



- Reconstructing unchanged representations to capture a stable difference representation for cross-view images
- interact cross-view features: multi-head token-wise matching
- learn view-invariant representations: cross-view contrastive alignment
- > Rebuilding the "after" image with the full representations of caption and "before" image to improve captioning quality
- model "hallucination" via the caption and "before"
- match "hallucination" with "after": cross-view contrastive alignment

2. Approach Overview



- Stage 1: cross-view feature extraction
- Stage 2: self-supervised cross-view representation reconstruction (SCORER)
- Stage 3: difference representation modeling and caption generation
- Stage 4: cross-modal backward reasoning (CBR)

Self-supervised Cross-view Representation Reconstruction for Change Captioning Yunbin Tu¹, Liang Li², Li Su¹, Zheng-Jun Zha³, Chenggang Yan⁴ and Qingming Huang¹ ¹University of Chinese Academy of Sciences, ²Institute of Computing Technology, CAS, ³University of Science and Technology of China, ⁴Hangzhou Dianzi University



PCL w/ Pre-training (AAAI 2022) [38] 51.2 36.2 71.7 128.9 -	Mathad			D	М	Total	C	S	 	Sem	antic Ch	ange
$\frac{M-VAM+RAF(ECCV 2020) [25]}{DUDA+(CCVP 2021) [7]} \underbrace{51.3}{37.8} \underbrace{70.4}{7.5} \underbrace{115.8}{30.7} \underbrace{30.7}{1.5} \underbrace{42.9}{42.9} \underbrace{22.7}{2.5} \underbrace{94.}{2.9} \underbrace{2010 [7]}{PUDA+(CCVP 2021) [7]} \underbrace{51.2}{51.2} \underbrace{37.7}{70.5} \underbrace{115.4}{31.1} \underbrace{49.9}{42.9} \underbrace{22.7}{36.2} \underbrace{69.8}{4.64} \underbrace{116} \underbrace{116}{N^{2}CC(CV 2021) [11]} \underbrace{52.4}{37.5} \underbrace{37.5}{1.14.2} \underbrace{31.0}{31.1} \underbrace{49.9}{42.9} \underbrace{52.7}{36.2} \underbrace{69.8}{4.66} \underbrace{116} \underbrace{116}{N^{2}CC(CV 2021) [12]} \underbrace{52.4}{37.5} \underbrace{37.5}{1.14.2} \underbrace{31.0}{31.1} \underbrace{49.9}{42.9} \underbrace{52.7}{36.2} \underbrace{69.8}{4.66} \underbrace{116} \underbrace{116}{N^{2}CC(CV 2021) [12]} \underbrace{51.1}{52.4} \underbrace{40.6}{73.3} \underbrace{71.2}{12.8} \underbrace{32.2} \underbrace{52.7}{36.4} \underbrace{69.7}{114.2} \underbrace{116} \underbrace{20.6} \underbrace{32.9}{1.6} \underbrace{52.7}{36.4} \underbrace{69.7}{1.14.2} \underbrace{116} \underbrace{116}{N^{2}CC(CV 2021) [12]} \underbrace{51.4}{32.4} \underbrace{40.6}{33.0} \underbrace{73.1} \underbrace{12.6}{2.7} \underbrace{32.9}{32.7} \underbrace{36.4} \underbrace{69.7}{1.16} \underbrace{116} \underbrace{60.7}{1.16} \underbrace{116} \underbrace{116}{N^{2}CC(CV 2021) [12]} \underbrace{52.4}{32.4} \underbrace{33.3} \underbrace{12.16}{2.2} \underbrace{32.9}{2.7} \underbrace{36.4} \underbrace{69.7}{1.16} \underbrace{116} \underbrace{116}{N^{2}CC(CV 