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

Jialang Xu, Yang Luo, Xinyue Chen, Chunbo Luo

xujialang@std.uestc.edu.cn {luoyang, c.luo}@uestc.edu.cn chenxinyue@stu.scu.edu.cn

2021 TORONTO Canada
June 6-11, 2021
Metro Toronto Convention Centre

Poster Number: 3406

1. Background

Change detection (CD) plays a vital role in monitoring and analyzing temporal changes in Earth observation, aiming to identify significant changes between multiple images taken at different periods of the same geographical area.

The increasing availability of **very-high-resolution (VHR) remote sensing images** brings promising opportunities for monitoring finer changes. The fine image details and complex texture features conveyed in VHR images introduce new challenges for CD, which has led to the rising of **deep learning-based CD methods.**

2. Challenges

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

- **Weak capability of feature extraction**: In real-world remote sensing tasks, images have abundant independent noises. This fact indicates that a feature extraction module for the CD task should have the ability to abstract high representative features from those images of strong noise and changes.
- **Limited effect of feature fusion**: How to concatenate pre-change features F1 and post-change features F2 obtained from bi-temporal images poses a feature fusion problem. Simple fusion methods cannot efficiently deal with irrelevant features and the heterogeneity problem caused by different informative features.
- **Defective loss function**: Per-pixel loss, widely used in deep learning-based CD methods, having harsh optimization objectives and only consider the pixel-level local information.

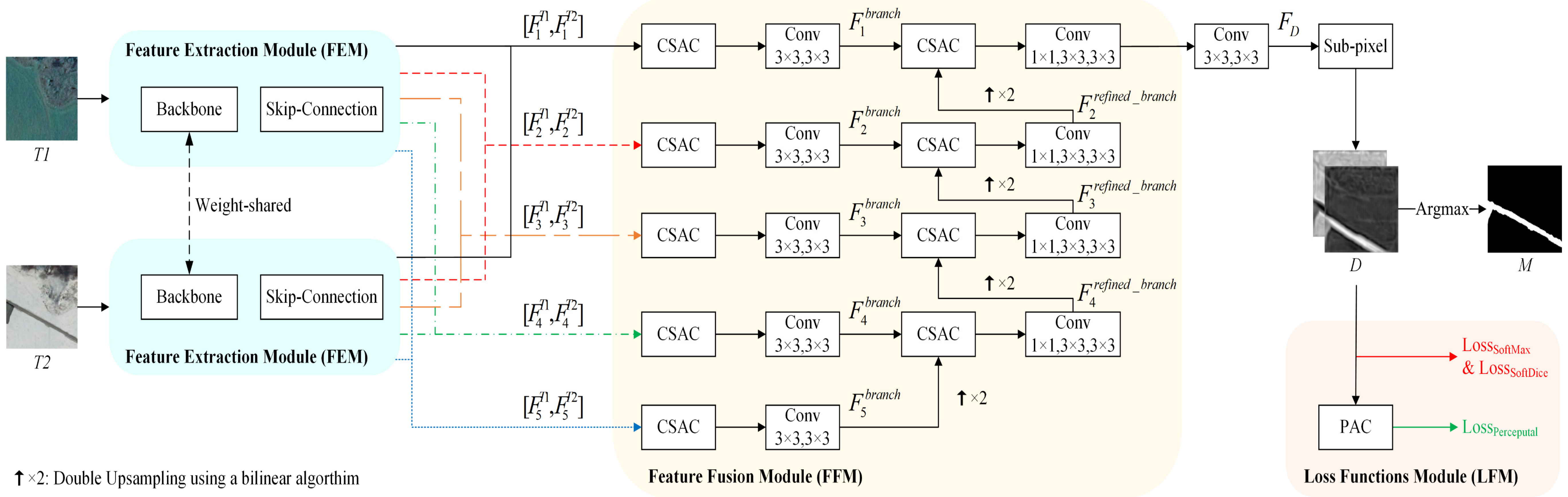
3. Motivation

To address those challenges, we propose an **adaptive multi-scale and multi-level features fusion network**.

