COREGISTRATION OF SATELLITE IMAGE TIME SERIES THROUGH ALIGNMENT OF ROAD NETWORKS

Andres F. Pérez
Manitoba Learning and AI Research (MLAIR)
Department of Electrical and Computer Engineering
University of Manitoba
perezmaf@myumanitoba.ca

Pooneh Maghoul
Department of Civil, Geological and Mining Engineering
Polytechnique Montréal
pooneh.maghoul@polymtl.ca

Ahmed Ashraf
Manitoba Learning and AI Research (MLAIR)
Department of Electrical and Computer Engineering
University of Manitoba
ahmed.ashraf@umanitoba.ca

ABSTRACT

Due to climate change, thawing permafrost affects transportation infrastructure in northern regions. Tracking deformations over time of these structures can allow identifying the most vulnerable sections to permafrost degradation and implement climate adaptation strategies. The Sentinel-2 mission provides data well-suited for multitemporal analysis due to its high temporal resolution and multispectral coverage. However, the geometrical misalignment of Sentinel-2 imagery makes this analysis challenging. Towards the goal of estimating the deformation of linear infrastructure in northern Canada, we propose an automatic subpixel coregistration algorithm for satellite image time series based on the matching of binary masks of roads produced by a deep learning model. We demonstrate the feasibility of achieving subpixel coregistration through alignment of roads on a small dataset of high-resolution Sentinel-2 images from the town of Gillam in northern Canada. This is the first step towards training a road deformation prediction model.

1 INTRODUCTION

Permafrost is soil or rock that remains at or below $0\,^{\circ}\mathrm{C}$ for at least two consecutive years (Liu et al., 2022). Climate warming has adversely affected permafrost areas, resulting in an accelerated thawing process and an increase in thickness of the active layer (Bishop et al., 2011). According to the latest report on climate change (Bush & Lemmen, 2019), Canada has been warming at about twice the rate of the entire globe. In fact, between 1948 and 2016, the average annual temperature has increased by close to $1.7\,^{\circ}\mathrm{C}$, and $2.3\,^{\circ}\mathrm{C}$ in the northern region. These effects have implications for transportation infrastructure, which in turn may also affect communities access to transportation services for commuting, healthcare, and for the movement of goods and resources.

While all transportation systems get impacted by the climate, northern transportation systems in Canada experience some of the greatest impacts on account of degrading permafrost (Palko & Lemmen, 2017). Permafrost thawing (ground settlement, slope instability, drainage issues, cracking) causes damages to roads, railways, airport taxiways, and runways (Palko & Lemmen, 2017). Tracking changes over time in the surface deformation of transportation infrastructure can allow us to monitor the quality of the underlying permafrost by correlating these changes to geophysical and hydrological data.

The Sentinel-2 (S2) satellite imagery mission, conducted by the European Space Agency (ESA), provides high-resolution imagery of Earth's surface suitable for multitemporal analyses of small geographic features down to subpixel size (Radoux et al., 2016). However, the multitemporal coregistration accuracy of S2 Level 1C data products is currently of $12\,\mathrm{m}$ (Gascon et al., 2017a) which results in coregistration errors within S2 time series of more than one full pixel in the visible (VIS)

and near-infrared (NIR) bands. These inaccuracies are expected to be reduced after the activation of the Global Reference Image (GRI) (Enache & Clerc, 2023), nevertheless there are no plans for reprocessing past data (Gascon et al., 2017b; Storey et al., 2016). Such settings are insufficient for multitemporal analysis and claim for better methods to eliminate pixel and subpixel inconsistencies.

Traditional approaches to address the problem of image coregistration in remote sensing provide low residual offsets after registration, which are unsuitable to analyze changes in small features such as roads and railways. For instance, Rufin et al. (2021) achieved a mean coregistration precision of $4.4\,\mathrm{m}$, while Stumpf et al. (2018) suggests that through an affine transformation, a residual Root Mean Square Error (RMSE) of approximately $3\,\mathrm{m}$ is possible.

In this work, instead of performing matching in the space of remote sensing images, we propose a multitemporal image co-registration method using the Enhanced Cross-Correlation (ECC) criterion Evangelidis & Psarakis (2008) to match road masks. These masks are represented as binary images produced by a road segmentation deep learning model trained in a supervised manner. The contributions of this work can be summarized as follows:

1. We propose an adaptation of a SOTA deep learning road extraction architecture (Zhou et al., 2018) able to process S2 data and generate accurate road classification masks.
2. We introduce a curated dataset of high-resolution optical S2 imagery from the region of interest with pixel-level labels of roads.
3. We propose an image alignment algorithm for binary images of detected road networks for the coregistration of multitemporal satellite image time series.

