

Specify and Edit: Overcoming Ambiguity in Text-Based Image Editing

Ekaterina Iakovleva*¹, Fabio Pizzati*^{2,3}, Philip Torr³, Stéphane Lathuilière^{1,4}

¹ LTCI, Télécom Paris, Institut Polytechnique de Paris

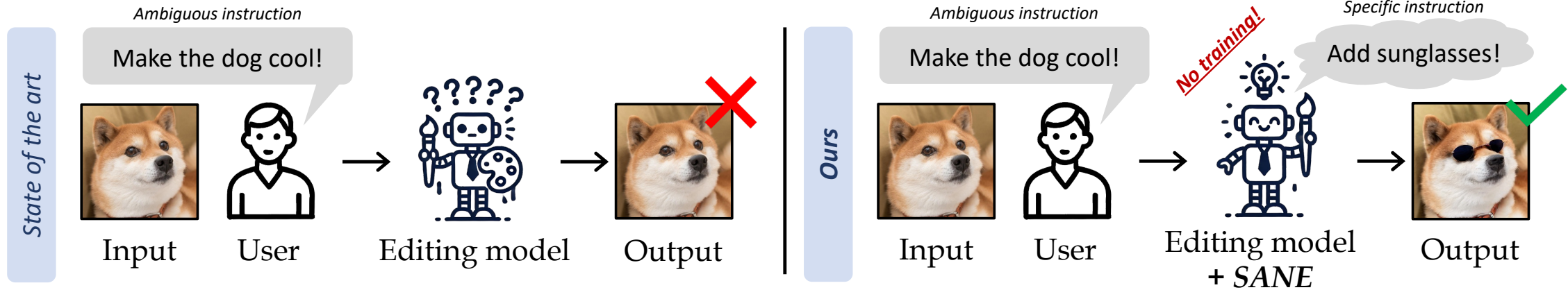
² MBZUAI

³ University of Oxford

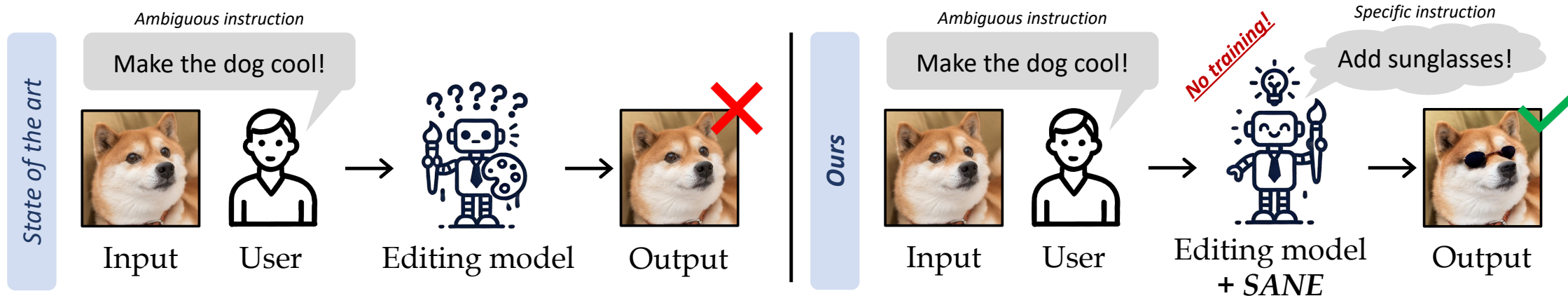