2021) [24]} \underbrace{52.4}{32.4} \underbrace{52.7}{33.1} \underbrace{122.6}{32.2} \underbrace{32.9}{52.7} \underbrace{36.4} \underbrace{69.7}{1.16} \underbrace{116} \underbrace{50.6} \underbrace{50.7}{1.16} \underbrace{116} \underbrace{50.6} \underbrace{50.7}{1.16} \underbrace{116} \underbrace{50.6} \underbrace{50.7}{1.16} \underbrace{116} \underbrace{50.7}{1.16} \underbrace{50.7} \underbrace{70.7}{1.16} \underbrace{118} \underbrace{70.7}{1.16} \underbrace{70.7}{1.17} $	PCL w/ Pre-training (A	AAI 2022	2) [38]	51.2	36.2	к 71.7	128.9	-	D -	-	- -	-
$\frac{DUDA (ICCV 2019) [22]}{DUDA+(CVPR 2021) [7]} = \frac{47.3}{51.2} \frac{33.9}{37.7} \frac{-112.3}{70.5} \frac{24.5}{115.4} \frac{42.9}{31.1} \frac{29.7}{9.4} \frac{-9.4}{9.6} \frac{9.43}{9.6} \frac{56.4}{9.6} \frac{101}{101} R^{3} ket4s SP (EMNLP 2021) [31]} \frac{54.7}{52.4} \frac{33.9}{35.7} \frac{-114.2}{112.0} \frac{31.0}{22.6} \frac{52.7}{52.6} \frac{52.7}{56.2} \frac{56.8}{116} \frac{101}{101} \frac{101}{102} \frac{52.4}{102} \frac{37.7}{33.1} \frac{112.2}{12.2} \frac{31.0}{32.6} \frac{-2.6}{2.6} \frac{-2.6}{1.6} \frac{-2.6}{1.7} \frac{-2.6}{1.7} \frac{-2.6}{1.6} \frac{-2.6}{1.6} \frac{-2.6}{1.6} \frac{-2.6}{1.6} \frac{-2.6}{1.7} \frac{-2.6}{1.7} \frac{-2.6}{1.6} \frac{-2.6}{1.6} \frac{-2.6}{1.7} \frac{-2.6}{1.7} \frac{-2.6}{1.6} \frac{-2.6}{1.6} \frac{-2.6}{1.7} \frac{-2.6}{1.7} \frac{-2.6}{1.7} \frac{-2.6}{1.6} \frac{-2.6}{1.6} \frac{-2.6}{1.7} -2.6$	M-VAM+RAF (ECC	V 2020) [25]	51.3	37.8	70.4	115.8	30.7	-	-	-	-
	DUDA (ICCV 20)19) [22]		47.3	33.9	-	112.3	24.5	42.9	29.7	-	94.6
VACC (ICCV 2021) [11] S24 37.5 - 114.2 31.0 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -	$R^{3}Net+SSP$ (EMNL)	P 2021) [7]	31]	54.7	39.8	70.5 73.1	113.4	32.6	49.9 52.7	34.3 36.2	69.8	116.6
SRDR1+AVS (ACL 2021) [24] 52.4 40.2 73.3 122.2 32.9 52.7 36.4 69.7 14 MCCFormers-D (ICCV 2021) [24] 52.4 38.3 - 121.6 26.8	VACC (ICCV 20 SGCC (ACM MM)21) [11] 2021) [15	5]	52.4 51.1	37.5 40.6	- 73.9	114.2 121.8	31.0 32.2	-	-	-	-
$\frac{ FDC (TMM 2022) [9]}{NCT (TMM 2023) [30]} = \frac{49.2}{55.1} \frac{32.5}{40.2} \frac{69.1}{78.1} \frac{118.7}{124.1} \frac{32.9}{32.5} \frac{53.3}{13.3} \frac{36.5}{16.5} \frac{70.7}{17.7} \frac{118}{122} \frac{124.1}{32.9} \frac{32.5}{33.3} \frac{54.4}{32.6} \frac{37.4}{17.7} \frac{71.5}{122} \frac{122.5}{55.8} \frac{40.1}{40.1} \frac{73.8}{73.8} \frac{124.1}{24.1} \frac{32.9}{32.6} \frac{54.1}{37.4} \frac{37.4}{71.5} \frac{71.5}{122} \frac{122.5}{56.3} \frac{41.2}{41.2} \frac{74.5}{74.5} \frac{126.6}{33.0} \frac{33.0}{54.4} \frac{54.1}{37.6} \frac{37.4}{71.7} \frac{71.7}{122}$	SRDRL+AVS (ACL MCCFormers-D (ICC	2021) [3 2V 2021) [2] [24]	54.9 52.4	40.2 38.3	73.3	122.2 121.6	32.9 26.8	52.7	36.4	69.7 -	- 114.2
$V_{ARD-Trains}(TIP 2023)[28] = 55.4 + 40.1 73.8 + 12.4 + 32.6 + 32.6 + 32.6 + 54.1 + 37.6 + 71.7 + 122 + 32.6 + 32.6 + 54.1 + 37.6 + 71.7 + 122 + 32.6 + 32.6 + 32.6 + 32.6 + 54.1 + 37.4 + 71.5 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 32.6 + 33.3 + 34.4 + 37.6 + 71.7 + 122 + 31.2 + $	IFDC (TMM 20 NCT (TMM 20)22) [9]		49.2	32.5 40.2	69.1 73.8	118.7 124 1	-	47.2	29.3	63.7 70.7	105.4 118.4