2 DATA

To validate our method we introduce a dataset of satellite images acquired from the Copernicus Sentinel-2 earth observation mission. The dataset consists of 2051 high-resolution S2 satellite images from the town of Gillam, Manitoba in northern Canada. The corresponding region of interest was buffered by $0.5\,\mathrm{km}$ and subsequently split into smaller more manageable tiles of $2.5\,\mathrm{km} \times 2.5\,\mathrm{km}$ (see Appendix A for details). The images were obtained for each tile for the period between January and December of 2020 and only those with a maximum cloud coverage of 80% were downloaded. We used only the VIS and NIR bands corresponding to the highest resolution of 10 m resulting in images of size 250×250 pixels.

The dataset images were annotated for roads based on OpenStreetMap (OSM) vector data (OSM codes 5111-5115, 5121-5124, 5131-5135 and 5141) where we would only need to adjust the road centerline annotations to a reference base image. The base image is a mosaic for the region of interest composed of summer images from July 2020 since they exhibit the greatest color contrast and thus allow for better distinction of geographical features. The curated road centerline annotations were buffered by $5\,\mathrm{m}$ and rasterized at a resolution of $2.5\,\mathrm{m}$ per pixel. These annotations were used for all of the images of the time series under the assumption that the roads in the region of interest do not change, or change very slightly, during the time interval of the time series.

3 METHODS

3.1 ROAD NETWORK EXTRACTION

Having a robust road network extraction model is critical to obtaining accurate road masks and thus to the performance of our algorithm. As basis for our method we adopt the D-LinkNet (Zhou et al., 2018) architecture, winner of the CVPR DeepGlobe 2018 Road Extraction Challenge (Demir et al., 2018). This network combines the power of dilated convolutions (Yu et al., 2017) and residual connections (He et al., 2016) with the popular U-Net (Ronneberger et al., 2015) architecture for semantic segmentation. To cope with the low spatial resolution of the S2 images we follow a strategy similar than Ayala et al. (2021) and apply bilinear interpolation to the input images with Gaussian smoothing to avoid aliasing effects. This way we upsample the images from their raw size of 250×250 pixels to a size of 1024×1024 pixels. Upsampled images combined with the ground truth masks rasterized at $2.5\,\mathrm{m}$ per pixel allow to produce high accuracy road masks from S2 images.

3.2 MULTITEMPORAL COREGISTRATION

We address the task of multitemporal SITS coregistration as an image alignment problem of binary images using the ECC criterion. The main concept of our SITS coregistration algorithm is illustrated in Figure 1. Given a SITS, we put all images in the same reference frame by aligning the detected road masks against a reference road mask and use the obtained transformation matrices to warp the satellite images (see Appendix C for algorithm pseudocode). The reference road mask corresponds to the satellite image closest in time to the one in the base image mosaic used for annotation.

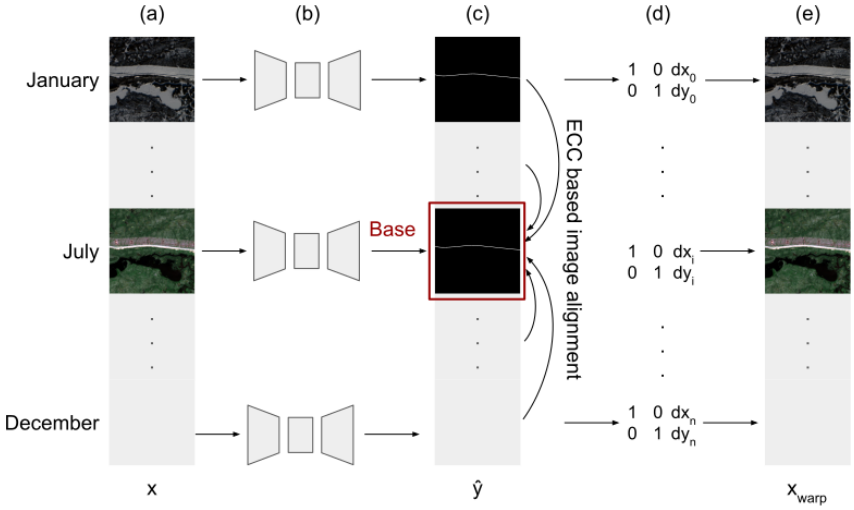


Figure 1: SITS Coregistration. We forward a time series of satellite images (a) through a road extraction model (b) and obtain a time series of road masks (c) from which we pick a reference road mask. The resulting road masks are aligned to the reference road mask and we obtain a set of warp matrices (d) that we use to obtain warped satellite images from the time series (e).

