How Modular Should Neural Module Networks Be for Systematic Generalization?

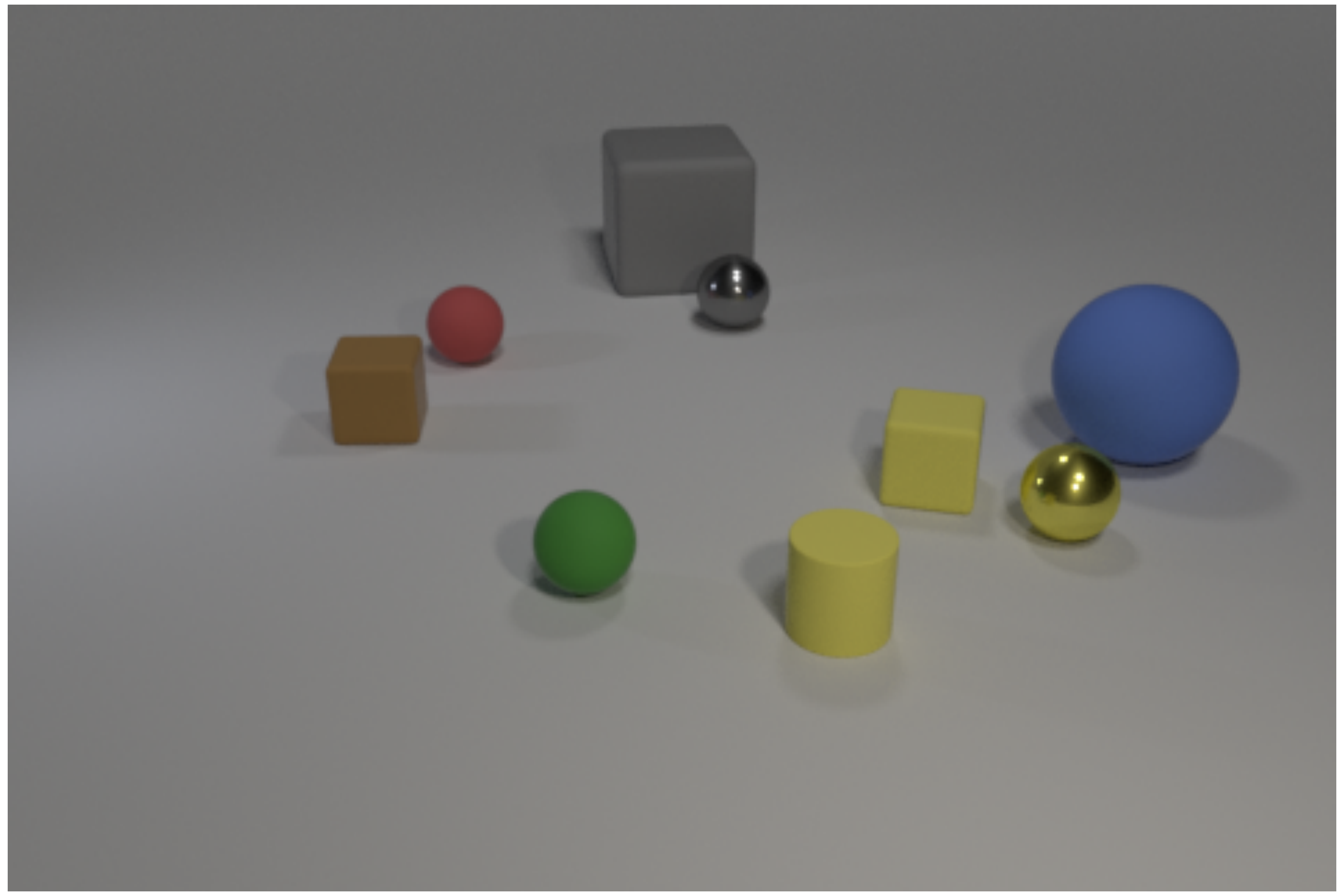
Vanessa D'Amario[1,3], Tomotake Sasaki[2,3], Xavier Boix[1,3]

1

2 FUJITSU

3 CENTER FOR Brains Minds+ Machines

Visual Question Answering (VQA)

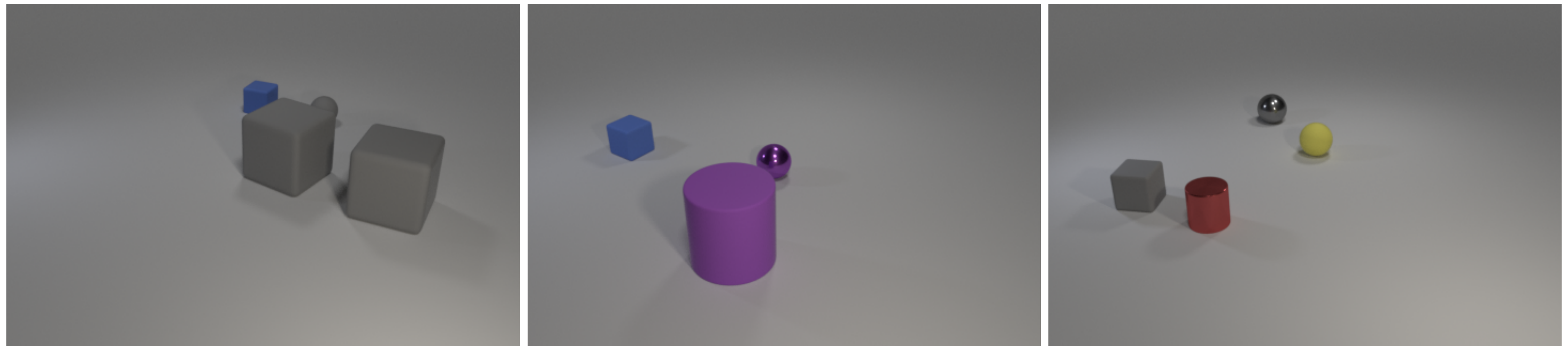


Q1: How many gray rubber cubes are the same size as the yellow block?

Q2: There is a rubber thing that is the same color as the cylinder; what shape is it?

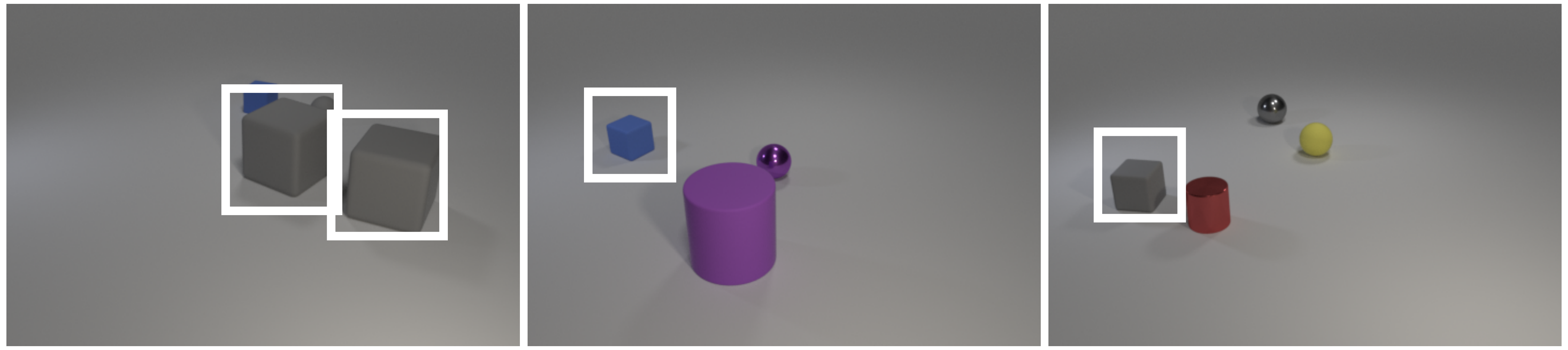
Q3: The matte ball that is the same size as the gray rubber object is what color?

Systematic generalization



Condition A

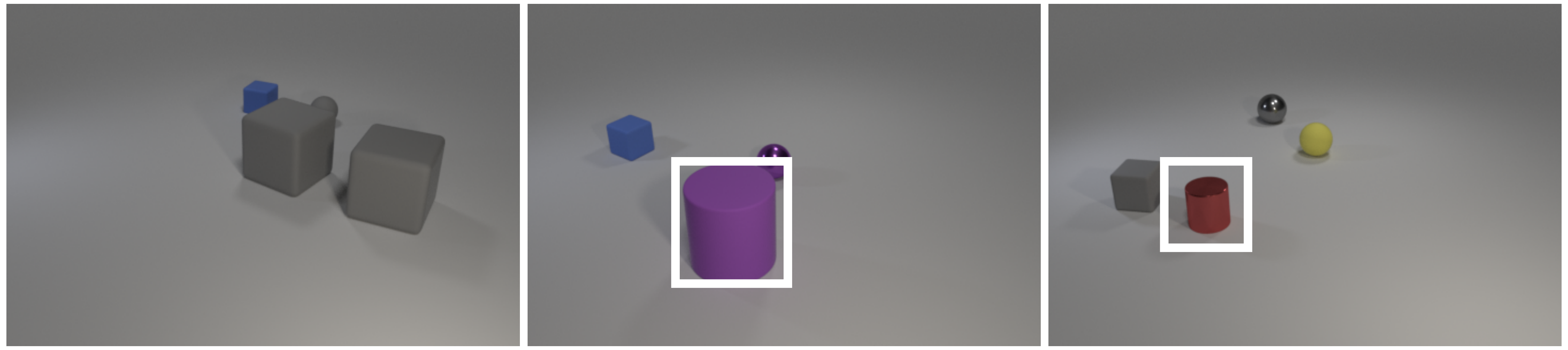
Systematic generalization



Condition A

Cubes are **gray**, **blue**, brown, or yellow.

Systematic generalization

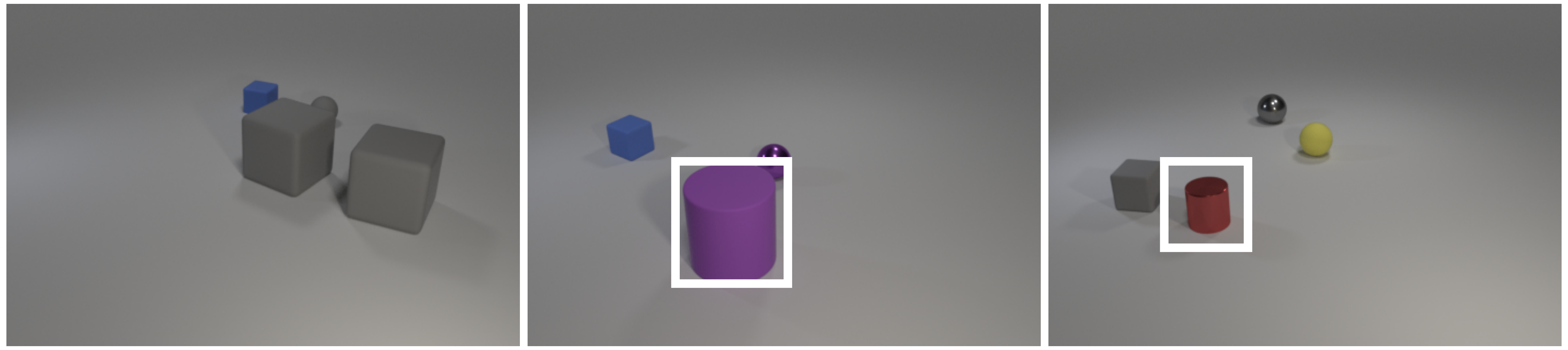


Condition A

Cubes are gray, blue, brown, or yellow.

Cylinders are **red**, green, **purple**, or cyan.

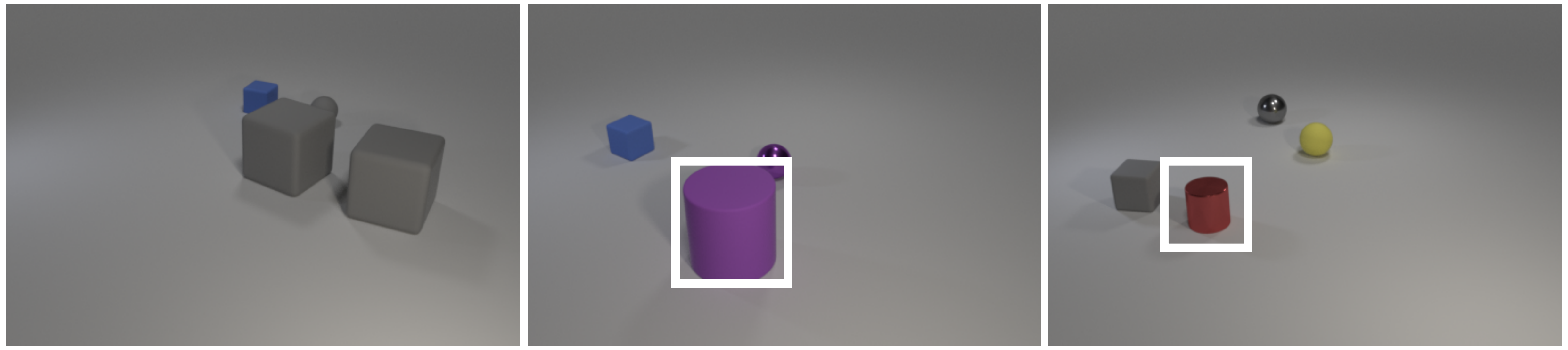
Systematic generalization



Condition A

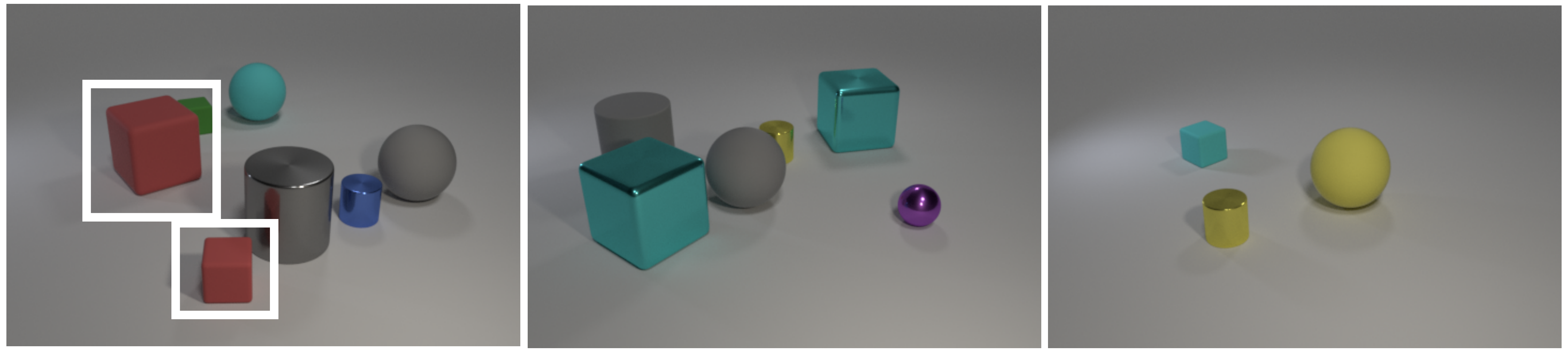
Cubes are gray, blue, brown, or yellow.
Cylinders are **red**, green, **purple**, or cyan.
Spheres are all colors.

Systematic generalization



Condition A

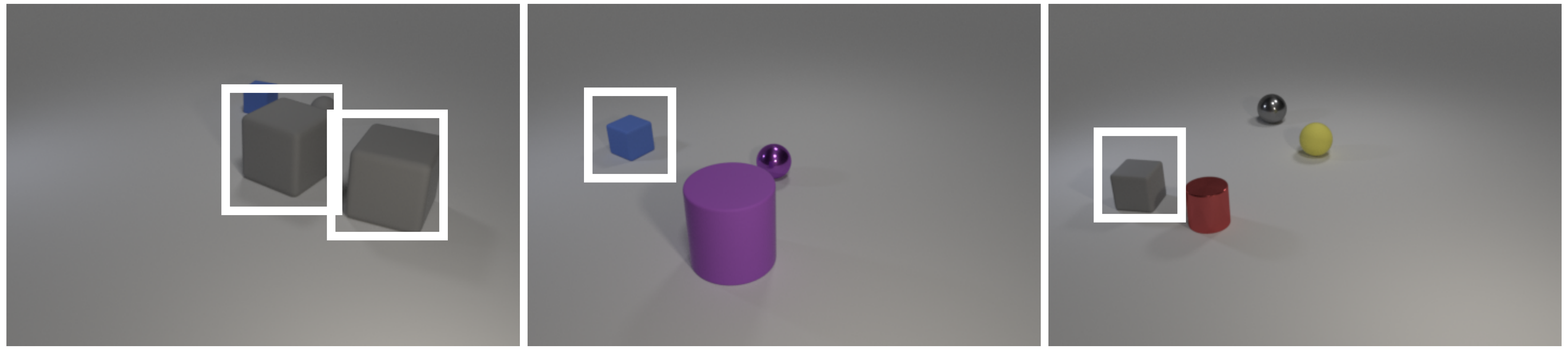
Cubes are gray, blue, brown, or yellow.
Cylinders are **red**, green, **purple**, or cyan.
Spheres are all colors.



Condition B

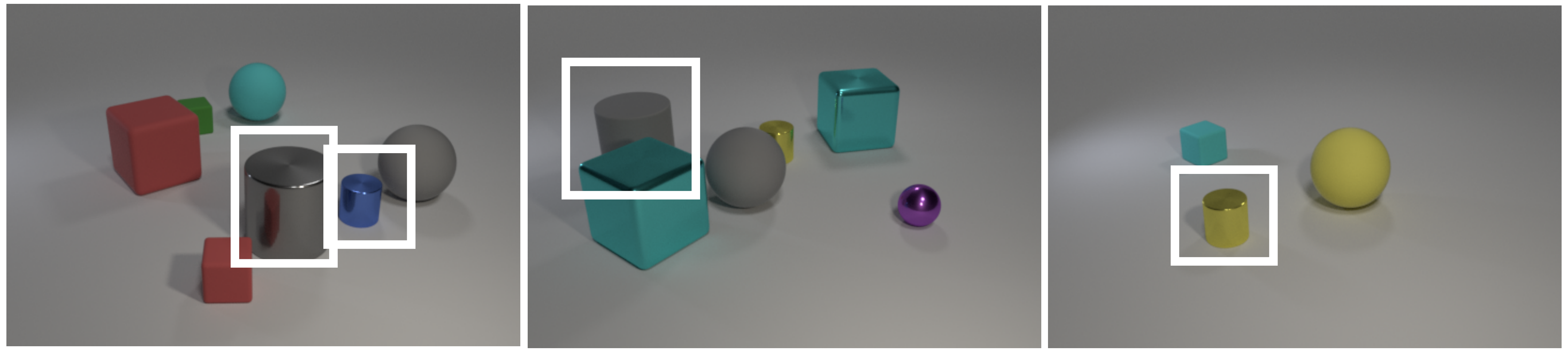
Cylinders are gray, blue, brown, or yellow.
Cubes are **red**, green, purple, or cyan.
Spheres are all colors.

