MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

# Rotation-Equivariant Neural Networks for Cloud Removal from Satellite Images

Lohit, Suhas; Marks, Tim K.

TR2025-009 January 08, 2025

**Abstract**

In this paper, we aim to recover a cloud-free optical image from a cloudy optical image and aligned synthetic aperture radar (SAR) image using a deep neural network. In contrast to previous approaches, we make the observation that satellite image features generally have no preferred orientation. This insight can be incorporated into the design of the neural architecture by making the network layers obey the geometric constraint that changing the orientation of an input image should only change the orientation of the corresponding output image, without otherwise affecting the quality or details of the reconstruction. We build a multimodal rotation-equivariant neural network, called EquiCR (Equivariant Cloud Removal), that encodes this geometric prior exactly. When trained on the public SEN12MSCR dataset, we observe improvements in reconstructed image quality using EquiCR, compared to using deep learning without built-in rotation equivariance. Interestingly, EquiCR results in greater improvements over the baseline method in the more difficult cases—when the amount of cloud cover is high or when the training dataset is small.

*Asilomar Conference on Signals, Systems, and Computers (ACSSC) 2025*



Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

# Rotation-Equivariant Neural Networks for Cloud Removal from Satellite Images

Suhas Lohit
Mitsubishi Electric Research Laboratories
Cambridge, MA
Email: slohit@merl.com

Tim K. Marks
Mitsubishi Electric Research Laboratories
Cambridge, MA
Email: tmarks@merl.com

***Abstract*—In this paper, we aim to recover a cloud-free optical image from a cloudy optical image and aligned synthetic aperture radar (SAR) image using a deep neural network. In contrast to previous approaches, we make the observation that satellite image features generally have no preferred orientation. This insight can be incorporated into the design of the neural architecture by making the network layers obey the geometric constraint that changing the orientation of an input image should only change the orientation of the corresponding output image, without otherwise affecting the quality or details of the reconstruction. We build a multimodal rotation-equivariant neural network, called EquiCR (Equivariant Cloud Removal), that encodes this geometric prior exactly. When trained on the public SEN12MSCR dataset, we observe improvements in reconstructed image quality using EquiCR, compared to using deep learning without built-in rotation equivariance. Interestingly, EquiCR results in greater improvements over the baseline method in the more difficult cases—when the amount of cloud cover is high or when the training dataset is small.**

## I. Introduction

Nearly 70% of the earth's surface is covered by clouds on average [1]. This hinders important remote sensing applications that use optical images in areas such as disaster management, agriculture, and ecological monitoring. Thus, being able to effectively remove clouds from satellite imagery is of vital importance. While optical images are affected by the presence of clouds, radar can penetrate through clouds and provide some information about edges and materials that lie beneath the clouds, providing important side information for remote sensing applications.

Recently, with the help of the Sentinel 1 and 2 missions [2], large amounts of Synthetic Aperture Radar (SAR) and multispectral data are easily available. By carefully combining the data from different times and using multimodal registration, large datasets such as SENMS12CR [3], [4] and SENMS12CR-TS [5] are now publicly available that come with triplets of aligned images (SAR, cloudy optical, and cloud-free optical images). By treating the cloud-free optical imagaes as ground truth, we can use these datasets to train large supervised deep learning models. Even with large datasets, the problem remains challenging, especially when there is significant cloud cover. Since the cloudy and cloud-free images in the datasets were captured at different times, the cloud-free images do not provide a perfect ground truth for the cloudy images, which adds to the challenge of the problem.

