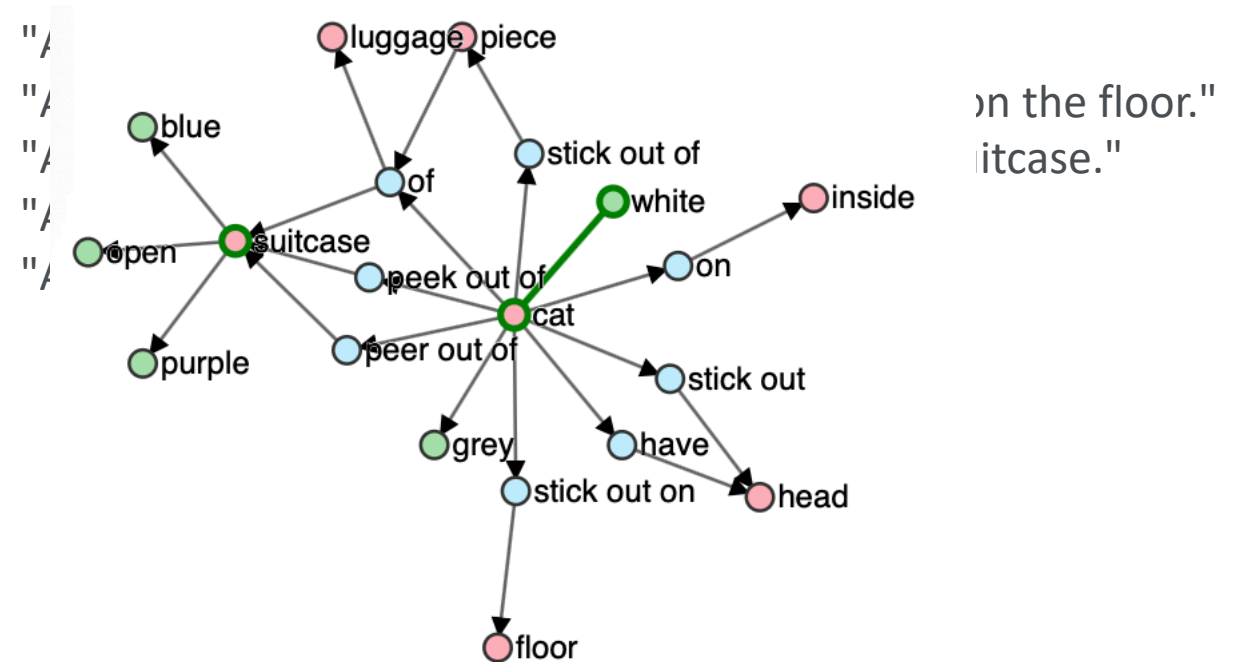
"A cat is peeking out of a blue suitcase."

# AIISPICE

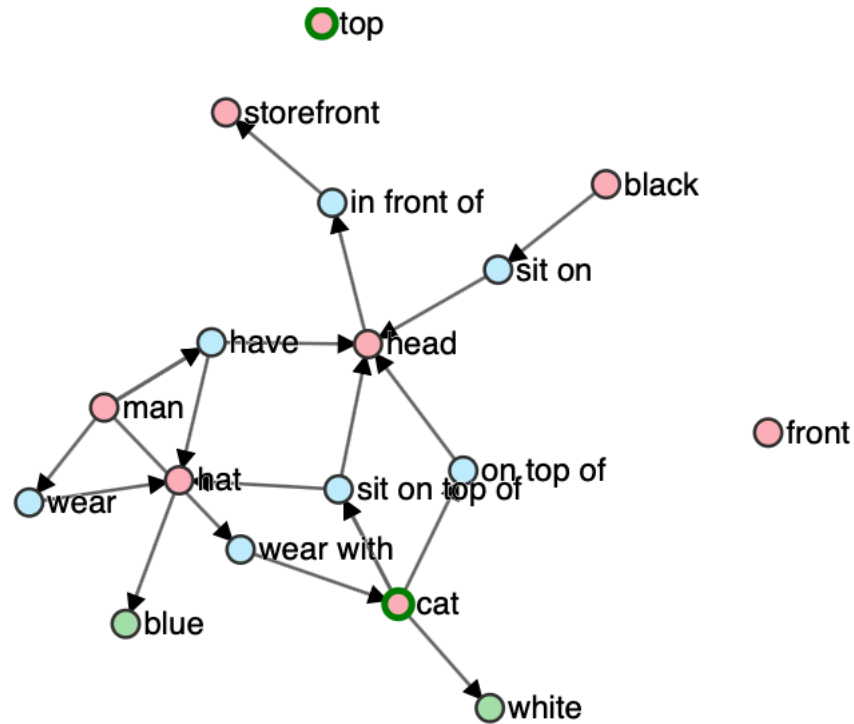
## Generated captions



## Reference captions



## Generated captions



"Black and white cat sitting on a man's head in front of a storefront."

