

ReaSCAN: Compositional Reasoning in Language Grounding

Zhengxuan Wu (Stanford University), Elisa Kreiss (Stanford University)
Desmond C. Ong (National University of Singapore)
Christopher Potts (Stanford University)



Zhengxuan Wu



Elisa Kreiss



Desmond C. Ong



Christopher Potts



Project Website



School of
Computing

Compositional Generalization in Language

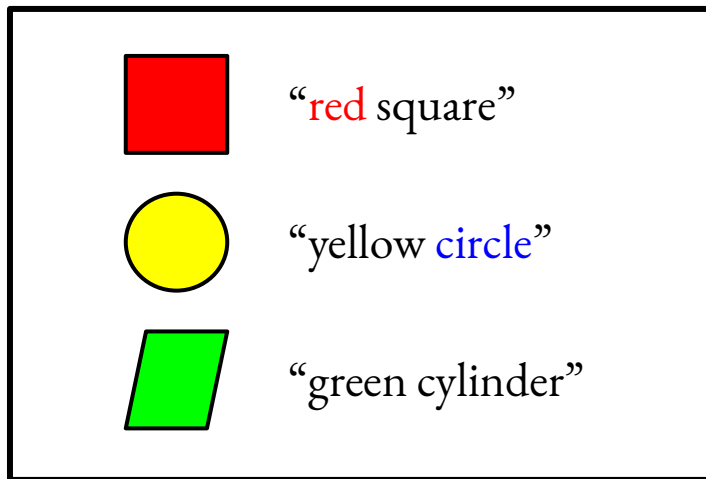
- An abstract example.
- An real-world example.

The ability to understand novel composites of known concepts.

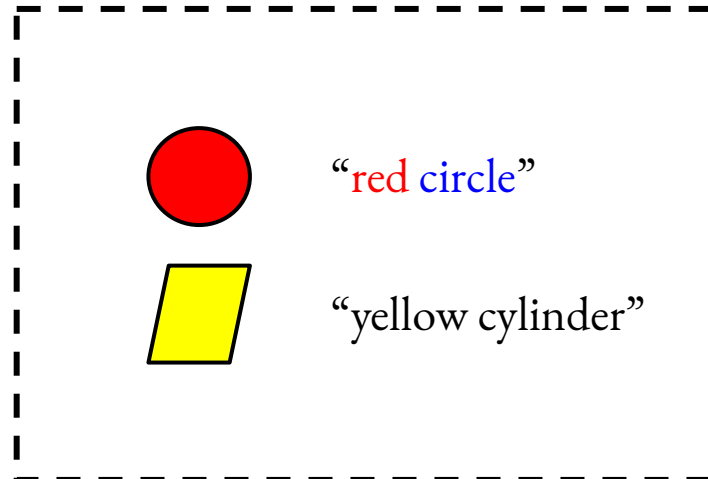
Compositional Generalization in Language

- An abstract example.
- An real-world example.

The ability to understand novel composites of known concepts.



Known Combinations of Concepts



Novel Combinations of Concepts

Compositional Generalization in Language

- An abstract example.
- An real-world example.

The ability to understand novel composites of known concepts.



Known Combinations of Concepts

“hand me the toy that is in the same color as the brush”

Novel Combinations of Concepts

“hand me the bag that is in the same color as the toy”

ReaSCAN Compared with Previous Benchmarks

- Generalization Splits & Grounded.
- Syntax Learning & Complexity.
- Extensibility.

Benchmarks	Generalization		Syntax Learning	Active Distractors	Scalable Difficulty
	Splits	Grounded			
SCAN (Lake & Baroni, 2018)	😊		😊		
gSCAN (Ruis et al., 2019)	😊	😊			
ReaSCAN (Ours)	😊	😊	😊	😊	😄

ReaSCAN Compared with Previous Benchmarks

- Generalization Splits & Grounded.
- Syntax Learning & Complexity.
- Extensibility.

Benchmarks	Generalization		Syntax Learning	Active Distractors	Scalable Difficulty
	Splits	Grounded			
SCAN (Lake & Baroni, 2018)	😊		😊		
gSCAN (Ruis et al., 2019)	😊	😊			
ReaSCAN (Ours)	😊	😊	😊	😊	😍

ReaSCAN Compared with Previous Benchmarks

- Generalization Splits & Grounded.
- **Syntax Learning & Complexity.**
- Extensibility.