Systematic generalization



Condition A

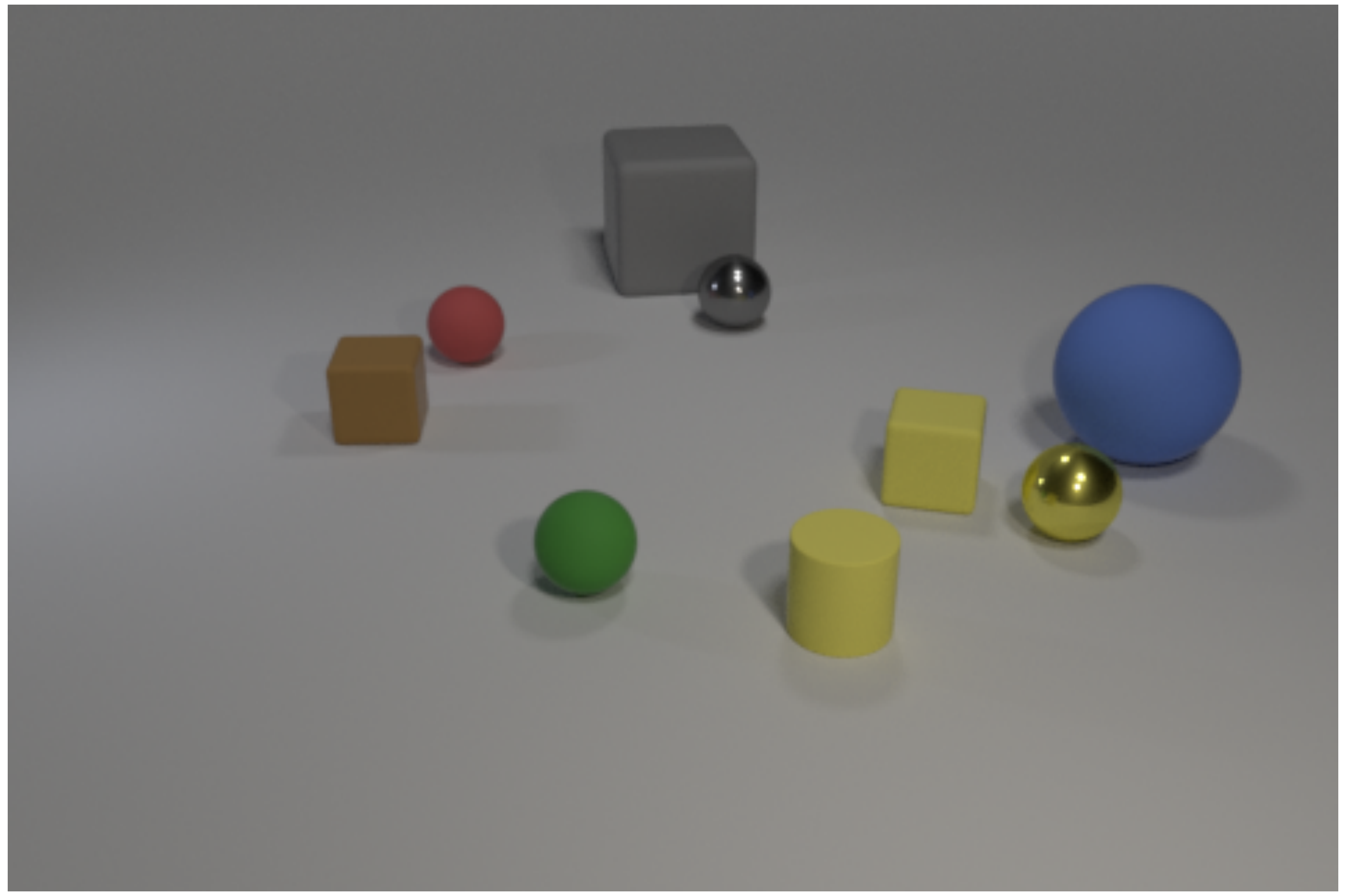
Cubes are **gray**, **blue**, brown, or yellow.
Cylinders are red, green, purple, or cyan.
Spheres are all colors.



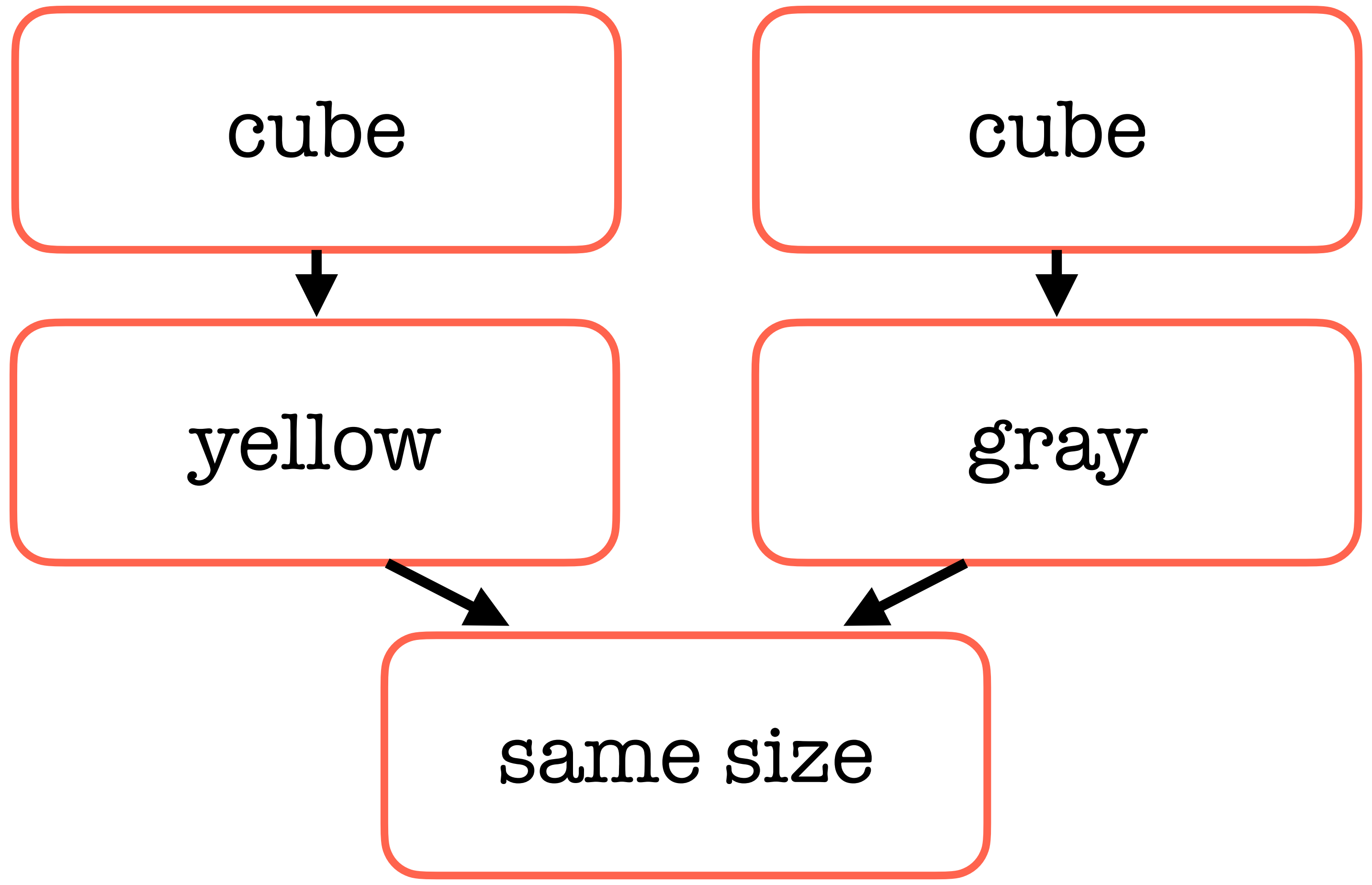
Condition B

Cylinders are **gray**, **blue**, brown, or yellow.
Cubes are red, green, purple, or cyan.
Spheres are all colors.

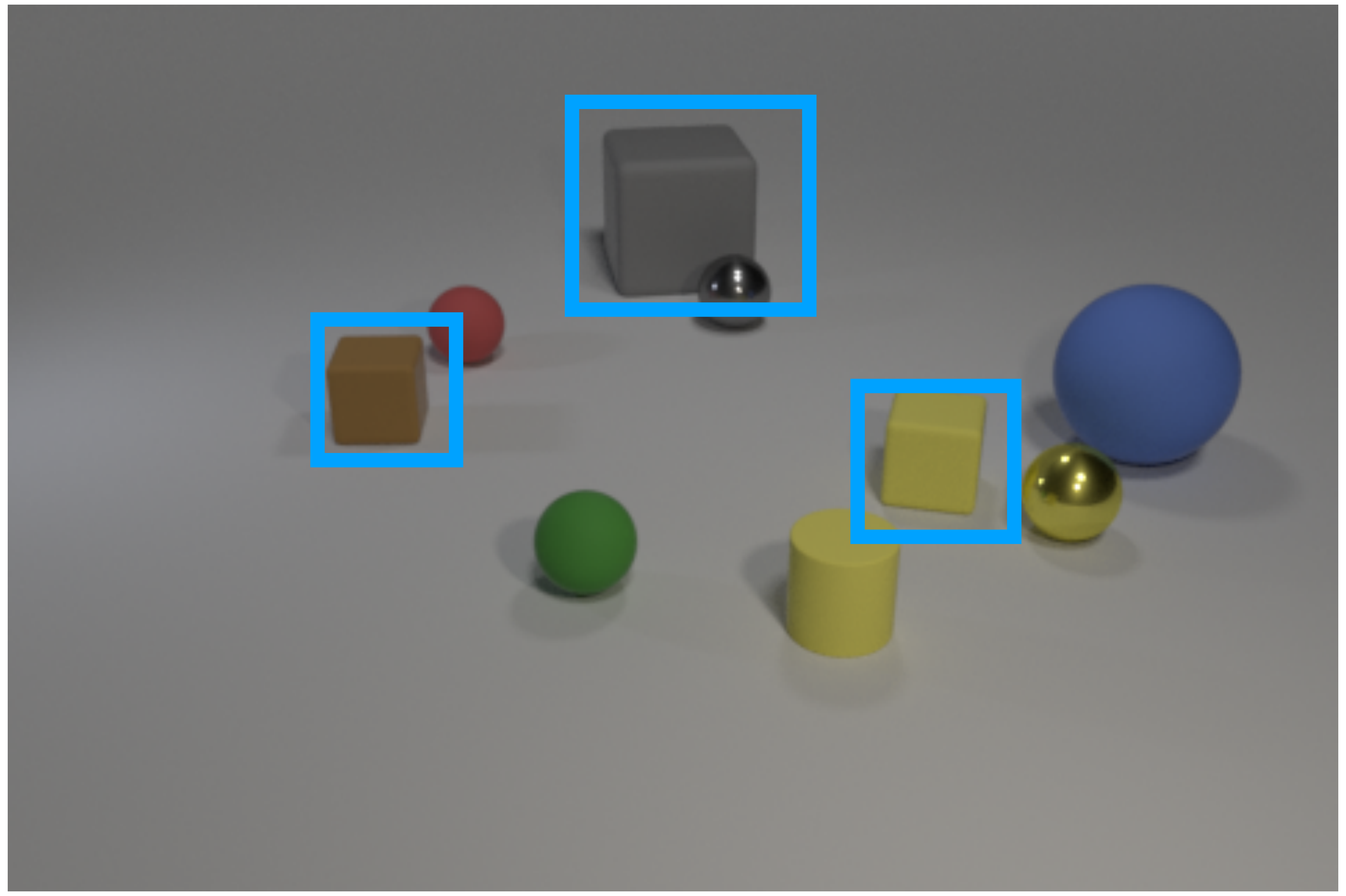
Modular approach to systematic generalization



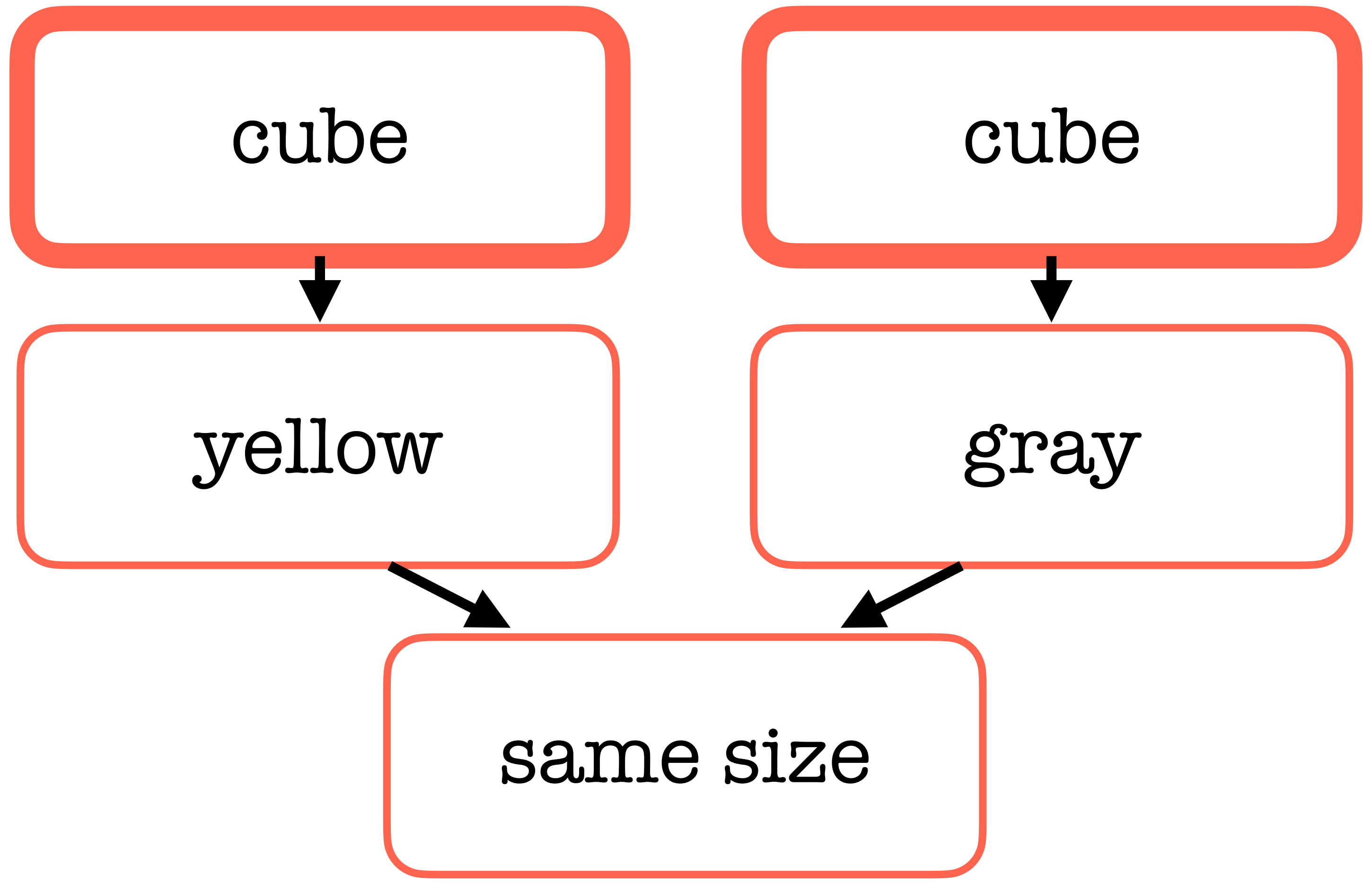
Q: Is the gray cube the same size as the yellow cube?



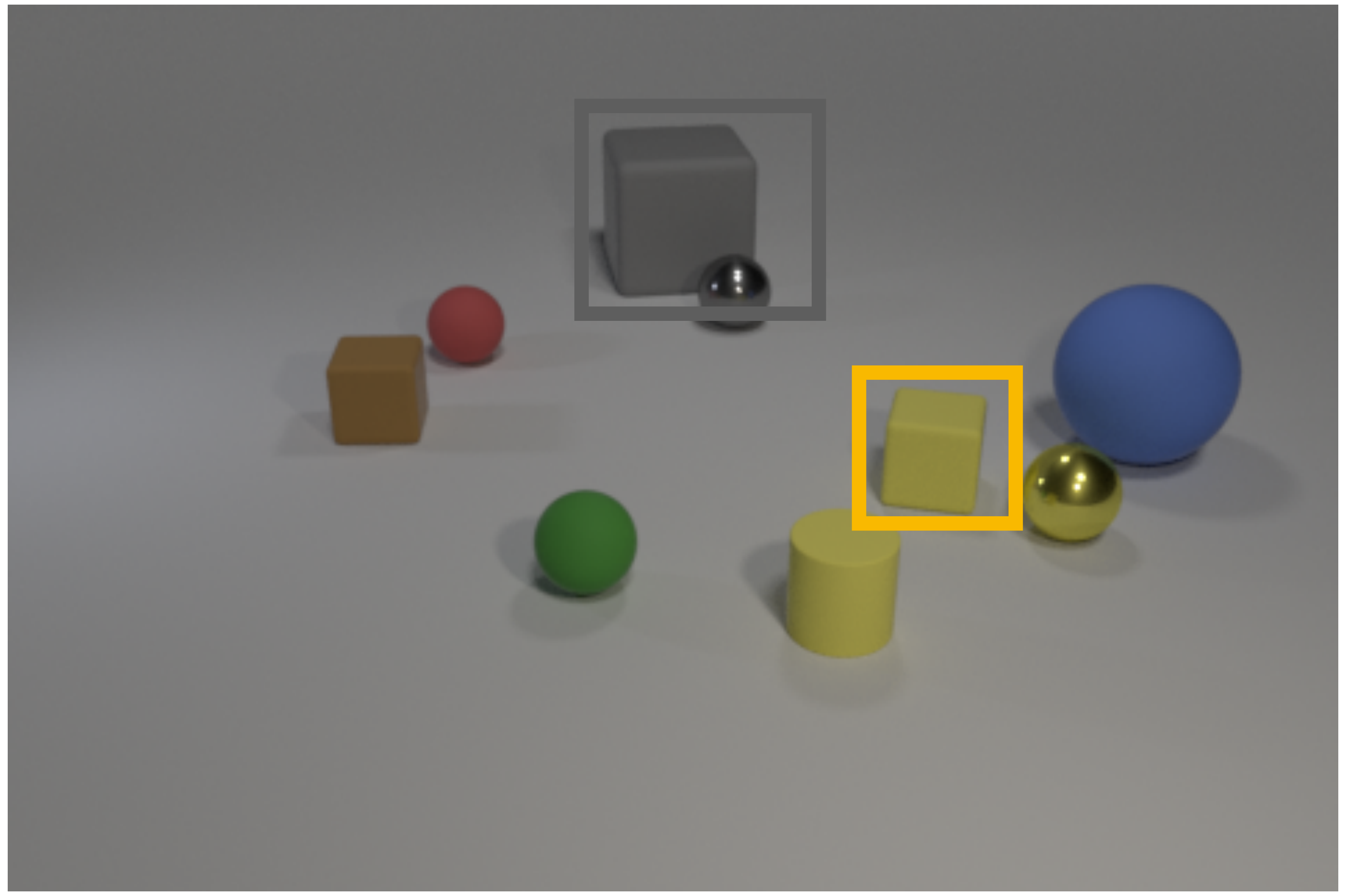
Modular approach to systematic generalization



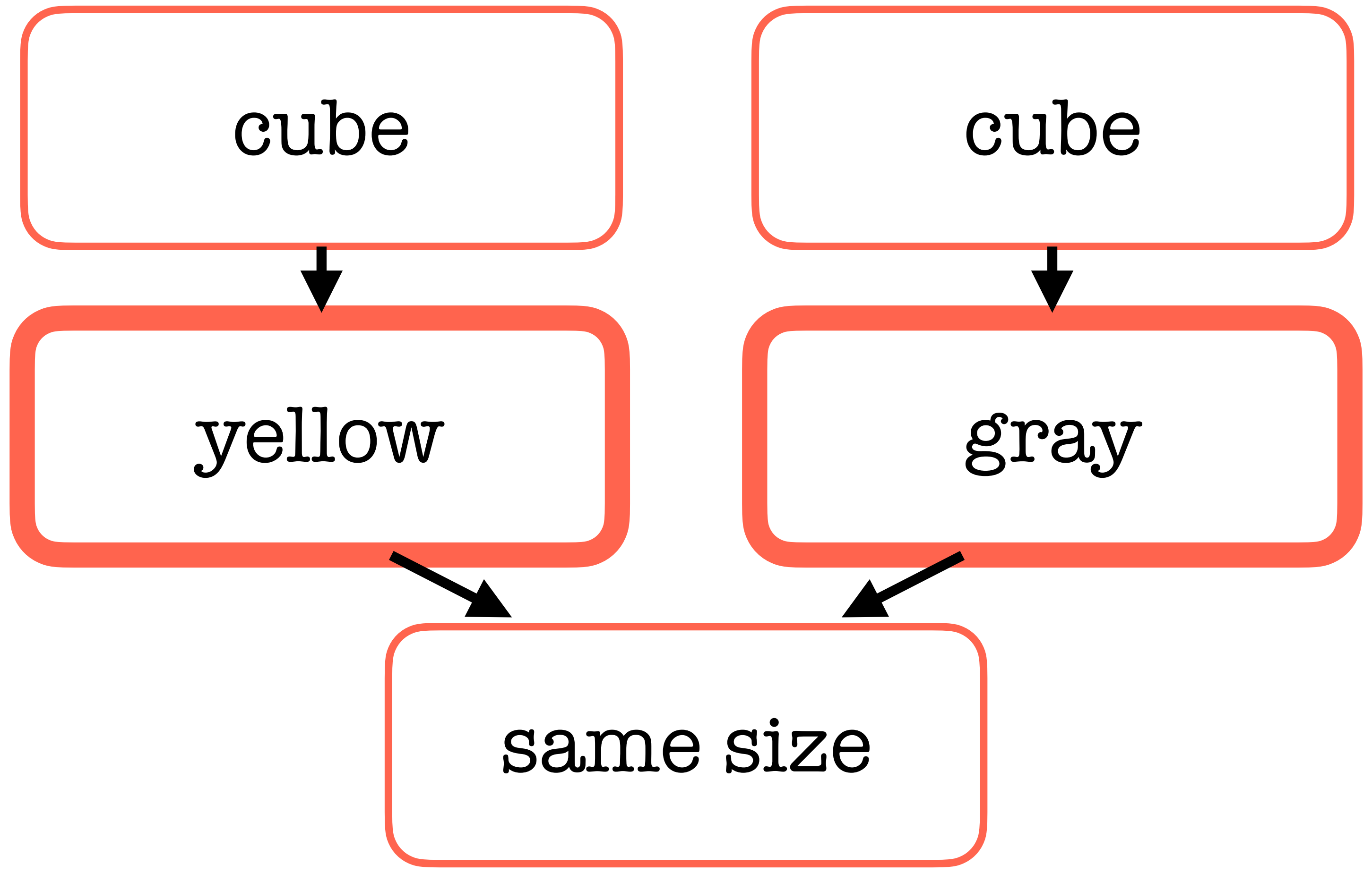
Q: Is the gray cube the same size as the yellow cube?



