A Change Detection Reality Check

Isaac Corley[1,2], Caleb Robinson[2], Anthony Ortiz[2]
[1]*University of Texas at San Antonio (UTSA)*
[2]*Microsoft AI for Good Research Lab*

The State of Change Detection

New architectures for Change Detection (CD) are being **proposed** almost **weekly**

Nearly all of them either:

- Have **no open-source code**
- Have untested experimental code **-- likely to have bugs**
- **Use new training methods independent of the architecture** but don't repeat experiments with prior works
- **Omit comparisons to better performing architectures** (whether purposefully or not)

As a research community we need to **do better**

This is not unique to CD. Other fields of ML suffer from the same problem:

- Metric Learning (Musgrave et al., 2020)
- Image Classification (Bello et al., 2021)
- Deep RL (Henderson et al., 2018)
- Point Cloud Classification (Uy et al., 2019)
- Video Recognition (Du et al., 2021)

Experiments

We re-benchmark U-Net (Ronneberger et al., 2015) on the LEVIR-CD and WHU-CD datasets

We modify the U-Net architecture to Siamese variants with pretrained backbones:

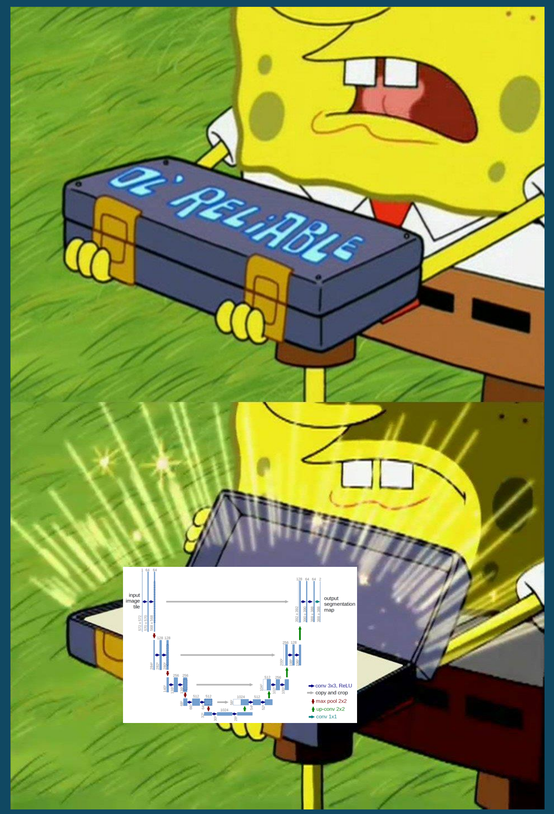
- *U-Net SiamConc* – concatenated features
- *U-Net SiamDiff* – difference of features

Our results in indicate that **U-Net does not require modifications or significant hyperparameter tuning to achieve SOTA results**

U-Net (2015) is **still** a **state-of-the-art** model when **properly benchmarked** for **change detection**

ARXIV

GITHUB



Microsoft | AI for Good Lab

UTSA
The University of Texas at San Antonio™

Benchmarks

Model	Backbone	Precision	Recall	F1
FC-EF (Daudt et al., 2018)	-	86.91	80.17	83.40
FC-Siam-Conc (Daudt et al., 2018)	-	91.99	76.77	83.69
FC-Siam-Diff (Daudt et al., 2018)	-	89.53	83.31	86.31
DTCDSCN (Liu et al., 2020)	SE-Resnet34	88.53	86.83	87.67
STANet (Chen & Shi, 2020)	ResNet-18	83.81	**91.00**	87.26
CDNet (Chen et al., 2021a)	ResNet-18	91.60	86.50	89.00
BIT (Chen et al., 2021b)	ResNet-18	89.24	89.37	89.31
ChangeFormer (Bandara & Patel, 2022b)	MiT-b1	92.59	89.68	91.11
Tiny-CD (Codegoni et al., 2023)	EfficientNet-b4	92.68	89.47	91.05
ChangerVanilla (Fang et al., 2023)	ResNet-18	92.66	89.60	91.10
ChangerEx (Fang et al., 2023)	ResNet-18	92.97	90.61	**91.77**
U-Net (Ronneberger et al., 2015)	EfficientNet-b4	92.69	87.16	89.25
U-Net (Ronneberger et al., 2015)	ResNet-50	91.97	89.78	90.38
U-Net SiamConc	ResNet-50	92.87	89.48	90.41
U-Net SiamDiff	ResNet-50	**93.21**	89.50	90.46

LEVIR-CD Benchmark Results

Model	Backbone	F1	Pre.	Rec.	IoU
Averaged Over 10 Seeds					
ChangeFormer	MiT-b1	75.65 ± 1.58	77.06 ± 3.22	74.67 ± 1.97	61.60 ± 2.05
TinyCD	EfficientNet-b4	78.53 ± 1.28	80.15 ± 2.49	77.56 ± 2.13	65.52 ± 1.72
BIT	ResNet-18	72.67 ± 2.69	70.30 ± 6.36	76.84 ± 4.53	58.06 ± 3.24
U-Net	ResNet-50	81.85 ± 1.32	83.72 ± 2.65	80.39 ± 2.32	69.96 ± 1.83
U-Net SiamConc	ResNet-50	81.33 ± 1.08	79.30 ± 2.78	**84.19 ± 1.55**	69.40 ± 1.52
U-Net SiamDiff	ResNet-50	**82.02 ± 1.48**	**83.82 ± 3.80**	80.92 ± 2.51	**70.29 ± 2.00**
Best Seed					
ChangeFormer	MiT-b1	77.75	82.60	78.57	64.22
TinyCD	EfficientNet-b4	78.53	80.15	77.56	65.52
BIT	ResNet-18	77.68	78.58	82.13	64.34
U-Net	ResNet-50	**84.17**	**88.65**	83.08	**73.23**
U-Net SiamConc	ResNet-50	82.75	83.69	**86.56**	71.15
U-Net SiamDiff	ResNet-50	84.01	88.56	85.63	73.02

WHU-CD Benchmark Results

Path Forward

We need to **accurately measure true progress in the change detection field**. To do so, we should consider the following steps:

- Utilize a standardized *evaluation harness* for *benchmarking new architectures*
- Create *larger* and more *geographically diverse benchmark datasets*
- Create a unified *leaderboard* for *change detection performance*
- Encourage all proposed methods to open-source their code/weights