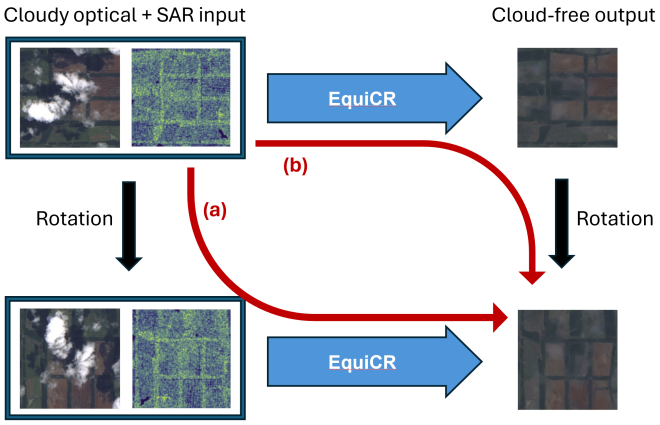


Fig. 1: As features in satellite images do not have canonical or preferred orientations, we build an equivariant cloud removal network, **EquiCR**, that has the property that the quality and details of reconstruction are the same even when the input is rotated. This is achieved by enforcing the constraint that (a) applying a rotation on both the SAR and cloudy optical image and then passing through EquiCR must yield the same result as (b) feeding the SAR and cloudy image through EquiCR first and then rotating the output image.

Several works employ deep learning for cloud removal, using methods such as deep ResNets [6], adversarial learning [4], or diffusion models [7]. These methods take in a cloudy optical image and a corresponding SAR image and feed them through a series of neural network layers to produce an estimated cloud-free image. In this paper, we make the observation that features in satellite images can appear in any orientation. Thus, the output of the neural network should be of the same quality irrespective of the rotation applied to the input image. In terms of network architecture, this requirement translates to having constraints on each of the layers that they are *rotation-equivariant*. This means that a rotation applied to the input images should be reflected exactly in the output estimated image, as illustrated in Figure 1.

We build a multimodal rotation-equivariant deep neural network for cloud removal, which we call EquiCR. The proposed model shows improved results, both qualitatively and

quantitatively, compared to a conventional neural network that does not have built-in rotation-equivariance. Additionally, we show that the improvements due to EquiCR are greater in the more difficult cases: when the training dataset is small, and when the amount of cloud cover is higher.

## II. Related Work

### A. Deep learning techniques for cloud removal

Several works using deep learning for cloud removal employ image-to-image translation such as pix2pix [8], [9] or CycleGAN [4]. The model DSen2-CR [6] is a simple convolutional neural network with residual connections that map SAR and cloudy optical images to an estimate of the cloudless optical image, and it has been shown to outperform other deep learning methods including U-Nets [6], [10]. However, unlike our architecture, this method is not equivariant to rotations of the image. We use DSen2-CR [6] as the baseline method for our work. Other works such as UncertainTS [11] are designed to work with a sequence of multimodal inputs taken at multiple times to enable easier cloud removal. In this paper, we focus on the more challenging cloud removal task in which there is a single snapshot in each modality. More recent work uses diffusion models [7] for performing cloud removal. However, such models are computationally slow due to the reverse diffusion sampling process.

### B. Equivariant neural networks

We start with a brief introduction to group theory. A set $\mathcal{G}$ with a binary operation $\cdot$ is called a group if the following properties hold:

1) Closure: $\forall\ g, h \in \mathcal{G},\ g \cdot h \in \mathcal{G}$
2) Associativity: $\forall\ g_1, g_2, g_3 \in \mathcal{G}, g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3$
3) Existence of identity: $\exists\ e \in \mathcal{G}$ such that $g \cdot e = e \cdot g = g, \forall\ g \in \mathcal{G}$
4) Existence of inverse: For each $g \in \mathcal{G}, \exists\ g^{-1}$ such that $g \cdot g^{-1} = g^{-1} \cdot g = e$.

We are interested in group elements that act on other objects such as images and neural network features. This is called group action. For a deep neural network represented as a function $f$ that maps input images to output images, let $\rho_g(\mathbf{x})$ denote the way a group element $g$ acts on an input image $\mathbf{x}$, and let $\rho'_g(f(\mathbf{x}))$ denote how the same group element $g$ acts on the output image[1]. The function $f$ is said to be equivariant to $\mathcal{G}$ if $f(\rho_g(\mathbf{x})) = \rho'_g(f(x)), \forall\ g \in \mathcal{G}$.

