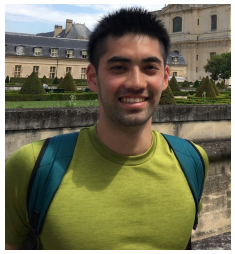
What is Learned in Visually Grounded Neural Syntax Acquisition

ACL 2020

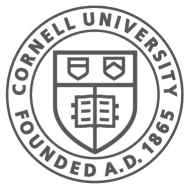


Noriyuki Kojima

Hadar Averbuch-Elor

Sasha Rush

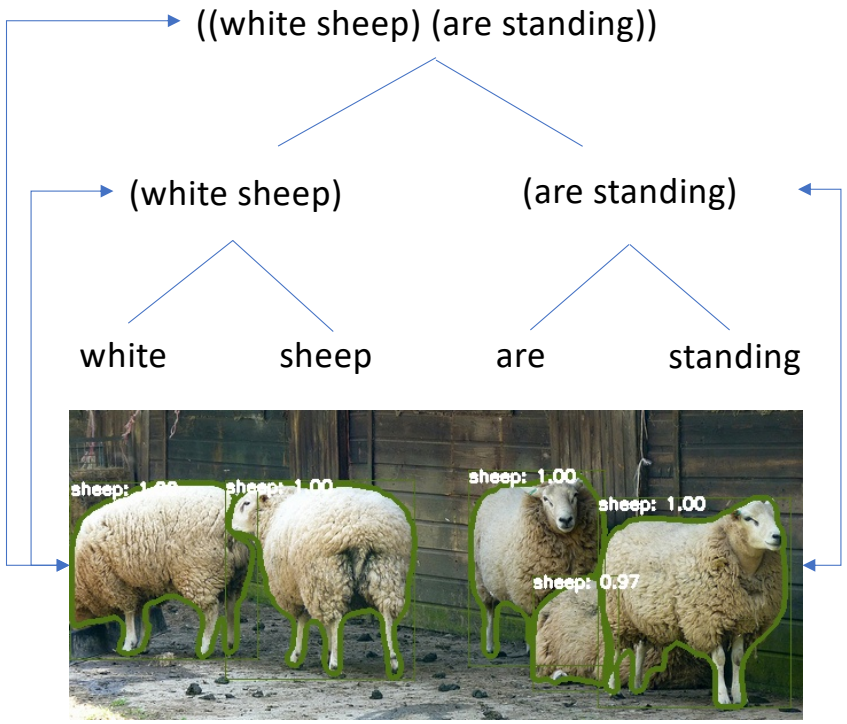
Yoav Artzi



Cornell CIS
Computer Science

CORNELL
TECH

Unsupervised Syntactic Parsing



Text-Based Syntax Acquisition
Shen et al., 2018a, 2019; Kim et al., 2019; Havrylov et al., 2019; Drozdov et al., 2019, inter alia

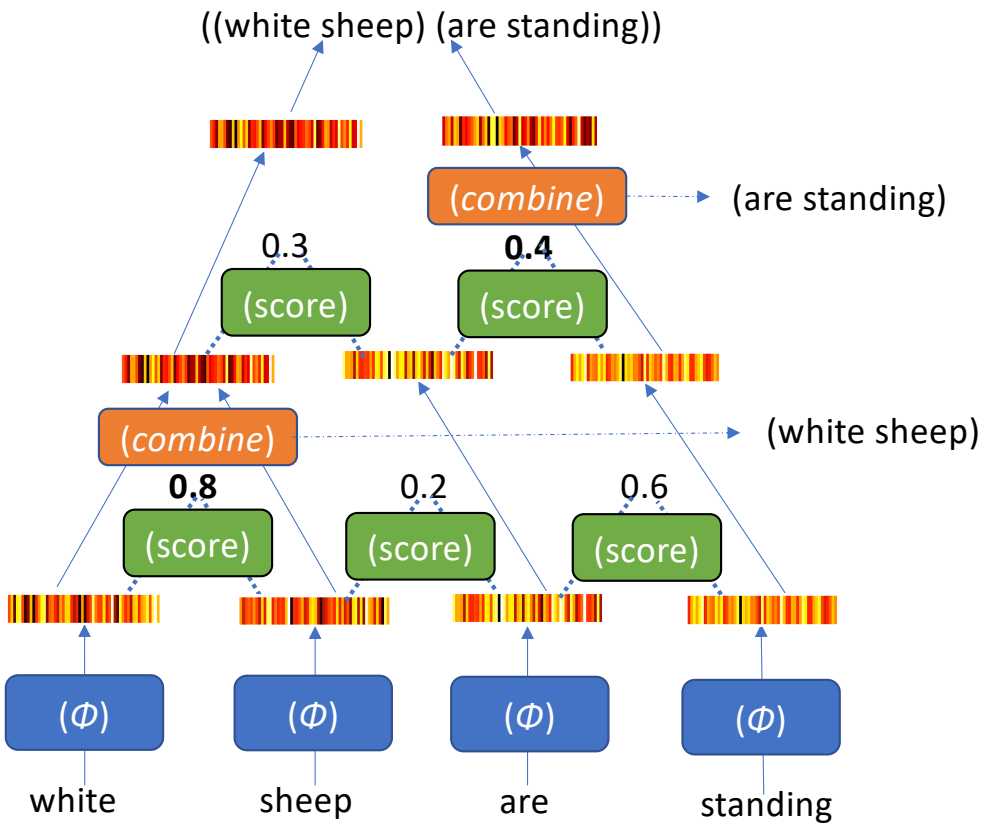
Visually Grounded Neural Syntax Learner (VG-NSL)
Shi et al., 2019

