CoDeRNet: Selective Cross-Task Routing under Heterogeneous Supervision for Change Detection and Captioning

Eunki Cho, Hyeon Bae Kim, and Seong Tae Kim



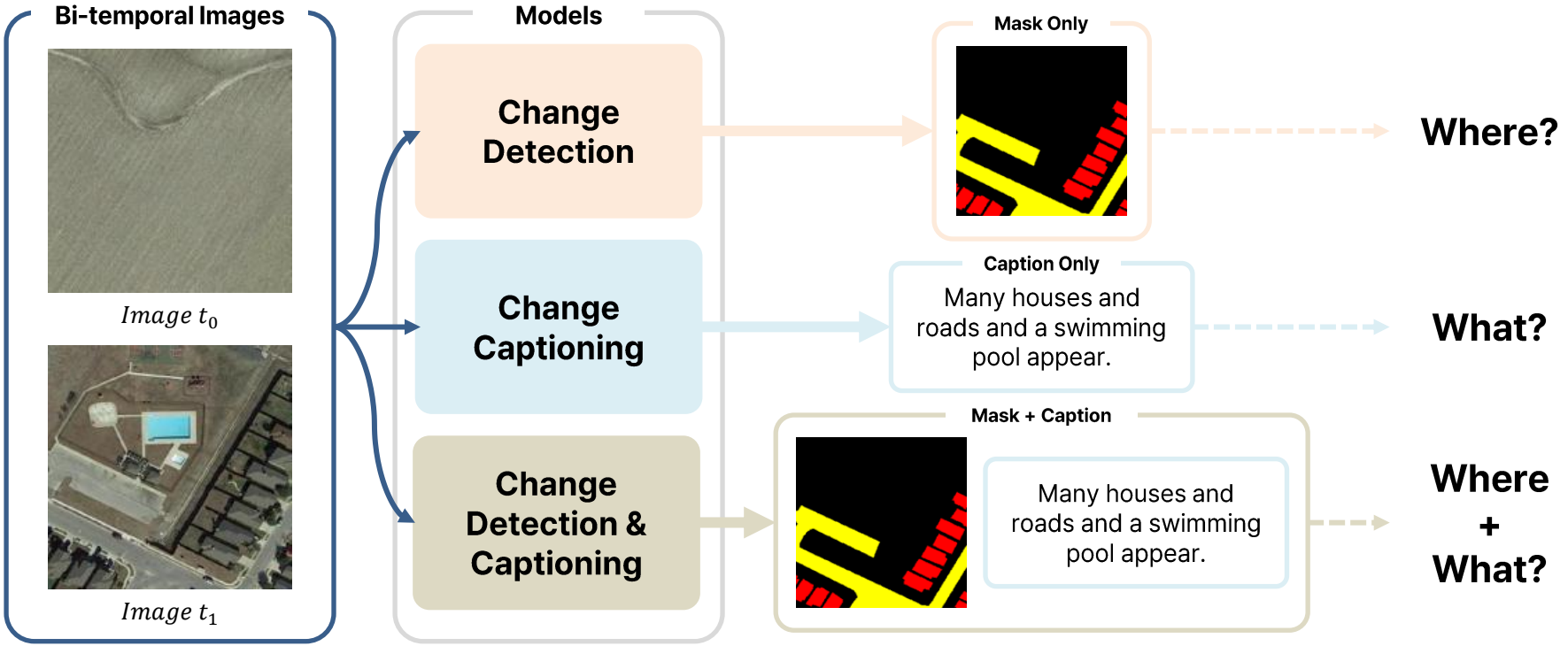


Introduction

AMI LAB

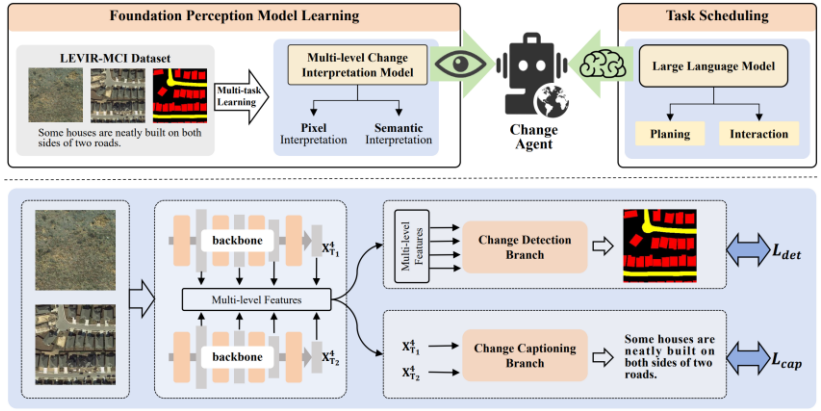
KYUNG HEE UNIVERSITY

➢ What is Change Detection & Captioning?