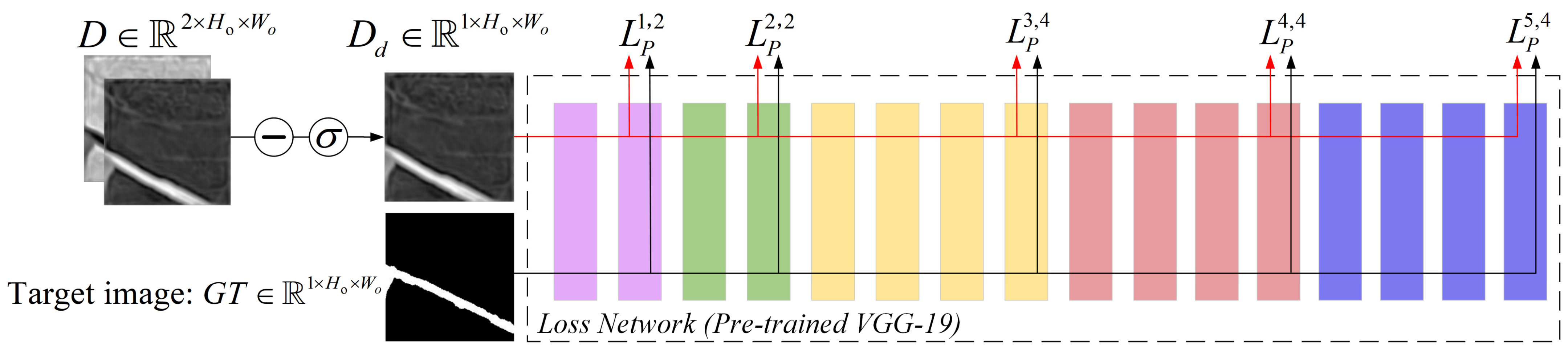
- **For the feature extraction**: We propose a feature extraction module consisting of a deeper pre-trained network SE-ResNet50 and a skip-connection process to acquire highly representative multi-scale and multi-level features efficiently.
- **For the feature fusion**: we develop an effective feature fusion module enhanced by the channel and spatial attention component, to extract distinct features and regions that provide accurate representations of changes and overcome the heterogeneity problem between different features.
- **For the loss function**: we design a loss function module with a perceptual auxiliary component for the CD model to optimize the prediction results in both pixel space and feature space, which makes the model not only pay attention to the relationship between local pixels but also capture the global perceptual difference and structural information. This module combines the benefits of per-pixel loss and perceptual loss, thus simplifying the optimization objectives and leading to promising performance.

4. Proposed Method

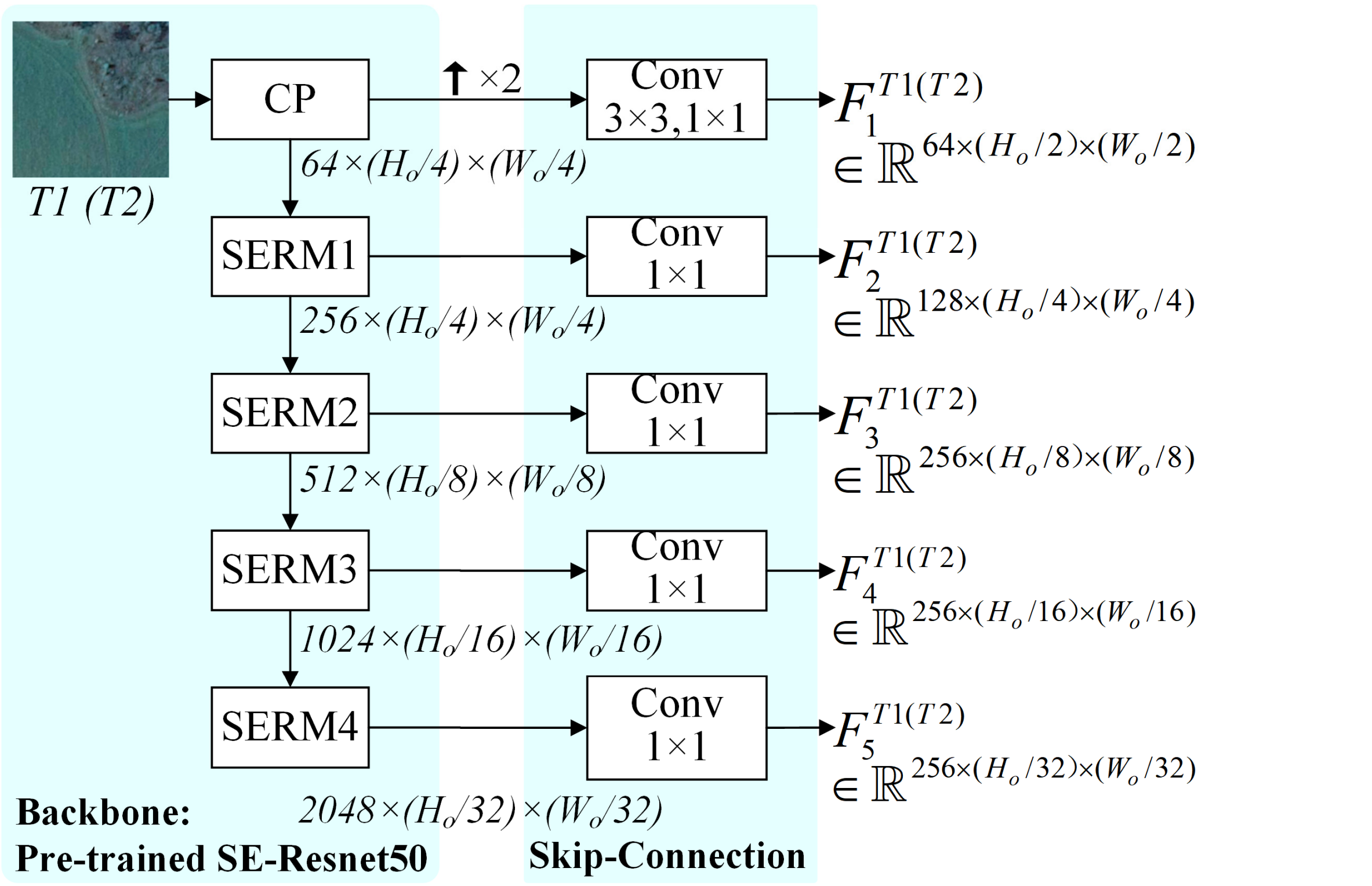
➢ Adaptive Multi-scale and Multi-level Features Fusion Network (AFFN)