Our Work

Question: What does VG-NSL learn?

- Reduce the expressivity of the architecture
- Our significantly less expressive architectures learn similar models
- Observe that VG-NSL largely models noun concreteness

VG-NSL

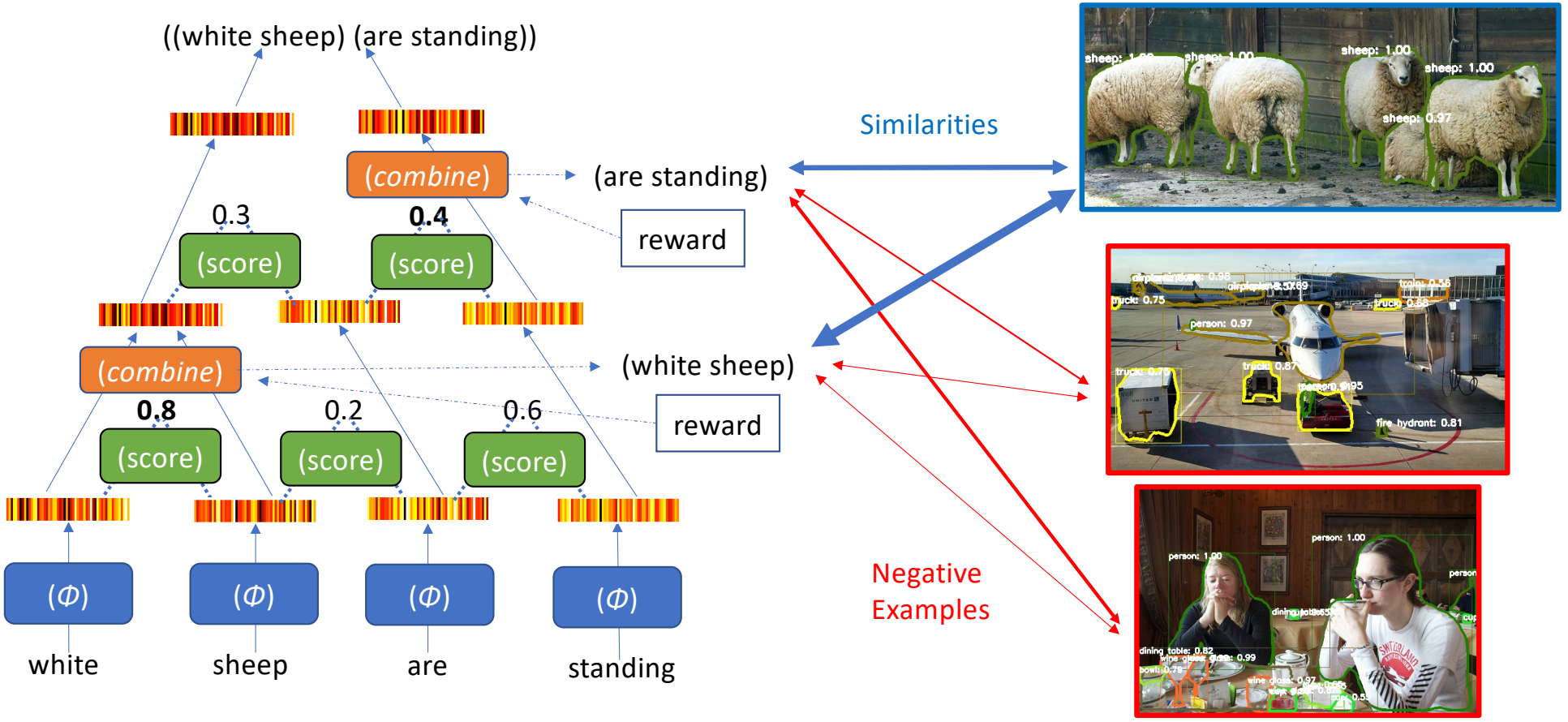


(Φ) Token embeddings

(score) Scoring function