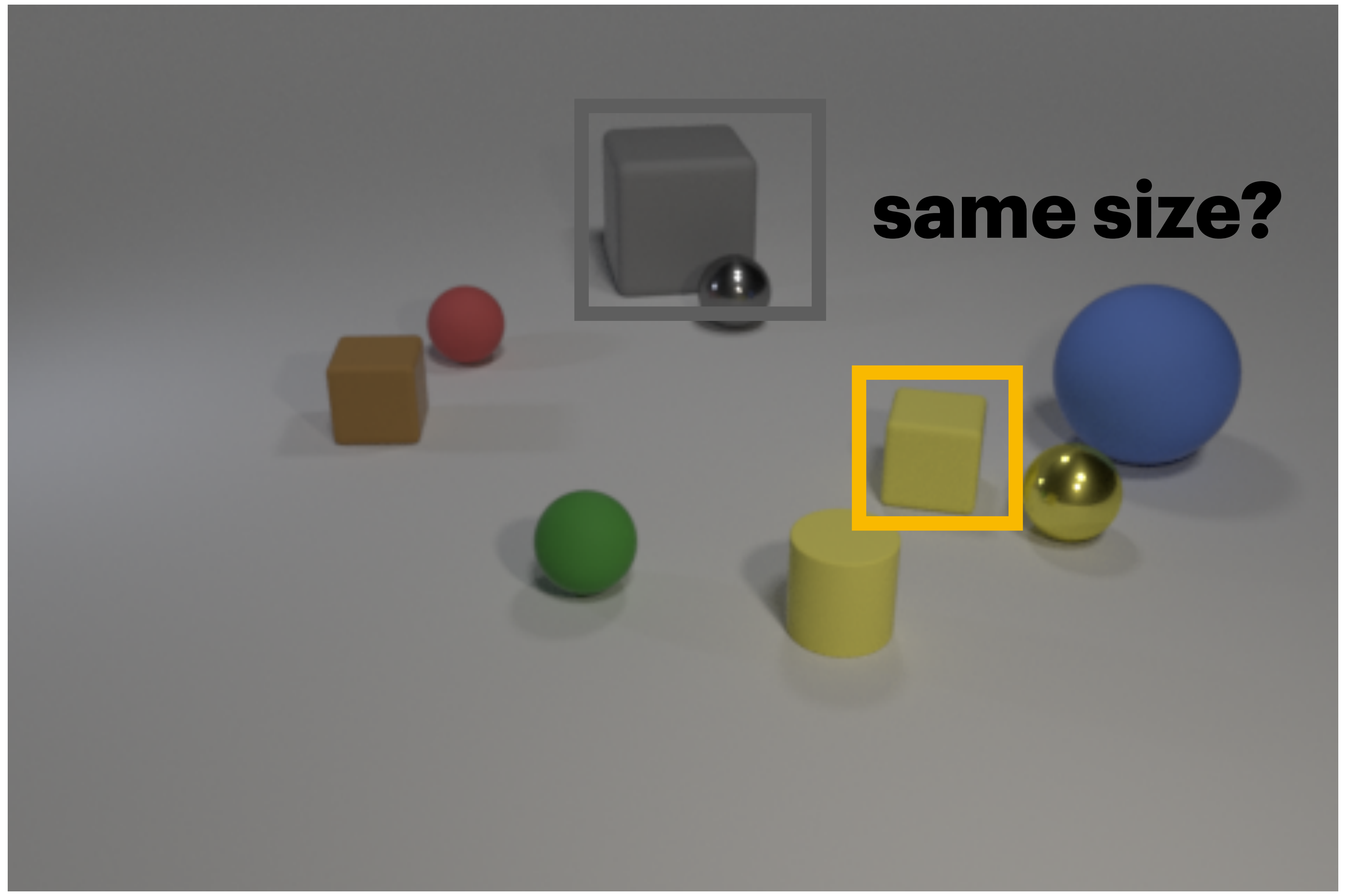
Modular approach to systematic generalization



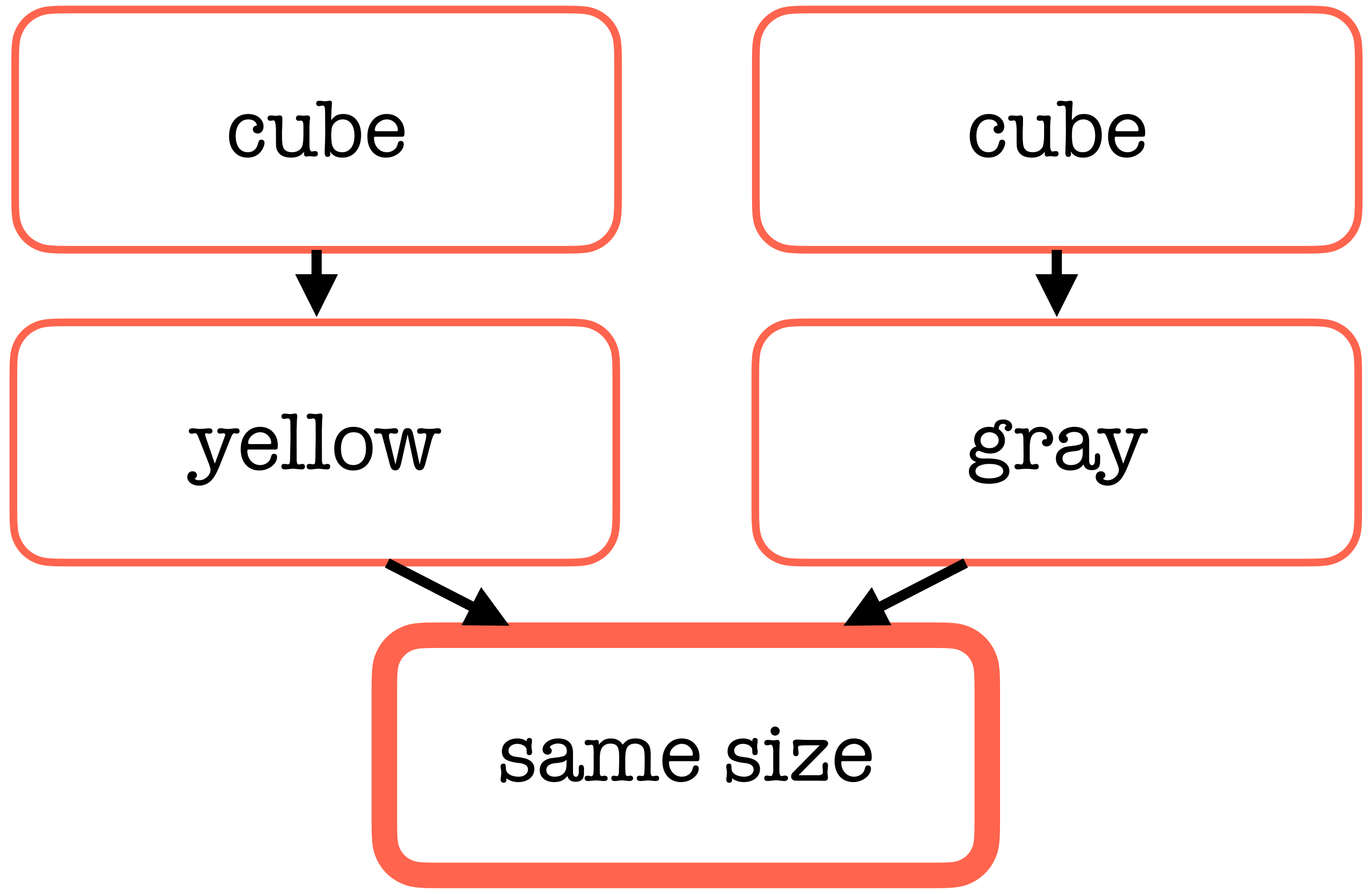
Q: Is the gray cube the same size as the yellow cube?



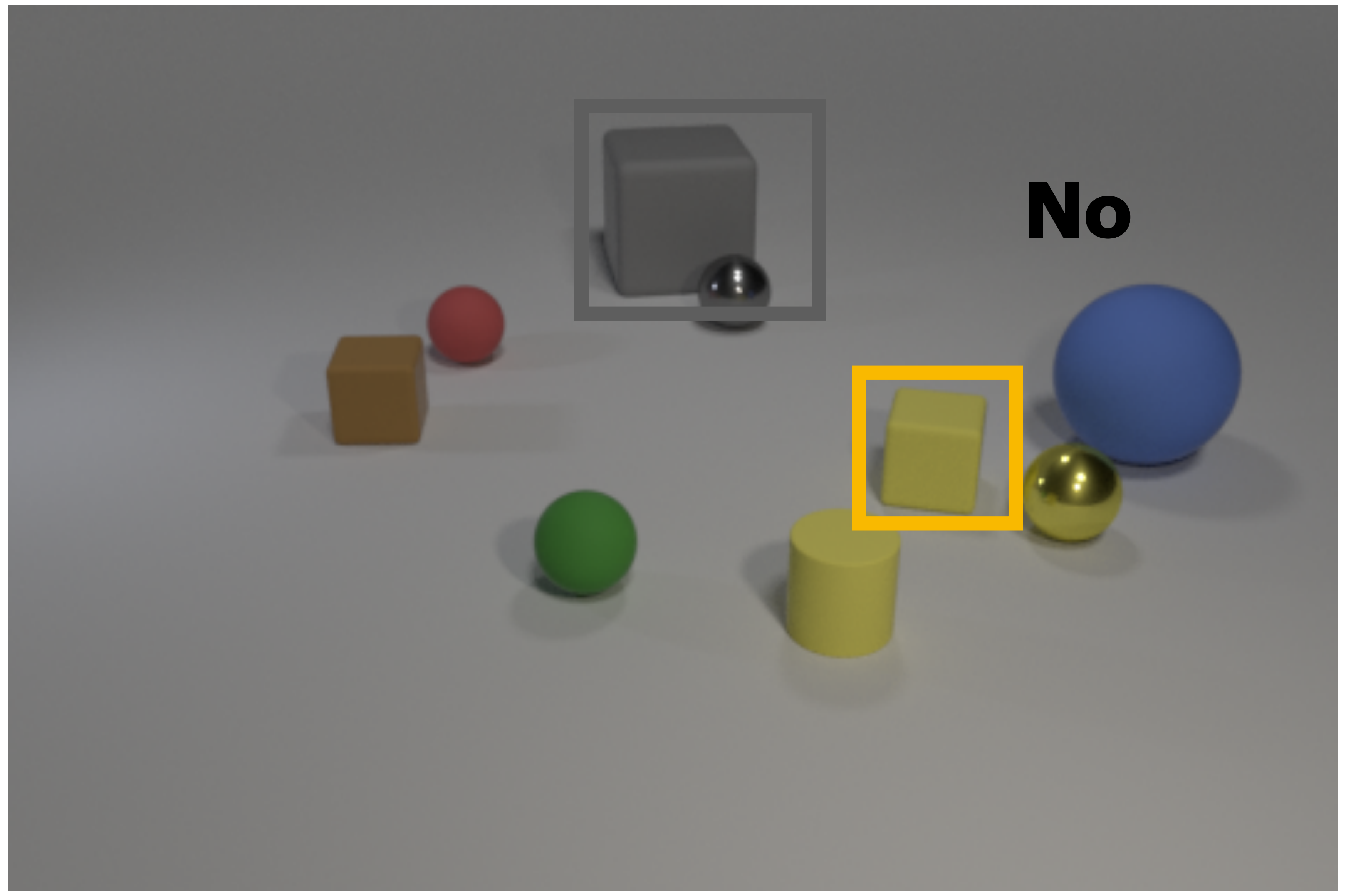
Modular approach to systematic generalization



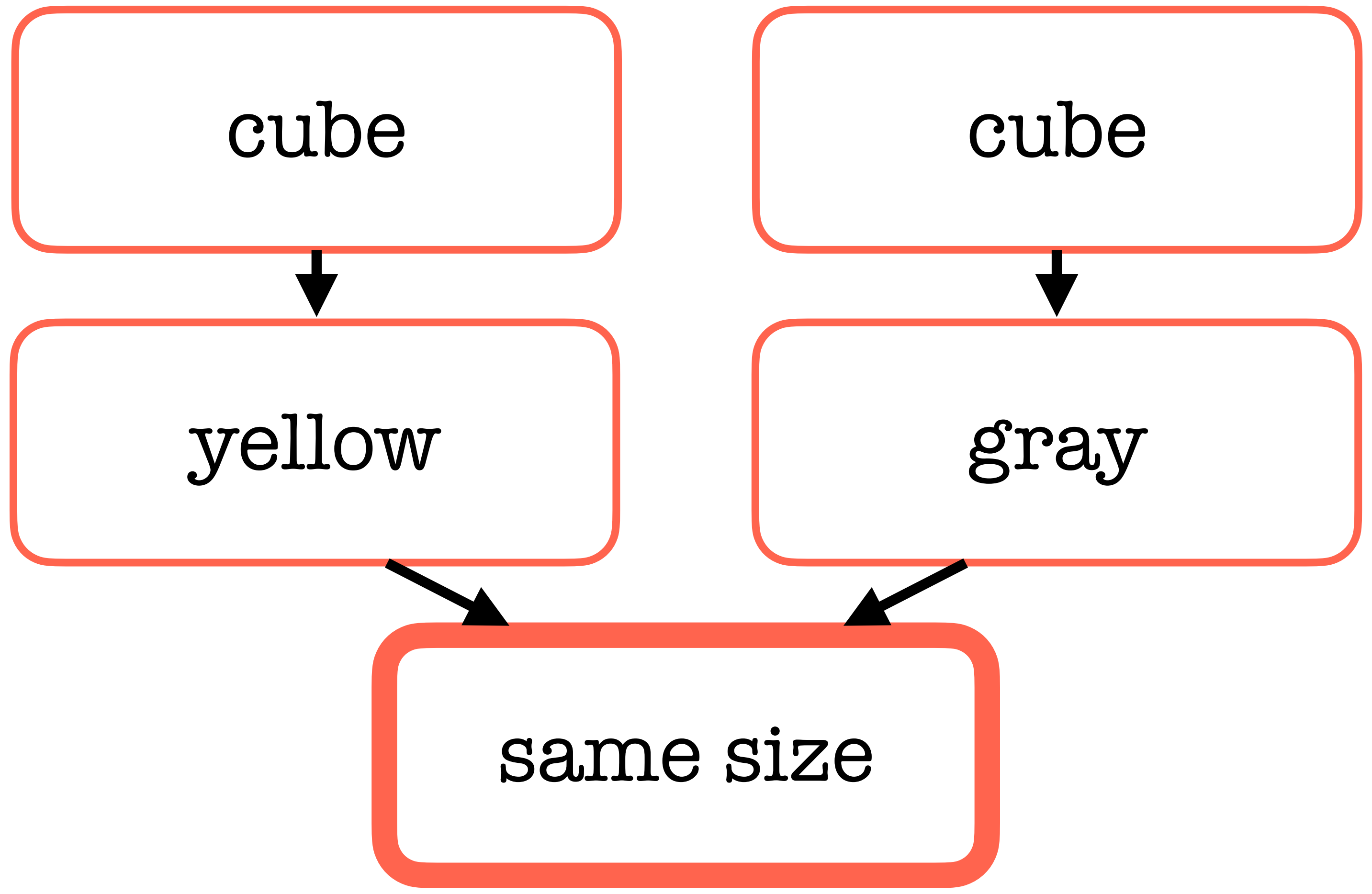
Q: Is the gray cube the same size as the yellow cube?



Modular approach to systematic generalization

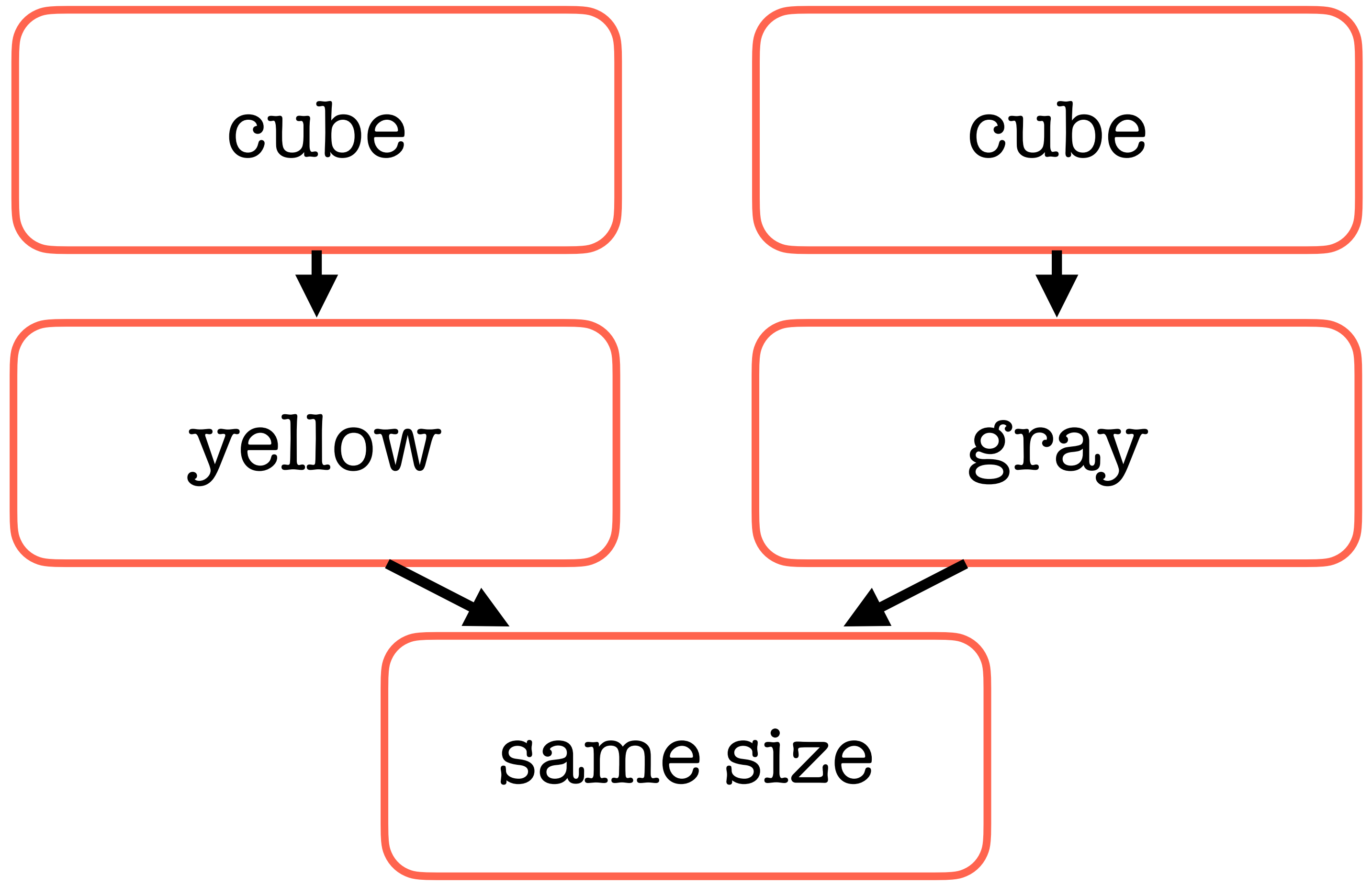


Q: Is the gray cube the same size as the yellow cube?