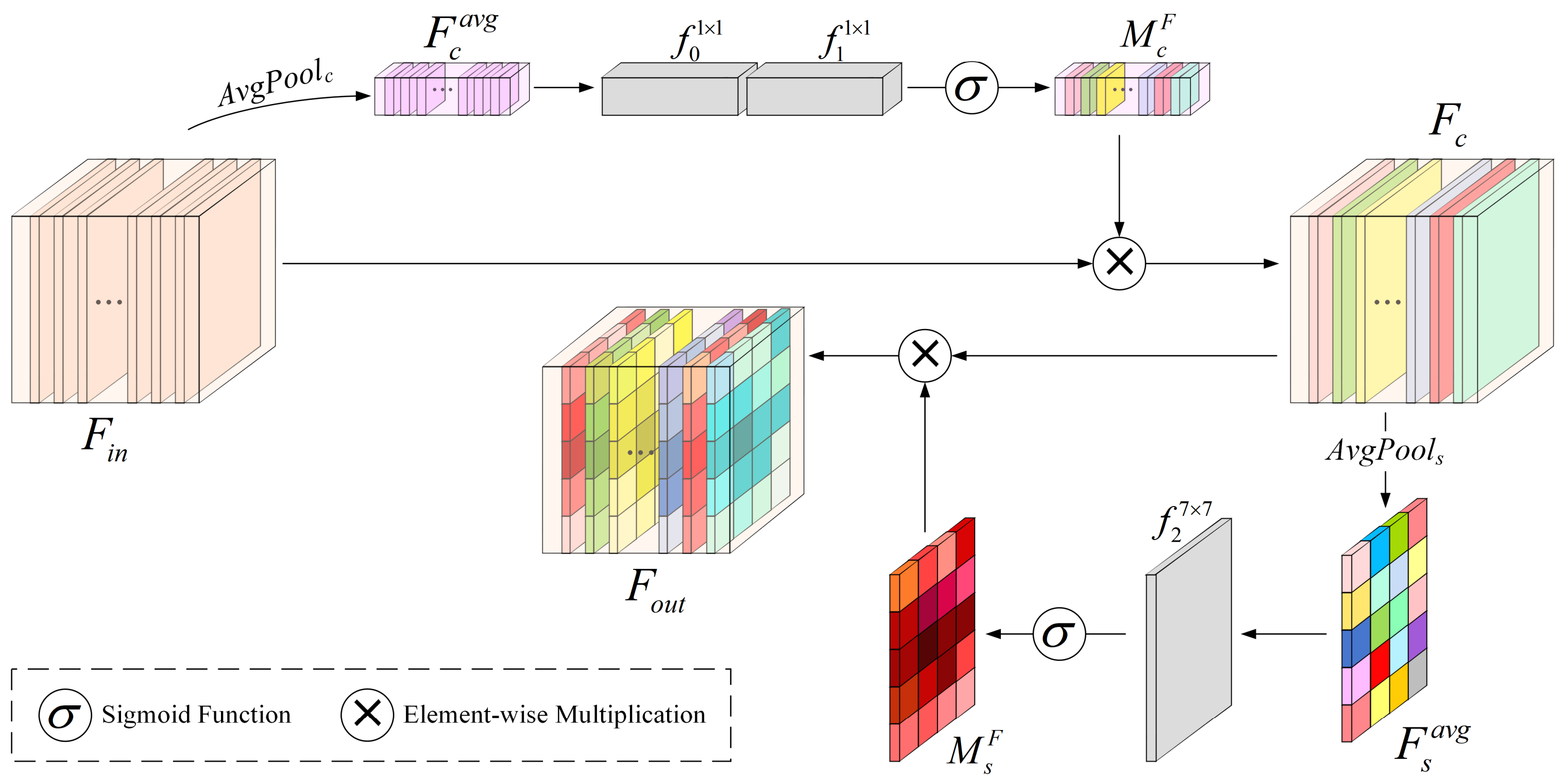
➢ Perceptual Auxiliary Component (PAC)