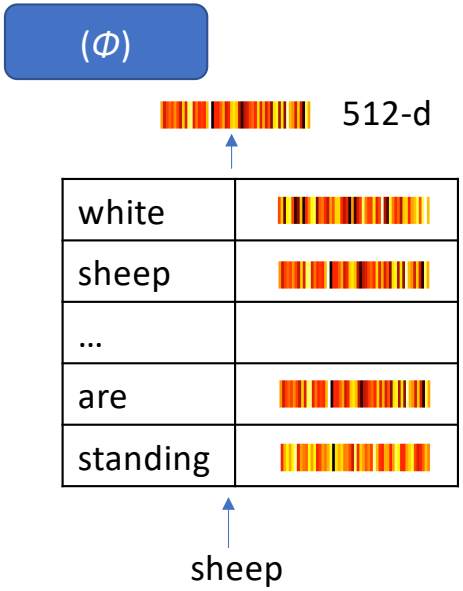
(*combine*) Combination function

VG-NSL

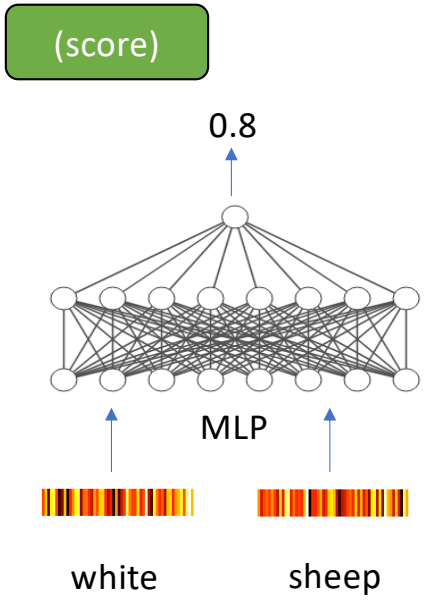

((white sheep) (are standing))
(combine)
(are standing)
reward
0.3
0.4
(score)
(score)
(combine)
(white sheep)
reward
0.8
0.2
0.6
(score)
(score)
(score)
(Φ)
(Φ)
(Φ)
(Φ)
white
sheep
are
standing
Similarities
Negative Examples