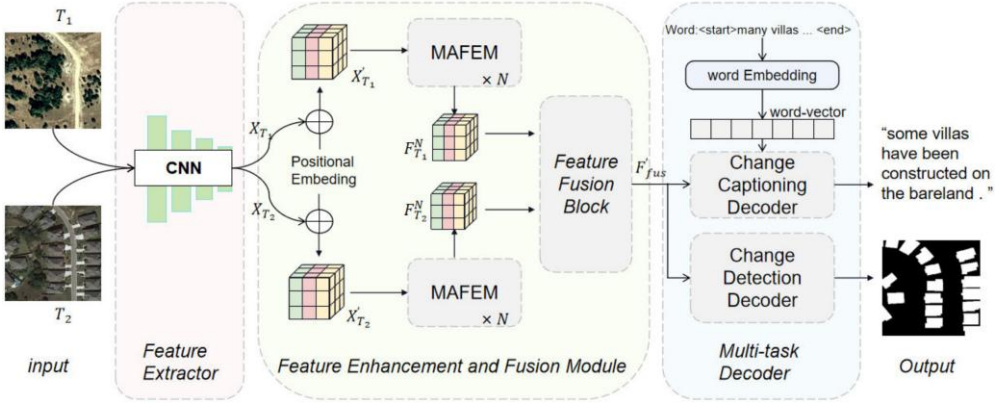


Introduction

➢ Beyond Joint Learning: Cross-Task Routing



< Figure reproduced from [1] >

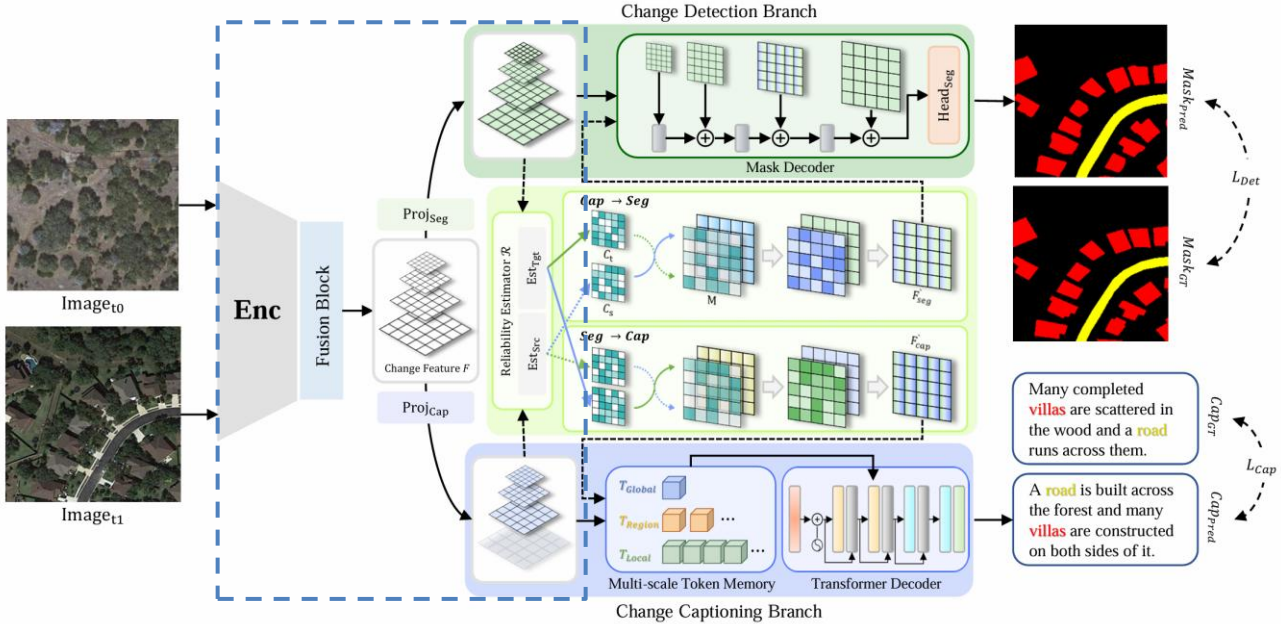


< Figure reproduced from [2] >

[1] Liu, Chenyang, et al. "Change-agent: Toward interactive comprehensive remote sensing change interpretation and analysis." IEEE Transactions on Geoscience and Remote Sensing 62 (2024): 1-16.
[2] Shi, Jingye, et al. "A multitask network and two large-scale datasets