With the aid of a previously trained road extraction model, we generate for each satellite image a binary road mask that indicates whether or not a pixel corresponds to a road. Due to the geometrical misalignment of the satellite images, the generated binary images are not directly comparable to each other and require to be put under a common reference frame. We pick one of these road masks to use as a reference and find the geometric transform (warp matrix) between every other road mask and the reference road mask in terms of the ECC criterion. The obtained transformation matrices are finally used to correct the offsets of the images in the time series. We consider only translation motion at a maximum range of $50\,\mathrm{m}$ otherwise we fallback to an identity transformation.

3.3 RESULTS AND DISCUSSION

Our goal is to put misaligned images from a time series in the same frame of reference to enable accurate further analysis. To evaluate how well our method addresses this problem we perform two sets of experiments, 1) we investigate the robustness of our road extraction model under different settings and 2) we perform a quantitative analysis of our coregistration algorithm in comparision with a standard coregistration method in terms of image alignment and road detection quality.

In the first set of experiments, we train a D-LinkNet network with two different backbone networks in a fully supervised manner. Following previous works Lian et al. (2020), we use a combination of Dice and Binary Cross-Entropy loss to address the class imbalance due to the low presence of roads on the images. To assess the robustness of our method, we randomly split the dataset into 5 folds and evaluate the models for each of them and across multiple runs. In all cases, we trained for 300 epochs with batches of 8 elements using the Adam optimizer with a learning rate of 2×10^{-4} and a threshold of 0.5 on the output binary masks.

To measure the road detection accuracy of each model we calculate Precision, Recall and Intersection Over Union (IoU). Table 1 reports the mean and standard deviation for each metric and fold across five runs. We observe that the additional parameters and residual connections of ResNet-34 allow it to capture more complex representations of the images and achieve a higher IoU. With either backbone in three out of five folds the model achieves a high recall but low precision, i.e. it returns a large amount of false positives. This is indicative that we have trained our models to detect as roads pixels which are not roads, e.g. embankments or deposits of snow along the road. We hypothesize that a higher multitemporal alignment would result in a higher road detection accuracy.

Backbone	**Metric**	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**
ResNet-18	IoU	57.28 ± 0.95	48.86 ± 1.09	52.11 ± 1.70	44.30 ± 0.32	48.10 ± 1.01
	Precision	$\mathbf{71.89 \pm 0.21}$	67.75 ± 0.97	$\mathbf{66.72 \pm 1.39}$	61.88 ± 0.78	$\mathbf{60.05 \pm 1.53}$
	Recall	$\mathbf{72.69 \pm 1.14}$	61.81 ± 1.37	$\mathbf{68.30 \pm 1.94}$	58.68 ± 0.43	$\mathbf{66.93 \pm 1.42}$
ResNet-34	IoU	61.56 ± 0.53	53.94 ± 1.18	58.63 ± 0.15	52.39 ± 0.97	54.76 ± 0.64
	Precision	$\mathbf{72.12 \pm 0.78}$	71.02 ± 1.52	$\mathbf{70.97 \pm 0.99}$	69.23 ± 0.30	$\mathbf{65.17 \pm 0.89}$
	Recall	$\mathbf{78.73 \pm 0.55}$	68.20 ± 2.45	$\mathbf{75.54 \pm 1.05}$	67.39 ± 1.53	$\mathbf{74.41 \pm 0.83}$

Table 1: Performance of our road extraction model measured across five folds and using five runs per fold.

For the second set of experiments, we run our algorithm on the road masks produced by the models with ResNet-34 backbone. As baseline, we use AROSICS (Stumpf et al., 2018) to sequentially match pairs of images from last to first in the NIR band, as it is less susceptible to seasonal changes. Table 2 shows the performance of both methods in terms of IoU and Structural Similarity Index (SSIM) to assess visual consistency[1]. Our method achieves an increase in road detection accuracy across all folds ($\sim$6 % in IoU) at the cost of a small decrease in visual consistency ($\sim$2 % in SSIM). Conversely, the baseline prioritizes visual consistency and registers an increase in visual consistency ($\sim$2 % in SSIM) at the expense of a decrease ($\sim$5 % in IoU) in road detection accuracy.