Benchmarks	Generalization		Syntax	Active	Scalable
	Splits	Grounded	Learning	Distractors	
SCAN (Lake & Baroni, 2018)	😊		😊		
gSCAN (Ruis et al., 2019)	😊	😊			
ReaSCAN (Ours)	😊	😊	😊	😊	😄

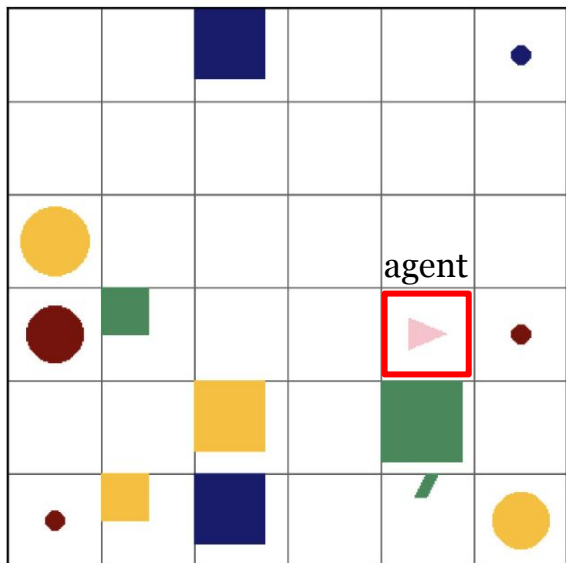
ReaSCAN Compared with Previous Benchmarks

- Generalization Splits & Grounded.
- Syntax Learning & Complexity.
- **Extensibility.**

Benchmarks	Generalization		Syntax Learning	Active Distractors	Scalable Difficulty
	Splits	Grounded			
SCAN (Lake & Baroni, 2018)	😊		😊		
gSCAN (Ruis et al., 2019)	😊	😊			
ReaSCAN (Ours)	😊	😊	😊	😊	😄

ReaSCAN Overview

- Address limitations of previous works.
- More scalable and complex set-up.



RTURN, LTURN, RTURN, RTURN,
LTURN, WALK, LTURN, RTURN,
RTURN, LTURN, WALK

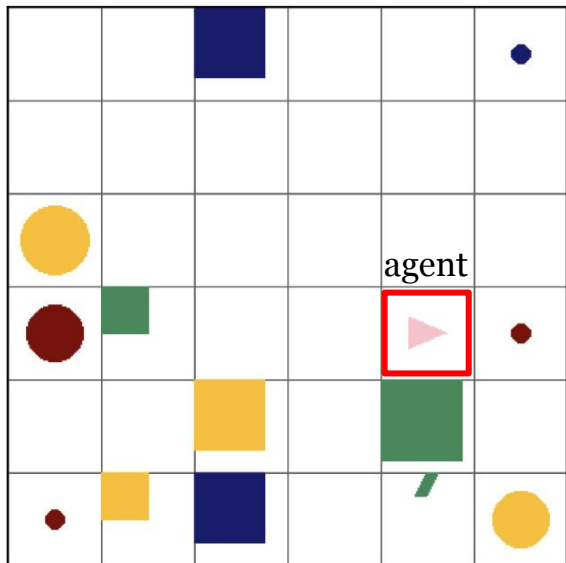
gSCAN Command: “pull the cylinder cautiously”

Our Command: “pull the object that is in the same column as the big green square cautiously”

- Addressing the issue where not all attributes are needed to identify the referent target in test splits.
- Understanding inter-object relations (six) and syntax of commands are necessary for resolving the target.
- Active distractor sampling strategies.
- Commands with extensible relative clauses with conjunctions.

ReaSCAN Overview

- Address limitations of previous works.
- **More scalable and complex set-up.**



RTURN, LTURN, RTURN, RTURN,
LTURN, WALK, LTURN, RTURN,
RTURN, LTURN, WALK

gSCAN Command: “pull the cylinder cautiously”

Our Command: “pull the object that is in the same column as the big green square cautiously”

- Addressing the issue where not all attributes are needed to identify the referent target in test splits.
- Understanding inter-object relations (six) and syntax of commands are necessary for resolving the target.
- Active distractor sampling strategies.
- Commands with extensible relative clauses with conjunctions.

ReaSCAN's Testing Split

- **Overview.**
- Testing splits A.
- Testing splits B & C.

Split	Description	Adapted from gSCAN?
A1	Novel color modifier	✓
A2	Novel color attribute	✓
A3	Novel size modifier	✓
B1	Novel co-occurrence of objects	
B2	Novel co-occurrence of relations	
C1	Novel conjunctive clause length	
C2	Novel relative clauses	

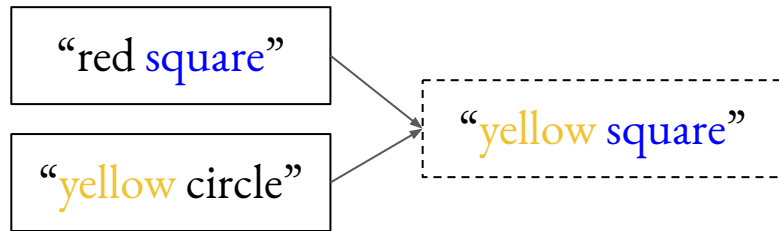
ReaSCAN's Testing Split

- Overview.
- **Testing splits A.**
- Testing splits B & C.