$D \in \mathbb{R}^{2\times H_o \times W_o}$
$D_d \in \mathbb{R}^{1\times H_o \times W_o}$
$L_P^{1,2}$
$L_P^{2,2}$
$L_P^{3,4}$
$L_P^{4,4}$
$L_P^{5,4}$
Target image: $GT \in \mathbb{R}^{1\times H_o \times W_o}$
Loss Network (Pre-trained VGG-19)

➢ Feature Extraction Module (FEM)


T1 (T2)
CP
↑×2
Conv 3×3,1×1
$64\times(H_o/4)\times(W_o/4)$
SERM1
Conv 1×1
$256\times(H_o/4)\times(W_o/4)$
SERM2
$512\times(H_o/8)\times(W_o/8)$
SERM3
$1024\times(H_o/16)\times(W_o/16)$
SERM4
$2048\times(H_o/32)\times(W_o/32)$
$F_1^{T1(T2)} \in \mathbb{R}^{64\times(H_o/2)\times(W_o/2)}$
$F_2^{T1(T2)} \in \mathbb{R}^{128\times(H_o/4)\times(W_o/4)}$
$F_3^{T1(T2)} \in \mathbb{R}^{256\times(H_o/8)\times(W_o/8)}$
$F_4^{T1(T2)} \in \mathbb{R}^{256\times(H_o/16)\times(W_o/16)}$
$F_5^{T1(T2)} \in \mathbb{R}^{256\times(H_o/32)\times(W_o/32)}$
Backbone: Pre-trained SE-Resnet50
Skip-Connection

➢ Channel and Spatial Attention Component (CSAC)