Method	**Metric**	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**
None	SSIM	74.90	74.53	72.75	72.72	73.34
Ours	IoU	$\mathbf{67.49 \pm 1.61}$	$\mathbf{59.98 \pm 1.61}$	$\mathbf{64.63 \pm 0.51}$	$\mathbf{60.40 \pm 1.06}$	$\mathbf{61.77 \pm 0.49}$
	SSIM	73.31 ± 0.96	74.62 ± 0.39	71.43 ± 0.35	70.36 ± 0.85	70.89 ± 0.53
Baseline	IoU	53.43 ± 0.57	49.78 ± 0.95	53.24 ± 0.18	47.28 ± 0.74	53.33 ± 0.33
	SSIM	**76.95**	**76.21**	**74.86**	**74.71**	**75.24**

Table 2: Road detection accuracy and image alignment quality after coregistration.

These results are not surprising given that our method favors alignment of detected structures and relies on the accuracy of the road extraction model to keep visual consistency. Oppositely, methods like AROSICS favor visual consistency and disregard inconsistencies in smaller structures. Our initial hypothesis that increased multitemporal alignment would result in a higher road detection accuracy was correct, nevertheless the results also indicate a trade off between road detection accuracy and visual consistency.

4 CONCLUSIONS

With the end goal of estimating the deformation of linear infrastructure due to permafrost degradation, in this work we presented a novel method for satellite images coregistration to enable accurate multitemporal analyses. We introduced a dataset of S2 images from the town of Gillam in northern Canada and adapted a SOTA road network extraction architecture to operate on S2 imagery. Through a quantitative analysis we proved that a higher road detection accuracy is possible through the alignment of detected structures at the cost of a decrease in visual consistency.

Aligning binary road masks has several advantages over aligning remote sensing images; namely, they are less sensitive to seasonal variability; they have less clutter easing the alignment process;

[1]Mean SSIM between pairs of images for all the time series.

and finally, they are more consistent across time. In spite of this, one of the main challenges of our method remains the need of a highly accurate classification model. Investigating the effect of inaccurate annotations in the task of roads segmentation constitutes an interesting research direction.

On the perspective of climate change, the proposed method can serve as the basis for a multimodal fusion model due to its agnostic nature to the satellite sensor and geographical features used for matching. Such fusion model would further facilitate the ultimate goal of training a road deformation prediction model. In this sense, we intend to evolve this work into a learnable model capable of handling misaligned images and still producing accurate detection results useful for downstream multitemporal analyses.

5 ACKNOWLEDGMENTS

The authors acknowledge the financial support of the New Frontiers in Research Fund - Exploration Grant [NFRF-2018-00966]. We also thank Sentinel Hub for providing us with access to the Sentinel-2 L1C data used in this study through the Network of Resources (NoR) sponsorship program. Finally, we would like to extend our thanks to Dr. Heather D. Couture from Pixel Scentia Labs for her support in the production of this article through the Climate Change AI (CCAI) mentorship program.

REFERENCES

Christian Ayala, Carlos Aranda, and Mikel Galar. Towards fine-grained road maps extraction using sentinel-2 imagery. *Isprs Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 5 (3), 9-14*, 2021.

Michael P Bishop, Helgi Björnsson, Wilfried Haeberli, Johannes Oerlemans, John F Shroder, and Martyn Tranter. *Encyclopedia of snow, ice and glaciers*. Springer Science & Business Media, 2011.

Elizabeth Bush and Donald Stanley Lemmen. *Canada's changing climate report*. Government of Canada, Ottawa, ON, 2019.

Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

Silvia Enache and Sebastien Clerc. Sentinel-2 L1C Data Quality Report. `https://sentinel.esa.int/documents/247904/4868341/OMPC.CS.DQR.001.12-2022+-+i83r0+-+MSI+L1C+DQR+January+2023.pdf`, 2023. [Online; accessed 2-February-2023].

Georgios D. Evangelidis and Emmanouil Z. Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–1865, 2008. doi: 10.1109/TPAMI.2008.113.