for change detection and captioning in remote sensing images." IEEE Transactions on Geoscience and Remote Sensing 62 (2024): 1-17.

Methods

➢ CoDeRNet: Change Feature Representation



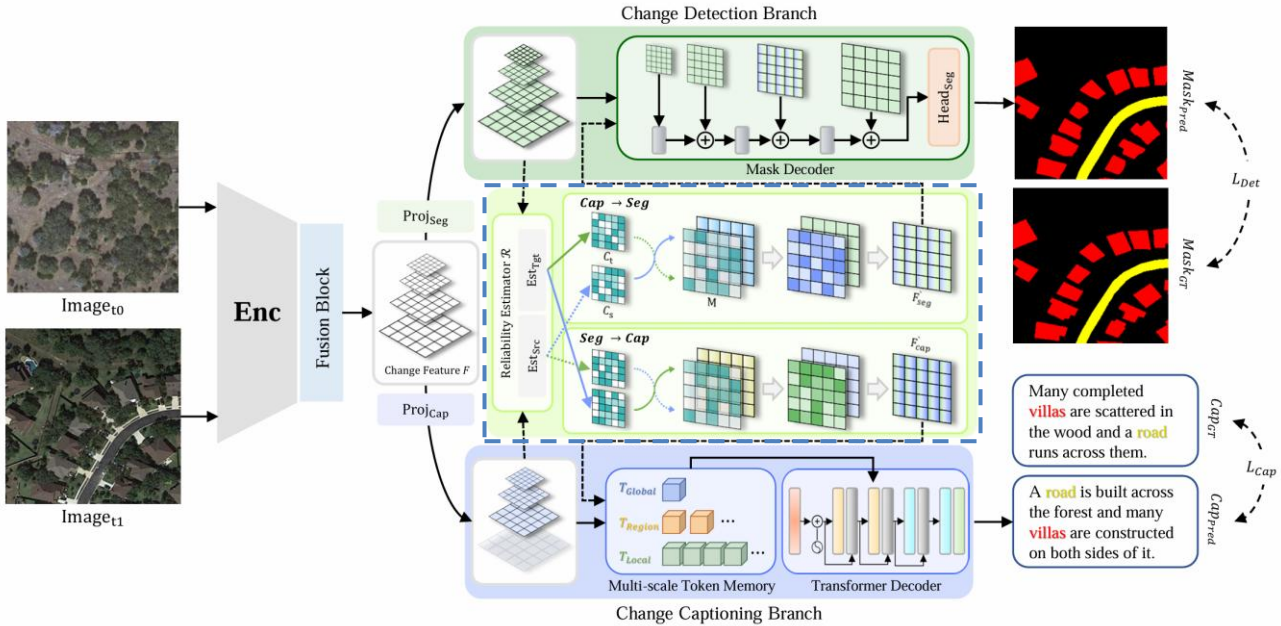
- **Shared Change Representation**
  Construct a shared change representation by combining bi-temporal features and their differences.
- $F^{(s)} = Conv_{1\times1}\left(\left[F_{t0}^{(s)}, F_{t1}^{(s)}, F_{|\Delta|}^{(s)}, F_{\Delta}^{(s)}\right]\right)$

