Research Project Poster

How do Transformer models perform in urban change detection with limited satellite datasets, and what strategies can enhance their accuracy for this task?

Current Research

Change Detection (CD) can be defined as a task of extracting natural or artificial changes from a specific land area using multiple images (usually obtained from satellites) from different timeframes.

Figure 1. Pre-Change Image, set A

Figure 2. Post-Change Image, set B

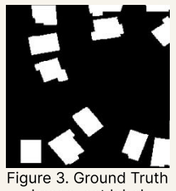

Figure 3. Ground Truth Image, set label

Challenges:
- weather
- illumination
- sensors calibration
- orbital parameters

Figure 4. Difference in representation of the same area.

Based on the Transformer model introduced in 2017, researchers have created a **Visual Image Transformer (ViT)** which splits the image into tokens similarly to NLP.

Bitemporal Image Transformer (BIT)
- Bitemporal image expressed into few tokens.
- Transformer encoder models token-based space-time contexts.
- Baseline model used for comparison to many other models.

Visual Change Transformer (VcT)
- More recent and complicated model.
- Proven to achieve higher accuracy.
- Using additional aspects, such as Graph Neural Network (GNN) to model the token space.

Methodology

Resolution Reduction: downscaling the images to contain less pixels

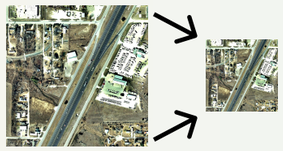

Figure 5. Resolution Reduction example.

Image Subset Sampling: selecting certain portion of the images to represent the set

Figure 6. Image Subset Sampling example.

Experiments conducted on BIT included 8 by size, 8 by resolution and 1 baseline. Due to difficulties with running high resolution images on VcT, only by size datasets were used for VcT experiments, as visible in Table 1 and Table 2.

BIT		
Category	**By size**	**By resolution**
Category 1	20%	205x205
Category 2	30%	307x307
Category 3	40%	410x410
Category 4	50%	512x512
Category 5	60%	614x614
Category 6	70%	717x717
Category 7	80%	819x819
Category 8	90%	922x922
Category 9	100%, 1024x1024 (baseline)	

Table 1. BIT experiments.

VcT	
Category	**By size**
Category 1	20%
Category 2	30%
Category 3	40%
Category 4	50%
Category 5	60%
Category 6	70%
Category 7	80%
Category 8	90%
Category 9	100% (baseline)

Table 2. VcT experiments.

Experiment parameters include:
- 200 epochs (BIT), 30 epochs (VcT)
- NVIDIA A100 GPU in Google Colab
- Training/Validation/Testing

Evaluation:
- Time
- GPU RAM
- Overall Accuracy

Conclusions

Results can be summarized as follows:
- OA > 0.5 - model is learning anything
- 30 epochs enough
- VcT better
- time increases linearly
- more time (epochs) - model eventually starts to learn
- **smaller datasets** can be effectively utilized to provide acceptable results

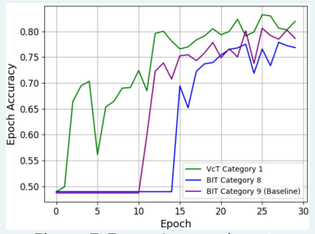

Figure 7. Example experiments epoch accuracy over time.

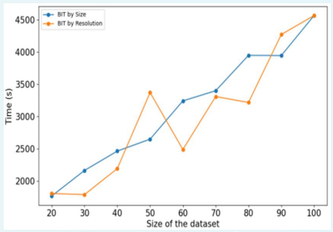

Figure 8. Time of execution for each BIT dataset.

Future work can explore how smaller datasets can still achieve effective results in collaborative AI, where humans and machine learning models cooperate for optimal outcomes. Additionally, while many studies produce accurate results, their high costs limit practical application. Cost-effective methods using limited data having accurate results are needed for practical use.

Ethical considerations in remote change detection, particularly in military contexts, encompass privacy concerns related to spying software and the potential for misuse. Balancing security needs with respect for individual rights is essential in navigating these ethical challenges.

Author
Jan Bryczkowski
J.M.Bryczkowski@student.tudelft.nl

Supervisor
Jan van Gemert
Delft University of Technology

Supervisor
Desislava Petrova-Antonova
GATE Institute, Sofia University

Affiliations
EEMCS
Delft University of Technology