VG-NSL

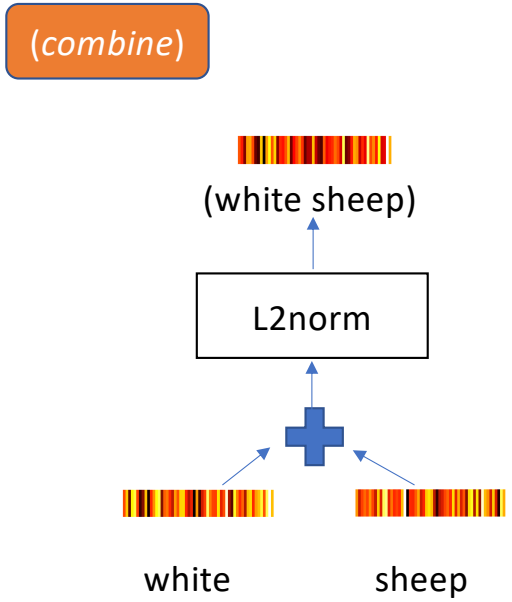
Token embeddings



Scoring function

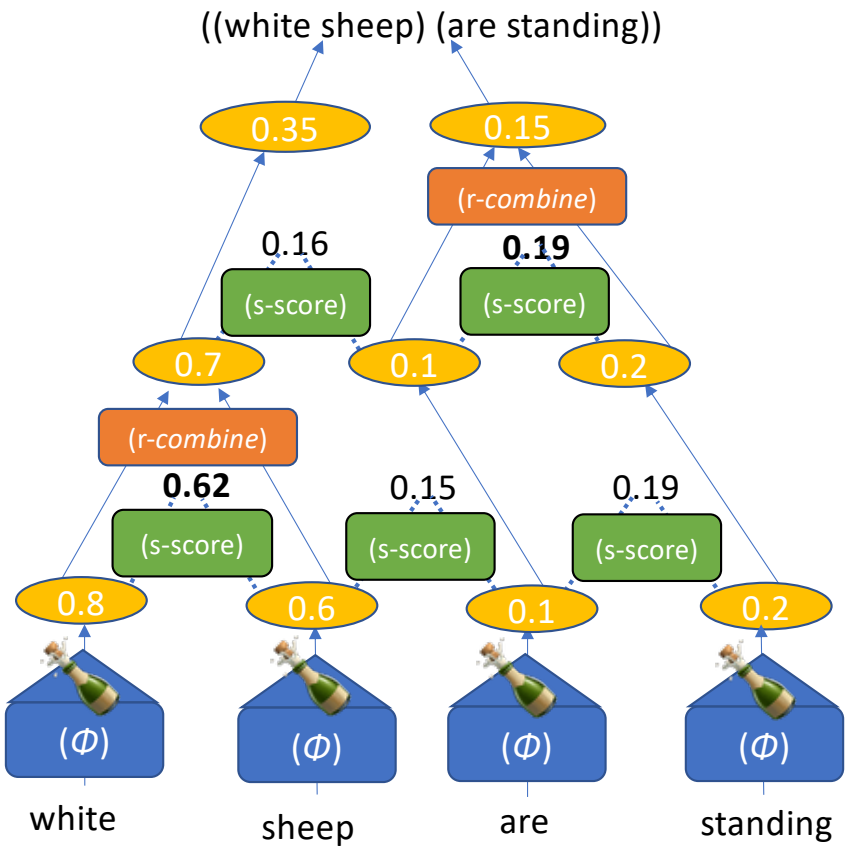


Combination function



Our approach: simplify each module to constrain the model

Our Simplified Variants



Embedding bottleneck: reduce the dimensionality of token embeddings from 512-d to 1-d/2-d

(s-score) **Simplified scoring**: parameterized weighted-sum

U · 0.8 + V · 0.6 ⟶ 0.62

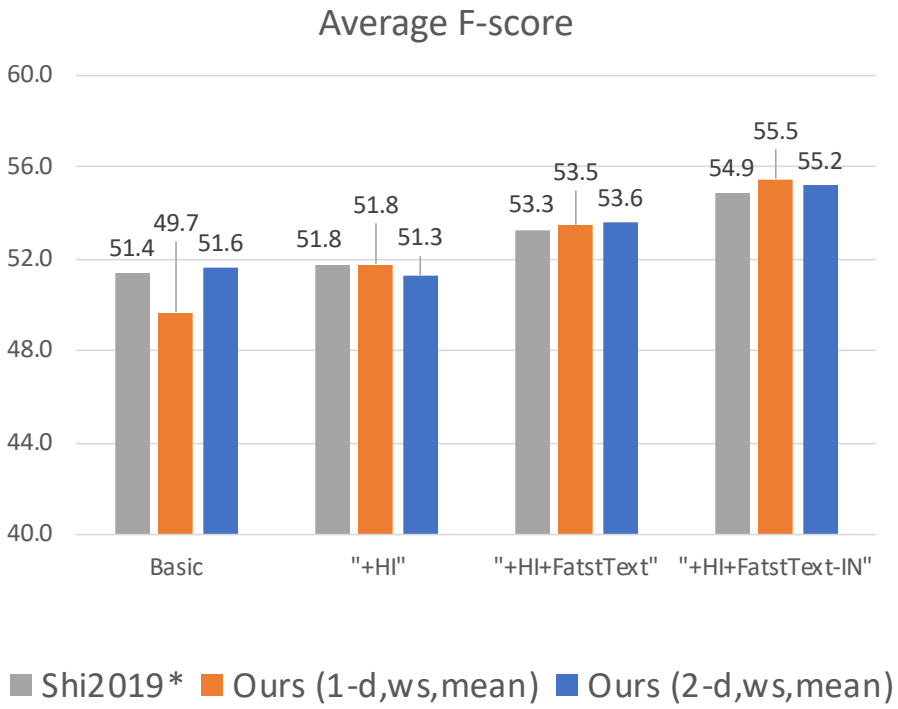
(r-combine) **Simplified combine**: mean pooling

