



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA



Project page

# LOOKING AT WORDS AND POINTS WITH ATTENTION: A BENCHMARK FOR TEXT-TO-SHAPE COHERENCE

Amaduzzi A., Lisanti G., Salti S., Di Stefano L.

University of Bologna - Department of Computer Science and Engineering

# ICCV23

PARIS

## GPT2Shape



"This is a professional table color is silver"

"Silver rectangular table looks like metal material,"

"This student desk is grey and has four legs. It is thin and has slotted sections for supplies.,"

"grey desk"

"Long rectangle shaped school desk"

**Describe geometry, shape, color and other key features in less than 40 words, in 5 different ways**

"This sleek silver desk has a rectangular shape with four thin legs and slotted sections for supplies."

"This desk is grey with a rectangular shape and four legs. It has a metal-like finish and slots for supplies."

"This grey desk has a rectangular shape and four legs. It has a silver finish and slots for organizing supplies."

"This rectangular desk is grey with a silver finish and four thin legs. It has slots for supplies."

"This metal-like desk is grey with a rectangular shape and four legs. It has a silver finish and slots for supplies."

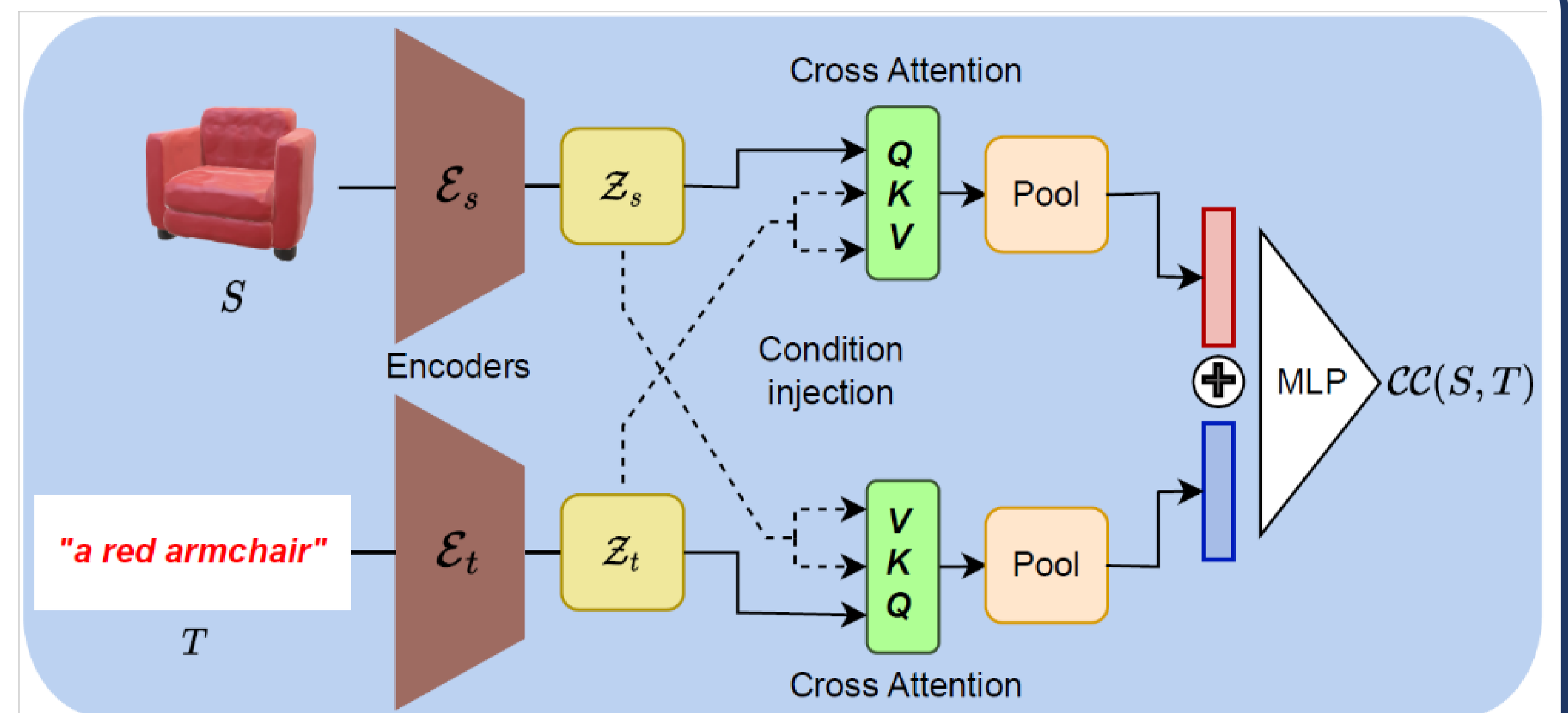
Text2Shape descriptions

Request to GPT-3

## GPT2Shape vs Text2Shape

Dataset	reference (↑)	cannot decide (↓)	distractor (↓)
Text2Shape	65.65 %	24.95 %	9.40 %
GPT2Shape	<b>75.77 %</b>	<b>16.21 %</b>	<b>8.02 %</b>

## CrossCoherence



## HST: human-validated test set

Test set	ShapeGlot	CrossCoherence	CLIP R-precision
HST chairs	5.36 %	<b>17.26 %</b>	9.87 %
full HST	-	<b>16.85 %</b>	9.38 %

Method	Acc. on chairs (↑)	Acc. on full HST (↑)
ShapeGlot	70.21 %	-
CLIP-Similarity	77.79 %	77.06 %
CrossCoherence	<b>81.04 %</b>	<b>80.45 %</b>

## Evaluation of generative models

Method	CrossCoherence	CLIP R-precision	
Point-E	4.92 %	5.95 %	
Shap-E	<b>7.09 %</b>	<b>20.41 % (STF)</b>	<b>15.78 % (NeRF)</b>
Liu et al.	5.09 %	5.83 %	

Method	CrossCoherence	CLIP-Similarity	CLIP-Similarity
Point-E	27.33 %	24.36 %	27.04 %
Shap-E	<b>38.02 %</b>	<b>65.47 % (STF)</b>	<b>57.18 % (NeRF)</b>
Liu et al.	34.65 %	10.18 %	15.78 %

## Results

	Point-E	Shap-E	Liu et al.
A tan tear drop shaped wooden table with a grey cubicle base.			
A brown color rectangular wooden table with four black legs.			
A tall, black-upholstered high chair with four brown wooden legs.			
This armchair has a square back, tan leather upholstery, and light brown wooden arms and legs.			

**CLIP-Similarity    CrossCoherence**

	Shap-E
Rectangular wooden table with <b>two legs</b> , light brown in color <b>with 10 drawers, 5 on each side.</b>	
Rectangular steel chair <b>with armrests</b> and four legs, white <b>wooden finish.</b>	
A brown color wooden table with oval shaped top and <b>four legs.</b>	
This is a tall, brown, rectangular table <b>with a door that opens on the end.</b> It is made of wood.	

**CLIP R-precision    CrossCoherence**