- **Branch-Specific Feature Interface**
  Project shared change features into task-specific spaces for detection and captioning before routing.

✓ $F_{seg}^{(s)} = \phi_{seg}^{(s)}(F^{(s)}),\ F_{cap}^{(s)} = \phi_{cap}^{(s)}(F^{(s)})$

Methods

➢ CoDeRNet: Confidence-Decomposed Routing (CoDeR)



- **Routing Cue Estimation**

  Decompose cross-task transfer into target-side need and source-side reliability.

✓ $C_t = R_t(F_t),\ C_s = R_s(F_s)$

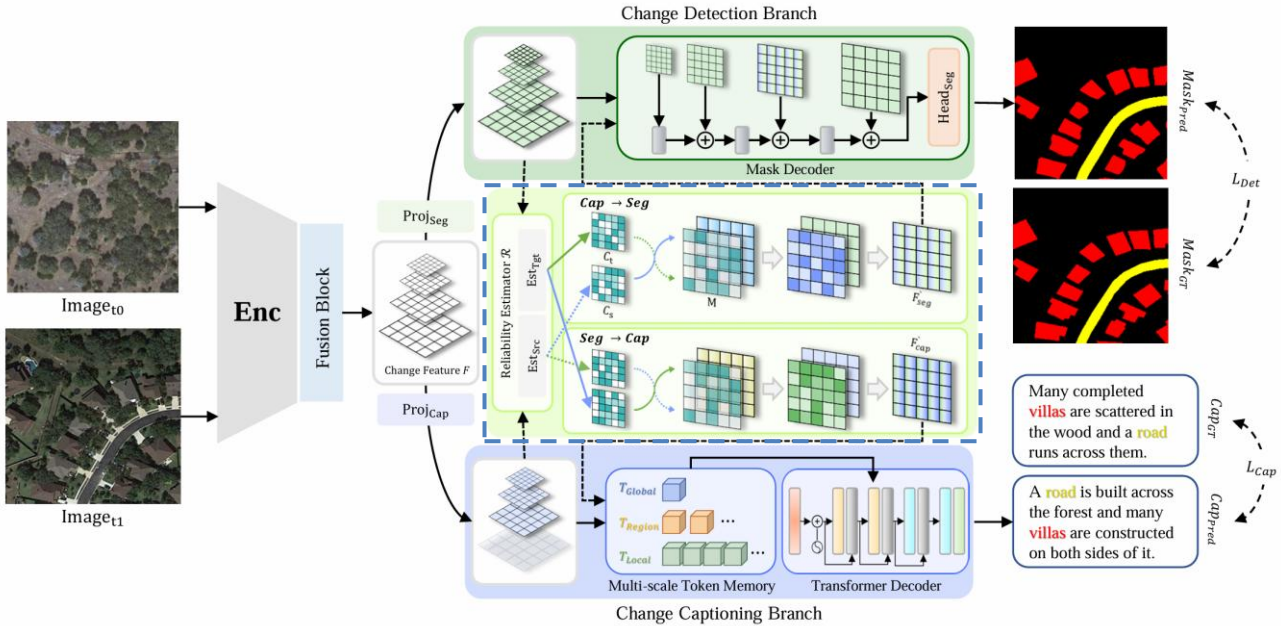
- **Routing Mask**

  Selectively modulate cross-task information transfer based on the learned routing cues.