⁴ Inria, Univ. Grenoble Alpes



Motivation: ambiguous instructions



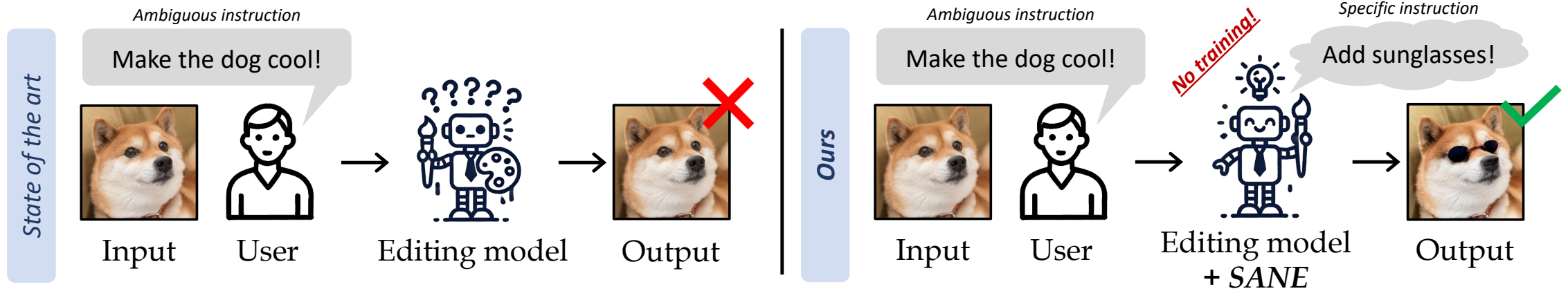
Motivation: ambiguous instructions



- What does it mean for a dog to be “cool”?



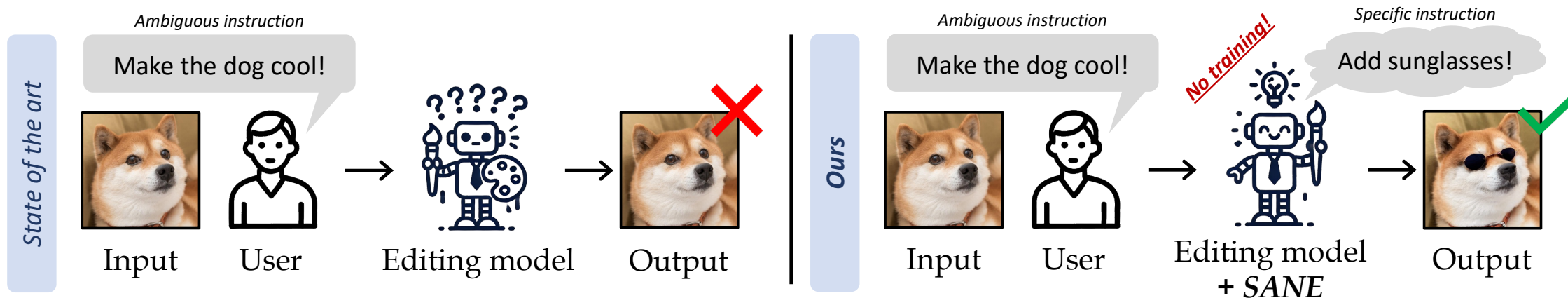
Motivation: ambiguous instructions



- What does it mean for a dog to be “cool”?
- Is there only one way to be cool?