SCORER(Ours) = 53.8 + 40.8 + 74.0 + 126.0 + 33.0 + 34.1 + 37.4 + 71.5 + 122 + 32.0 + 33.0 + 34.1 + 37.6 + 71.7 + 122 + 33.0 + 34.1 + 37.6 + 71.7 + 122 + 33.0 + 126.8 + 33.3 + 14 + 37.6 + 71.7 + 122 + 34.1 + 37.6 + 71.7 + 122 + 33.0 + 14.9 + 35.0 + 14.9 + 35.0 + 14.9 + 35.0 + 16.5 + 116.4 + 35.0 + 16.5 + 16.5 + 116.4 + 35.0 + 16.5 + 16.5 + 16.4 + 37.6 + 71.7 + 122.4 + 31.6 + 62.0 + 51.7 + 77.4 + 117.9 + 35.0 + 16.5	VARD-Trans (TIP	2023) [30] 2023) [28	3]	55.4	40.1	73.8	124.1	32.0	-	-	-	-
Ablation studyMethodBMRCSBMRCSubtraction50.234.167.1108.02857.348.474.7113.85R53.337.170.8119.130.461.150.776.4114.92SCORER54.337.571.5122.031.261.450.676.5116.42SCORER54.137.471.5122.431.260.751.276.9114.92SCORER+CBR54.437.671.7122.431.662.051.777.4117.93	SCORER (O SCORER+CBR	Durs) R (Ours)		55.8 56.3	40.8 41.2	74.0 74.5	126.0 126.8	33.0 33.3	54.1 54.4	37.4 37.6	71.5 71.7	122.0 122.4
$\begin{array}{c} \begin{array}{c} SCORER \\ RR+CBR \\ SCORER + CBR \\ SCORER + CBR \\ \hline 54.1 \\ 37.4 \\ 71.5 \\ \hline 122.4 \\ 31.2 \\ \hline 60.7 \\ 51.2 \\ 70.9 \\ \hline 114.9 \\ \hline 54.4 \\ 37.6 \\ 71.7 \\ 122.4 \\ \hline 31.6 \\ \hline 62.0 \\ 51.7 \\ 77.4 \\ \hline 117.9 \\ \hline 77.4 \\ $	Method Subtraction RR	B 50.2	M 34.1	R 67.1 70.8	C 108. 119.	0 2	S 28 5 0.4 6	в 57.3 51.1	M 48.4 50.7	R 74.7 76.4	C 113.8 114.9	S 34.
Subtraction 50.2 34.1 67.1 108.0 28 57.3 48.4 74.7 113.8 3 RR 53.3 37.1 70.8 119.1 30.4 61.1 50.7 76.4 114.9 3 SCORER 54.3 37.5 71.5 122.0 31.2 61.4 50.6 76.5 116.4 3 SCORER+CBR 54.1 37.4 71.5 122.4 31.2 60.7 51.2 76.9 114.9 3 SCORER+CBR 54.4 37.6 71.7 122.4 31.6 62.0 51.7 77.4 117.9 3	Method		M	R	C	0	S	B	M	R	C	S
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	RR	50.2	27.1	70.8	119.	1 3	$\begin{array}{c c} 20 \\ 0.4 \\ \end{array}$	57.5 51.1	50.7	76.4	114.9	34. 34.
SCORER+CBR54.437.671.7122.431.662.051.777.4117.93SCORER+CBR54.437.671.7122.431.662.051.777.4117.93Scorer \circ		53.3	37.1	70.0								
$ \sqrt{\text{isualization for shared object matching and caption gene} } \\ (2) + ($	SCORER RR+CBR	53.3 54.3 54.1	37.1 37.5 37.4	71.5 71.5	122. 122.	0 3 4 3	1.2 (1.2 (51.4 50.7	50.6 51.2	76.5 76.9	116.4 114.9	34. 34.
 <th>SCORER RR+CBR SCORER+CBR</th> <th>53.3 54.3 54.1 54.4</th> <th>37.1 37.5 37.4 37.6</th> <th>71.5 71.5 71.7</th> <th>122. 122. 122.</th> <th>0 3 4 3 4 3</th> <th>1.2 6 1.2 6 1.6 6</th> <th>51.4 50.7 52.0</th> <th>50.6 51.2 51.7</th> <th>76.5 76.9 77.4</th> <th>116.4 114.9 117.9</th> <th>34. 34. 35.</th>	SCORER RR+CBR SCORER+CBR	53.3 54.3 54.1 54.4	37.1 37.5 37.4 37.6	71.5 71.5 71.7	122. 122. 122.	0 3 4 3 4 3	1.2 6 1.2 6 1.6 6	51.4 50.7 5 2.0	50.6 51.2 51.7	76.5 76.9 77.4	116.4 114.9 117.9	34. 34. 35.



PARIS