- $M = (1 - C_t) \cdot C_S$

Methods

➢ CoDeRNet: Confidence-Decomposed Routing (CoDeR)



- **Selective Feature Transfer**

  Refine the target feature with task-beneficial source evidence through residual injection.

✓ $\hat{F_t} = F_t + M \odot g(F_s)$

- **Bidirectional Routing**

  Exchange complementary spatial and semantic cues between detection and captioning.

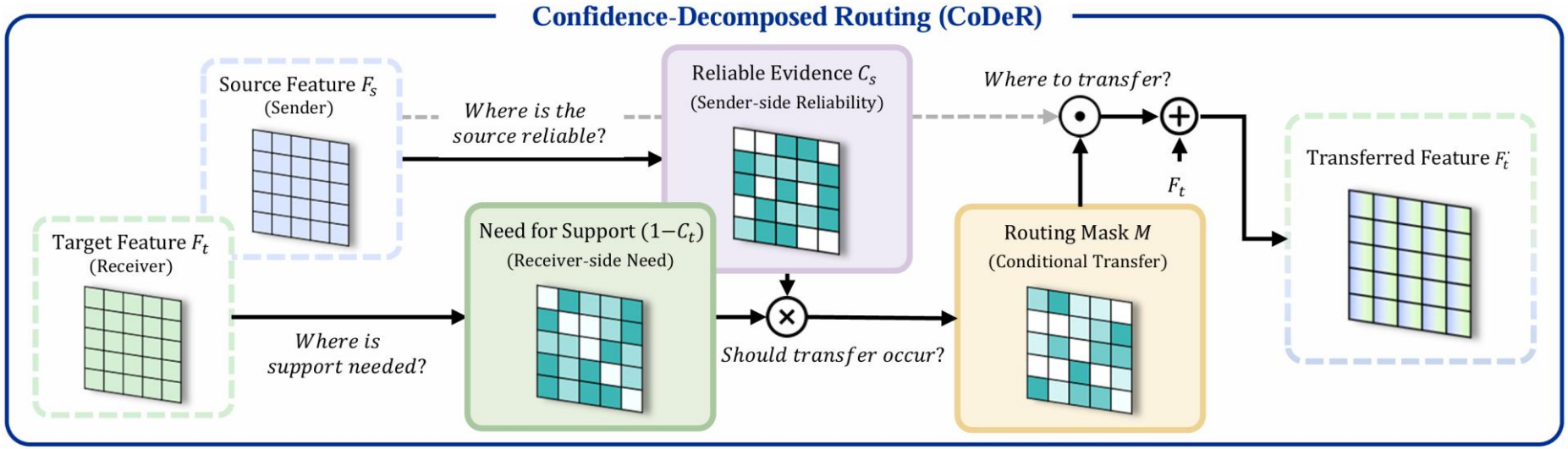
✓ $\hat{F_{seg}} = CoDeR_{cap \to seg}, \hat{F_{cap}} = CoDeR_{seg \to cap}$

Methods

AMI LAB

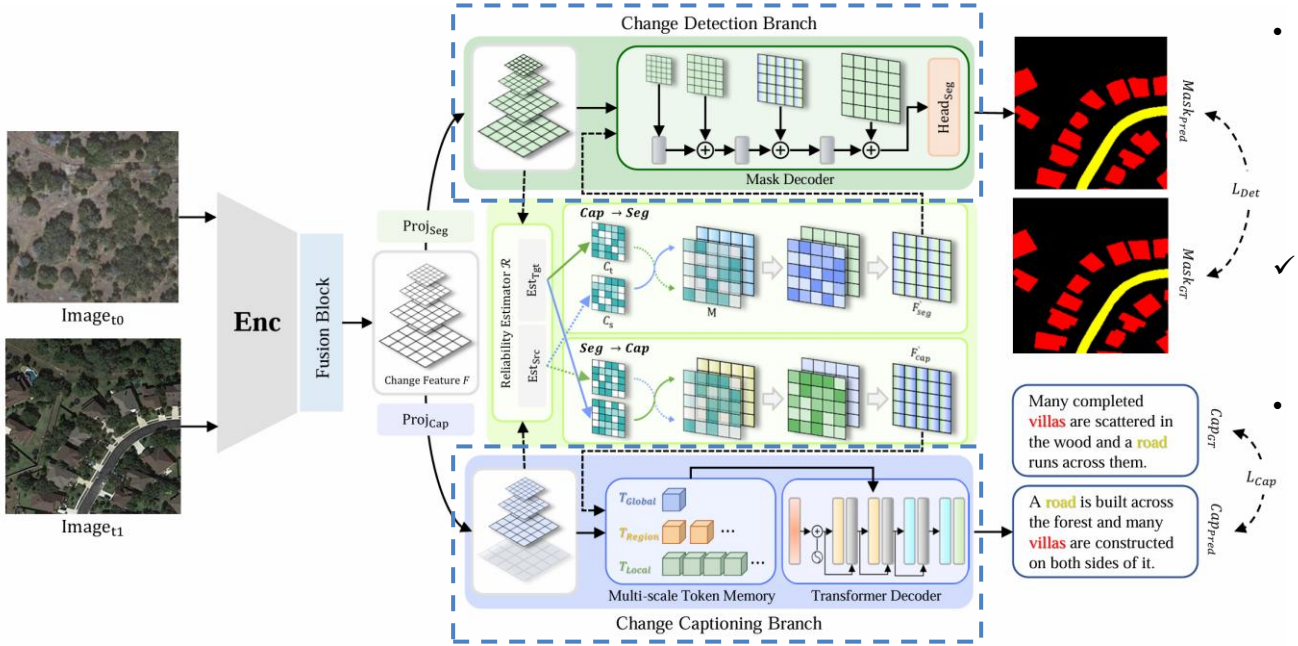
KYUNG HEE UNIVERSITY

➢ CoDeRNet: Confidence-Decomposed Routing (CoDeR)



Methods

➢ CoDeRNet: Task-Specific Decoding for Detection and Captioning



- **Segmentation Decoder**

  Aggregate multi-scale features to recover spatial details for pixel-wise change prediction.

✓ $\widehat{M} = D_{seg}\left(\left\{\widehat{F}_{seg}^{(s)}\right\}_{s\in S}\right)$

- **Captioning Decoder**

  Construct multi-scale token memory from global, regional, and local change features, and generate captions autoregressively.

$\mathcal{M} = [t_{glob}; T_{seg}; T_{loc}],\ \hat{p}_{t+1} = D_{cap}(y^*_{1:t}, \mathcal{M})$

Experiments

AMI LAB

KYUNG HEE UNIVERSITY

➢ Quantitative Results: LEVIR-MCI, WHU-CDC

- Bold: best, Underline: second-best.

Task	Method	Metrics							
		mIoU	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L	CIDEr-D
Change Detection	SNUNet [10]	82.53	-	-	-	-	-	-	-
	ChangeFormer [1]	84.35	-	-	-	-	-	-	-
	TFI-GR [17]	79.34	-	-	-	-	-	-	-
	A2Net [18]	80.82	-	-	-	-	-	-	-
	HANet [12]	71.23	-	-	-	-	-	-	-
	ChangeMamba [6]	87.61	-	-	-	-	-	-	-
Change Captioning	DUDA [31]	-	81.44	72.22	64.24	57.79	37.15	71.04	124.32
	MCCFormers [32]	-	82.16	72.95	65.42	59.41	38.26	72.10	128.34
	RSICCformer [21]	-	84.11	75.40	68.01	61.93	38.79	73.02	131.40
	Prompt-CC [23]	-	83.66	75.73	69.10	63.54	38.82	73.72	136.44
	Chg2Cap [3]	-	85.14	76.91	69.86	64.09	39.83	74.62	135.41
	Sparse Focus [35]	-	84.56	75.87	68.64	62.87	39.93	74.69	137.05
	SEN [48]	-	85.10	77.05	70.01	64.09	39.59	74.57	136.02
	MaskApproxNet [36]	-	85.90	77.12	70.72	64.32	39.91	75.67	137.71
Change Detection & Captioning	CTMTNet [34]	85.37	85.16	77.24	70.13	64.16	39.34	74.03	134.50
	MCINet [24]	86.43	86.68	78.75	71.74	65.95	40.80	75.96	140.29
	CoDeRNet (Ours)	**87.79**	**87.25**	**79.42**	**72.26**	**66.27**	**41.08**	**76.36**	**141.79**