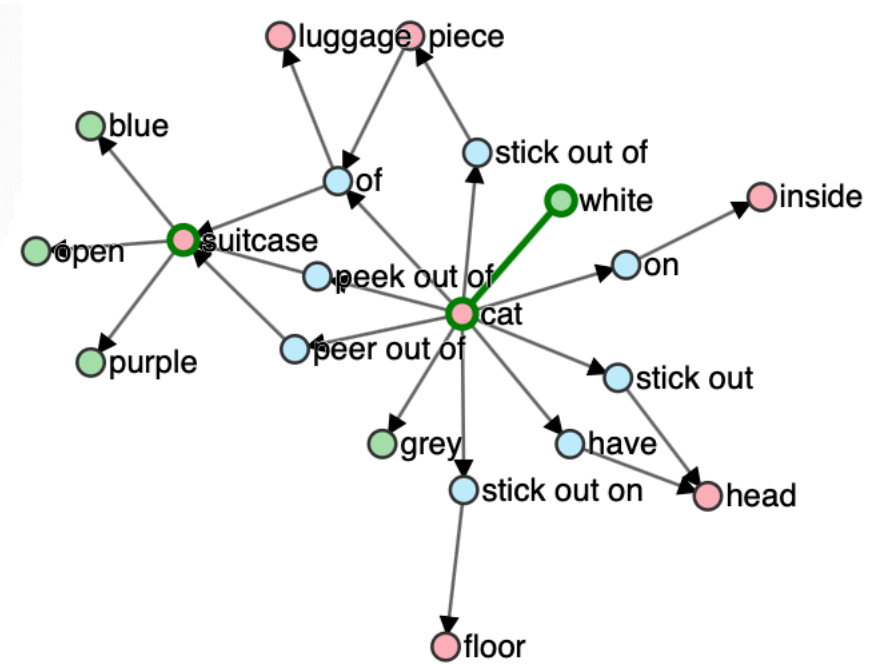
"A cat is sitting on top of a man's head."

"A cat is sitting on top of a man's hat."

"The cat is sitting on top of a man's head."

"A man wearing a blue hat with a cat on top of his head."

## Reference captions



"A cat peers out of an open suitcase."

"A cat sticking its head out of a piece of luggage on the floor."

"A grey and white cat on the inside of a purple suitcase."

"A cat peeking out of a partially open suitcase."

"A cat is peeking out of a blue suitcase."