5. Experimental Results

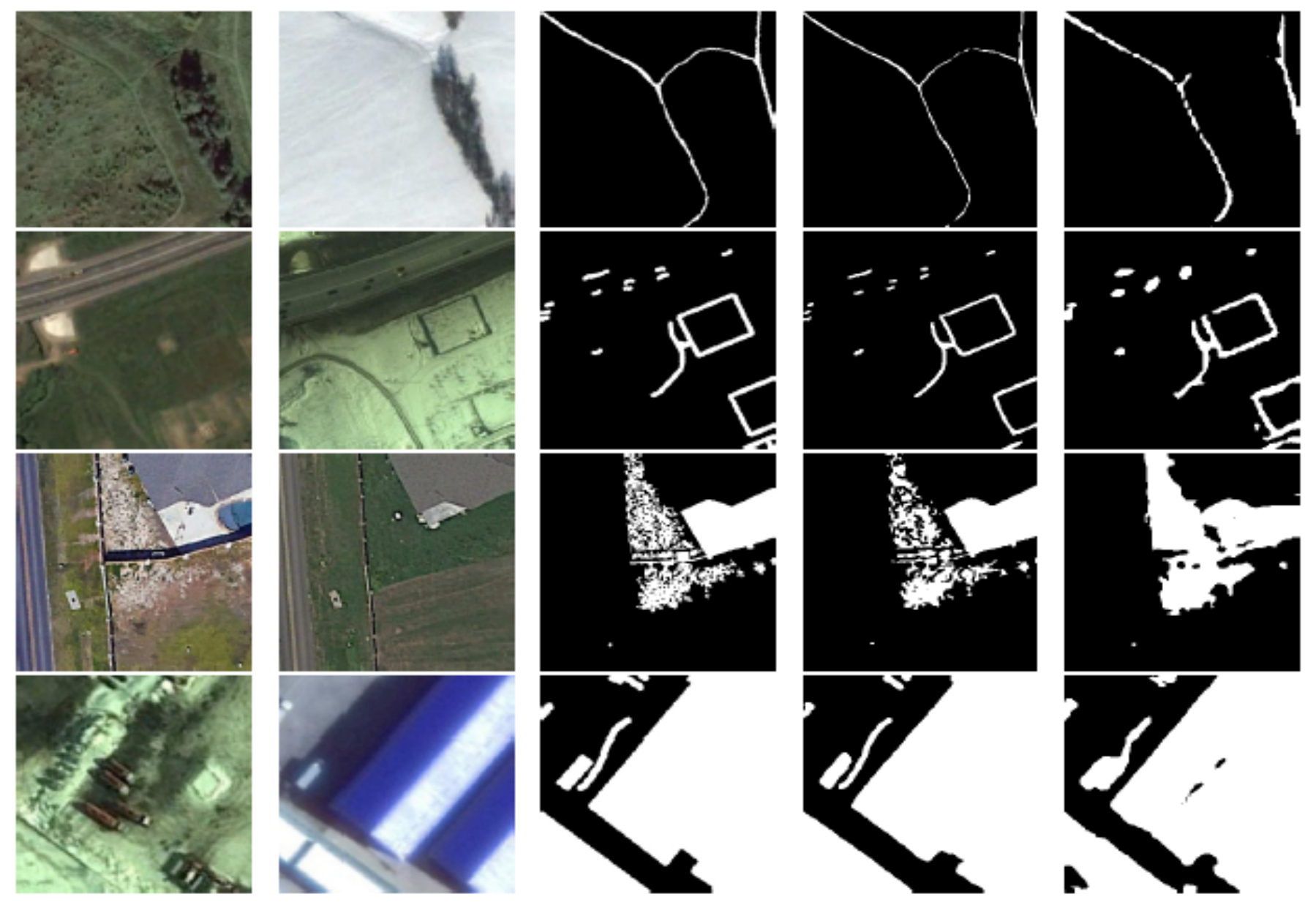
➢ Ablation Study

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**

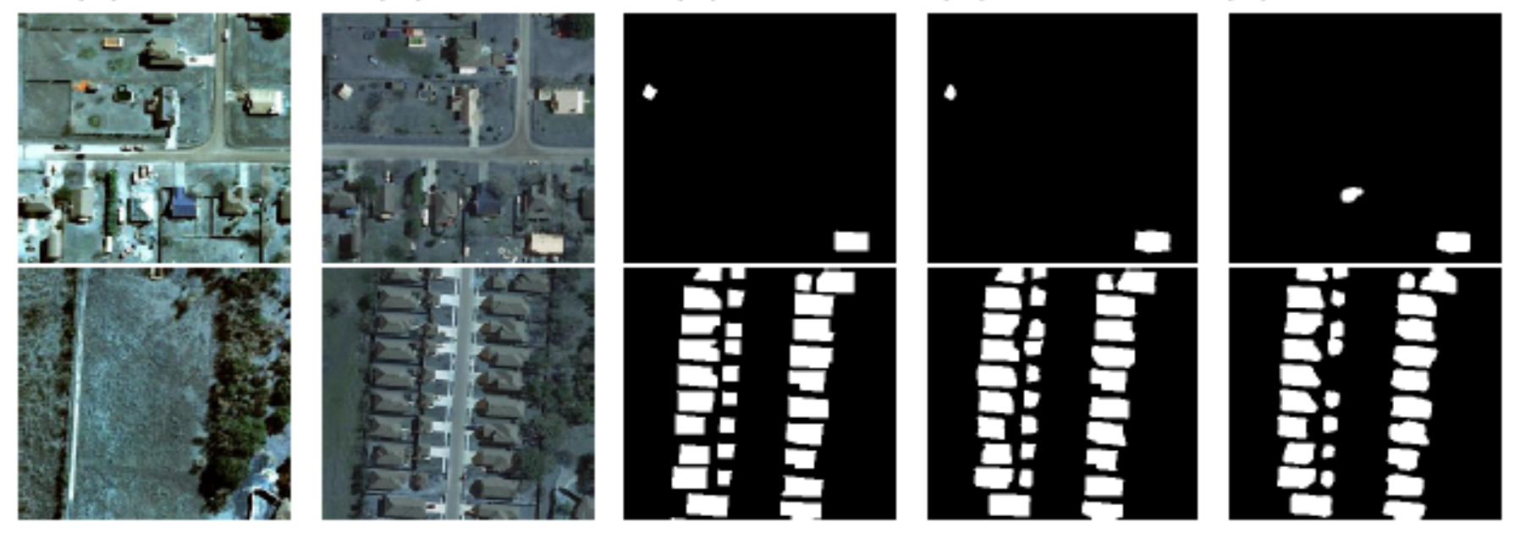
➢ Quantitative Results

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**

➢ Qualitative Results



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



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