Motivation: ambiguous instructions

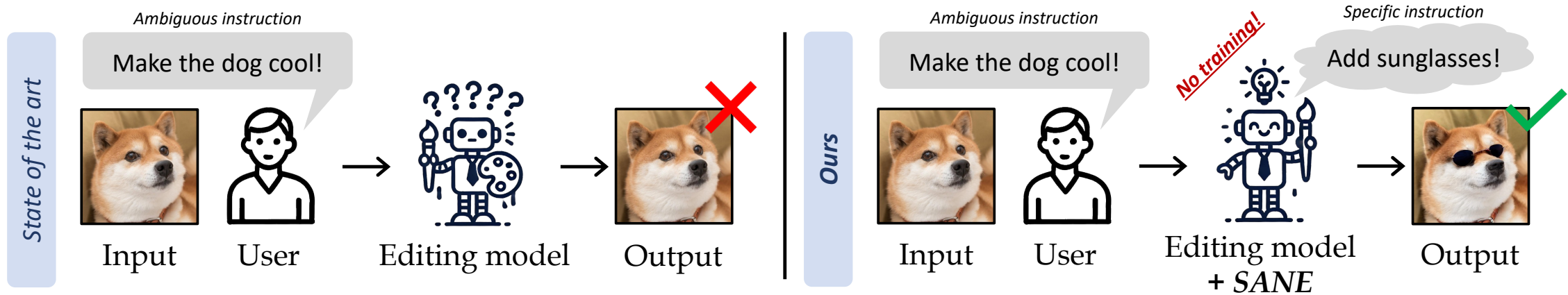


- What does it mean for a dog to be “cool”?
- Is there only one way to be cool?

Instructions can be **ambiguous**, i.e., subject to interpretation or have multiple realisations.



Motivation: ambiguous instructions

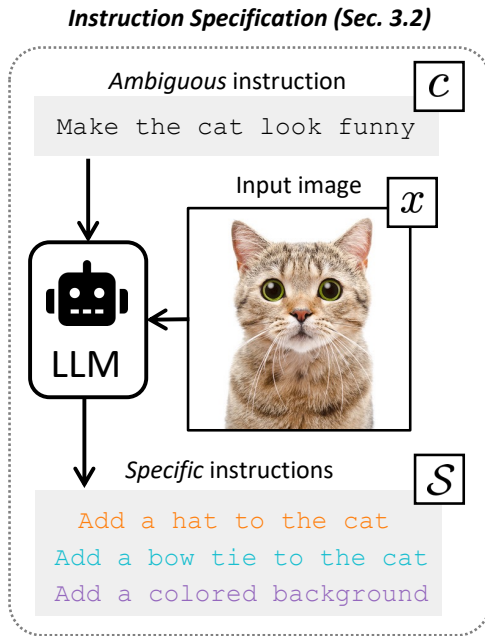


- What does it mean for a dog to be “cool”?
- Is there only one way to be cool?

Instructions can be **ambiguous**, i.e., subject to interpretation or have multiple realisations.
Can we reverse the associated abstractions through the process of **specification**?