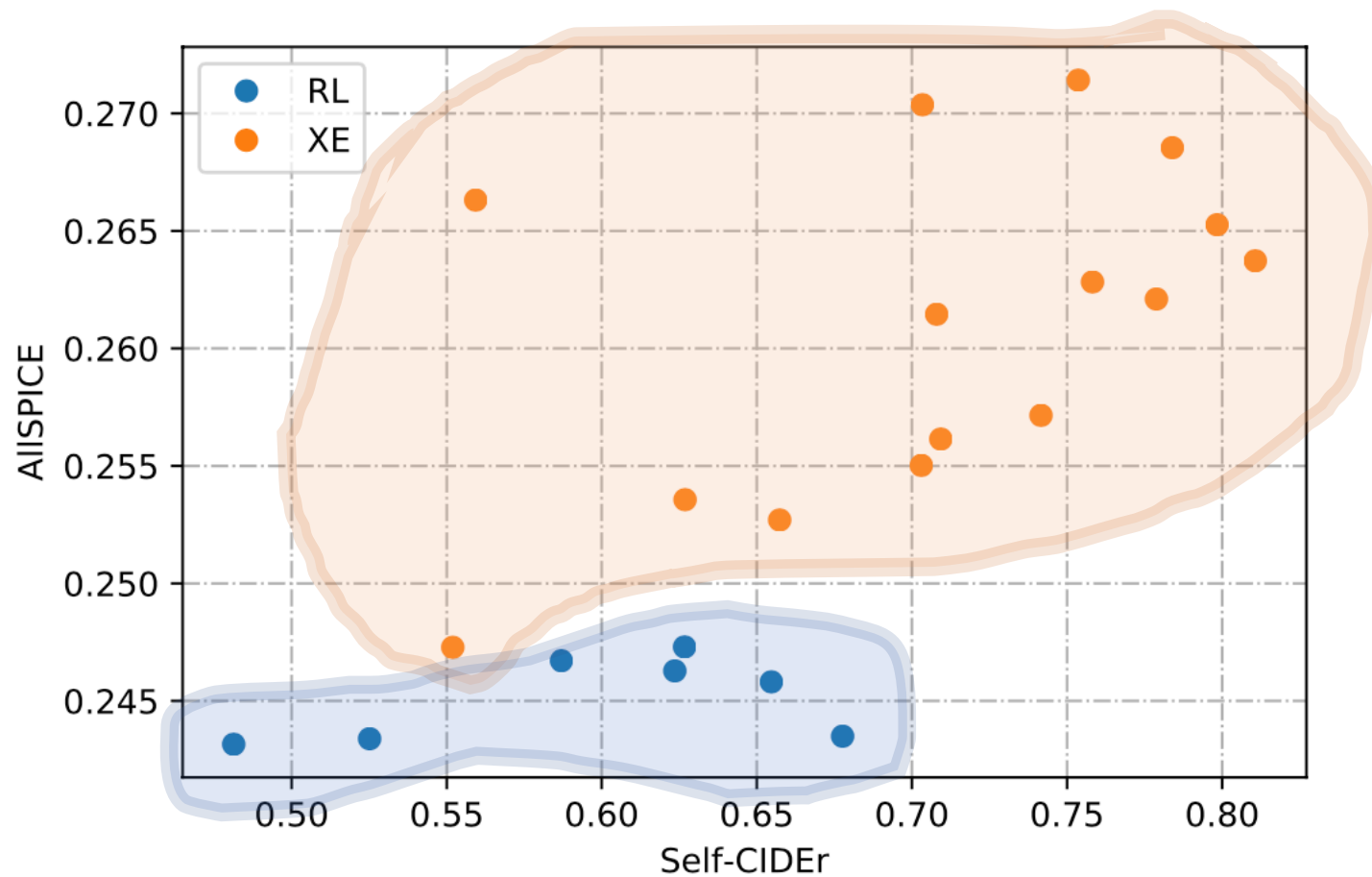
# Experiments

- Same model architecture
- Two training objectives:
  - Cross entropy loss
  - Optimize CIDEr using RL
- Sampling methods:
  - Random sampling, Top-K sampling, nucleus sampling
  - Beam search, diverse beam search



# Takeaways

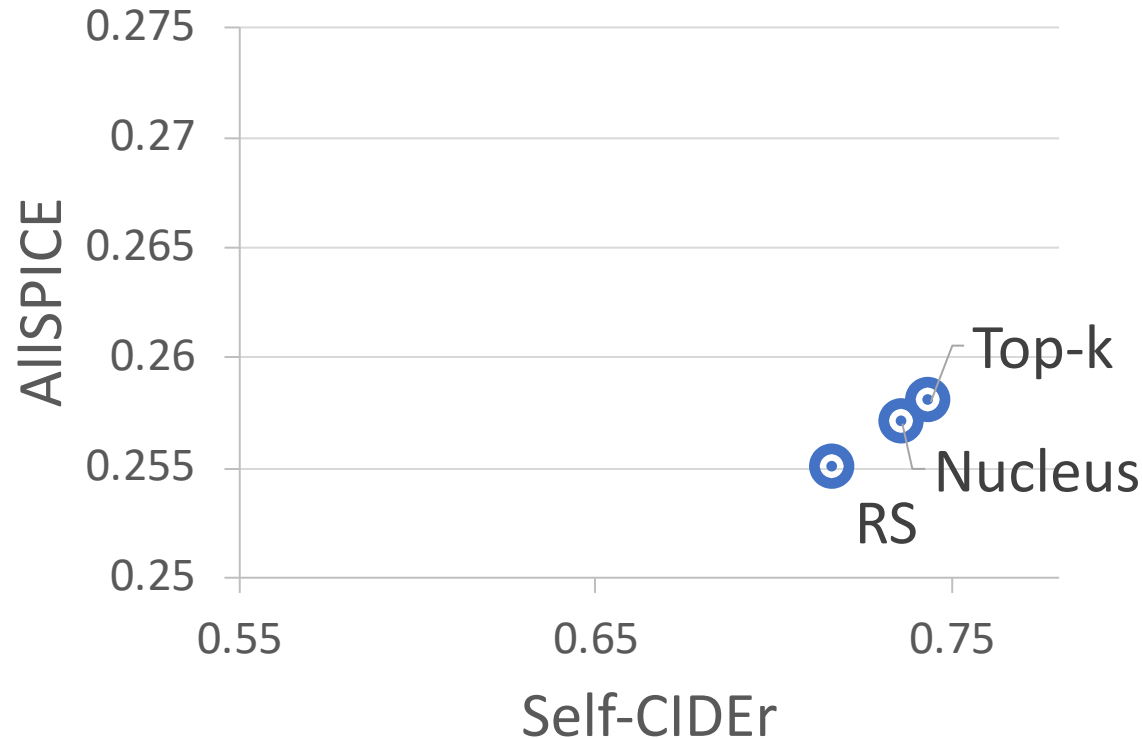
- Reinforcement learning hurts diversity.