Conventional convolutional neural networks (CNNs) used for image restoration techniques including cloud removal are, in principle, only equivariant to the translation group, because each individual convolutional layer is equivariant to the group of translations. In order to design networks to have equivariance to other transformations, a more general class of group convolutional neural networks (GCNNs) can be designed for the transformation group of interest [12]. This generally boils down to constraining the learned filters in each layer to obey the desired equivariance property [13] or by different weight sharing mechanisms, depending on the group [14]. It is important to note that designing exact equivariant layers is different from data augmentation. In general, not only does equivariance perform better than data augmentation, but it also provides guaranteed robustness, which is not possible with just data augmentation [15].

In the case of satellite images, based on the insight that there are no canonical orientations of these images or their features, we build a multimodal rotation-equivariant network. Our network is designed to take in SAR inputs and cloudy optical images such that if the input SAR and cloudy optical images are rotated, then all of the intermediate feature maps as well as the output of the network rotate by the same amount. We refer the reader to [16] for a longer introductory treatment of group equivariant neural networks.

## III. Proposed method

We now describe the rotation-equivariant network we build for cloud removal from satellite images. To test the effect of incorporating rotation equivariance, we closely follow the baseline architecture of DSen2-CR [6] except for the rotation-equivariance. The input to our neural network is a combination of a SAR image and the corresponding aligned cloudy optical image. We concatenate the two inputs in the channel dimension and feed the result to the network. The network consists of series of rotation-equivariant convolutional blocks, each of which consists of rotation-equivariant group convolutional layers, which are described next.

As mentioned earlier, conventional 2D convolutions are equivariant to the translation group. Recall that in a conventional CNN, given a single-channel input $\mathbf{x}$ and a single filter $\Psi$, the convolution operation[2] is defined as

$$[\mathbf{x} * \Psi](t) = \sum_{\tau \in \mathbb{Z}^2} \mathbf{x}(\tau)\Psi(t - \tau) \tag{1}$$

In this paper, we build our network using convolutions that are equivariant to the $p4$ group. For ease of presentation, we refer to these as rotation-equivariant convolutions. Each element of the $p4$ group is a composition of a translation $\tau$ and rotation $r \in \{0°, 90°, 180°, 270°\}$ acting on a square 2D image grid [12]. We choose this group as it provides a good trade-off between benefits of equivariance and computational complexity for the cloud removal problem, similar to what has been observed in some earlier studies for other applications [13]. The idea of convolution (or cross-correlation) in Equation (1) can be generalized to the $p4$ group. We first "lift" the input to the $p4$ group and then perform group convolutions in the subsequent layers. The lifting convolution in the first layer is given by

$$\mathbf{x}^{(1)} = [\mathbf{x} * \Psi^{(1)}](g) = \sum_{\tau \in \mathbb{Z}^2} \mathbf{x}(\tau)\Psi^{(1)}(g^{-1}\tau). \tag{2}$$

[1] $\rho_g(\mathbf{x})$ and $\rho'_g(f(\mathbf{x}))$ are technically known as group representations.

[2] In implementation, the cross-correlation operation is used instead of convolution, but this is a minor detail.