< Comparison on LEVIR-MCI >

Task	Method	Metrics							
		IoU	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L	CIDEr-D
Change Detection	SNUNet [10]	74.81	-	-	-	-	-	-	-
	ChangeFormer [1]	77.88	-	-	-	-	-	-	-
	TFI-GR [17]	77.60	-	-	-	-	-	-	-
	A2Net [18]	80.43	-	-	-	-	-	-	-
	HANet [12]	79.95	-	-	-	-	-	-	-
	ChangeMamba [6]	89.69	-	-	-	-	-	-	-
Change Captioning	DUDA [31]	-	79.04	69.53	61.57	55.64	34.29	68.98	121.85
	MCCFormers [32]	-	81.12	75.04	69.95	65.34	42.11	78.52	147.09
	RSICCformer [21]	-	80.05	74.24	69.61	66.54	42.65	73.91	133.44
	Prompt-CC [23]	-	81.12	73.96	67.22	61.45	36.99	71.88	134.50
	Chg2Cap [3]	-	77.27	71.33	65.10	58.44	39.95	78.67	160.64
	SparseFocus [35]	-	81.17	72.90	66.06	60.27	37.34	72.63	134.64
	SEN [48]	-	80.60	74.64	67.69	61.97	36.76	71.70	133.57
	MaskApproxNet [36]	-	81.34	75.68	71.16	67.73	43.89	75.41	135.31
Change Detection & Captioning	CTMTNet [34]	81.54	83.56	77.66	72.76	69.00	45.39	79.23	149.40
	MCINet [24]	88.23	**87.82**	**82.84**	**78.66**	**75.49**	49.03	82.54	159.57
	CoDeRNet (Ours)	**90.25**	87.44	82.38	77.85	74.17	**49.04**	**83.58**	**163.37**