SANE: Specify and Edit

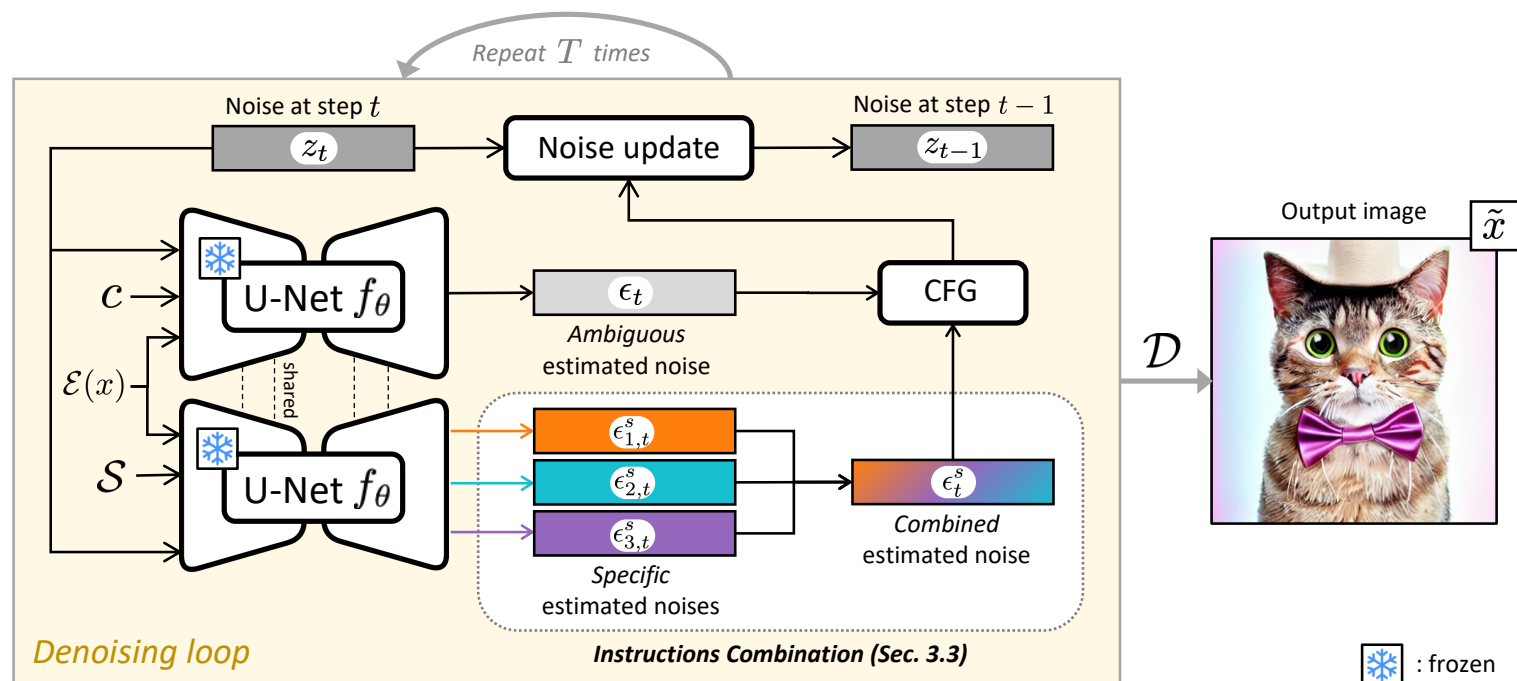
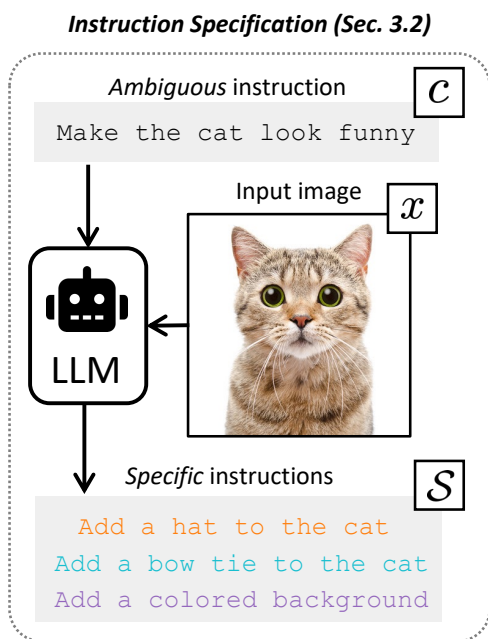


We propose:

1. to **decompose** an ambiguous instruction into a set of specific instructions with a **LLM**;



SANE: Specify and Edit



We propose:

1. to **decompose** an ambiguous instruction into a set of specific instructions with a **LLM**;
2. to **combine** these instructions using our novel training-free pipeline.