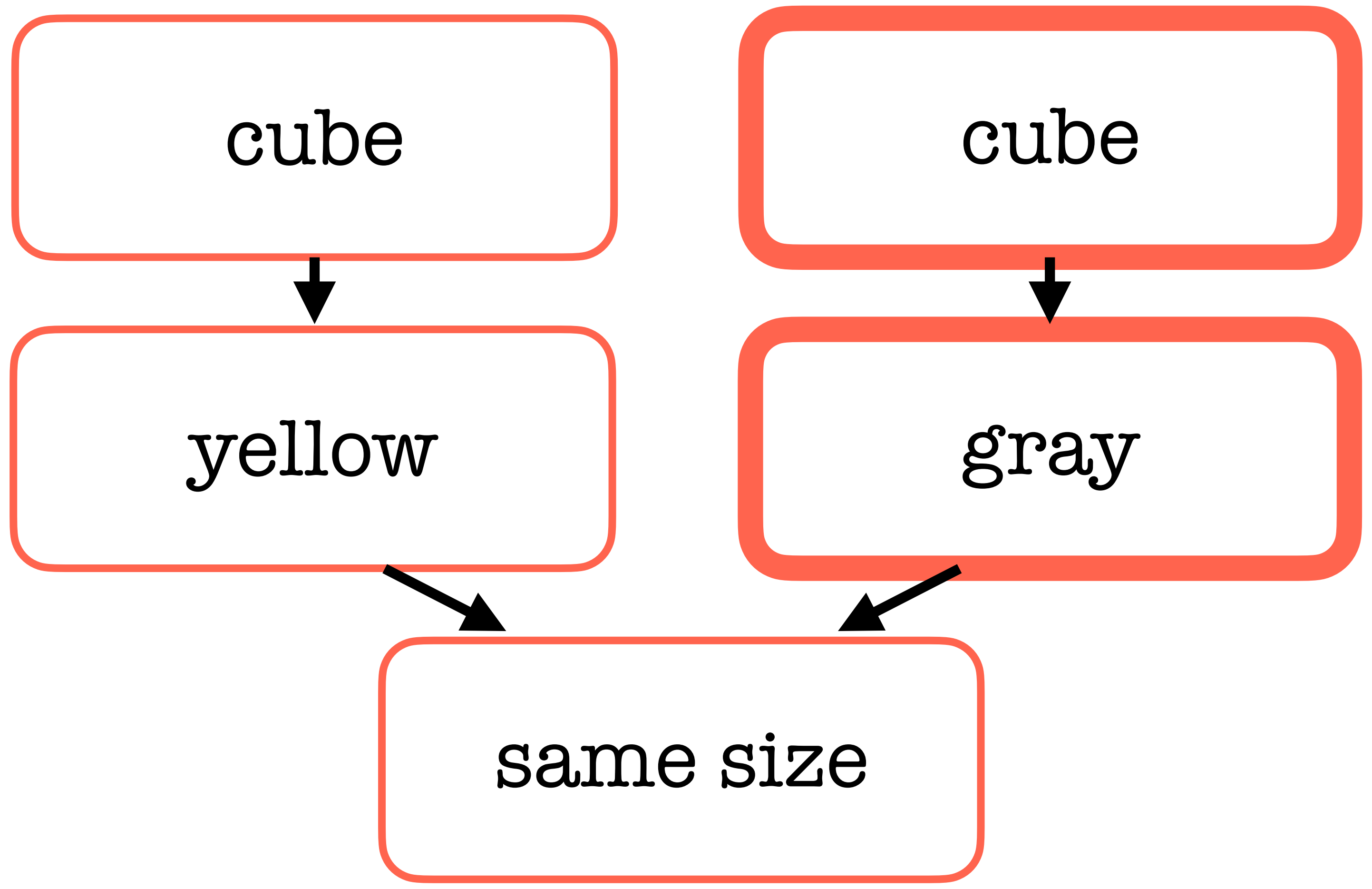
Modular approach to systematic generalization

Is the gray cube the same size as the yellow cube?



Modular approach to systematic generalization

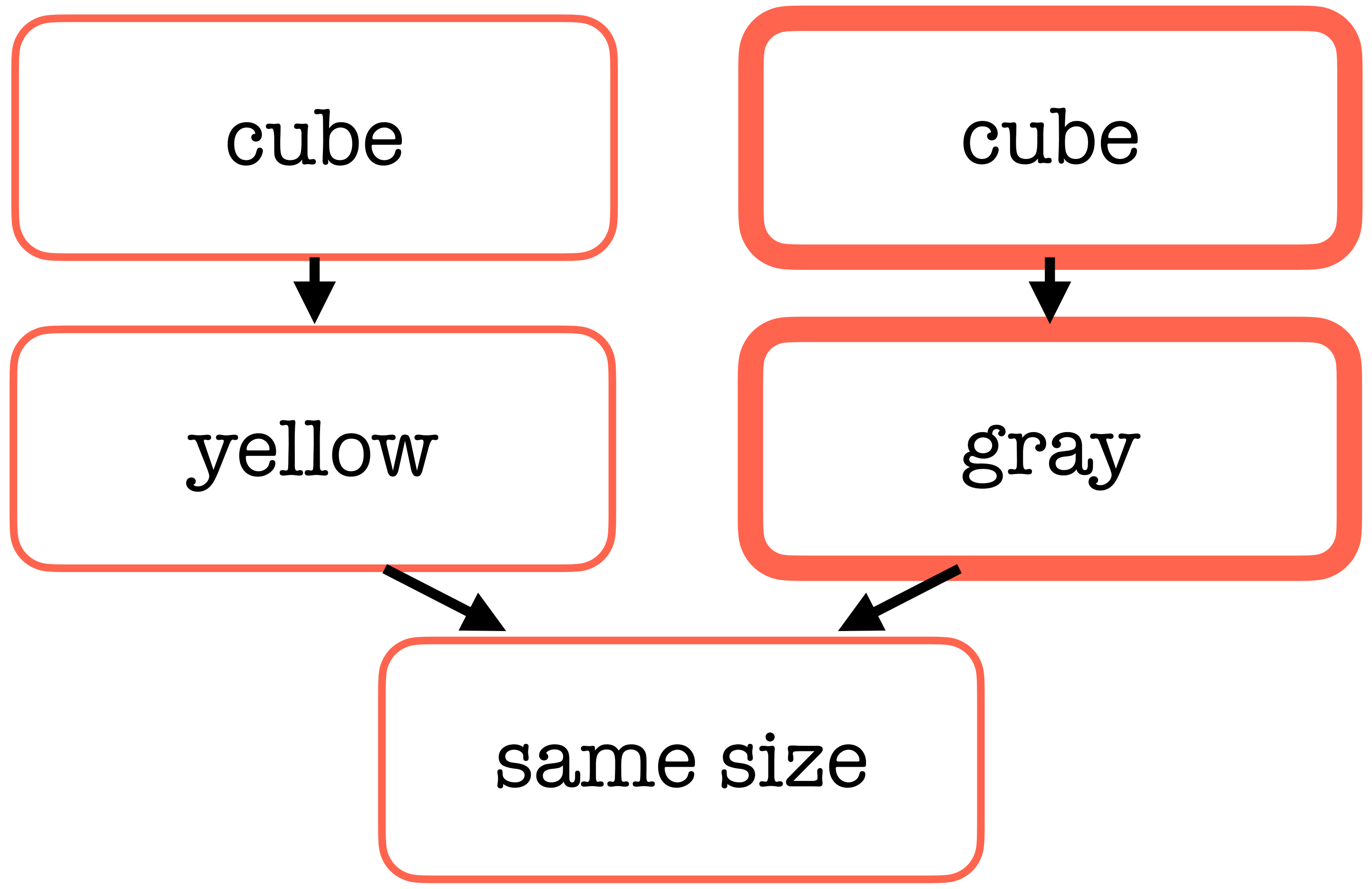
Is the **gray cube** the same size as the yellow cube?



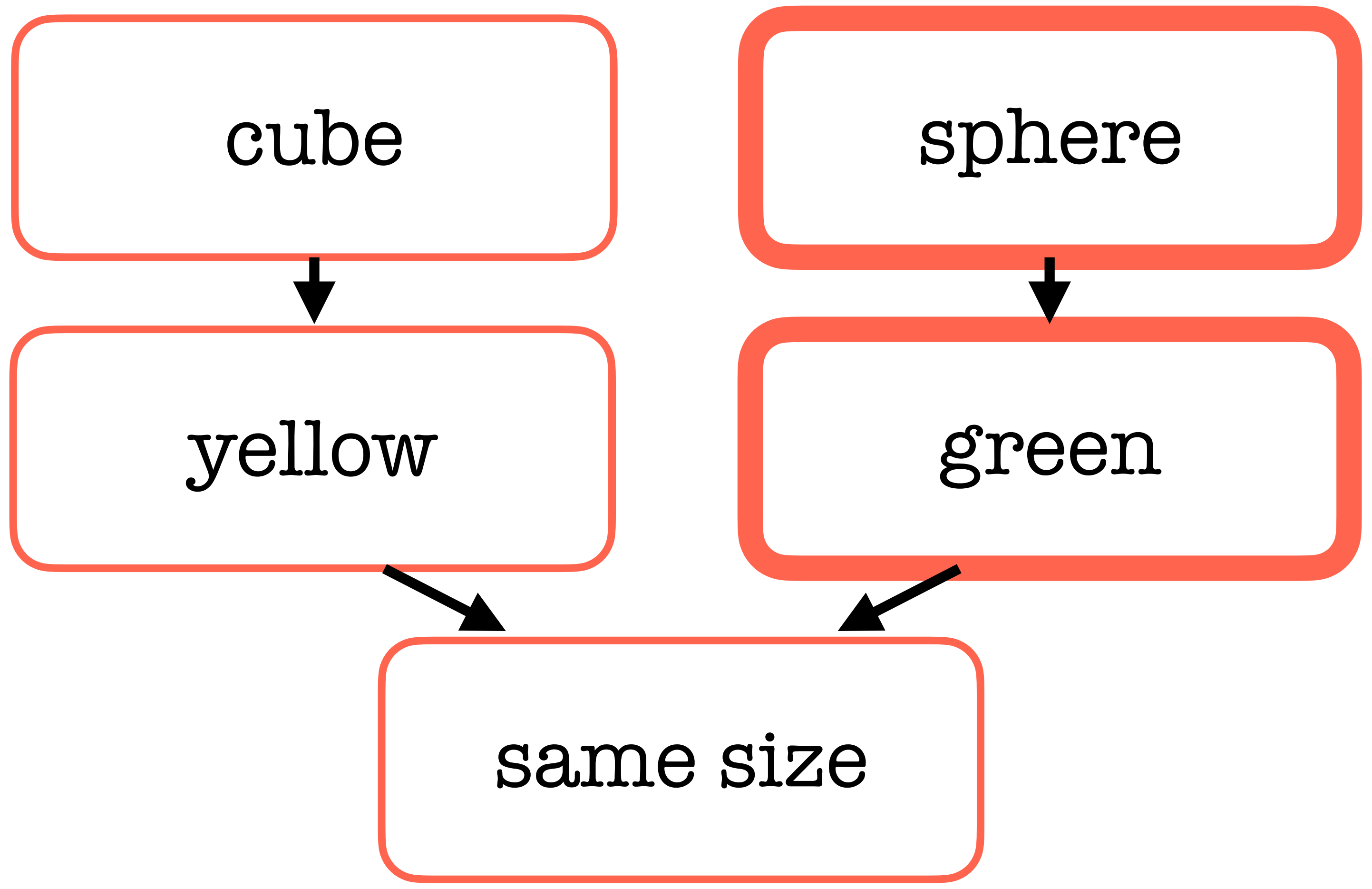
Is the **green sphere** the same size as the yellow cube?

Modular approach to systematic generalization

Is the **gray cube** the same size as the yellow cube?



Is the **green sphere** the same size as the yellow cube?



Modular approach to systematic generalization

- Bahdanau et al. (2019), *Systematic generalization: what is required and can it be learned?*
- Bahdanau et al. (2020), *CLOSURE: Assessing systematic generalization of CLEVR models*
- Purushwalkam et al. (2019), *Task-driven modular networks for zero-shot compositional learning*
- Madan et al. (2021), *When and how do CNNs generalize to out-of-distribution category-viewpoint combinations?*

Modules in Neural Module Networks

World

spheres, cubes / yellow, blue

Modules in Neural Module Networks

World

spheres, cubes / yellow, blue

Library

shared module	module per group of sub-tasks		one module per sub-task			
all [<sphere, cube, yellow, blue>]	shape [<sphere, cube>]	color [<yellow, blue>]	sphere	cube	yellow	blue

Modules in Neural Module Networks

World

spheres, cubes / yellow, blue

Library

shared module	module per group of sub-tasks	one module per sub-task
all [<sphere, cube, yellow, blue>]	shape [<sphere, cube>]; color [<yellow, blue>]	sphere; cube; yellow; blue

Usage

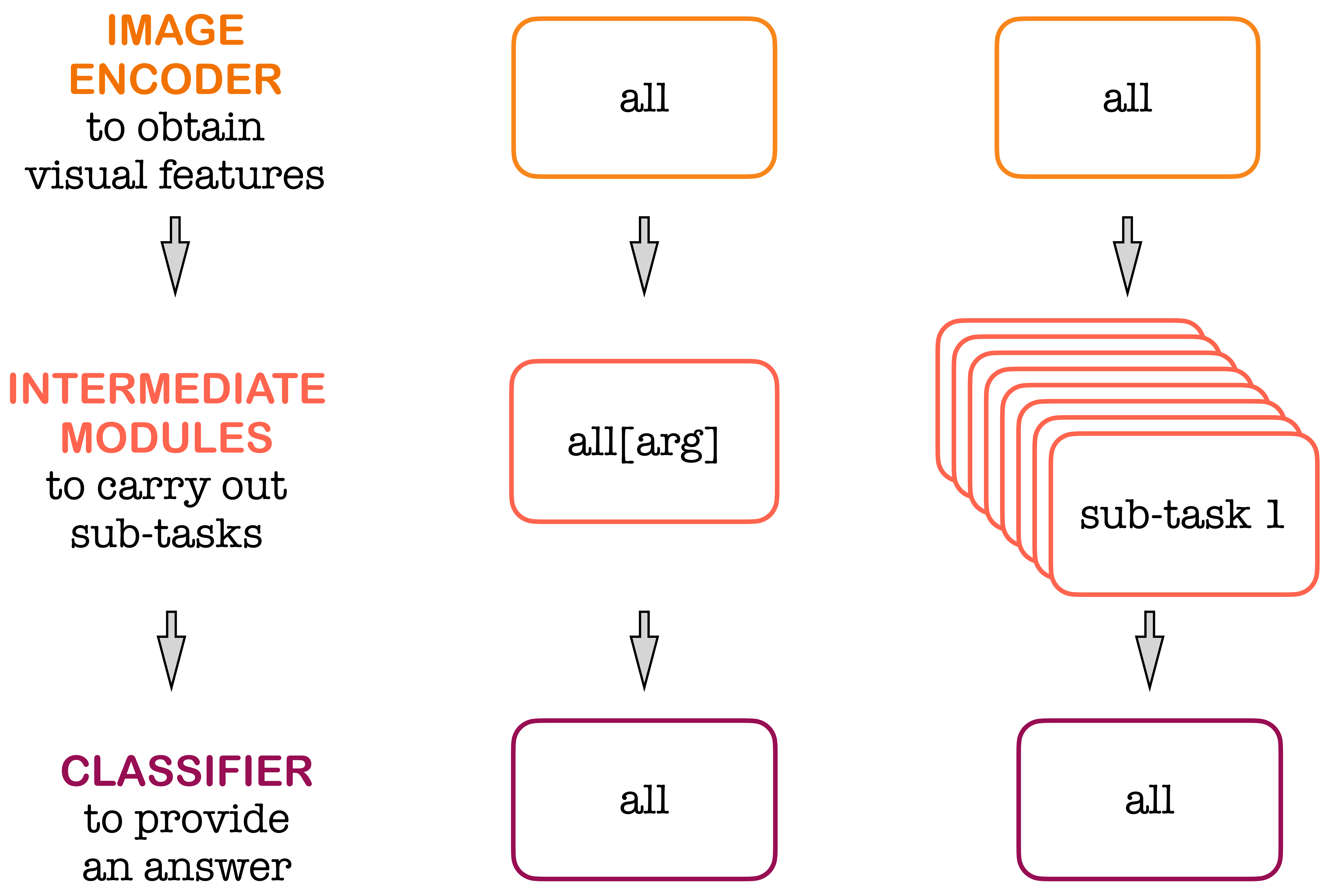
Question: Is this a yellow cube?

all [yellow] → all [cube]

color [yellow] → shape [cube]