TABLE I: The proposed rotation-equivariant EquiCR outperforms the baseline CNN DSen2-CR for cloud removal.

| Architecture | Rotation Equivariance? | Inputs | PSNR (dB) ↑ | MAE ↓ | SSIM ↑ | SAM (degrees) ↓ |
|---|---|---|---|---|---|---|
| DSen2-CR | ✗ | SAR + Optical | 31.84 | 0.0193 | 0.9309 | 5.58 |
| DSen2-CR | ✗ | Optical only | 31.08 | 0.0213 | 0.9283 | 6.43 |
| DSen2-CR | ✗ | SAR only | 29.32 | 0.0260 | 0.8815 | 7.50 |
| EquiCR (ours) | ✓ | SAR + Optical | **32.06** | **0.0190** | **0.9351** | **5.31** |
| EquiCR (ours) | ✓ | Optical only | 31.64 | 0.0201 | 0.9313 | 5.91 |
| EquiCR (ours) | ✓ | SAR only | 29.60 | 0.0262 | 0.8808 | 7.42 |

TABLE II: Effect of training dataset size on performance. Full, $10^4$ and $10^3$ refer to the training set size.

| Architecture | Rotation Equivariance? | PSNR (dB) ↑ | | | MAE ↓ | | | SSIM ↑ | | | SAM (degrees) ↓ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Full | $10^3$ | $10^2$ | Full | $10^3$ | $10^2$ | Full | $10^3$ | $10^2$ | Full | $10^3$ | $10^2$ |
| DSen2-CR | ✗ | 31.84 | 29.42 | 26.90 | 0.0193 | **0.0257** | 0.0349 | 0.9309 | 0.8938 | 0.7778 | 5.58 | **7.96** | 11.07 |
| EquiCR (ours) | ✓ | **32.06** | **29.57** | **27.69** | **0.0190** | **0.0257** | **0.0315** | **0.9351** | **0.9127** | **0.8770** | **5.31** | 8.36 | **10.27** |

where $g$ is an element of the $p4$ group. Note that the output of the first layer is now a feature defined on the $p4$ group. The rest of the intermediate layers $l > 1$ are group convolutions mapping feature maps defined on the $p4$ group to other feature maps on $p4$ group. This is given by

$$\mathbf{x}^{(l+1)} = [\mathbf{x}^{(l)} * \Psi^{(l)}](g) = \sum_{h \in p4} \mathbf{x}^{(l)}(h)\Psi^{(l)}(g^{-1}h) \qquad (3)$$

For simplicity, the above equations consider each layer to have a single-channel input and single-channel output, which can be easily generalized to multiple input and output channels. Additionally, pointwise nonlinearities such as Rectified Linear Units (ReLUs), and optionally some normalization layers and residual connections that maintain the required $p4$-equivariance, can also be included in the group convolutional neural network architecture.

Finally, to create an equivariant output given features on the $p4$ group, pooling is performed along the rotation dimension:

$$\mathbf{y}(\tau) = \sum_{r \in \{0,90,180,270\}} \mathbf{x}^{(L)}(\tau, r). \qquad (4)$$

We use the $\ell_1$ loss between the estimated cloud-free image $\hat{\mathbf{y}}$ and the ground-truth $\mathbf{y}$ to train the parameters of the network, using mini-batch gradient descent:

$$\text{Loss} = \frac{1}{B}\sum_{j=1}^{B} ||\mathbf{y}^{(j)} - \hat{\mathbf{y}}^{(j)}||_1 \,, \qquad (5)$$

where $B$ is the number of examples in a batch.

## IV. Experimental details and results

### A. Dataset details

For our experiments, we use the SEN12MSCR dataset [4] and we follow their recommended training, validation, and test splits. The training, validation, and test sets contain 114056, 7176, and 7899 examples, respectively, belonging to 169 different regions of interest around the globe and captured during different seasons and years. Each example in the dataset is a triplet consisting of (a) a SAR image, which is a 2-channel image that consists of $\sigma_0$ backscatter coefficients in VV and VH polarization, (b) a multispectral cloudy optical image, which is a multispectral image with 13 bands ranging from 443 nm to 2190 nm, and (c) a cloud-free multispectral image for the same region. The input and output patches are $256 \times 256$.