 seen  unseen

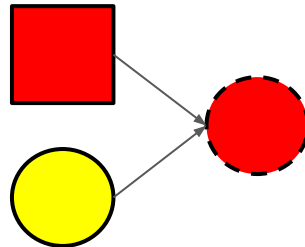
A1 & A3: Novel Modifier

Testing splits contain unseen composites of size, color and shape expressions in command.



A2: Novel Attribute

Testing splits contain unseen composites of size, color and shape for the referent target.



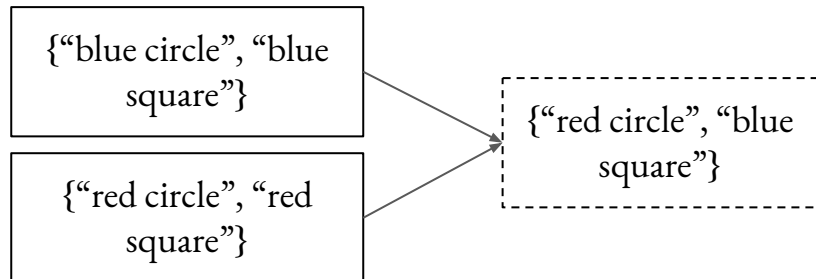
ReaSCAN's Testing Split

- Overview.
- Testing splits A.
- Testing splits B & C.

 seen  unseen

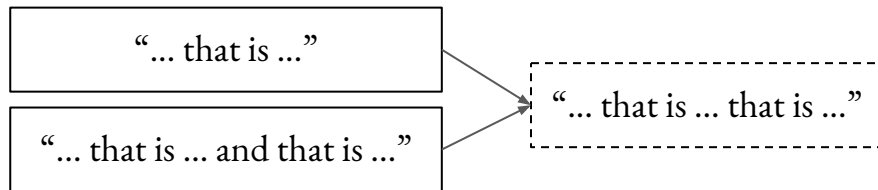
B: Novel Co-occurrence of Concept

Testing splits contain unseen co-occurrence of object expressions or object relations.



C: Novel Clause Structure

Testing splits contain commands with one more conjunctive clauses, or 2 recursive relative clauses.



Model's Performance Results

- Overview.
- gSCAN comparison.

Splits	Exact Match% ($\pm\Delta_{\text{split-rand}}$ %)
	GCN-LSTM
Random (Seen)	99.0% (−0.00%)
A1*: novel color modifier	92.3% (−6.78%)
A2*: novel color attribute	42.1% (−57.5%)
B2: novel co-occurrence of relations	52.8% (−46.6%)
C1: novel conjunctive clause length	57.0% (−42.4%)
C2: novel relative clauses	22.1% (−77.7%)

*gSCAN has similar splits but these are solved nearly perfectly by the GCN-LSTM (Gao et al. 2020) model.

Model's Performance Results

- Overview.
- gSCAN comparison.

Splits	Exact Match% ($\pm \Delta_{\text{split-rand}}$ %)
	GCN-LSTM
Random (Seen)	99.0% (−0.00%)
A1*: novel color modifier	92.3% (−6.78%)
A2*: novel color attribute	42.1% (−57.5%)
B2: novel co-occurrence of relations	52.8% (−46.6%)
C1: novel conjunctive clause length	57.0% (−42.4%)
C2: novel relative clauses	22.1% (−77.7%)

*gSCAN has similar splits but these are solved nearly perfectly by the GCN-LSTM (Gao et al. 2020) model.

Model's Performance Results

- Overview.
- **gSCAN comparison.**

Splits	Exact Match% ($\pm \Delta_{\text{split-rand}}$ %)
	GCN-LSTM
Random (Seen)	99.0% (−0.00%)
A1*: novel color modifier	92.3% (−6.78%)
A2*: novel color attribute	42.1% (−57.5%)
B2: novel co-occurrence of relations	52.8% (−46.6%)
C1: novel conjunctive clause length	57.0% (−42.4%)
C2: novel relative clauses	22.1% (−77.7%)

*gSCAN has similar splits but these are solved nearly perfectly by the GCN-LSTM (Gao et al. 2020) model.

Accept the ReaSCAN Challenge!

7 compositional generalization testing splits, and provide scripts for customized splits!

Open-source data generation framework.

ReaSCAN-v1.0 can be downloaded in our project landing page.

<https://reascan.github.io/>



Project Website