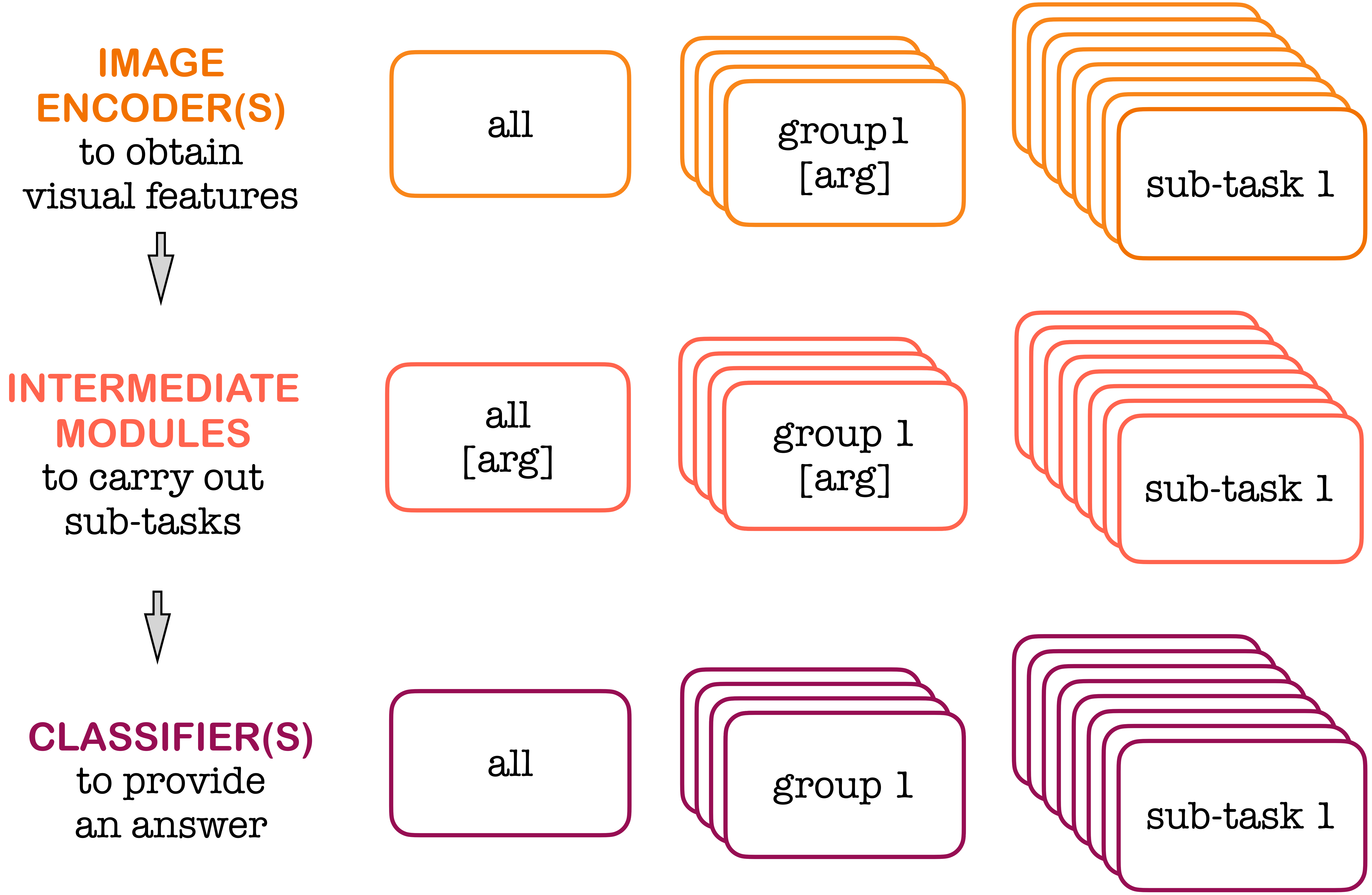
yellow → cube

Three stages library

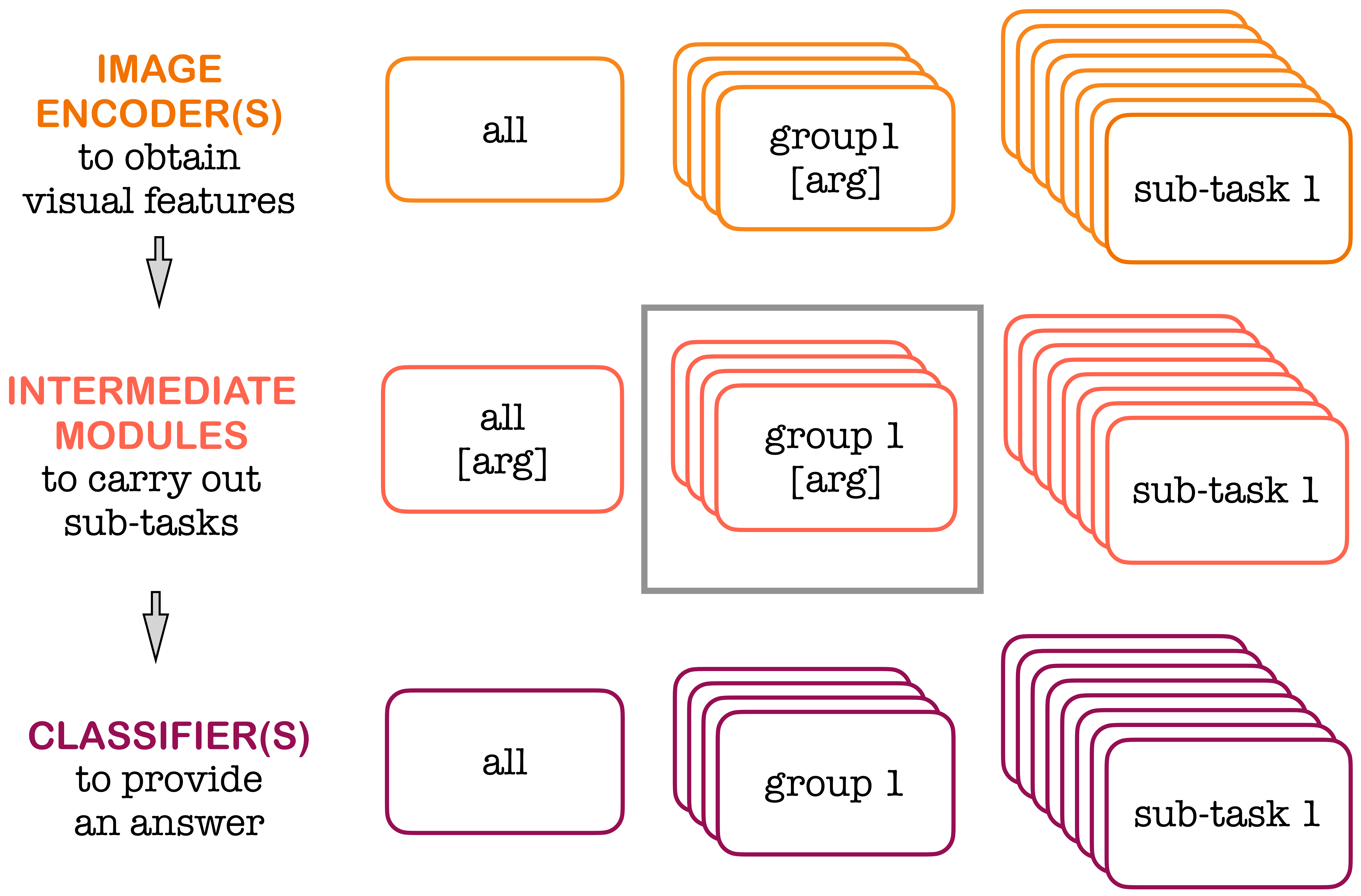


How Modular Should Neural Module Networks Be for Systematic Generalization?

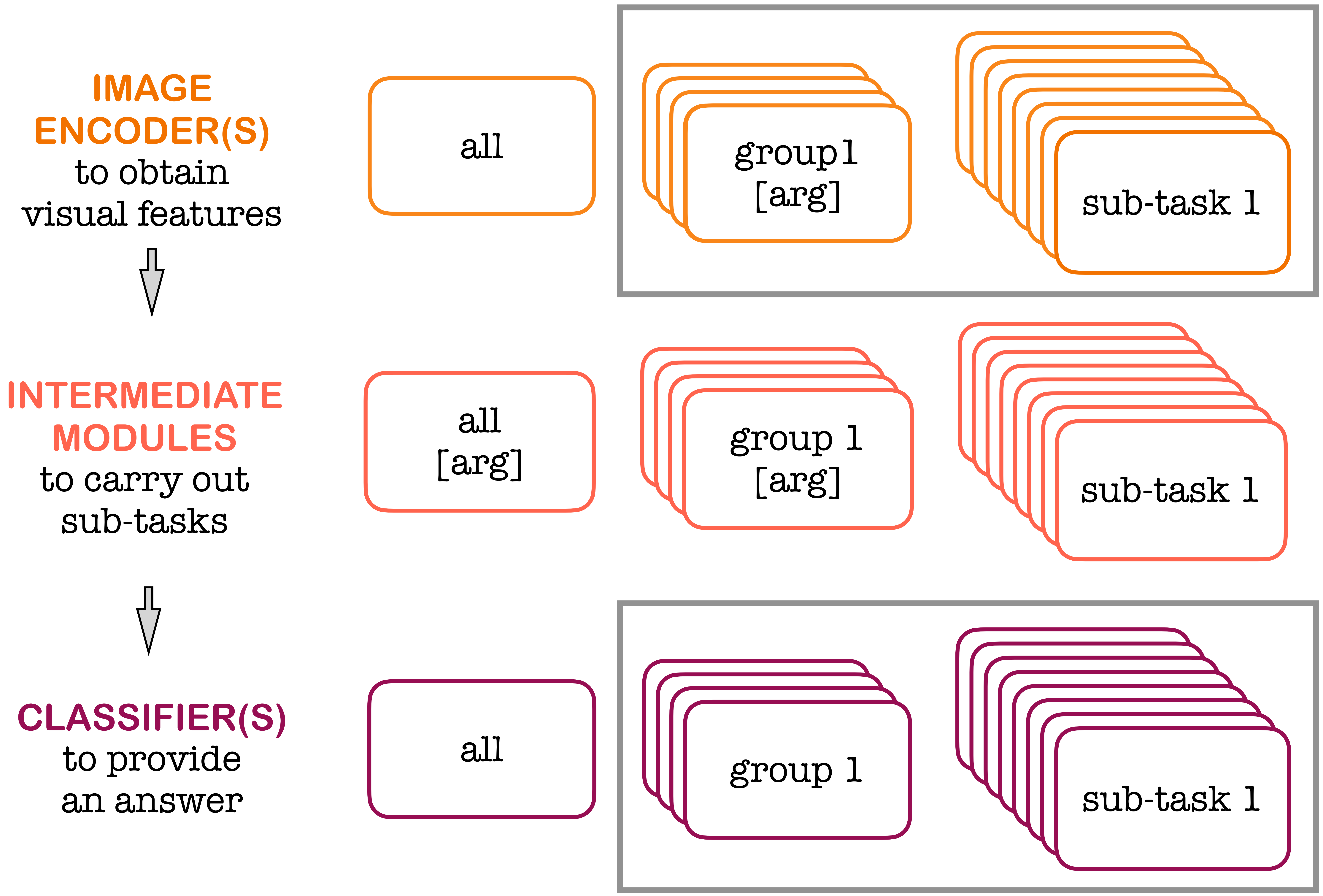
Libraries with different degrees of modularity



Libraries with different degrees of modularity



Libraries with different degrees of modularity



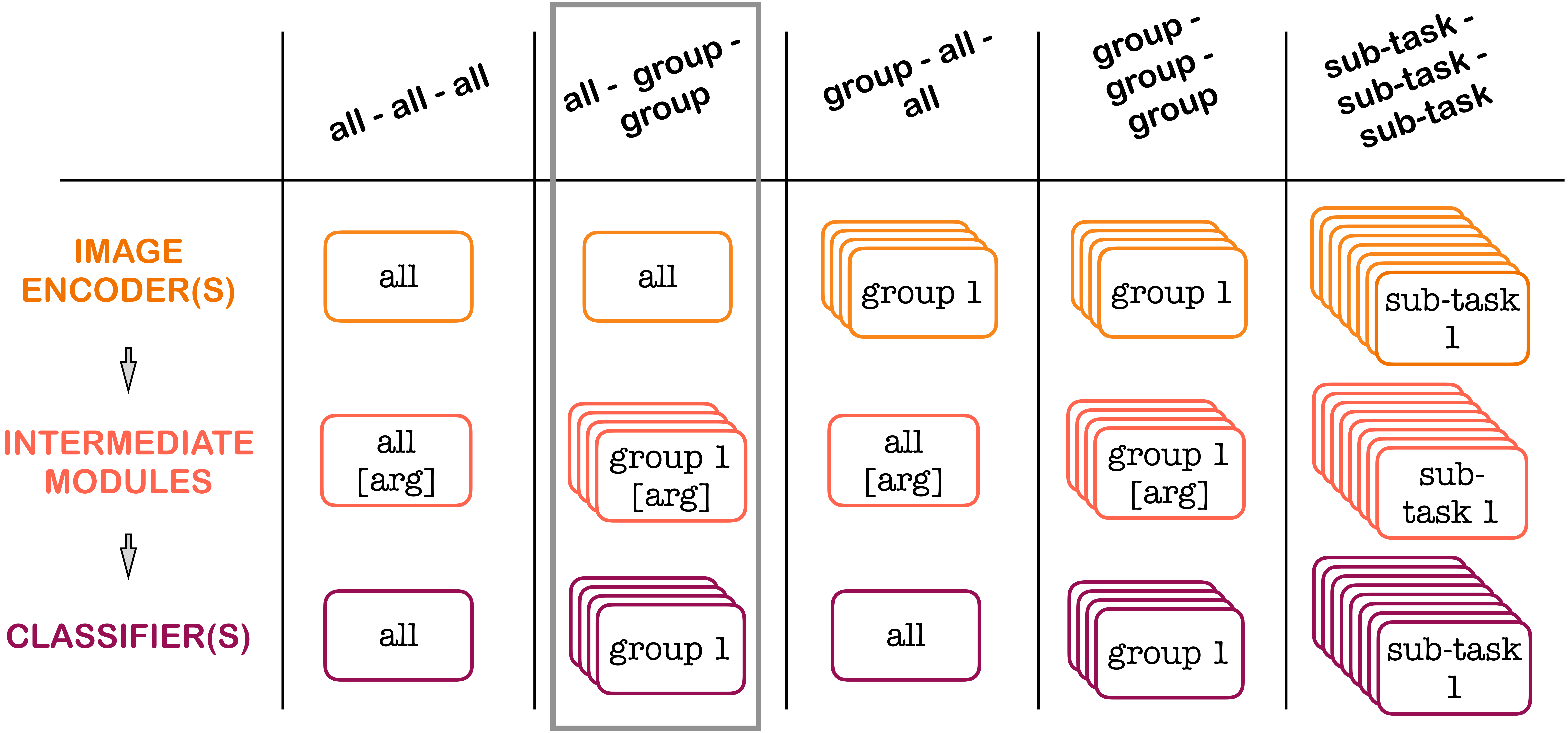
Main libraries used for this study

	all - all - all	all - group - group	group - all - all	group - group - group	sub-task - sub-task - sub-task
IMAGE ENCODER(S)	all	all	group 1	group 1	sub-task 1
INTERMEDIATE MODULES	all [arg]	group 1 [arg]	all [arg]	group 1 [arg]	sub-task 1
CLASSIFIER(S)	all	group 1	all	group 1	sub-task 1

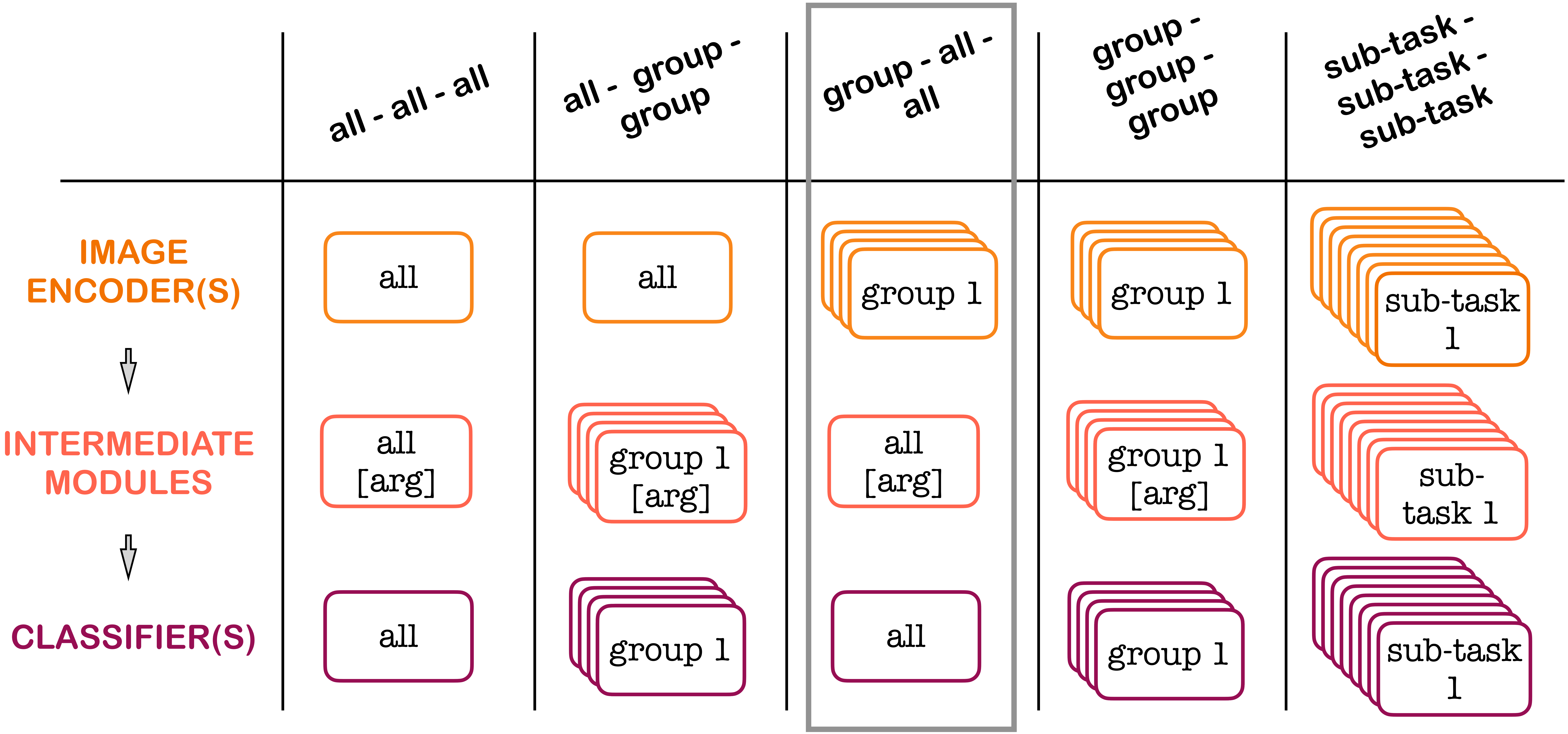
Main libraries used for this study

	all - all - all	all - group - group	group - all - all	group - group - group	sub-task - sub-task - sub-task
IMAGE ENCODER(S)	all	all	group 1	group 1	sub-task 1
INTERMEDIATE MODULES	all [arg]	group 1 [arg]	all [arg]	group 1 [arg]	sub-task 1
CLASSIFIER(S)	all	group 1	all	group 1	sub-task 1