Experiments

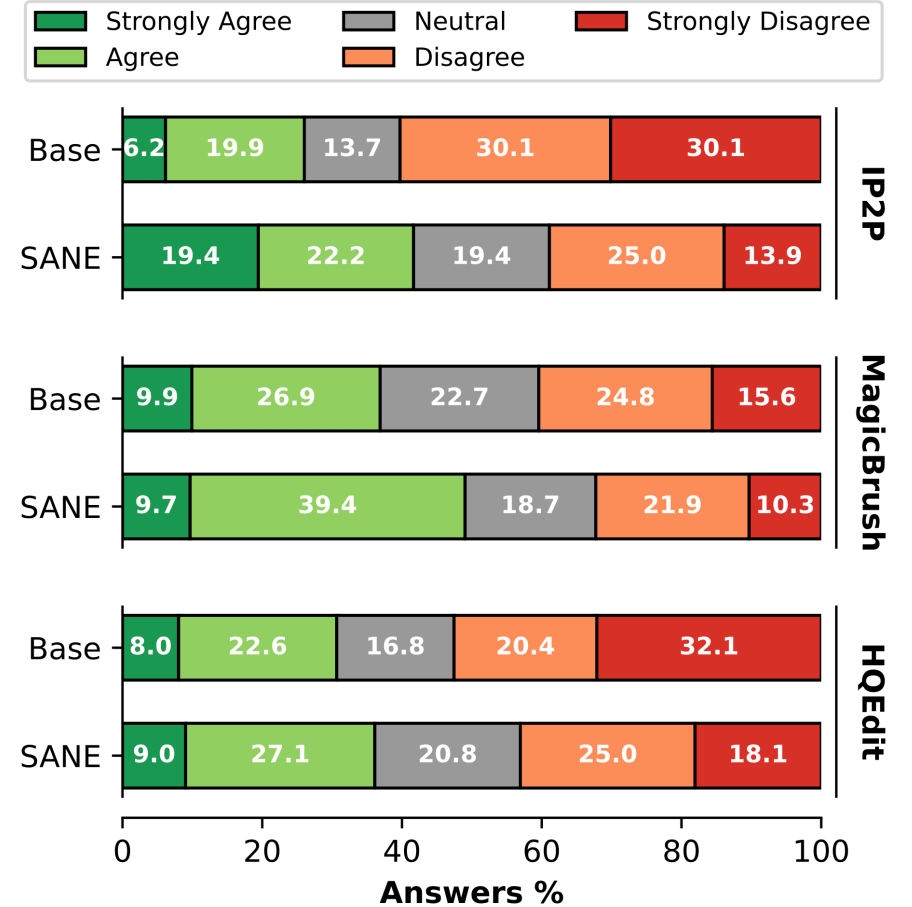
<i>EMU-Edit</i>					<i>IP2P data</i>			
Method	N	CLIP _d $\uparrow$	CLIP _i $\uparrow$	CLIP _{Δ} $\uparrow$	CLIP _d $\uparrow$	CLIP _i $\uparrow$	CLIP _{Δ} $\uparrow$	
<i>IP2P</i>	Baseline	-	0.2923	0.8810	0.1203	0.2903	0.9027	0.1724
	Rewrite (long)	-	0.2934	<u>0.8730</u>	0.0981	0.2856	<u>0.8965</u>	0.1466
	Rewrite (short)	-	0.2960	0.8482	0.1204	0.2875	0.8764	0.1652
	SANE	1	0.2961	0.7779	0.1705	0.2886	0.8196	0.1999
		2	<u>0.2962</u>	0.7654	<u>0.1785</u>	<u>0.2910</u>	0.8107	0.2063
		3	0.2968	0.7531	0.1858	0.2935	0.8101	<u>0.2057</u>
<i>MB</i>	Baseline	-	0.2888	0.7858	0.1618	0.2855	0.7934	0.2028
	Rewrite (long)	-	0.2750	0.7456	0.1204	0.2769	0.7716	0.1690
	Rewrite (short)	-	0.2878	0.7443	0.1487	0.2748	0.7868	0.1711
	SANE	1	0.2948	0.7994	<u>0.1661</u>	0.2848	0.7970	0.1992
		2	<u>0.2952</u>	<u>0.8134</u>	0.1669	<u>0.2870</u>	0.8068	0.1996
		3	0.3006	0.8209	0.1655	0.2878	<u>0.8056</u>	<u>0.1998</u>
<i>HQEdit</i>	Baseline	-	0.2675	0.6501	0.1417	<u>0.2785</u>	0.7035	<u>0.1848</u>
	Rewrite (long)	-	0.2788	0.6536	0.1233	0.2722	0.6972	0.1691
	Rewrite (short)	-	0.2839	0.6350	0.1353	0.2716	0.6838	0.1786
	SANE	1	0.2725	0.6546	<u>0.1458</u>	0.2734	0.6997	0.1759
		2	0.2789	<u>0.6802</u>	0.1433	0.2782	<u>0.7057</u>	0.1842
		3	<u>0.2823</u>	0.6870	0.1474	0.2818	0.7136	0.1855