### B. Network architecture details

We use the network from DSen2-CR as our baseline method as it is simple to train, effective, and allows for demonstrating the utility of adding rotation-equivariance priors into the network. We change the number of filters in each layer from 256 in DSen2-CR to 156 so as to keep the total number of trainable parameters about the same for the baseline network ($18.9 \times 10^6$ parameters) and the proposed rotation-equivariant network ($18.7 \times 10^6$ parameters). As in DSen2-CR, we use 16 residual convolutional blocks, each with 4 convolutional layers, a residual connection, and $3 \times 3$ filters. We use a batch size of 8 for training the network, an initial learning rate of $3 \times 10^{-4}$, and Adam optimizer, as used in the baseline. All of these hyperparameters are inherited from DSen2-CR and not tuned for our method. We use the `escnn` package to build EquiCR [17], [13].

### C. Metrics to quantify performance

Following [4], we use four commonly used metrics to evaluate the methods. Denoting the multispectral cloudless ground-truth image as $\mathbf{y}$ and denoting the network's estimate as $\hat{\mathbf{y}}$, both of size $N_H \times N_W \times N_C$ pixels, where $N_C$ is the number of multispectral channels. We denote $N = N_H \times N_W \times N_C$. The metrics are defined as follows.

1) Peak Signal-to-Noise Ratio (PSNR):

$$\text{PSNR} = -20 * \log(\sqrt{\frac{1}{N}\sum_{i,j,k}(\mathbf{y}_{i,j,k} - \hat{\mathbf{y}}_{i,j,k})^2}. \qquad (6)$$

2) Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{N}\sum_{i,j,k}|\mathbf{y}_{i,j,k} - \hat{\mathbf{y}}_{i,j,k}|. \qquad (7)$$

3) Structural Similarity Index Measure (SSIM): We use the implementation in `scikit image`[18] that computes

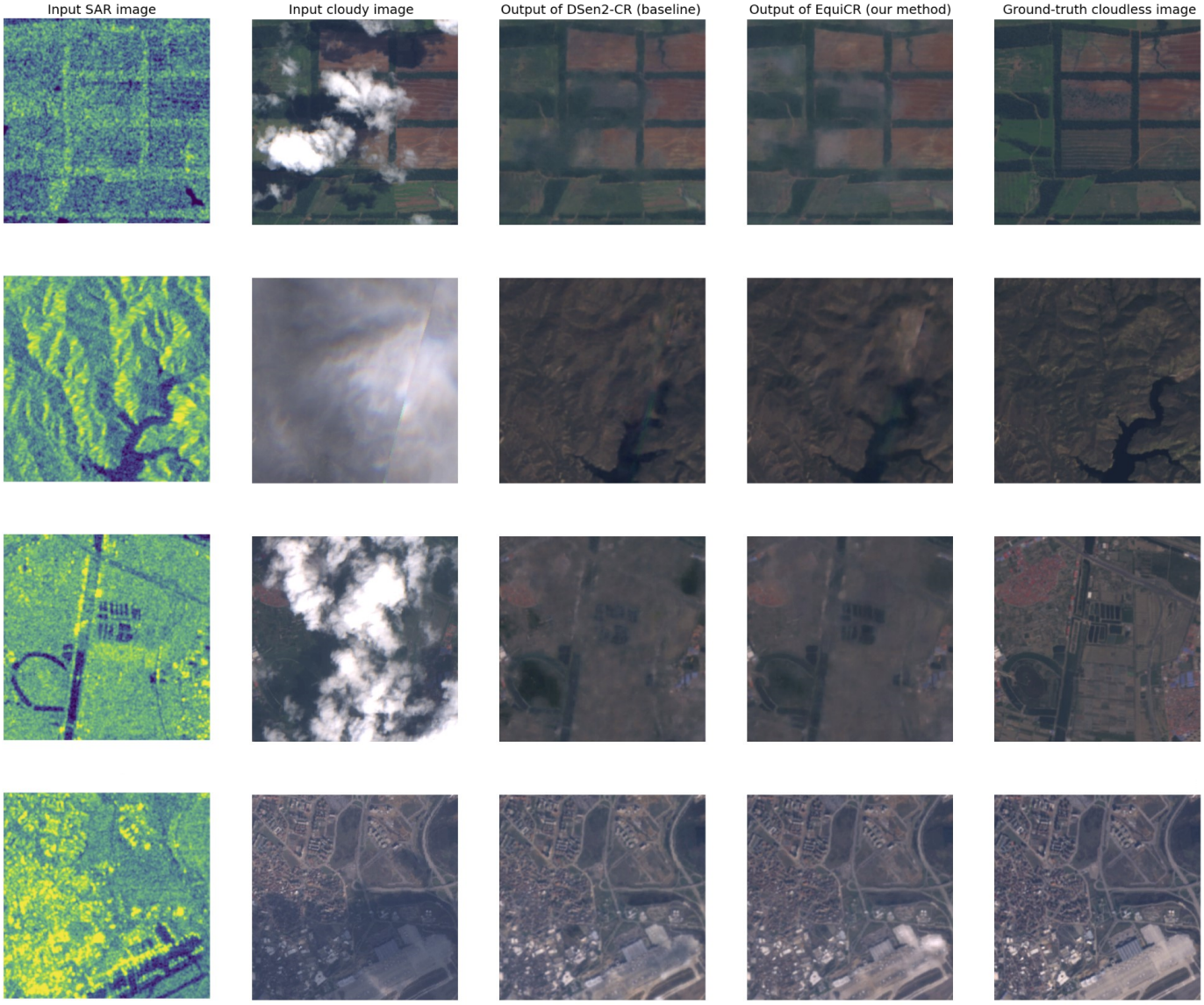


Fig. 2: Some qualitiative results on the SENMS12CR dataset. We see that using the proposed rotation-equivariant EquiCR model, the estimated cloudless image has better quality than the baseline DSEN2-CR output.

SSIM as the average over $M$ window-wise SSIMs as shown below. Windows are indexed by $(h, w)$:

$$\text{SSIM} = \frac{1}{M}\text{SSIM}_{win}(h, w), \tag{8}$$

$$\text{SSIM}_{win}(h, w) = \frac{(2\mu_{\mathbf{y}}\mu_{\hat{\mathbf{y}}} + c_1)(2\sigma_{\mathbf{y}\hat{\mathbf{y}}} + c_2)}{(\mu_{\mathbf{y}} + \mu_{\hat{\mathbf{y}}} + c_1)(\sigma_{\mathbf{y}} + \sigma_{\hat{\mathbf{y}}} + c_2)}. \tag{9}$$

4) Spectral Angle Mapper (SAM)[3]:

$$\text{SAM} = \frac{1}{N_H N_W}\sum_{i=1}^{N_H}\sum_{j=1}^{N_W}\cos^{-1}\left(\frac{\sum_{k=1}^{N_C}\mathbf{y}_{ijk}\hat{\mathbf{y}}_{ijk}}{\sqrt{(\sum_{k=1}^{N_C}\mathbf{y}_{ijk}^2)\odot(\sum_{k=1}^{N_C}\hat{\mathbf{y}}_{ijk}^2)}}\right) \tag{10}$$

[3]The expression for SAM in [4] is not correct. We use the correct expression for SAM based on the code of UnCRtainTS [19].

In these equations, the images have normalized intensities between 0 and 1. In Eq. (9), $\mu_{\mathbf{y}}$ and $\sigma_{\mathbf{y}}$ are the mean and standard deviation of an image window respectively. $c_1$ and $c_2$ are constants to avoid numerical errors in the division operation. In Eq. (10), $\odot$ denotes element-wise multiplication. Both the division in the fraction and $\arccos$ operation are also performed element-wise. The above metrics are computed for each test-set example, and the reported metrics are averaged over the full test set.

### D. Results

We present the main quantitative results in Table I. We observe that using rotation equivariance results in improved performance in terms of all quantitative metrics. Figure 2 shows qualitative comparisons between the reconstructions obtained from the baseline DSen2-CR network those from

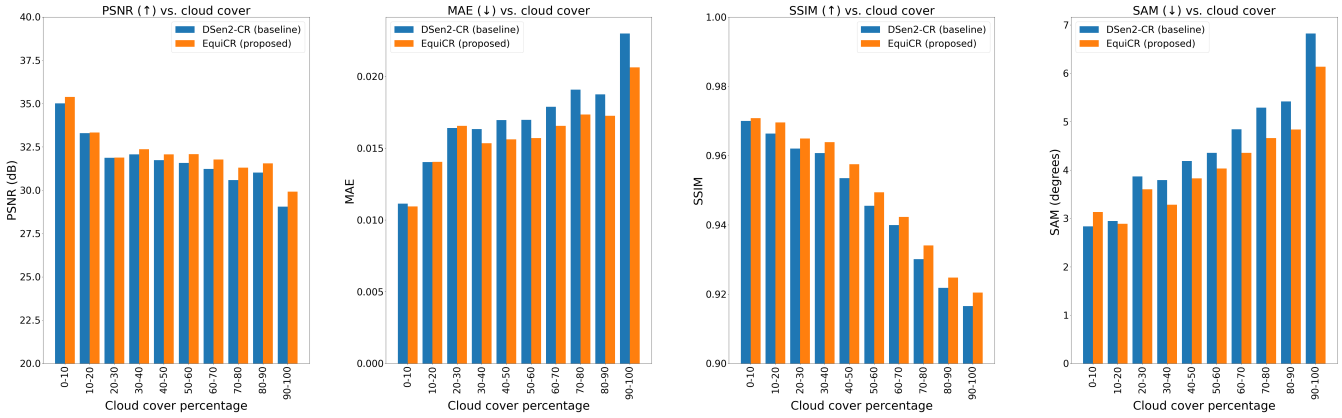


Fig. 3: Performance comparison between the baseline DSen2-CR network and the proposed EquiCR networks as a function of amount of cloud cover in the image. Our method (orange bars) outperforms the baseline (blue bars) generally across all cloud cover percentages. Interestingly, the performance gap generally increases when the amount of cloud cover is higher.

our proposed rotation-equivariant EquiCR by displaying just the red, green, and blue channels of the multispectral images. We generally observe improvements in reconstruction quality, especially when there is a large amount of cloud cover, and the features and edges tend to be sharper with our equivariant network. In some cases, rotation equivariance leads to better reconstruction of some colors and better removal of cloud shadows.

Table I also contains results when only optical images or only SAR images are available as inputs. As expected, using both optical and SAR images yields the best performance. However, the performance improvement of EquiCR over the baseline is greater when only optical images are used as input to the cloud removal networks.

In Table II, we examine the effect of dataset size on the performance of the conventional CNN baseline model and the proposed EquiCR model. We trained three sets of models—using (a) the full dataset, (b) 10000 training examples, or (c) 1000 training examples. We see that when the dataset size is small, the overall performance is lower, but the performance improvement of EquiCR over the baseline is greater when the training dataset size is small.

Finally, we test how the amount of cloud cover affects performance. We use the S2Cloudless cloud detector [20] to detect the binary masks of clouds and compute the percentage of cloud pixels in each of the test set cloudy images. We then compute the median quantitative metrics for 10 different levels of cloud cover, from 0–10% to 90–100%. The results are shown in Figure 3. As expected, the performance of both DSen2-CR and the proposed equivariant model gradually drop as the amount of cloud cover in the image goes up, because the task gets harder with more cloud cover. More interestingly, the rotation-equivariant model provides greater performance improvements over the baseline when the amount of cloud cover is higher.

