An Adaptive Multi-scale and Multi-level Features Fusion Network with Perceptual Loss for Change Detection

Jialang Xu, Yang Luo, Xinyue Chen, Chunbo Luo

School of Information and Communication Engineering,
University of Electronic Science and Technology of China

Supported by NSFC 61871096

IEEE International Conference on Acoustics, Speech and Signal Processing (2021)

Remote Sensing Change Detection

- **Change detection** is the task of identifying significant changes between multiple images taken at different periods of the same geographical area.
- It could be used in global resource monitoring, natural disaster assessments, urban settlements, and other remote sensing applications.
- We focus on the pixel-level **binary change detection task** of bi-temporal **very-high-resolution** images.

Pixel-level Binary Change Detection

T1 image

T2 image

Change map

Methods For Change Detection

- Early stages:

  Medium- and low-resolution images
  - Algebra-based methods
  - Transform-based methods
  - Post-classification methods

  → Rough change maps

- Nowadays:

  Very-high-resolution images → Fine image details and complex texture features

  State-of-the-art **deep learning-based methods** have achieved superior performances than others on very-high-resolution images.

Challenges and Motivation

State-of-the-art deep learning-based methods are limited by the following **3 constraints:**

- **Weak capability of feature extraction**:
  - VHR images have abundant noises

    → More powerful feature extraction backbone
- **Limited effect of feature fusion**:
  - Existed semantic gaps and irrelevant features

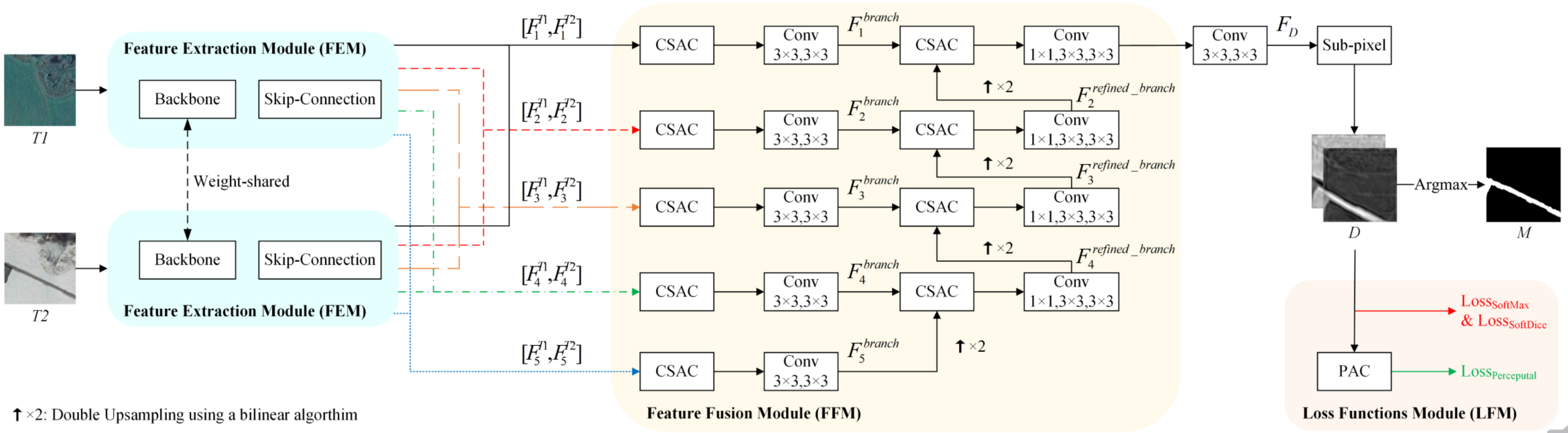
    → Introducing a channel and spatial attention mechanism
- **Defective loss function**:
  - Per-pixel loss has harsh optimization objectives and only considers the pixel-level local information → Models hard to converge well and the quality of change maps are poor

    → Introducing perceptual loss

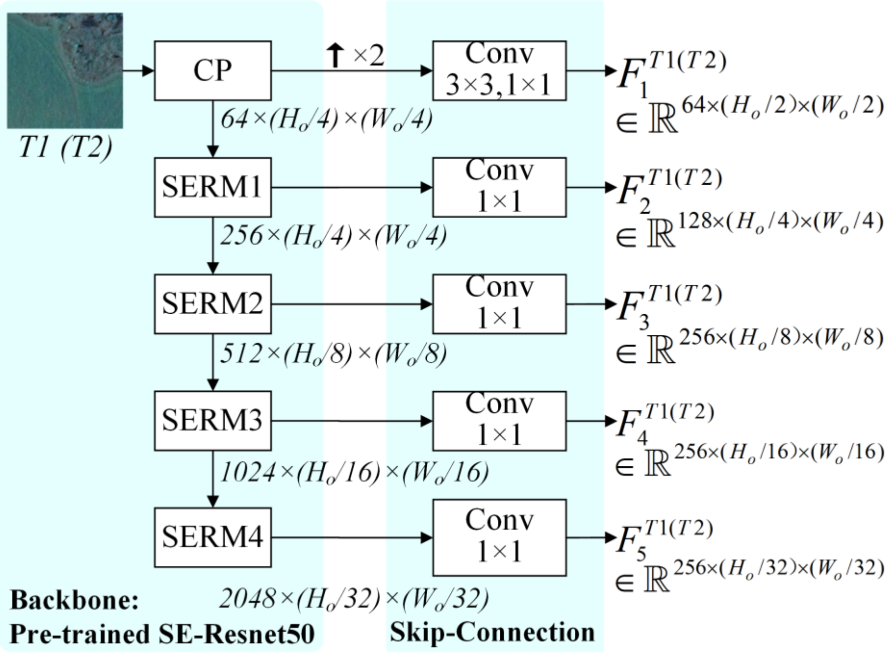
Adaptive Multi-scale and Multi-level Features Fusion Network

3 functional parts:

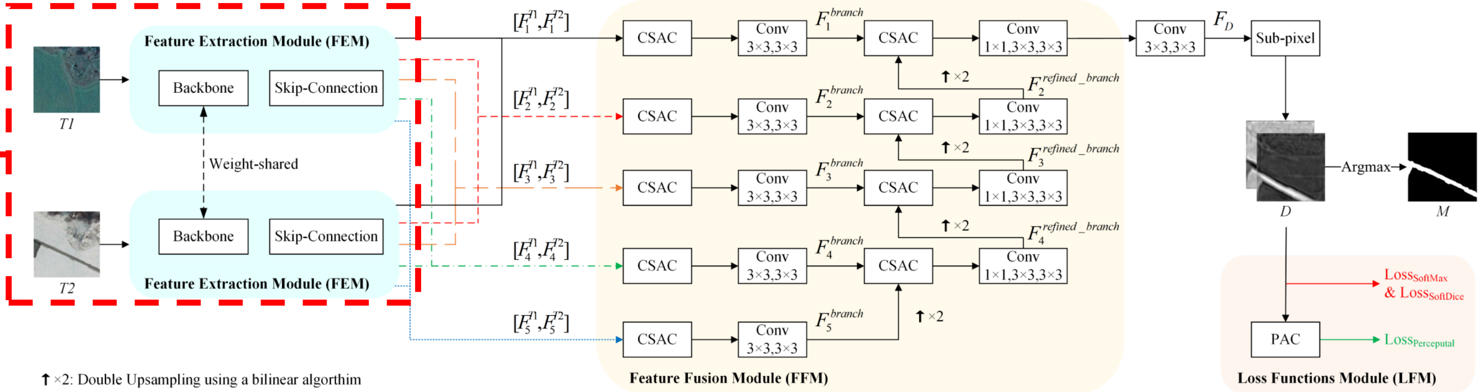
- Part-A. Feature extraction module (FEM)
- Part-B. Feature fusion module (FFM)
- Part-C. Loss function module (LFM)



Part-A. Feature Extraction Module

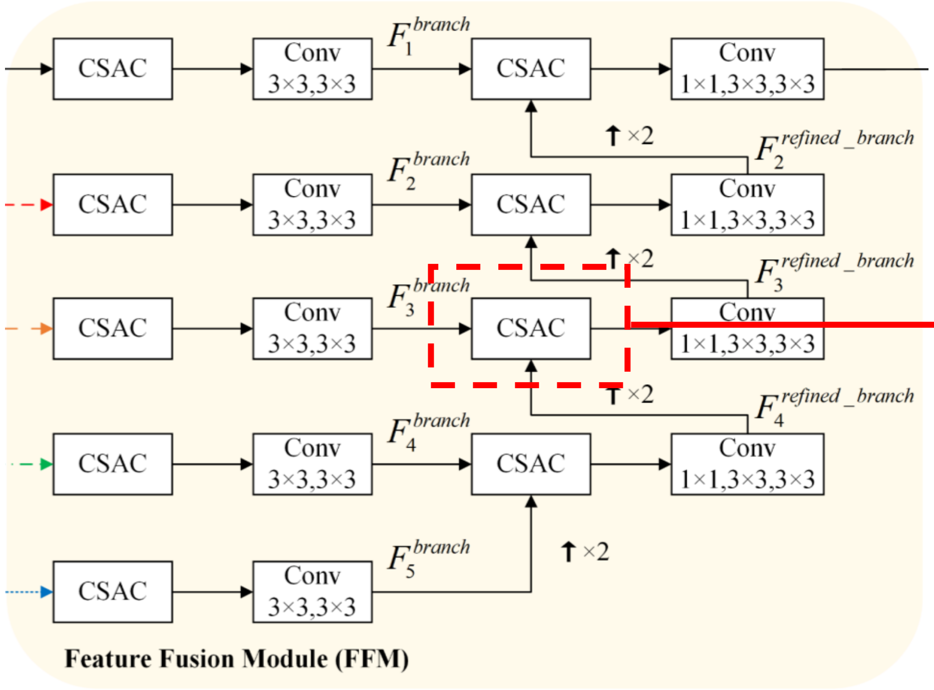


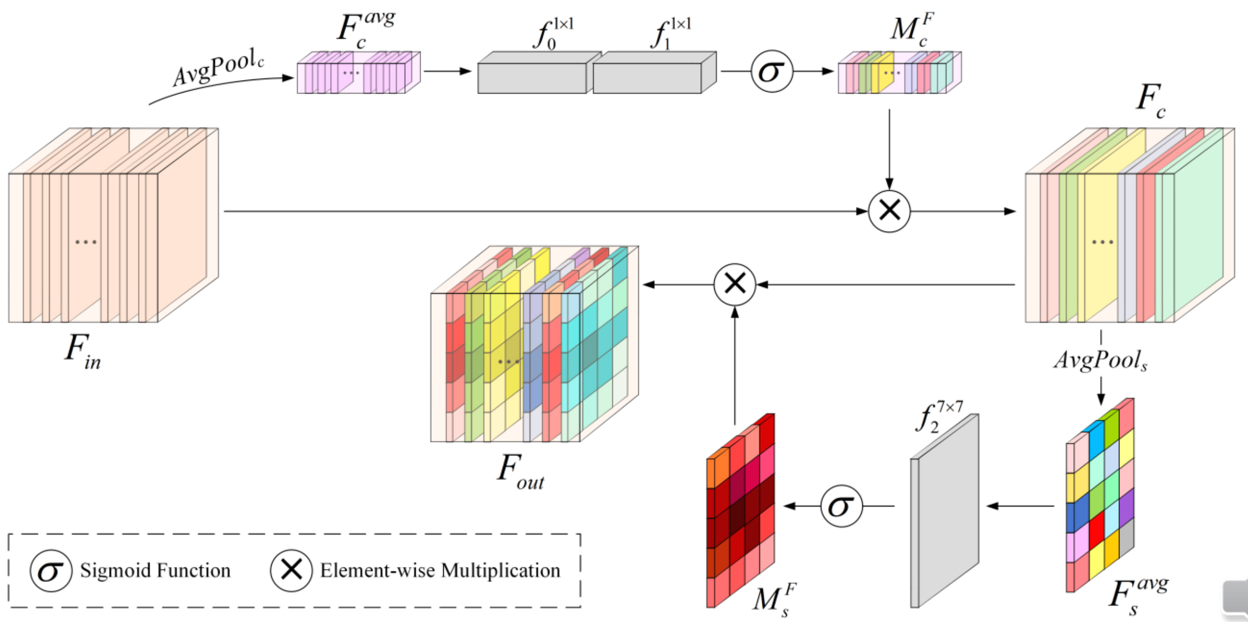
- **Backbone:** SE-ResNet50
  - Accelerate training process and enhance feature representativeness
- **Skip-Connection:** 1×1 convolutional layers
  - Reduce model complexity



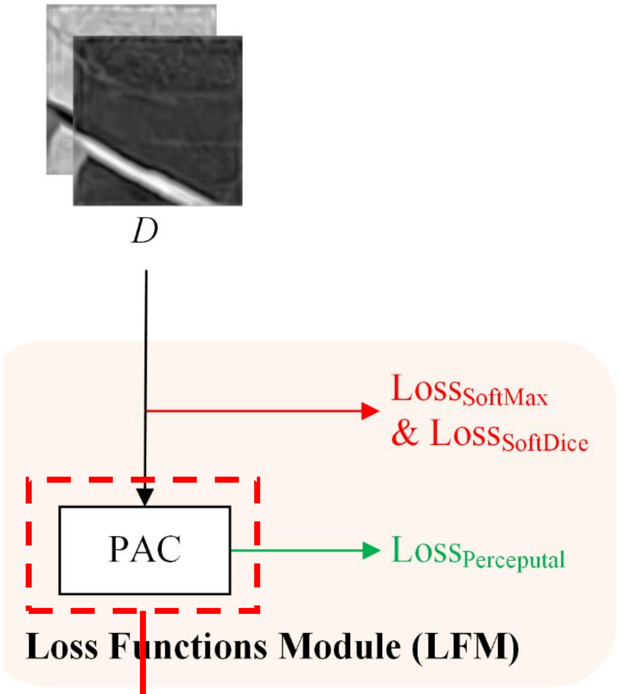
Part-B. Feature Fusion Module

- Incorporate high-level features with low-level features
  - Recover fine-grained details and better mask of change objects
- **Channel and spatial attention component** (CSAC) for each fusion node
  - Highlight salient features and fuse different feature in an adaptive weighted fusion manner



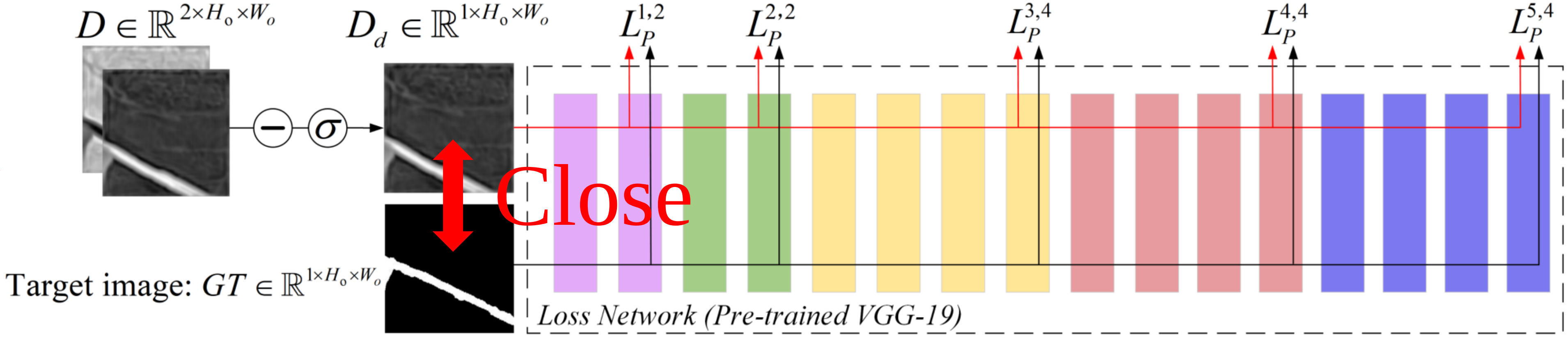


Part-C. Loss Function Module (LFM)



- SoftMax per-pixel cross-entropy loss
- Dice loss
- Perceptual loss: calculated by **Perceptual Auxiliary Component (PAC)**
  - Capture the global perceptual difference and structural information
  - Encourage results to be perceptually similar to the ground truth
  - $L_P\left(D_d, GT\right) = L_P^{1,2} + L_P^{2,2} + L_P^{3,4} + L_P^{4,4} + L_P^{5,4}$,

    where $L_P^{i,j} = \frac{1}{C^{i,j}H^{i,j}W^{i,j}} \left\| \phi_P^{i,j}\left(D_d\right) - \phi_P^{i,j}\left(GT\right) \right\|_2^2$

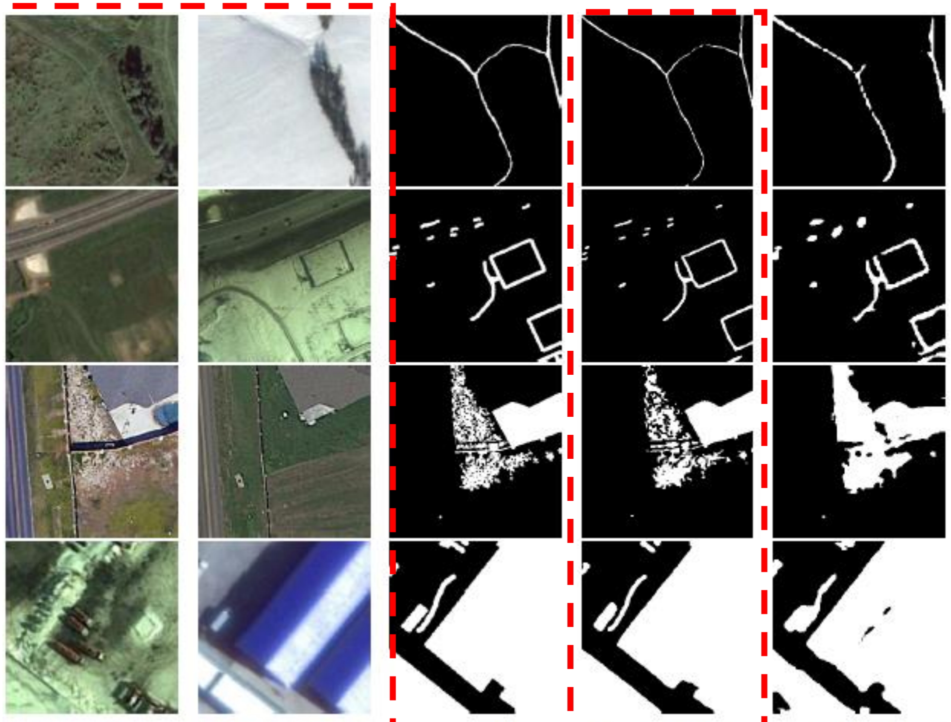


Experiments and Results

Comparison experiment

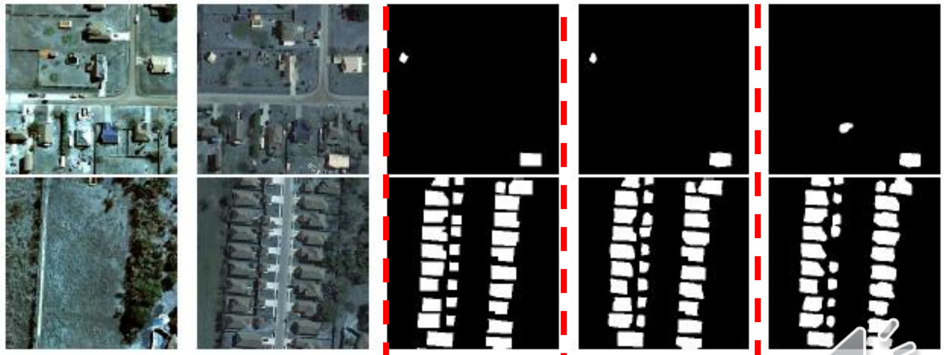
- **Quantitative Analysis**: State-of-the-art performance on two datasets
  - For the Season-varying dataset:
    - ↑ 2.61%-13.16% in Precision
    - ↑ 2.86%-15.97% in Recall
    - ↑ 5.68%- 14.55% in F1
    - ↑ 1.41%-3.57% in overall accuracy
  - For the LEVIR-CD dataset:
    - ↑ 0.9%-9.58% in F1
    - ↑ 0.09%-0.77% in overall accuracy
- **Qualitative Analysis:**
  - Fort the Season-varying dataset:
    - ✓ Small and thin area changes: more tiny changes.
    - ✓ Complex and large area changes: finer details and clearer boundaries
  - For the LEVIR-CD dataset:
    - ✓ Scattered building changes: right spatial location
    - ✓ Dense building changes: better conformed with geometric edges.

Dataset	Season-varying dataset				LEVIR-CD dataset			
Metrics	P(%)	R(%)	F1(%)	OA(%)	P(%)	R(%)	F1(%)	OA(%)
FC-Siam-conc [19]	84.41*	82.50*	82.44*	95.72*	**93.96**	71.87	81.44	98.33
FC-Siam-diff [19]	85.78*	83.64*	84.70*	95.75*	92.52	76.55	83.78	98.49
FCN-PP [18]	89.97	80.45	84.95	96.61	89.64	88.56	89.10	98.90
UNet++_MSOF [17]	89.54•	87.11•	87.56•	96.73•	92.07	86.01	88.94	98.91
IFN [9]	94.96•	86.08•	90.30•	97.71•	92.18	88.15	90.12	99.01
STANet [13]	89.17	93.56	91.31	97.88	83.80•	**91.00•**	87.30•	-
The proposed AFFN	**97.57**	**96.42**	**96.99**	**99.29**	92.59	89.51	**91.02**	**99.10**



(a) T1 (b) T2 (c) GT (d) AFFN (e) STANet

Season-varying dataset



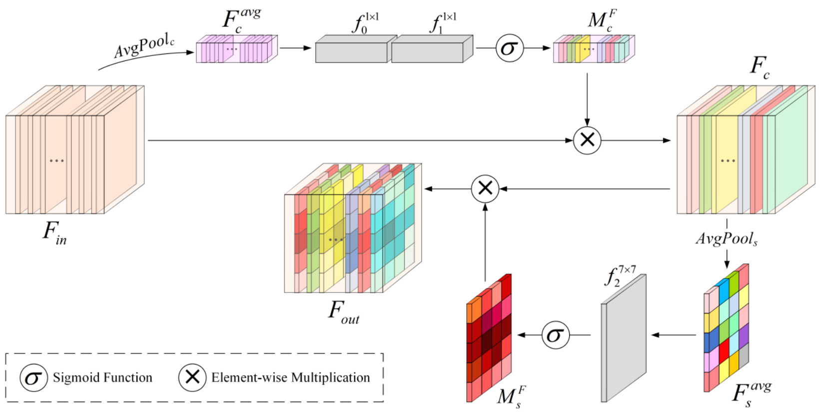
(a) T1 (b) T2 (c) GT (d) AFFN (e) IFN

LEVIR-CD dataset

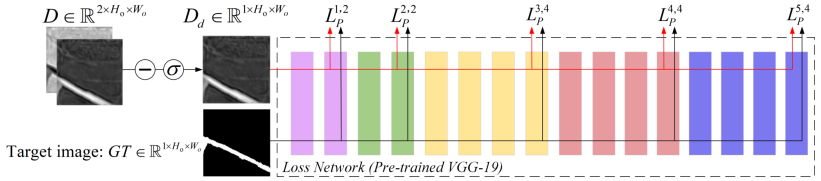
Experiments and Results

Ablation study: from the baseline framework (i.e. AFFN without CSAC and PAC) to add each key component (i.e. CSAC and PAC)

- **Effect of CSAC:** AFFN without PAC
  - +2.03% in R, +1.04% in F1, +0.23% in OA compared with the baseline
- **Effect of PAC:** AFFN without CSAC
  - +0.13% in P, +2.86% in R, +1.53% in F1, +0.35% in OA compared with the baseline
  - Using per-pixel loss auxiliary with perceptual loss has a clear advantage in change detection



CSAC



PSM

Framework	Key components		Evaluation metrics			
	CSAC	PAC	P(%)	R(%)	F1(%)	OA(%)
Baseline			96.75	92.89	94.78	98.78
AFFN without PAC	√		96.74	94.92	95.82	99.01
AFFN without CSAC		√	96.88	95.75	96.31	99.13
The proposed AFFN	√	√	**97.57**	**96.42**	**96.99**	**99.29**

Conclusion

- Key limitations of deep learning-based binary change detection methods lie in:
  - Feature extraction
  - Feature fusion
  - Loss function
- Powerful feature extraction backbone is helpful for change detection.
- Attention mechanism could be utilized in change detection to improve the effect of feature fusion.
- Perceptual loss works well in change detection tasks.

Thank You !