Experiments

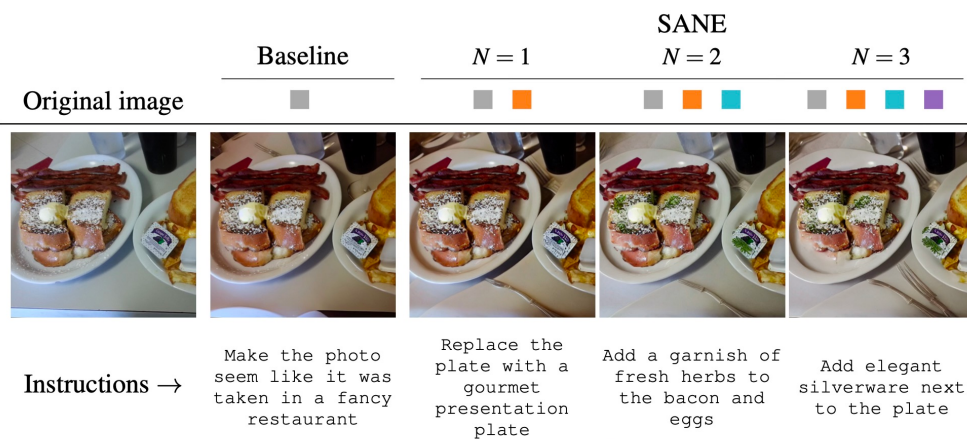
EMU-Edit					IP2P data			
Method	N	CLIP _d $\uparrow$	CLIP _i $\uparrow$	CLIP $_{\Delta}$ $\uparrow$	CLIP _d $\uparrow$	CLIP _i $\uparrow$	CLIP $_{\Delta}$ $\uparrow$	
IP2P	Baseline	-	0.2923	0.8810	0.1203	0.2903	0.9027	0.1724
	Rewrite (long)	-	0.2934	<u>0.8730</u>	0.0981	0.2856	<u>0.8965</u>	0.1466
	Rewrite (short)	-	0.2960	0.8482	0.1204	0.2875	0.8764	0.1652
	SANE	1	0.2961	0.7779	0.1705	0.2886	0.8196	0.1999
		2	<u>0.2962</u>	0.7654	<u>0.1785</u>	<u>0.2910</u>	0.8107	0.2063
		3	0.2968	0.7531	0.1858	0.2935	0.8101	<u>0.2057</u>
MB	Baseline	-	0.2888	0.7858	0.1618	0.2855	0.7934	0.2028
	Rewrite (long)	-	0.2750	0.7456	0.1204	0.2769	0.7716	0.1690
	Rewrite (short)	-	0.2878	0.7443	0.1487	0.2748	0.7868	0.1711
	SANE	1	0.2948	0.7994	<u>0.1661</u>	0.2848	0.7970	0.1992
		2	<u>0.2952</u>	<u>0.8134</u>	0.1669	<u>0.2870</u>	0.8068	0.1996
		3	0.3006	0.8209	0.1655	0.2878	<u>0.8056</u>	<u>0.1998</u>
HQEdit	Baseline	-	0.2675	0.6501	0.1417	<u>0.2785</u>	0.7035	<u>0.1848</u>
	Rewrite (long)	-	0.2788	0.6536	0.1233	<u>0.2722</u>	0.6972	0.1691
	Rewrite (short)	-	0.2839	0.6350	0.1353	0.2716	0.6838	0.1786
	SANE	1	0.2725	0.6546	<u>0.1458</u>	0.2734	0.6997	0.1759
		2	0.2789	<u>0.6802</u>	0.1433	0.2782	<u>0.7057</u>	0.1842
		3	<u>0.2823</u>	0.6870	0.1474	0.2818	0.7136	0.1855

“the objective set by the user instruction is achieved”



Experiments

InstructPix2Pix



MagicBrush



HQEdit



Code available:

