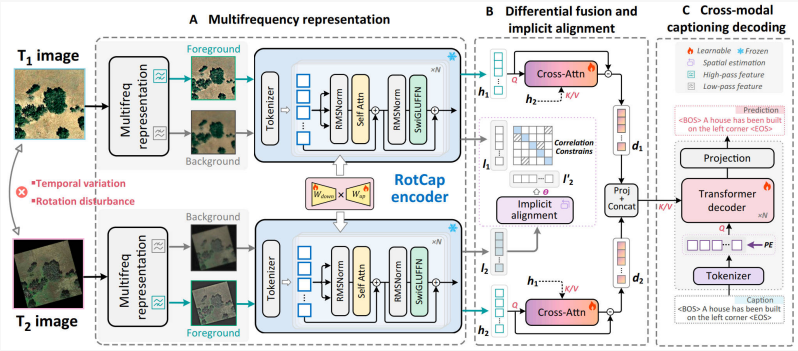
Rotation-tolerant remote sensing captioning

GA, UNITED STATES, April 28, 2026 /EINPresswire.com/ -- A new artificial intelligence method can describe changes in satellite images even when the images are rotated or spatially misaligned. The framework, called RotCap, improves remote sensing change captioning by separating stable background information from changed foreground objects, then aligning scene content before generating text. The advance could reduce dependence on manual image registration and improve large-scale monitoring of urban expansion, land-use change, and disaster impacts.



Pipeline of the proposed RotCap framework.

Remote sensing change detection is widely used to track construction, land transformation, and disaster damage, but most systems only identify changed pixels or objects rather than explain them in language. Change captioning goes further by converting visual differences between images taken at different times into human-readable text. However, this task is highly sensitive to real-world imaging disturbances such as rotation, scale variation, illumination differences, and background noise. Manual postprocessing can reduce these effects, but it is too time-consuming for large volumes of uncalibrated satellite data. Based on these challenges, it is necessary to carry out in-depth research on robust remote sensing change captioning under disturbance.

Researchers from Southwest Jiaotong University, the University of the Witwatersrand, and the Aerospace Information Research Institute of the Chinese Academy of Sciences published (DOI: 10.34133/remotesensing.1037) the study on March 3, 2026, in Journal of Remote Sensing. Their paper introduces RotCap, a disturbance-robust framework designed to generate accurate textual descriptions of scene changes from multitemporal satellite images, even when those images are affected by rotation and related imaging inconsistencies that often weaken conventional captioning systems.

RotCap is built around a self-supervised multifrequency representation strategy. It separates each image into high-frequency and low-frequency components, allowing the model to focus on changed foreground objects while preserving relatively stable background structure for alignment. A cross-attention mechanism captures change patterns between bitemporal foreground tokens, while a self-supervised correlation constraint aligns low-frequency

background tokens across time. On the LEVIR-CC-Rot15 dataset, RotCap improved BLEU-4 by 13.51% and CIDEr-D by 14.43% over the best competing method. On LEVIR-CC-Rot30, it achieved a 13.47% BLEU-4 gain over the next-best model. It also maintained strong performance on the original LEVIR-CC benchmark, showing good generalization in both disturbed and standard settings.

The researchers used the LEVIR-CC benchmark, which contains 11,163 pairs of 256 × 256 bitemporal images, each paired with five reference sentences describing building-related changes. To test robustness, they created two disturbed datasets by rotating the original images by 15° and 30° while preserving scene content through bilinear interpolation. RotCap first decomposes images in the frequency domain, then fuses the original, low-frequency, and high-frequency information into weighted representations. The low-frequency branch supports implicit alignment through a spatial estimation network and learnable transformation, while the high-frequency branch highlights changed objects through cross-attention and differential fusion. Visualization experiments showed that RotCap could correctly interpret new construction even in scenes with rotation, scale variation, pseudo-change, and background interference, where competing methods often misread the changes. Ablation studies further showed that both multifrequency representation and implicit alignment improved results, with balanced frequency weighting and a lightweight cross-attention design producing the strongest performance.

According to the paper's discussion, the study points toward a shift from traditional pixel-level change extraction to scene-level, disturbance-invariant interpretation. The authors suggest that robust captioning from unregistered bitemporal imagery may help build a more practical remote sensing analysis paradigm for complex real-world imaging conditions.

All experiments were conducted on an NVIDIA RTX 3090 GPU. The selected models were trained for 40 epochs with a batch size of 4 using the AdamW optimizer and an initial learning rate of 5×10^{-4}. RotCap used QLoRA optimization to reduce memory demands, with rank set to 8 and alpha set to 32. Performance was evaluated with BLEU-1/2/3/4, ROUGE-L, METEOR, and CIDEr-D. To directly test robustness to rotation disturbance, the team deliberately excluded random rotation augmentation during training.

The authors indicate that RotCap could support more efficient urban expansion monitoring, disaster dynamic assessment, and large-scale Earth observation workflows by reducing reliance on manual preprocessing. Its lightweight design and accelerated inference strategy also make it promising for real-time interpretation in edge-computing environments. Future work will explore more complex disturbances, including broader image corruption and noisy text labels, while integrating stronger foundation models to further improve robustness and semantic understanding.

DOI
10.34133/remotesensing.1037

Original Source URL
https://doi.org/10.34133/remotesensing.1037

Funding information
This work was supported in part by the National Natural Science Foundation of China under grant 62271418, the Natural Science Foundation of Sichuan Province under grants 2023NSFSC0030 and 2025ZNSFSC1154, the Postdoctoral Fellowship Program and China Postdoctoral Science Foundation under grants BX20240291 and 2025M770235, and the Fundamental Research Funds for the Central Universities under grant 2682025CX033.

Lucy Wang
BioDesign Research
email us here

This press release can be viewed online at: https://www.einpresswire.com/article/908770286

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.
© 1995-2026 Newsmatics Inc. All Right Reserved.