## V. Conclusion and Future Work

In this paper, we showed that by incorporating the geometric prior that satellite image features do not have a canonical orientation, we can boost performance relative to a conventional neural network. We designed a multimodal rotation equivariant network that produces superior cloud removal results, especially in harder cases that have high cloud cover or few training samples. We believe that applying these ideas to other types of restoration paradigms, such as diffusion models, and to other applications involving satellite images, are important directions for future research.

## References


[1] Michael D King, Steven Platnick, W Paul Menzel, Steven A Ackerman, and Paul A Hubanks, "Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites," *IEEE transactions on geoscience and remote sensing*, vol. 51, no. 7, pp. 3826–3852, 2013.

[2] European Space Agency, "Sentinel overview," https://sentinels.copernicus.eu/web/sentinel/missions, 2016, Available online: (accessed on 20 November 2024).

[3] Michael Schmitt, Lloyd Haydn Hughes, Chunping Qiu, and Xiao Xiang Zhu, "SEN12MS–a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," *arXiv preprint arXiv:1906.07789*, 2019.

[4] Patrick Ebel, Andrea Meraner, Michael Schmitt, and Xiao Xiang Zhu, "Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5866–5878, 2020.

[5] Patrick Ebel, Yajin Xu, Michael Schmitt, and Xiao Xiang Zhu, "SEN12MS-CR-TS: A remote-sensing data set for multimodal multi-temporal cloud removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[6] Andrea Meraner, Patrick Ebel, Xiao Xiang Zhu, and Michael Schmitt, "Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 333–346, 2020.

[7] Xuechao Zou, Kai Li, Junliang Xing, Yu Zhang, Shiying Wang, Lei Jin, and Pin Tao, "Diffcr: A fast conditional diffusion framework for cloud removal from optical satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[8] JD Bermudez, PN Happ, DAB Oliveira, and RQ Feitosa, "SAR to optical image synthesis for cloud removal with generative adversarial networks," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4, pp. 5–11, 2018.

[9] Claas Grohnfeldt, Michael Schmitt, and Xiaoxiang Zhu, "A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from sentinel-2 images," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 1726–1729.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[11] Patrick Ebel, Vivien Sainte Fare Garnot, Michael Schmitt, Jan Dirk Wegner, and Xiao Xiang Zhu, "UnCRtainTS: Uncertainty quantification for cloud removal in optical satellite time series," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 2086–2096.

[12] Taco Cohen and Max Welling, "Group equivariant convolutional networks," in *International conference on machine learning*. PMLR, 2016, pp. 2990–2999.

[13] Maurice Weiler and Gabriele Cesa, "General E(2)-equivariant steerable cnns," *Advances in neural information processing systems*, vol. 32, 2019.

[14] Siamak Ravanbakhsh, Jeff Schneider, and Barnabas Poczos, "Equivariance through parameter-sharing," in *International conference on machine learning*. PMLR, 2017, pp. 2892–2901.

[15] Bastiaan S Veeling, Jasper Linmans, Jim Winkens, Taco Cohen, and Max Welling, "Rotation equivariant CNNs for digital pathology," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*. Springer, 2018, pp. 210–218.

[16] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," *arXiv preprint arXiv:2104.13478*, 2021.

[17] Gabriele Cesa, Leon Lang, and Maurice Weiler, "A program to build E(N)-equivariant steerable CNNs," in *International Conference on Learning Representations*, 2022.

[18] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu, "scikit-image: image processing in python," *PeerJ*, vol. 2, pp. e453, 2014.

[19] Patrick Ebel, Vivien Sainte Fare Garnot, Michael Schmitt, Jan Wegner, and Xiao Xiang Zhu, "Code for UnCRtainTS: Uncertainty Quantification for Cloud Removal in Optical Satellite Time Series," "https://github.com/PatrickTUM/UnCRtainTS", 2023, Available online: (accessed on 20 November 2024).

[20] Anze Zupanc, "Improving cloud detection with machine learning," https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13, 2019, Available online: (accessed on 20 November 2024).