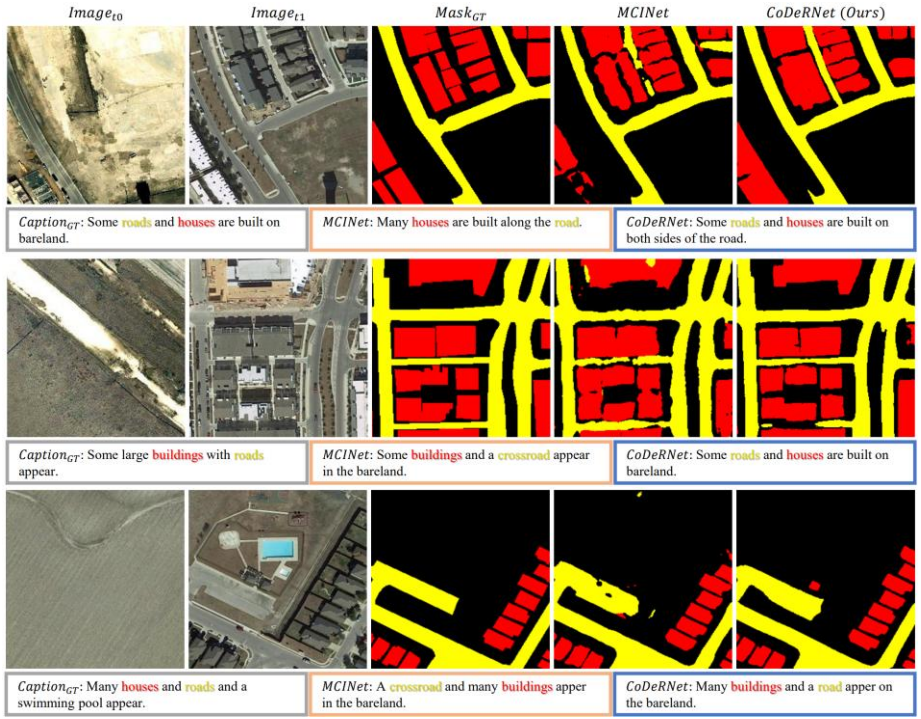
< Comparison on WHU-CDC >

Experiments

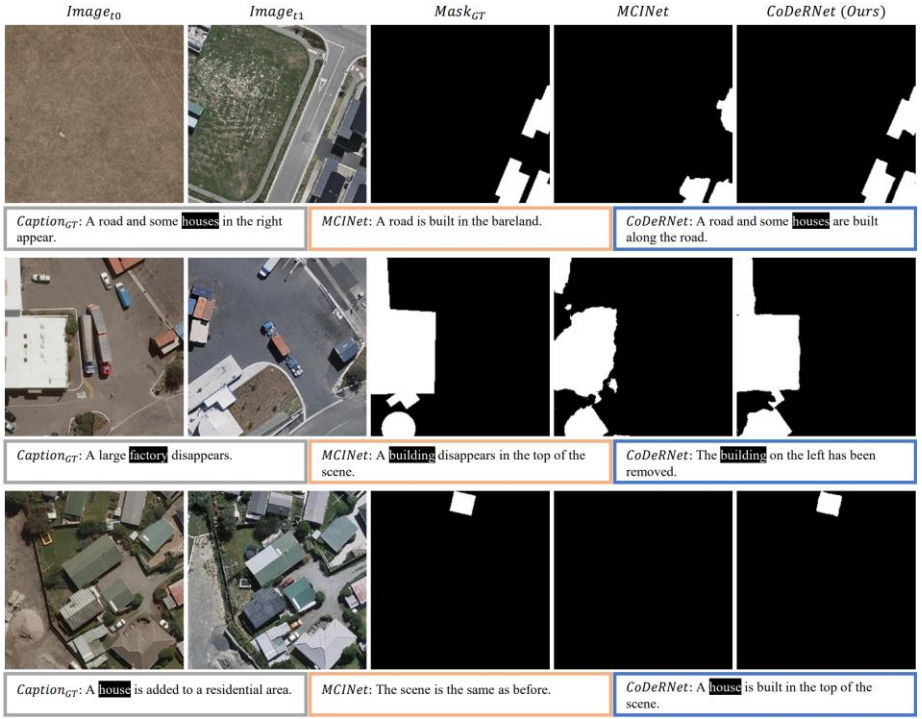
AMI LAB

KYUNG HEE UNIVERSITY

➢ Qualitative Results : LEVIR-MCI, WHU-CDC



< Comparison on LEVIR-MCI >



< Comparison on WHU-CDC >

Conclusion

- Limitations
  - ✓ Although CoDeRNet improves both tasks within a unified CDC framework, the gains are asymmetric, with captioning benefiting more strongly from detection cues than the reverse direction.
  - ✓ In addition, the routing cues are not explicitly supervised by region-language correspondence, which may limit semantic faithfulness in subtle or ambiguous change scenarios.

- Conclusion
  - ✓ We revisit change detection and captioning as a conditional information routing problem under heterogeneous supervision, rather than as simple joint learning.
  - ✓ CoDeRNet selectively exchanges complementary information by modeling receiver-side need and sender-side reliability while preserving task-specific representations.
  - ✓ Experiments show improved semantic description quality with strong localization accuracy, highlighting the value of selective routing for CDC and broader multi-task learning settings.