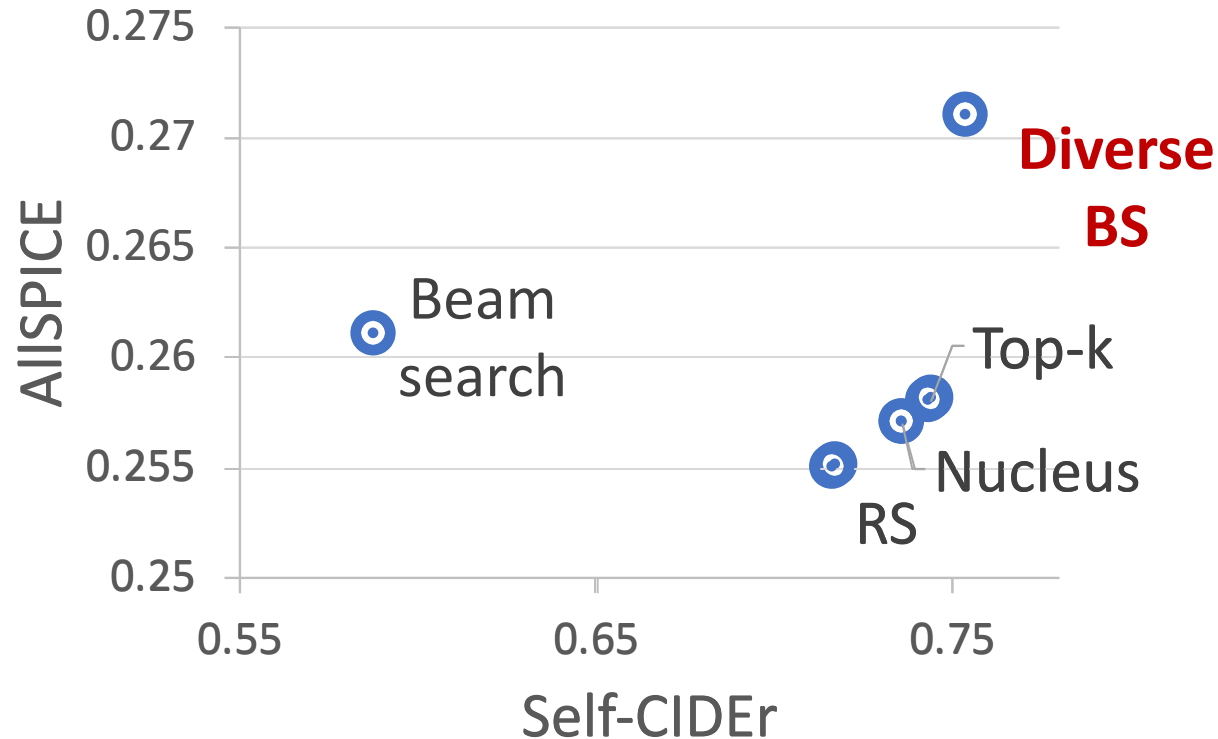
# Takeaways

- Random sampling is good with temperature lower than 1.
- Top-k and nucleus are marginally better, in **captioning**. (with the cost of one more hyperparameter)



# Takeaways

- If time allows, diverse beam search is the best.



- AllSPICE code available at:
- <https://github.com/ruotianluo/coco-caption>