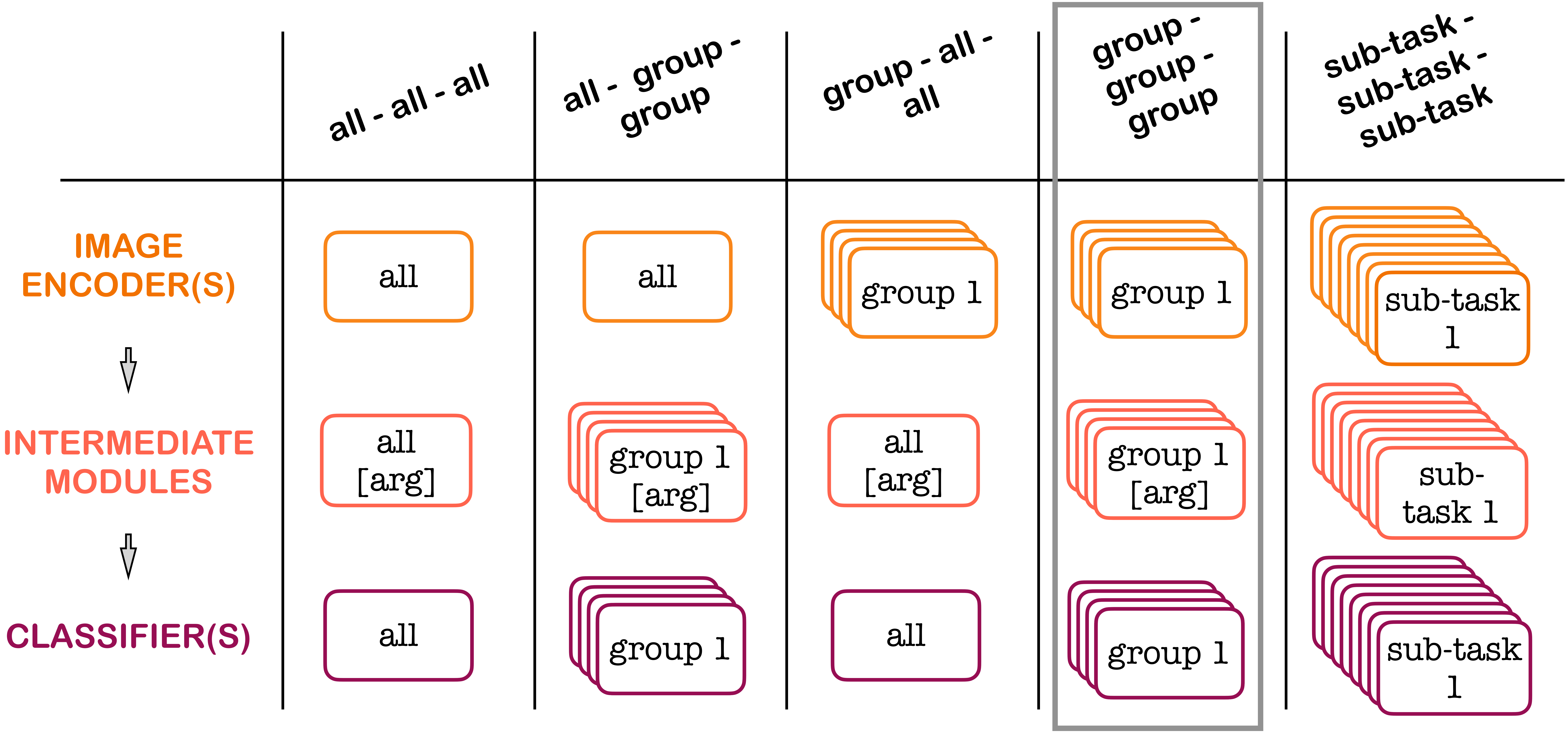
Main libraries used for this study



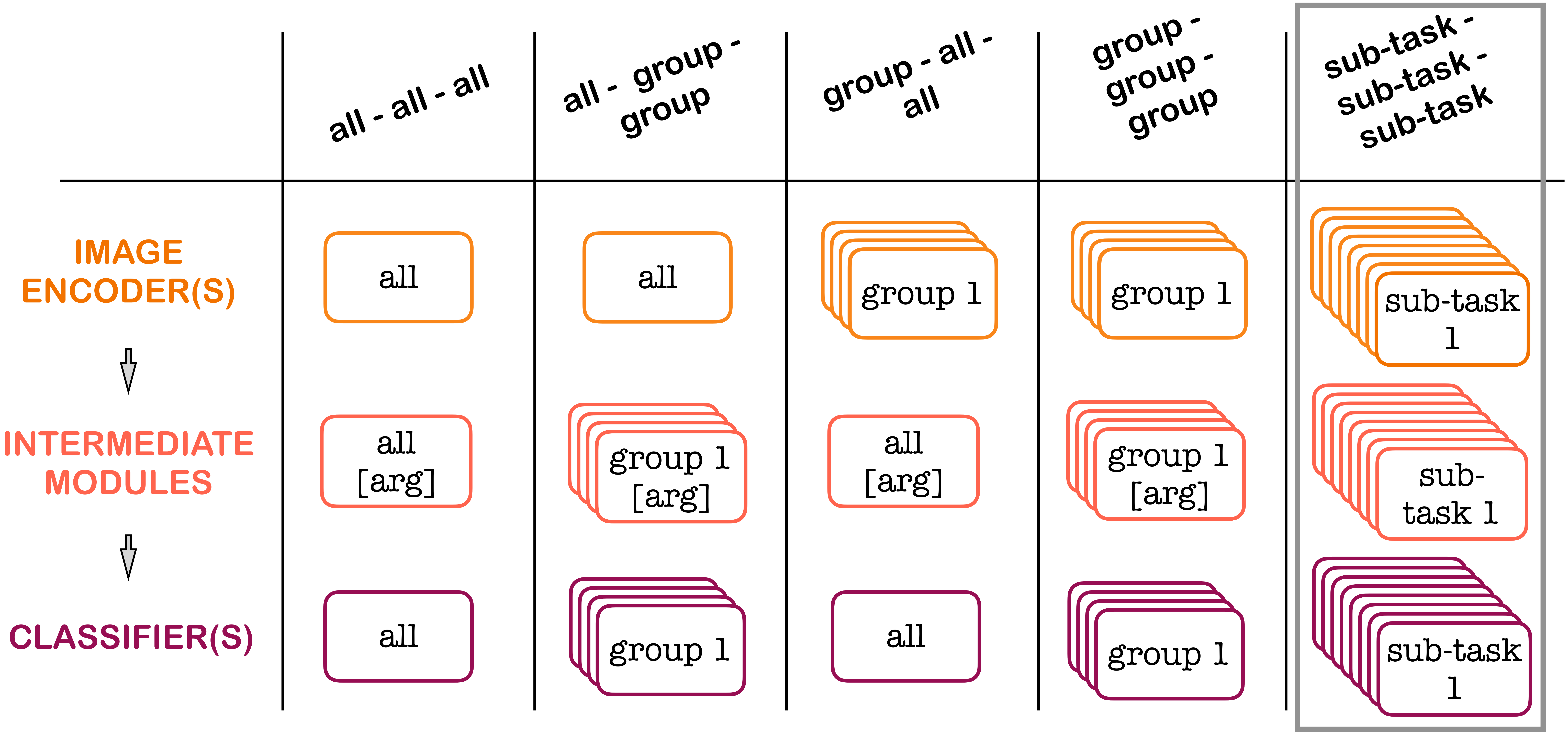
Main libraries used for this study



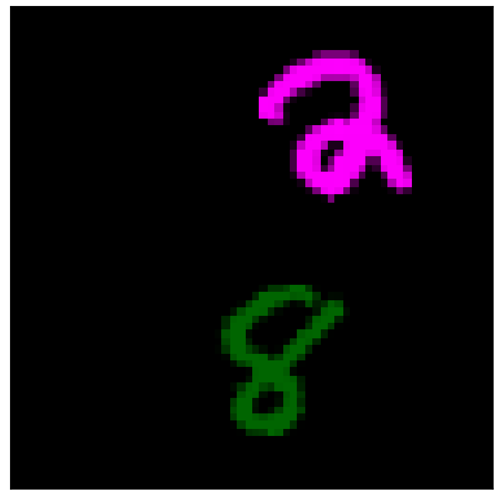
Main libraries used for this study



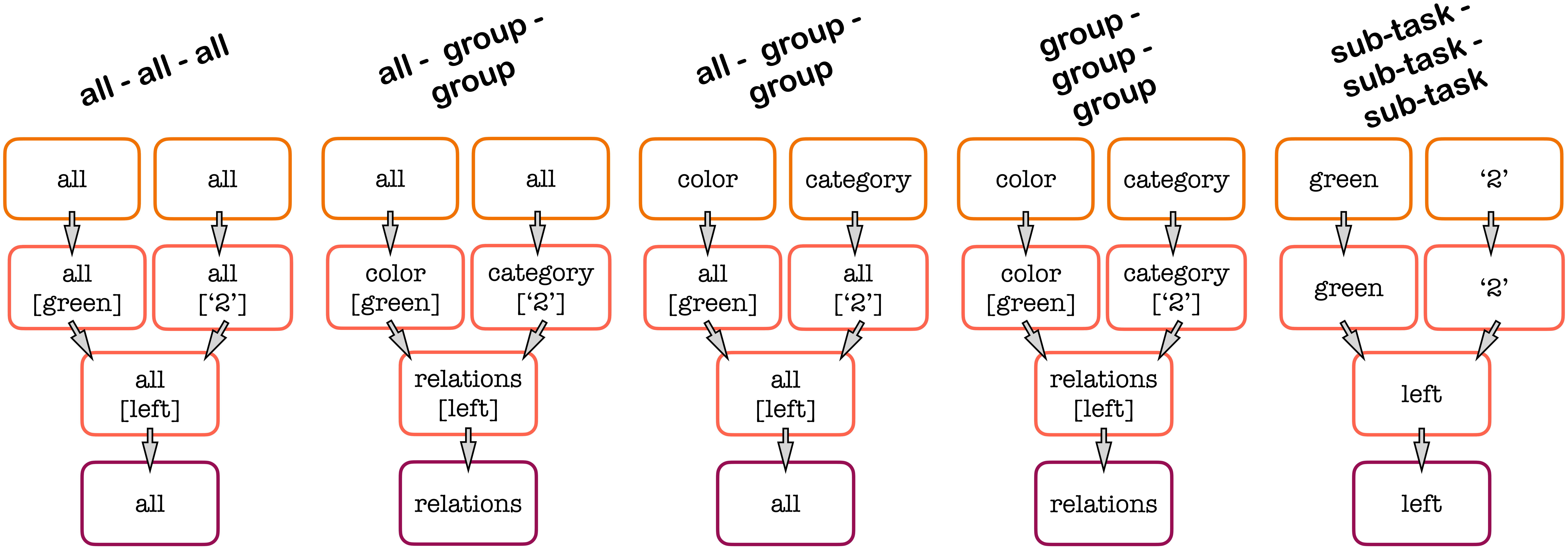
Main libraries used for this study



Library usage



Q: “Is the green object left of ‘2’?”



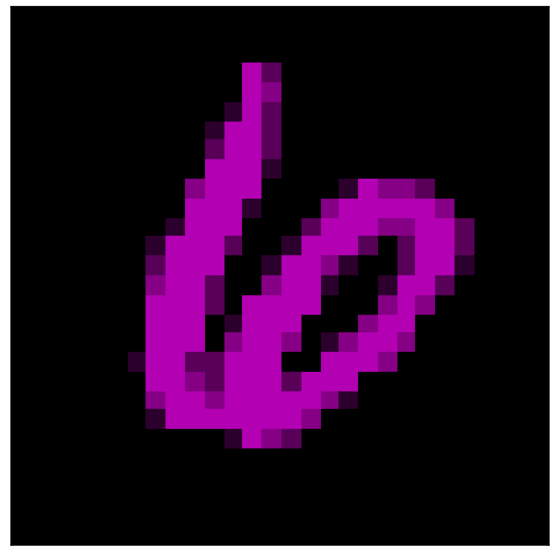
Experiment Outline

VQA-MNIST limited combinations of visual attributes

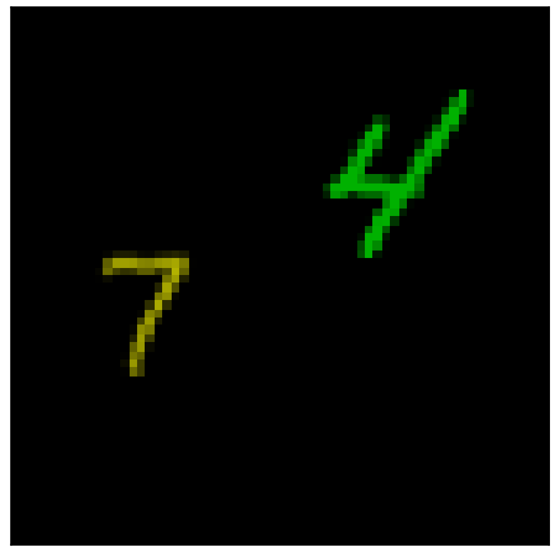
SQOOP limited co-occurrence of objects

CLEVR-CoGenT application

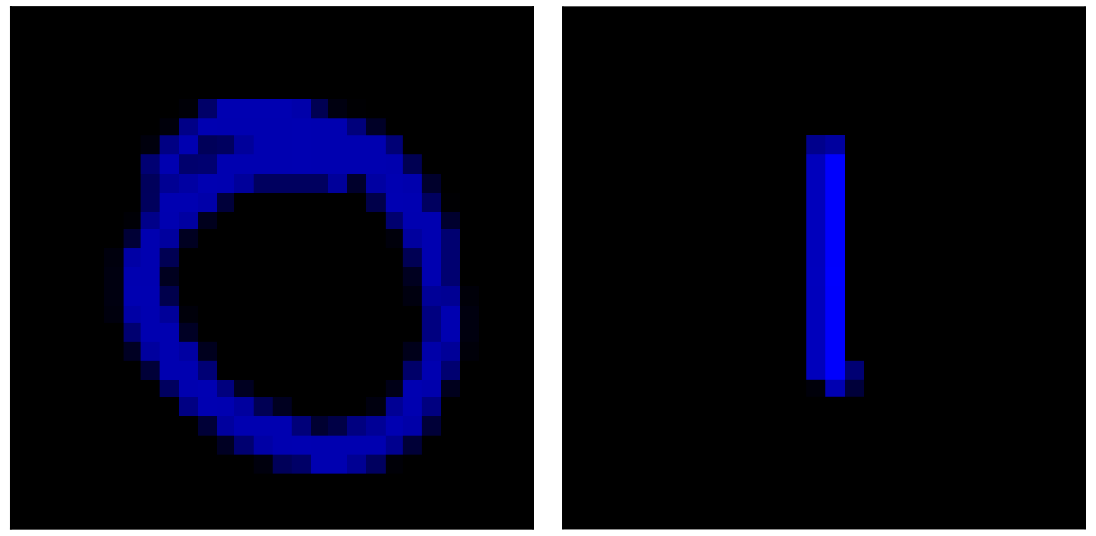
VQA-MNIST: Limited visual attributes



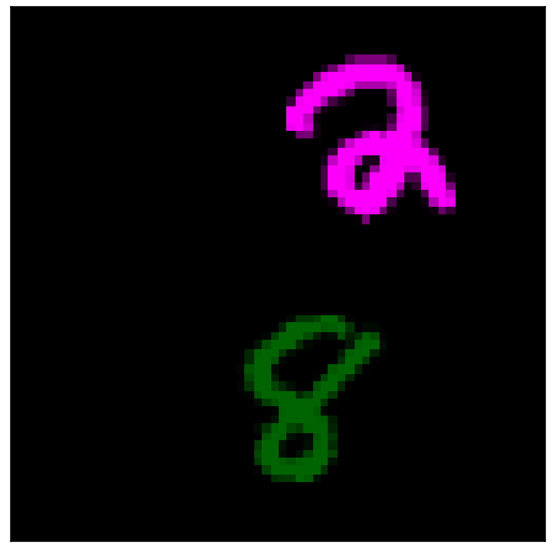
Is the object blue?
Is the object a ‘6’?
Is the object small?



Is there a green object?
Is there a bright object?
Is there a ‘4’?

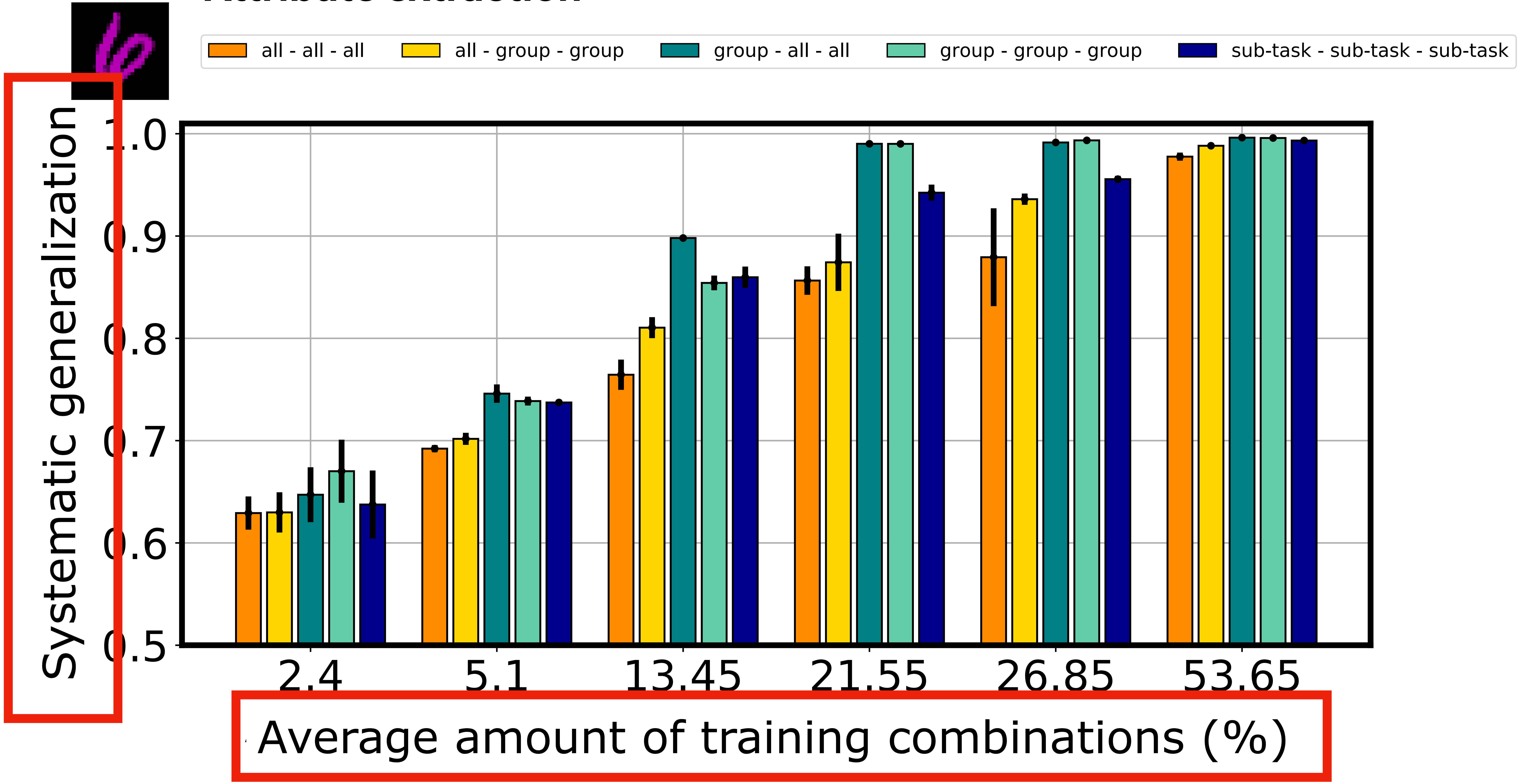


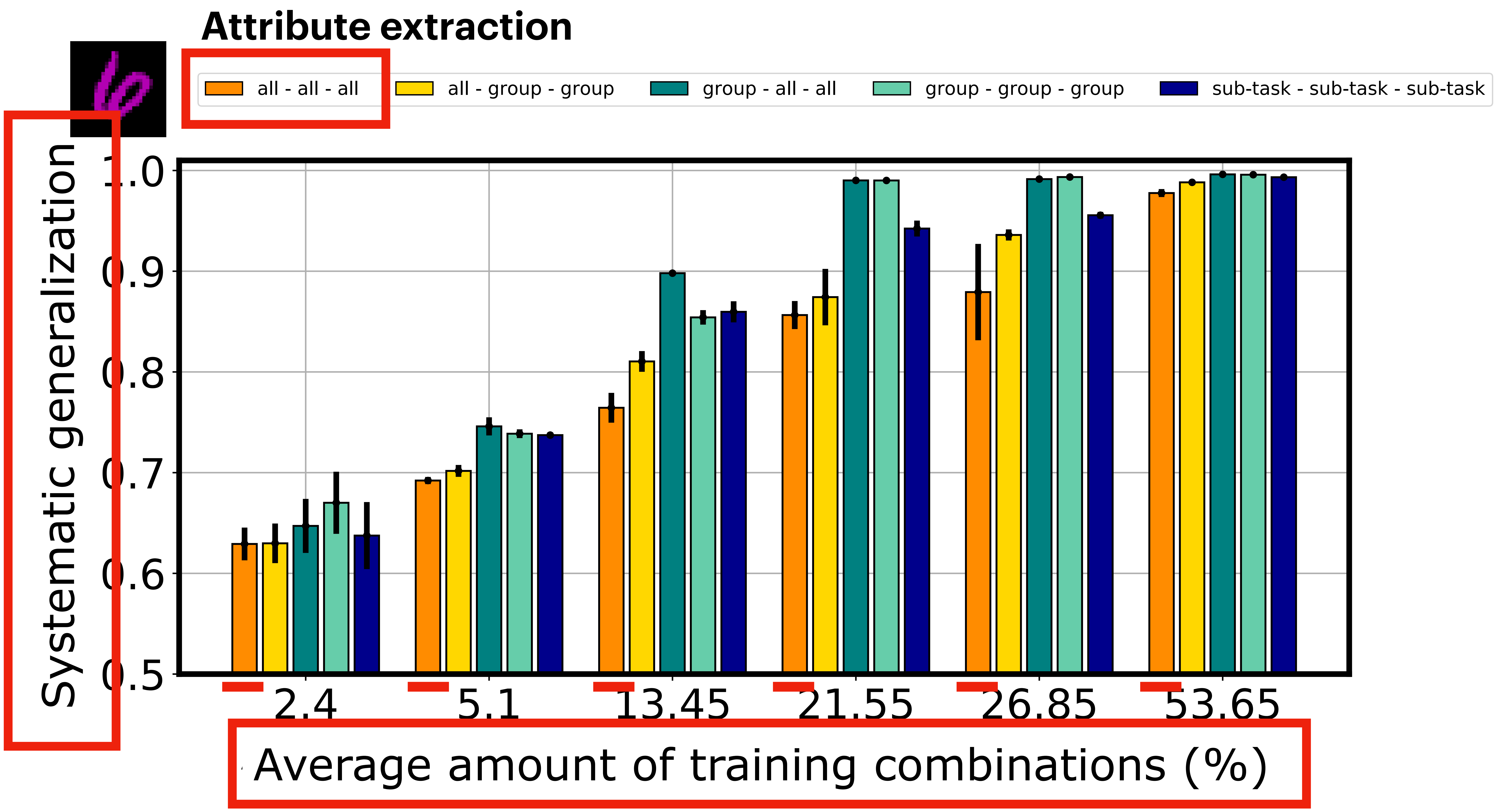
Are the two objects the same color/size/category/brightness?

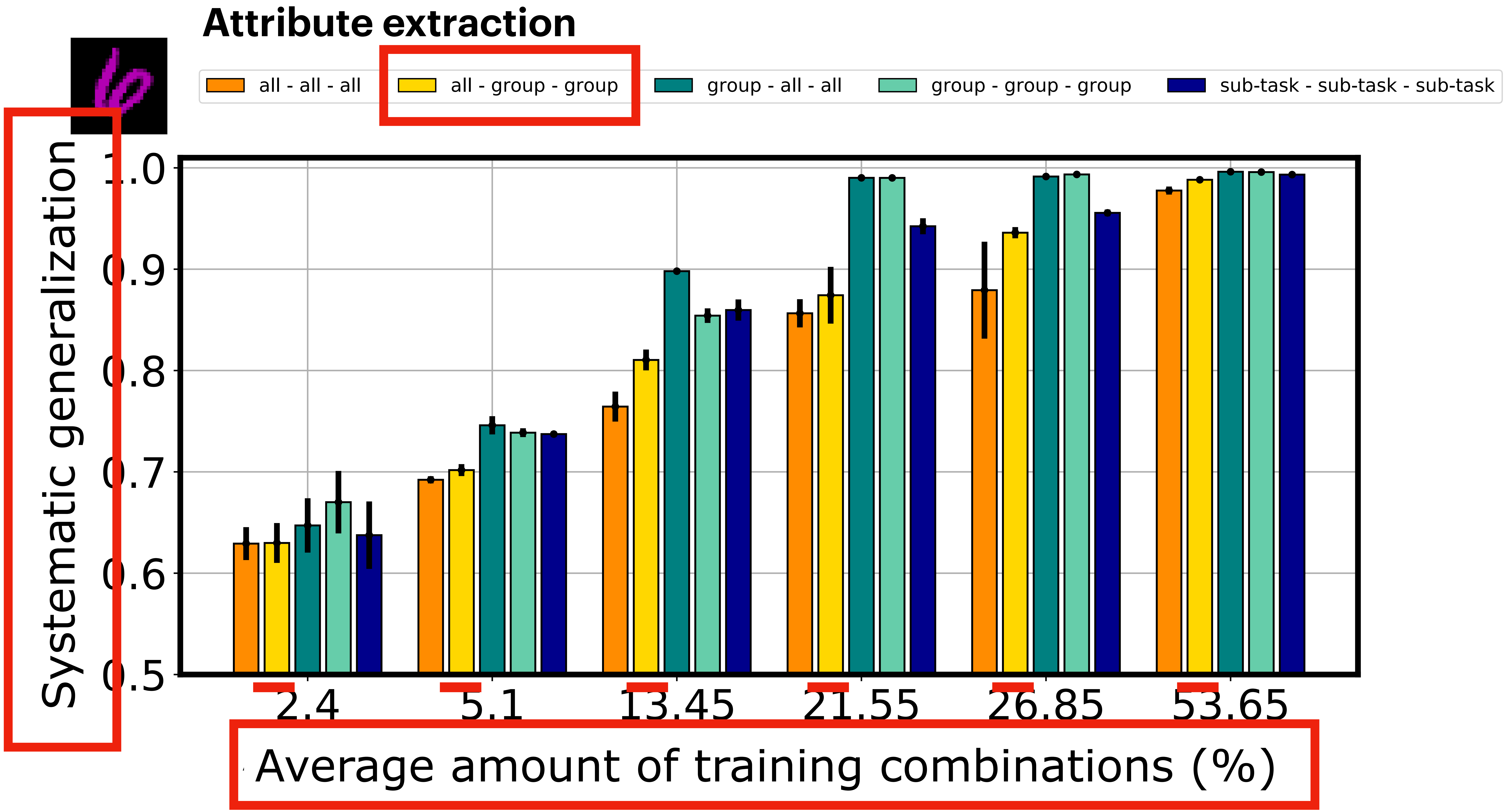


Is the green object left of ‘2’?
Is ‘8’ below the pink object?