Experiments

- Follow the experimental setup of Shi et al., 2019.
  - Data: MSCOCO (Lin et al., 2014)
  - Gold trees: Benepar (Kitaev and Klein, 2018)
  - Token embeddings: fastText (Joulin et al., 2016)

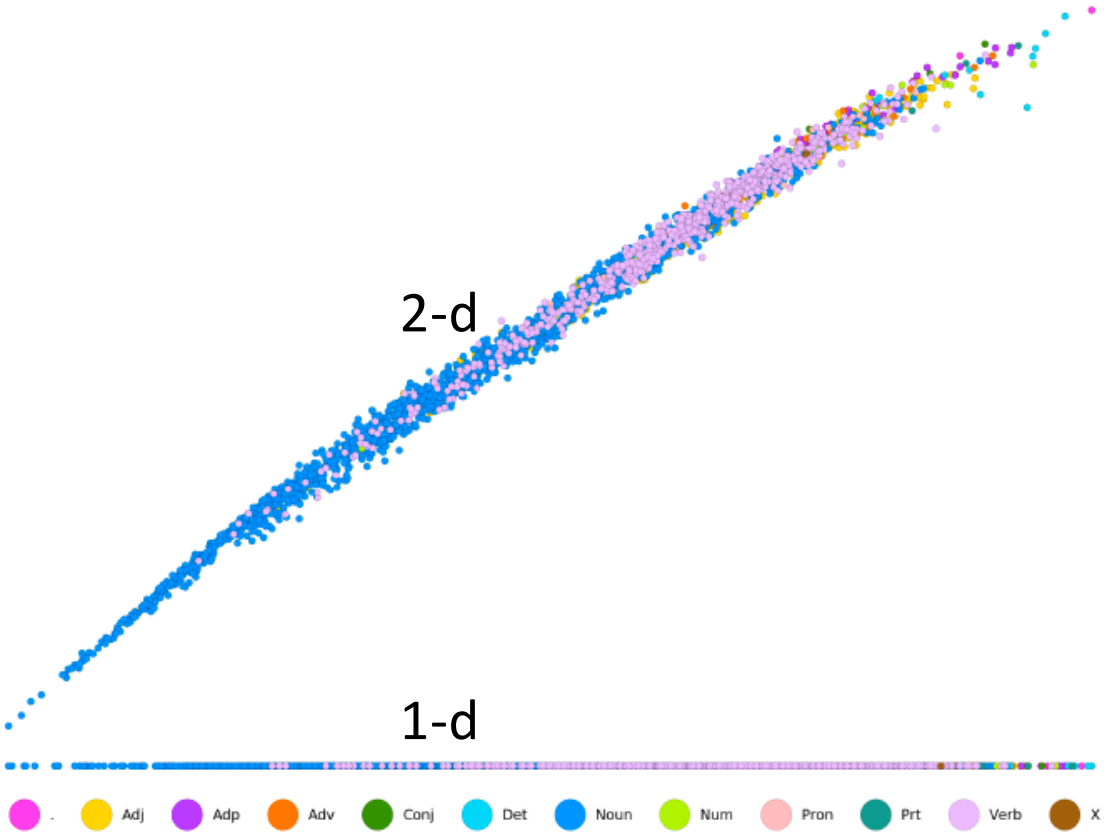
- Evaluate parsing performance using F score
  - Based on overlaps of constituents in the model predictions and gold trees

Parsing Performance



- Our variants consistently achieve comparable performance to VG-NSL across different training setup
- Our variants learn nearly identical models to VG-NSL