Ferran Gascon, Catherine Bouzinac, Olivier Thépaut, Mathieu Jung, Benjamin Francesconi, Jérôme Louis, Vincent Lonjou, Bruno Lafrance, Stéphane Massera, Angélique Gaudel-Vacaresse, Florie Languille, Bahjat Alhammoud, Françoise Viallefont, Bringfried Pflug, Jakub Bieniarz, Sébastien Clerc, Laëtitia Pessiot, Thierry Trémas, Enrico Cadau, Roberto De Bonis, Claudia Isola, Philippe Martimort, and Valérie Fernandez. Copernicus sentinel-2a calibration and products validation status. *Remote Sensing*, 9(6), 2017a. ISSN 2072-4292. doi: 10.3390/rs9060584. URL `https://www.mdpi.com/2072-4292/9/6/584`.

Ferran Gascon, Catherine Bouzinac, Olivier Thépaut, Mathieu Jung, Benjamin Francesconi, Jérôme Louis, Vincent Lonjou, Bruno Lafrance, Stéphane Massera, Angélique Gaudel-Vacaresse, Florie Languille, Bahjat Alhammoud, Françoise Viallefont, Bringfried Pflug, Jakub Bieniarz, Sébastien Clerc, Laëtitia Pessiot, Thierry Trémas, Enrico Cadau, Roberto De Bonis, Claudia Isola, Philippe Martimort, and Valérie Fernandez. Copernicus sentinel-2a calibration and products validation

status. *Remote Sensing*, 9(6), 2017b. ISSN 2072-4292. doi: 10.3390/rs9060584. URL https://www.mdpi.com/2072-4292/9/6/584.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

Renbao Lian, Weixing Wang, Nadir Mustafa, and Liqin Huang. Road extraction methods in high-resolution remote sensing images: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:5489–5507, 2020. doi: 10.1109/JSTARS.2020.3023549.

H. Liu, P. Maghoul, and A. Shalaby. Seismic physics-based characterization of permafrost sites using surface waves. *The Cryosphere*, 16(4):1157–1180, 2022. doi: 10.5194/tc-16-1157-2022. URL https://tc.copernicus.org/articles/16/1157/2022/.

K Palko and Donald Stanley Lemmen. Climate risks and adaptation practices for the canadian transportation sector 2016. pp. 27–64, 2017.

Julien Radoux, Guillaume Chomé, Damien Christophe Jacques, François Waldner, Nicolas Bellemans, Nicolas Matton, Céline Lamarche, Raphaël D'Andrimont, and Pierre Defourny. Sentinel-2's potential for sub-pixel landscape feature detection. *Remote Sensing*, 8(6), 2016. ISSN 2072-4292. doi: 10.3390/rs8060488. URL https://www.mdpi.com/2072-4292/8/6/488.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi (eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

Philippe Rufin, David Frantz, Lin Yan, and Patrick Hostert. Operational coregistration of the sentinel-2a/b image archive using multitemporal landsat spectral averages. *IEEE Geoscience and Remote Sensing Letters*, 18(4):712–716, 2021. doi: 10.1109/LGRS.2020.2982245.

James Storey, David P. Roy, Jeffrey Masek, Ferran Gascon, John Dwyer, and Michael Choate. A note on the temporary misregistration of landsat-8 operational land imager (oli) and sentinel-2 multi spectral instrument (msi) imagery. *Remote Sensing of Environment*, 186:121–122, 2016. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2016.08.025. URL https://www.sciencedirect.com/science/article/pii/S0034425716303285.

André Stumpf, David Michéa, and Jean-Philippe Malet. Improved co-registration of sentinel-2 and landsat-8 imagery for earth surface motion measurements. *Remote Sensing*, 10(2), 2018. ISSN 2072-4292. doi: 10.3390/rs10020160. URL https://www.mdpi.com/2072-4292/10/2/160.

Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. Dilated residual networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

Lichen Zhou, Chuang Zhang, and Ming Wu. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 192–1924, 2018. doi: 10.1109/CVPRW.2018.00034.

A Region Of Interest

Our main study region is the town of Gillam located in the province of Manitoba in northern Canada. The administrative region was obtained from OSM and subsequently buffered by $0.5\,\mathrm{km}$ and split into 455 tiles of $2.5\,\mathrm{km} \times 2.5\,\mathrm{km}$. The dataset images belong only to tiles where there is a road, i.e., images with a certain road presence ($0.1\,\%$).