Attribute extraction

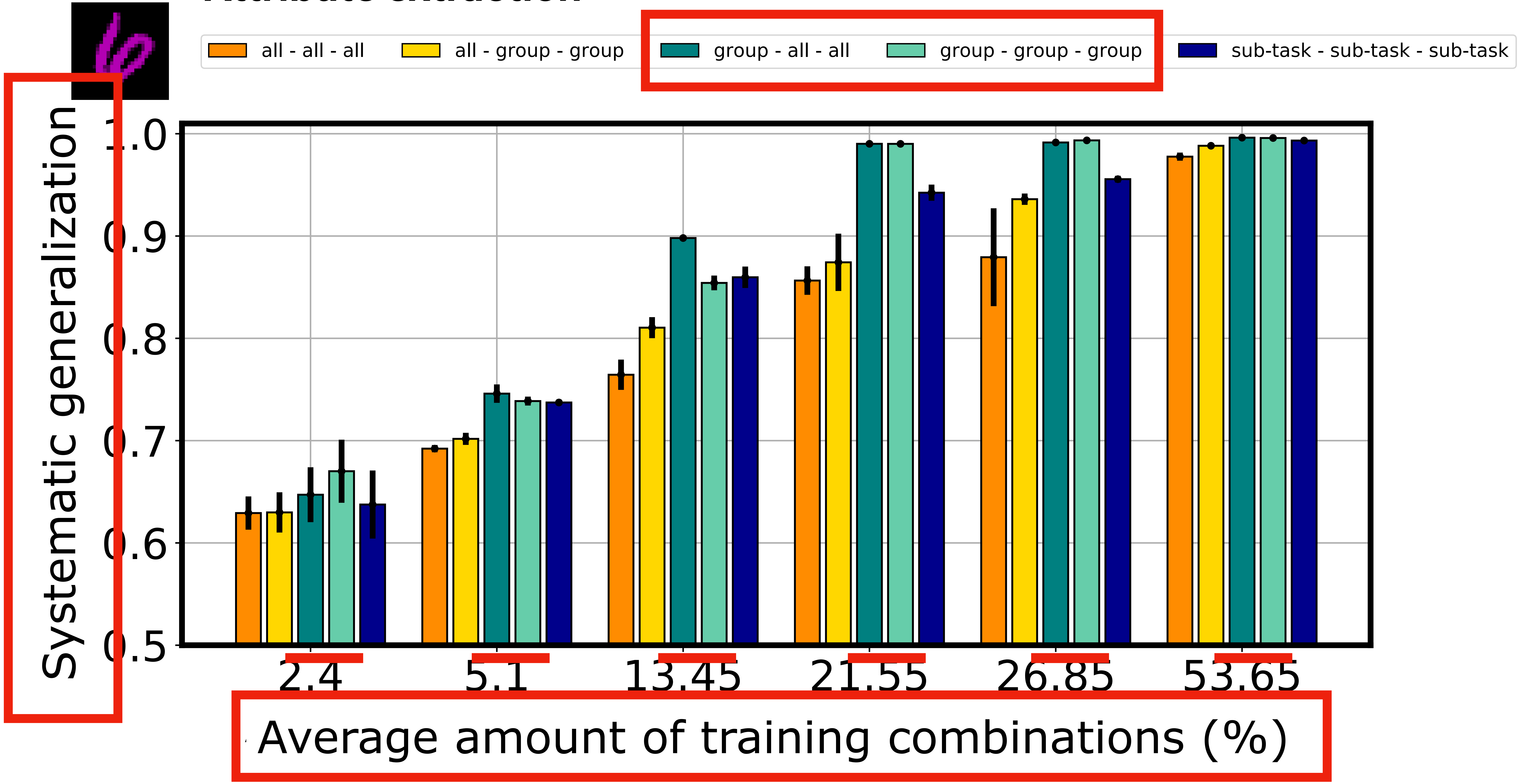


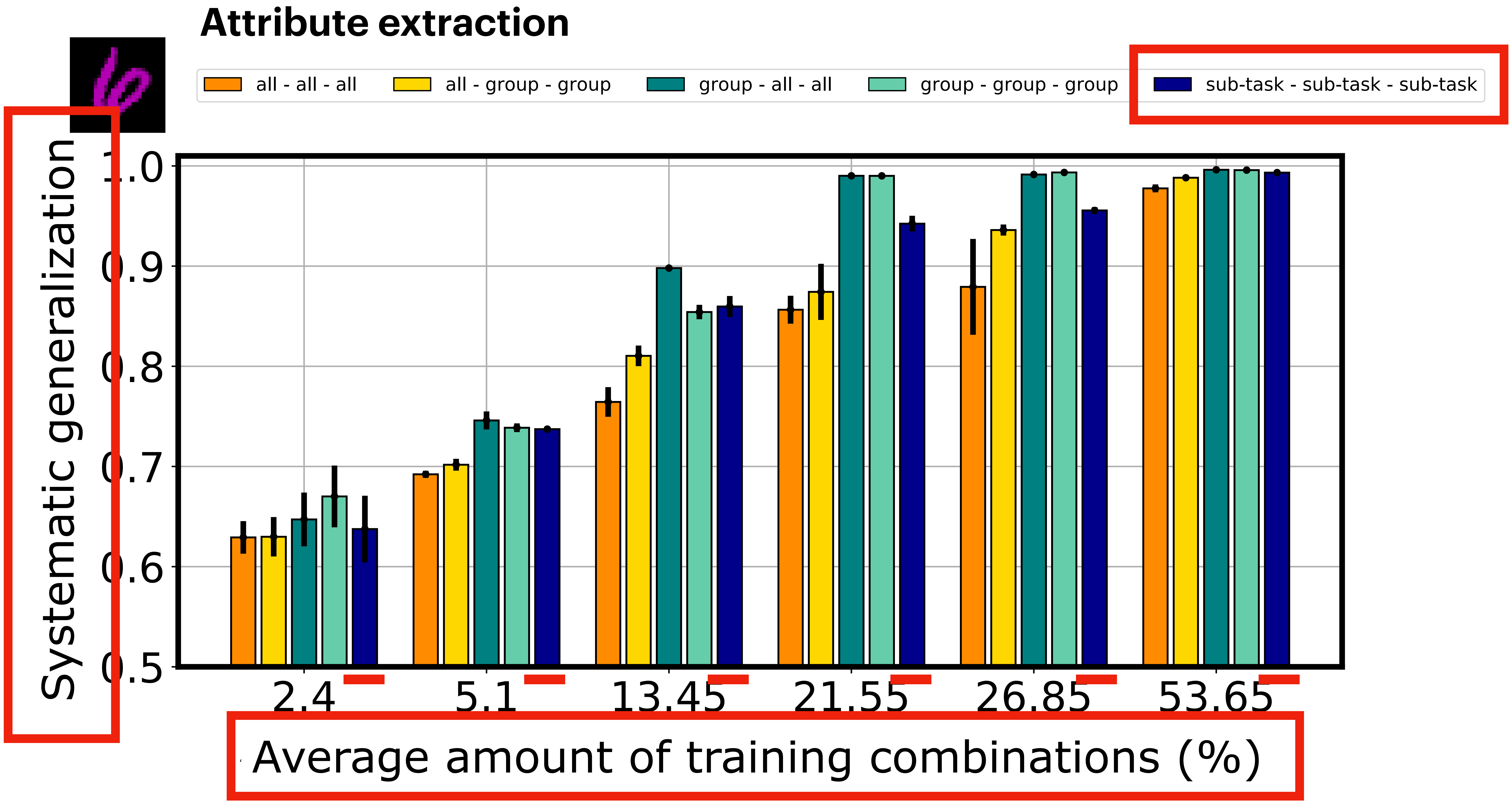

Attribute extraction
all - all - all
all - group - group
group - all - all
group - group - group
sub-task - sub-task - sub-task
Systematic generalization
1.0
0.9
0.8
0.7
0.6
0.5
2.4
5.1
13.45
21.55
26.85
53.65
Average amount of training combinations (%)

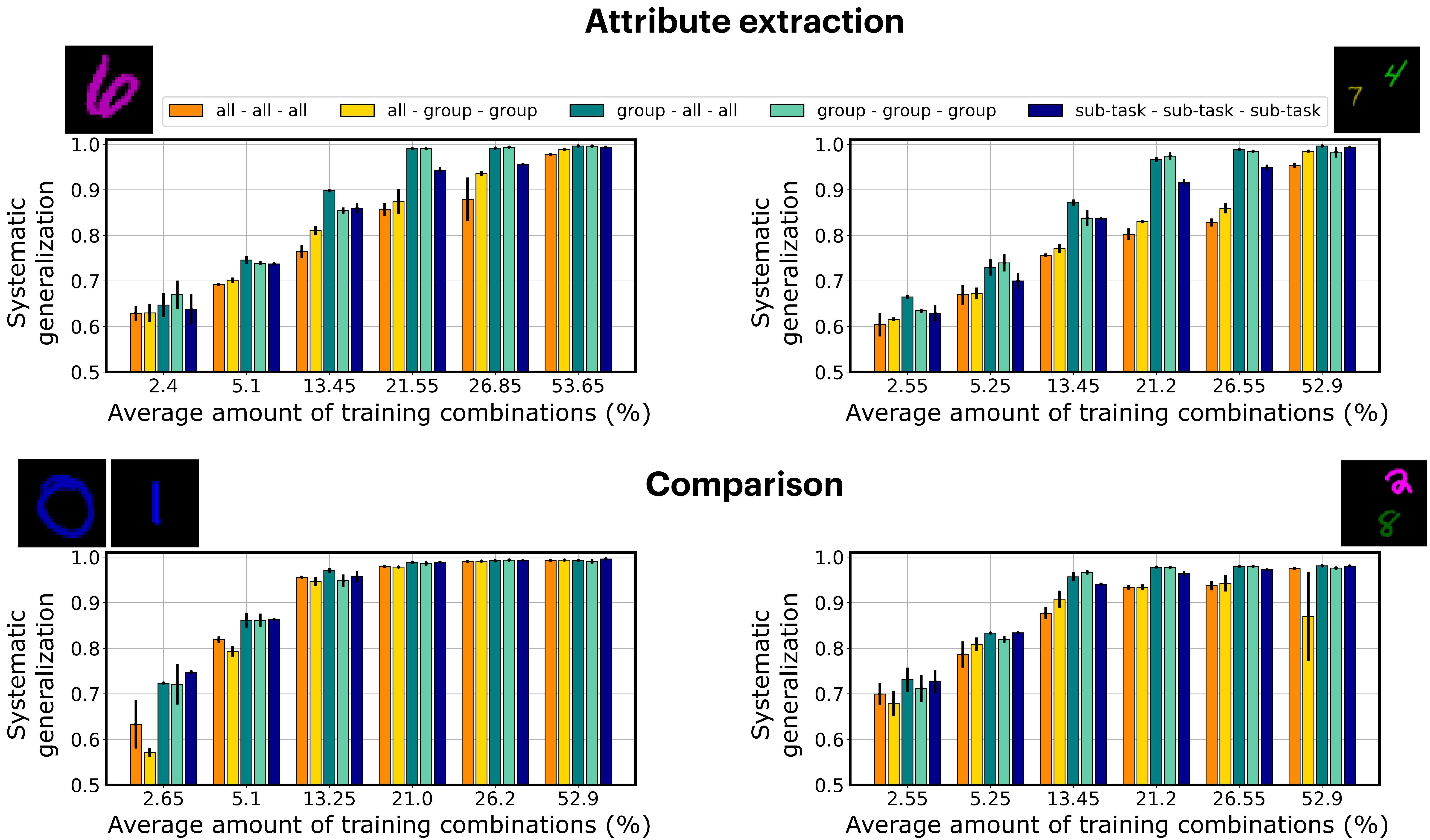


Attribute extraction
all - all - all
all - group - group
group - all - all
group - group - group
sub-task - sub-task - sub-task
Systematic generalization
1.0
0.9
0.8
0.7
0.6
0.5
2.4
5.1
13.45
21.55
26.85
53.65
Average amount of training combinations (%)

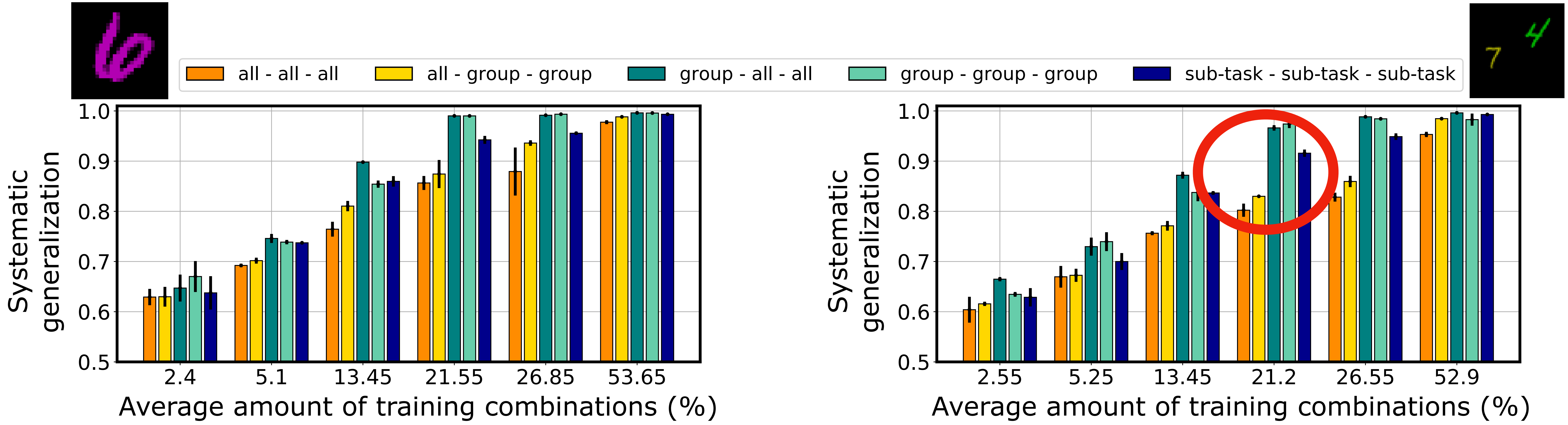
Attribute extraction



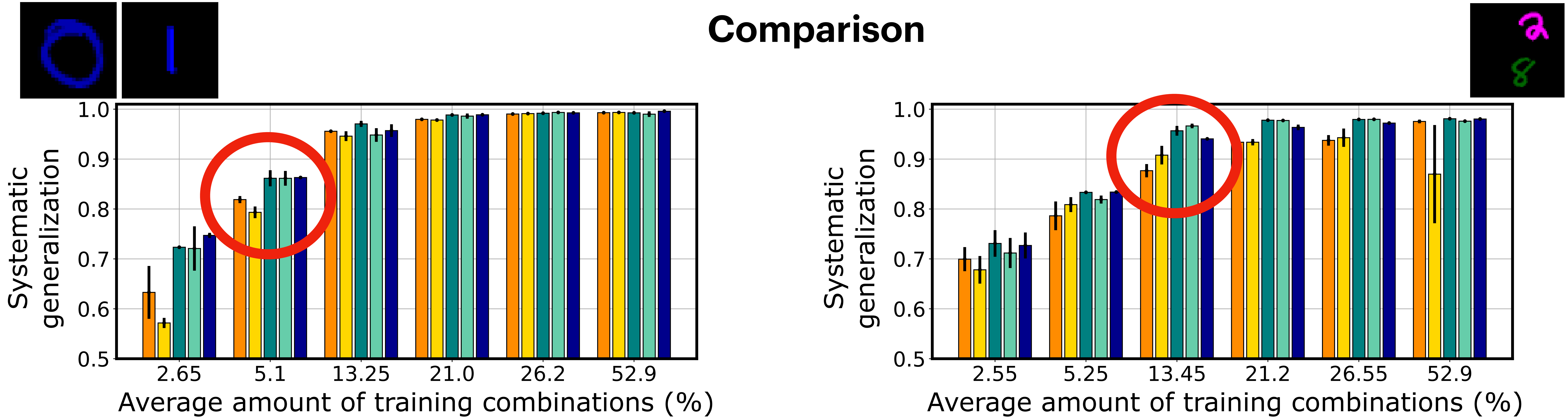

Attribute extraction
all - all - all
all - group - group
group - all - all
group - group - group
sub-task - sub-task - sub-task
Systematic generalization
1.0
0.9
0.8
0.7
0.6
0.5
2.4
5.1
13.45
21.55
26.85
53.65
Average amount of training combinations (%)


Attribute extraction
all - all - all
all - group - group
group - all - all
group - group - group
sub-task - sub-task - sub-task
Systematic generalization
1.0
0.9
0.8
0.7
0.6
0.5
2.4
5.1
13.45
21.55
26.85
53.65
Average amount of training combinations (%)
Systematic generalization
2.55
5.25
13.45
21.2
26.55
52.9
Average amount of training combinations (%)
Comparison
Systematic generalization
2.65
5.1
13.25
21.0
26.2
52.9
Average amount of training combinations (%)
Systematic generalization
2.55
5.25
13.45
21.2
26.55
52.9
Average amount of training combinations (%)