Noun Concreteness

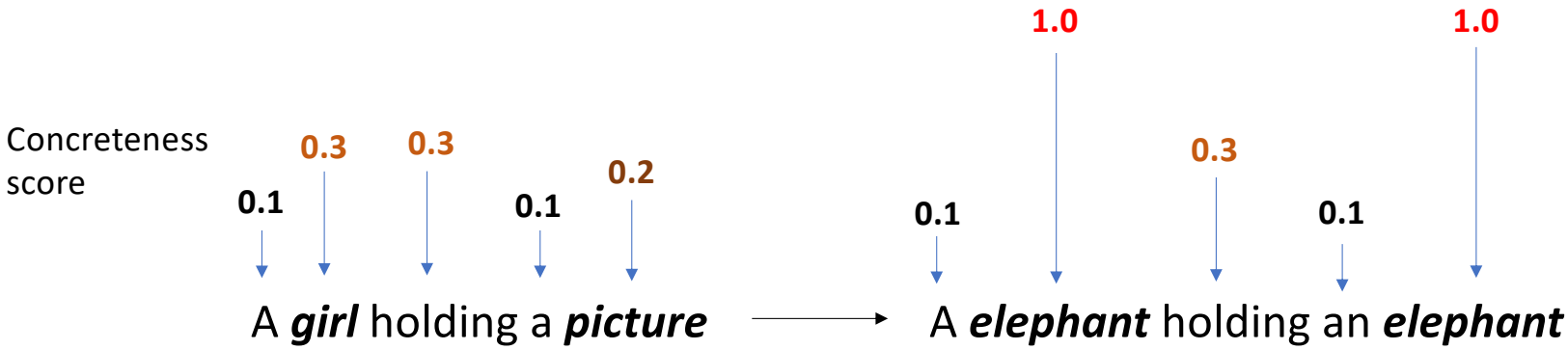


- The visualization of token embeddings shows a strong preference for separating nouns from other parts of speech

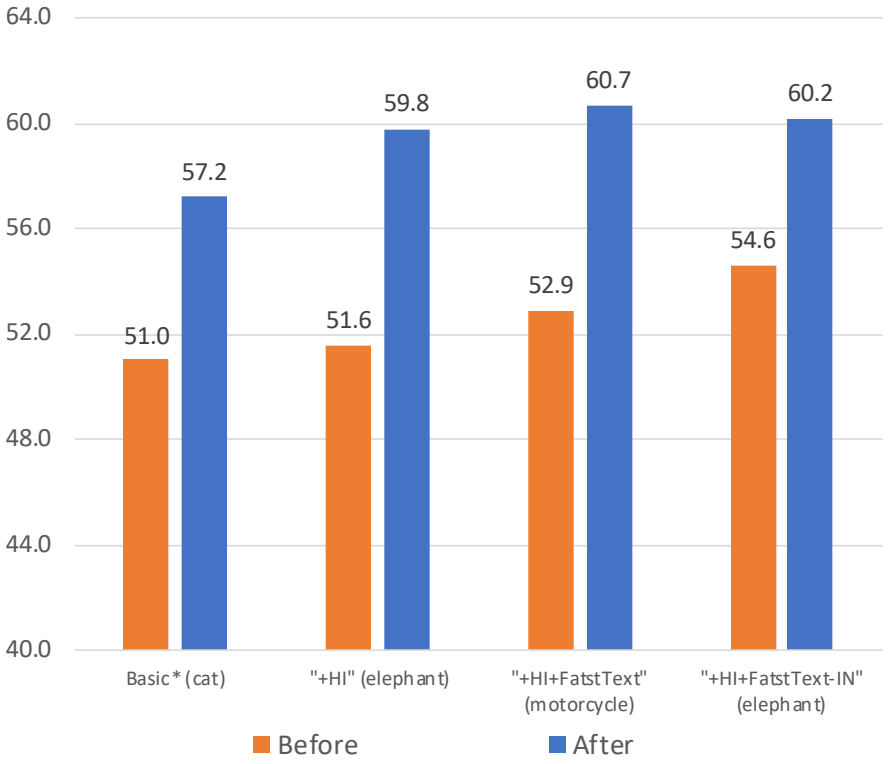
Noun Concreteness

Hypothesis: noun identification via concreteness plays a central role in VG-NSL performance

- Modify test-time captions to maximize the alignment between noun and concreteness



Noun Concreteness



- Parsing performance improves significantly
- Noun identification via concreteness provides an effective parsing strategy

Conclusion

- We introduce significantly less expressive variants of VG-NSL, maintaining similar performance and predictions
- We identify the key signal learned is noun concreteness
- Our method of analysis is general and applicable beyond parsing

Code: https://github.com/lil-lab/vgnsl_analysis_cleaning