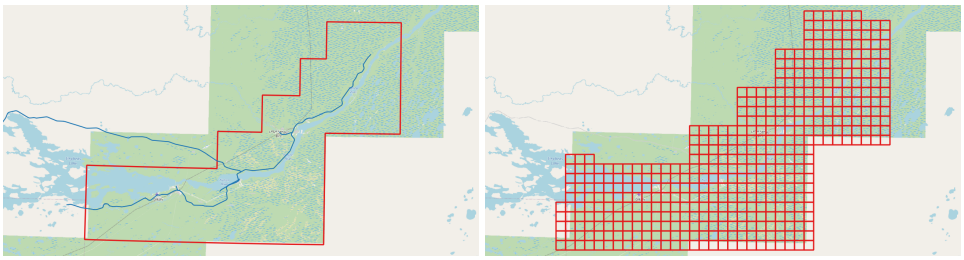
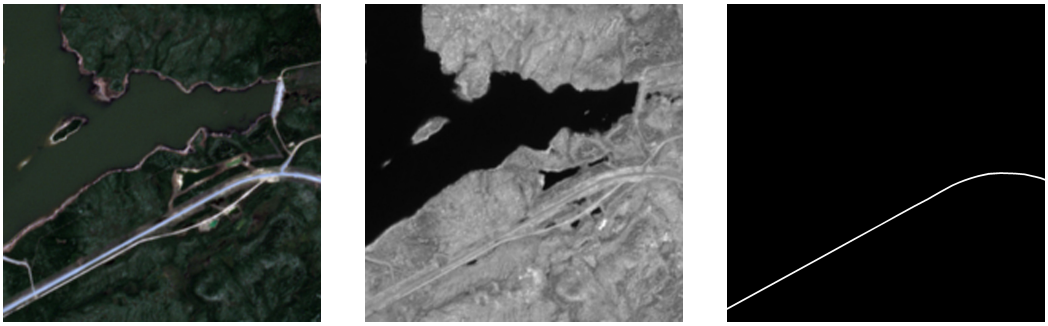


Figure 2: *Left:* Region of interest with roads annotated. *Right:* Buffered region of interest split into manageable smaller tiles.

B Dataset Images

Sentinel-2 images consist of 13 spectral bands at different spatial resolutions, only the VIS and NIR bands are at $10\,\mathrm{m}$. In our training experiments of the road extraction models, we only used the VIS bands corresponding to the RGB channels. This was due to the fact that the ResNet backbones were previously trained on the ImageNet dataset, which is made of only RGB images. However we downloaded as well the NIR band used for matching in the AROSICS baseline.



(a) RGB Image (b) NIR Image (c) Road Mask

Figure 3: Dataset sample. (a) True color RGB image. (b) Grayscale NIR image. (c) Binary road mask image.

C Coregistration Algorithm

Our coregistration algorithm was textually described in the main text, here we provide a detailed description of the algorithm pseudocode.

Algorithm 1 Satellite Image Time Series (SITS) Coregistration

Inputs: X, time series of satellite images
$model()$, road extraction model
$date_{ref}$, date of reference base image
Output: X_{warp}, time series of warped satellite images

function SITS-COREGISTRATION(X, $model()$, $date_{ref}$)
  $\hat{Y} \leftarrow \emptyset$ ▷ *Initialize time series of road masks*
  for $t = 0$ **to** $|X|$ **do**
    $x \leftarrow X(t)$ ▷ *Retrieve satellite image at time step t*
    $\hat{y} \leftarrow$ EXTRACT-ROADS(x, $model()$) ▷ *Extract road mask*
    $\hat{Y}(t) \leftarrow \hat{y}$ ▷ *Append extracted road mask*
  end for
  $\hat{X}_{warp} \leftarrow \emptyset$ ▷ *Initialize time series of warped satellite images*
  $\hat{y}_{ref} \leftarrow$ NEAREST($\hat{Y}$, $date_{ref}$) ▷ *Find reference road mask*
  for $t = 0$ **to** $|X|$ **do**
    $x, \hat{y} \leftarrow X(t), \hat{Y}(t)$ ▷ *Retrieve satellite image and road mask at time step t*
    $m \leftarrow$ REGISTER($\hat{y}$, $\hat{y}_{ref}$) ▷ *Register road mask to reference road mask*
    $x_{warp} \leftarrow$ WARP(x, m) ▷ *Warp satellite image*
    $X_{warp}(t) \leftarrow x_{warp}$ ▷ *Append warped satellite image*
  end for
  return X_{warp} ▷ *Return time series of warped satellite images*
end function