Attribute extraction



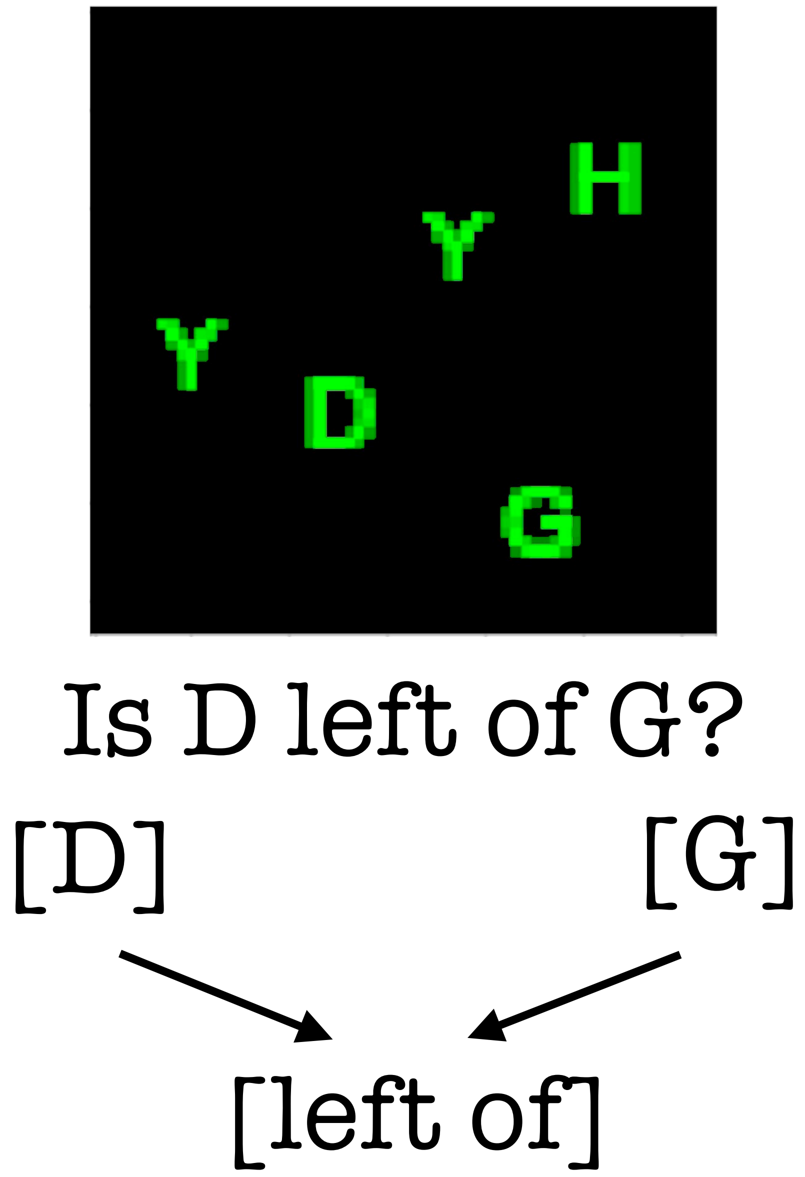
Comparison



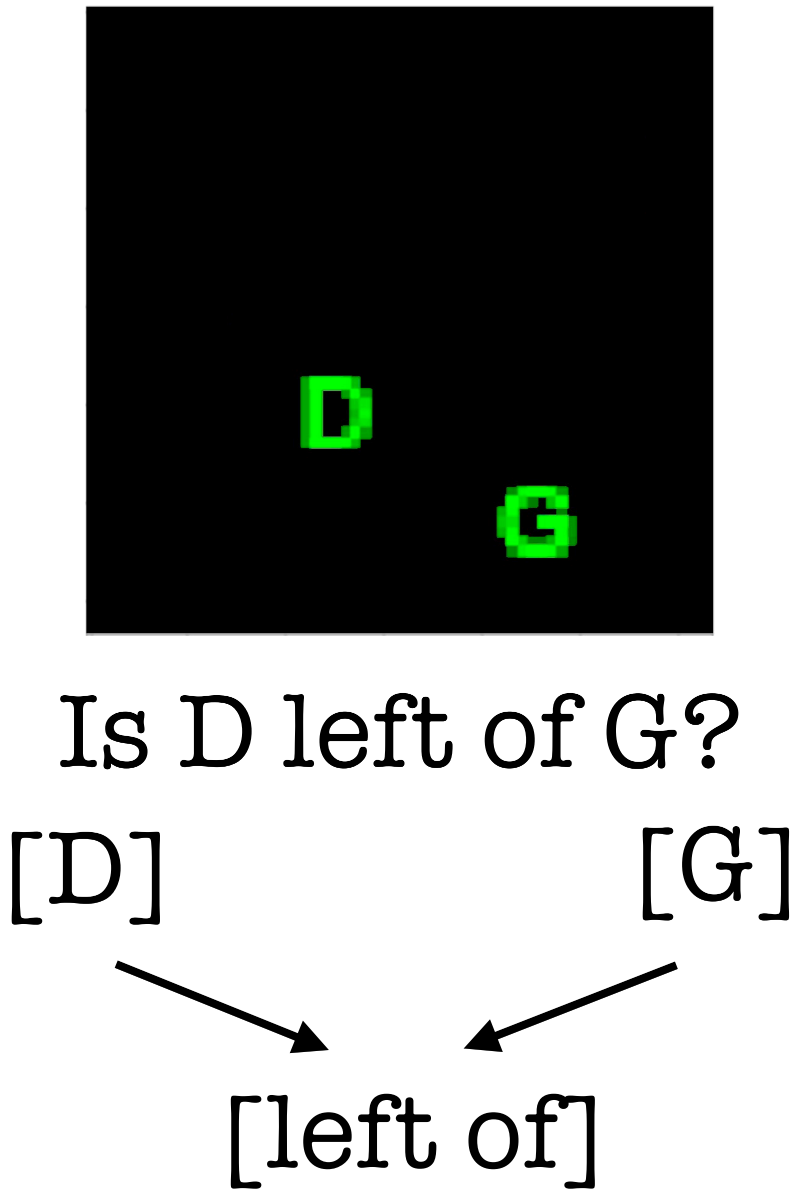
About library choice:
Tuning the degree of modularity,
specially at the image encoder stage,
improves systematic generalization

SQOOP: Limited co-occurrence of objects

Weaker bias



Stronger bias

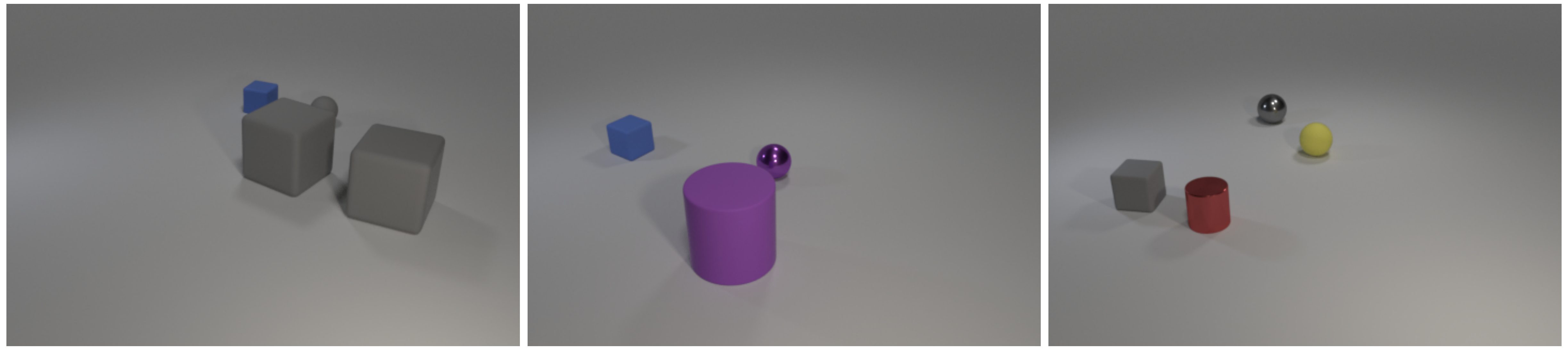


Bahdanau et al. (2019), *Systematic generalization: what is required and can it be learned?*

Systematic generalization performance (%) on SQOOP

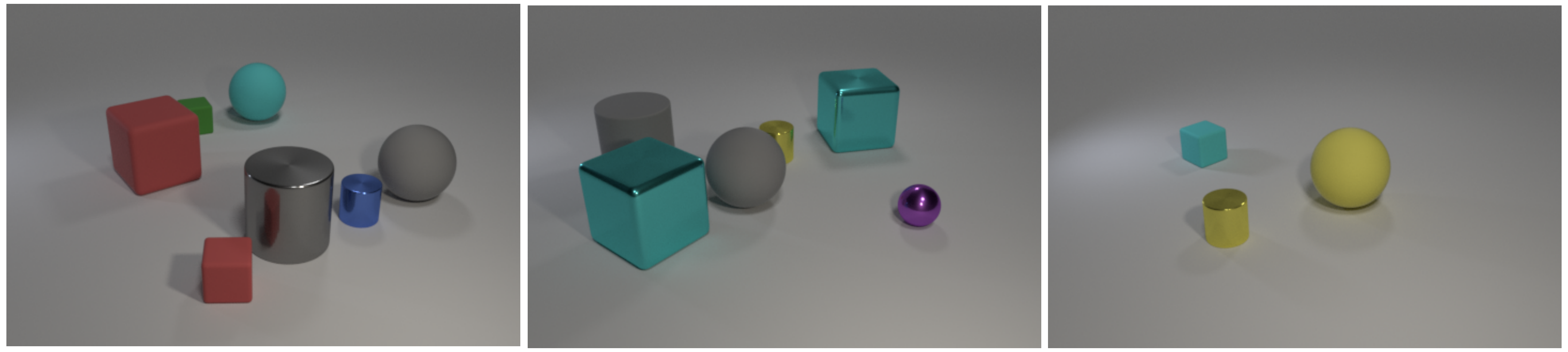
	all - all - all	*all - sub-task - all*
	$\mathbf{99.8 \pm 0.2}$	$\mathbf{99.96 \pm 0.06}$
	84 ± 2	$\boxed{\mathbf{88.5 \pm 0.5}}$

Application on CLEVR-CoGenT split



Condition A

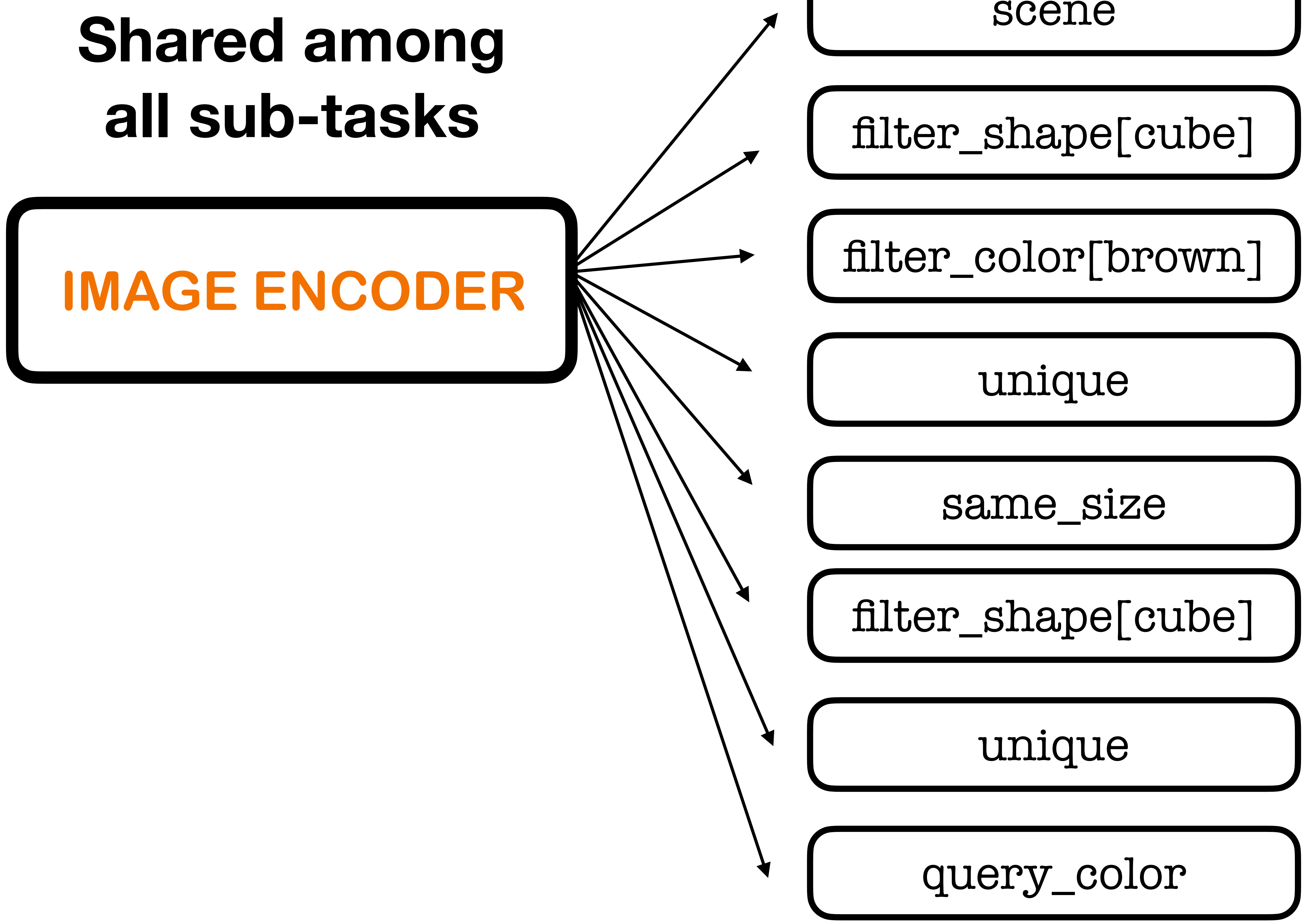
Cubes are gray, blue, brown, or yellow.
Cylinders are red, green, purple, or cyan.
Spheres are all colors.



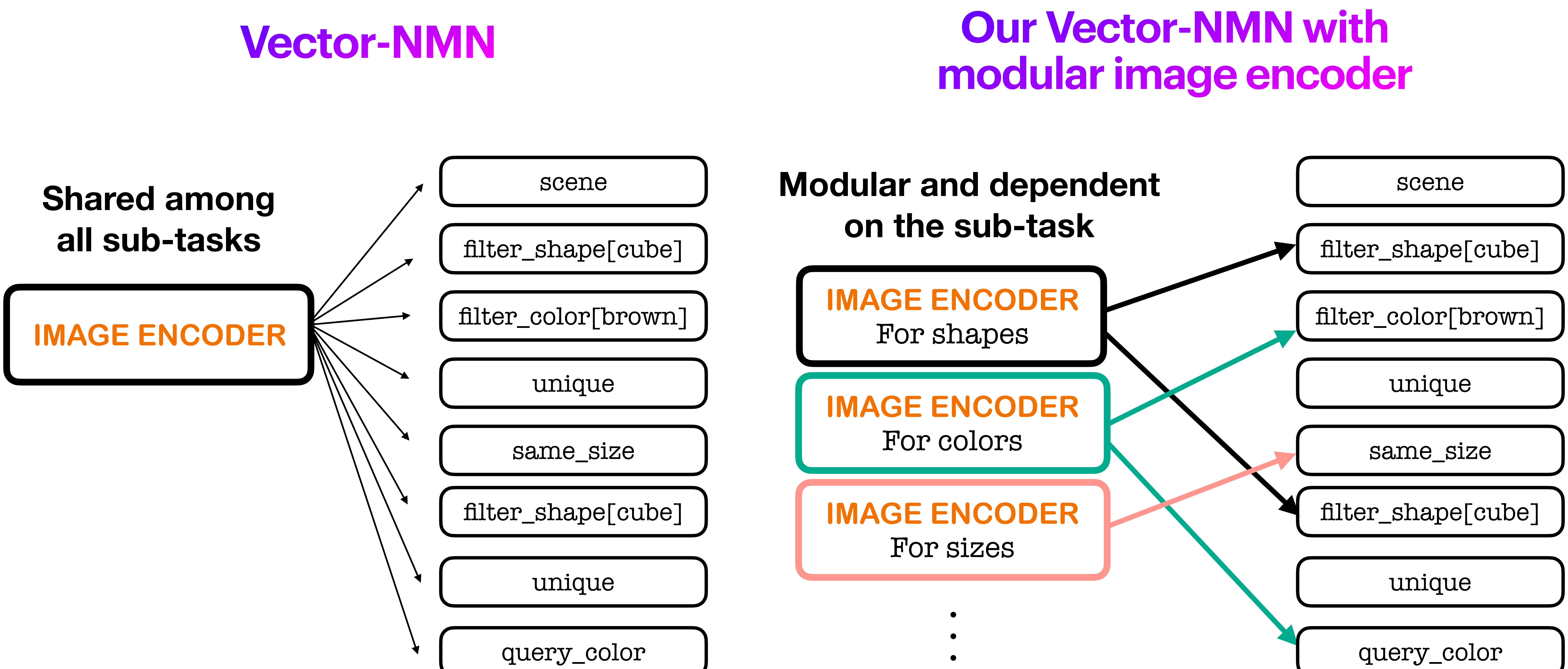
Condition B

Cylinders are gray, blue, brown, or yellow.
Cubes are red, green, purple, or cyan.
Spheres are all colors.

Vector-NMN



CLOSURE: Assessing systematic generalization of CLEVR models
Bahdanau et al. 2020


Vector-NMN
Our Vector-NMN with modular image encoder
Shared among all sub-tasks
IMAGE ENCODER
scene
filter_shape[cube]
filter_color[brown]
unique
same_size
filter_shape[cube]
unique
query_color
Modular and dependent on the sub-task
IMAGE ENCODER For shapes
IMAGE ENCODER For colors
IMAGE ENCODER For sizes
scene
filter_shape[cube]
filter_color[brown]
unique
same_size
filter_shape[cube]
unique
query_color

Systematic generalization performance (%) on CLEVR-CoGenT

	Tensor-NMN	Vector-NMN	Vector-NMN with modular image encoder (**ours**)
count	69.7 ± 0.8	70.4 ± 0.4	$\mathbf{71 \pm 1}$
equal_color	75.6 ± 0.8	74 ± 1	$\mathbf{80 \pm 1}$
equal_integer	82.7 ± 0.3	78 ± 2	$\mathbf{85 \pm 2}$
equal_material	74 ± 2	74.2 ± 0.7	$\mathbf{84 \pm 2}$
equal_shape	$\mathbf{91 \pm 2}$	89 ± 3	79 ± 2
equal_size	75 ± 1	75 ± 1	$\mathbf{88 \pm 2}$
exist	84.2 ± 0.4	$\mathbf{84.4 \pm 0.4}$	84.4 ± 0.5
greater_than	83.8 ± 0.6	83.6 ± 0.4	$\mathbf{89 \pm 1}$
less_than	80.7 ± 0.9	82.0 ± 0.5	$\mathbf{87 \pm 2}$
query_color	58 ± 1	60 ± 1	$\mathbf{67 \pm 4}$
query_material	84.1 ± 0.9	84.7 ± 0.4	$\mathbf{88.2 \pm 0.8}$
query_shape	37 ± 1	40 ± 3	$\mathbf{52 \pm 3}$
query_size	83.5 ± 0.6	84.7 ± 0.7	$\mathbf{89.5 \pm 0.5}$

Systematic generalization performance (%) on CLEVR-CoGenT

	Tensor-NMN	Vector-NMN	Vector-NMN with modular image encoder (**ours**)
`count`	69.7 ± 0.8	70.4 ± 0.4	$\mathbf{71 \pm 1}$
`equal_color`	75.6 ± 0.8	74 ± 1	$\mathbf{80 \pm 1}$
`equal_integer`	82.7 ± 0.3	78 ± 2	$\mathbf{85 \pm 2}$
`equal_material`	74 ± 2	74.2 ± 0.7	$\mathbf{84 \pm 2}$
`equal_shape`	$\mathbf{91 \pm 2}$	89 ± 3	79 ± 2
`equal_size`	75 ± 1	75 ± 1	$\mathbf{88 \pm 2}$
`exist`	84.2 ± 0.4	$\mathbf{84.4 \pm 0.4}$	84.4 ± 0.5
`greater_than`	83.8 ± 0.6	83.6 ± 0.4	$\mathbf{89 \pm 1}$
`less_than`	80.7 ± 0.9	82.0 ± 0.5	$\mathbf{87 \pm 2}$
`query_color`	58 ± 1	60 ± 1	$\mathbf{67 \pm 4}$
`query_material`	84.1 ± 0.9	84.7 ± 0.4	$\mathbf{88.2 \pm 0.8}$
`query_shape`	37 ± 1	40 ± 3	$\mathbf{52 \pm 3}$
`query_size`	83.5 ± 0.6	84.7 ± 0.7	$\mathbf{89.5 \pm 0.5}$

Systematic generalization performance (%) on CLEVR-CoGenT

	Tensor-NMN	Vector-NMN	Vector-NMN with modular image encoder (**ours**)
`count`	69.7 ± 0.8	70.4 ± 0.4	$\mathbf{71 \pm 1}$
`equal_color`	75.6 ± 0.8	74 ± 1	$\mathbf{80 \pm 1}$
`equal_integer`	82.7 ± 0.3	78 ± 2	$\mathbf{85 \pm 2}$
`equal_material`	74 ± 2	74.2 ± 0.7	$\mathbf{84 \pm 2}$
`equal_shape`	$\mathbf{91 \pm 2}$	89 ± 3	79 ± 2
`equal_size`	75 ± 1	75 ± 1	$\mathbf{88 \pm 2}$
`exist`	84.2 ± 0.4	$\mathbf{84.4 \pm 0.4}$	84.4 ± 0.5
`greater_than`	83.8 ± 0.6	83.6 ± 0.4	$\mathbf{89 \pm 1}$
`less_than`	80.7 ± 0.9	82.0 ± 0.5	$\mathbf{87 \pm 2}$
`query_color`	58 ± 1	60 ± 1	$\mathbf{67 \pm 4}$
`query_material`	84.1 ± 0.9	84.7 ± 0.4	$\mathbf{88.2 \pm 0.8}$
`query_shape`	37 ± 1	40 ± 3	$\mathbf{52 \pm 3}$
`query_size`	83.5 ± 0.6	84.7 ± 0.7	$\mathbf{89.5 \pm 0.5}$

Conclusions

Library choice:

- Tuning the degree of modularity improves systematic generalization
- Critical at the image encoder stage (for bias in the image)

New research questions

Other types of bias

Neural mechanisms for systematic generalization

How Modular Should Neural Module Networks Be for Systematic Generalization?

Vanessa D'Amario[1,3], Tomotake Sasaki[2,3], Xavier Boix[1,3]

1

2

3 CENTER FOR Brains Minds+ Machines