

CloudNet: A Deep Learning Approach for Mitigating Occlusions in Landsat-8 Imagery using Data Coalescence

Paahuni Khandelwal, Samuel Armstrong, Abdul Matin, Shrideep Pallickara, Sangmi Lee Pallickara

Department of Computer Science, Colorado State University, Fort Collins, Colorado, 80521

{Paahuni.Khandelwal, Sam.Armstrong, Abdul.Matin, Shrideep.Pallickara, Sangmi.Pallickara}@colostate.edu

Abstract—Multi-spectral satellite images that remotely sense the Earth’s surface at regular intervals are often contaminated due to occlusion by clouds. Remote sensing imagery captured via satellites, drones, and aircraft has successfully influenced a wide range of fields such as monitoring vegetation health, tracking droughts, and weather forecasting, among others. Researchers studying the Earth’s surface are often hindered while gathering reliable observations due to contaminated reflectance values that are sensitive to thin, thick, and cirrus clouds, as well as their shadows. In this study, we propose a deep learning network architecture, CloudNet, to alleviate cloud-occluded remote sensing imagery captured by Landsat-8 satellite for both visible and non-visible spectral bands. We propose a deep neural network model trained on a distributed storage cluster that leverages historical trends within Landsat-8 imagery while complementing this analysis with high-resolution Sentinel-2 imagery. Our empirical benchmarks profile the efficiency of the CloudNet model with a range of cloud-occluded pixels in the input image. We further compare our CloudNet’s performance with state-of-the-art deep learning approaches such as SpAGAN and Resnet. We propose a novel method, dynamic hierarchical transfer learning, to reduce computational resource requirements while training the model to achieve the desired accuracy. Our model regenerates features of cloudy images with a high PSNR accuracy of 34.28 dB.

Index Terms—remote sensing, deep learning, cloud removal, convolution network

I. INTRODUCTION

The proliferation of remote sensing technology has transformed geosciences. Satellite-based remote sensing has expanded the spatial extents, precision, resolutions, and frequencies at which these remote sensing operations are performed. These satellite systems may either be government-operated or powered by commercial satellite providers. These satellite systems acquire hyperspectral imagery in hundreds of narrow, contiguous spectral bands. The wide frequency range over which these measurements are performed facilitates analysis not only over the visible spectrum but also of atmospheric and terrestrial phenomena outside the visible spectrum.

However, while the number of bands and the resolutions at which remote sensing is performed have improved significantly, scientists still encounter a fundamental problem: cloud occlusion. Regardless of their thickness, clouds corrupt the reflectance signal across all optical frequency bands and adversely impact sensing behind the clouds. At any given moment, around 70% of the globe is covered with clouds [1]. We compare the number of non-cloudy images covering an area of 4.9km², which were collected over the year 2020 at different sites across CONUS (Contiguous U.S) based on their

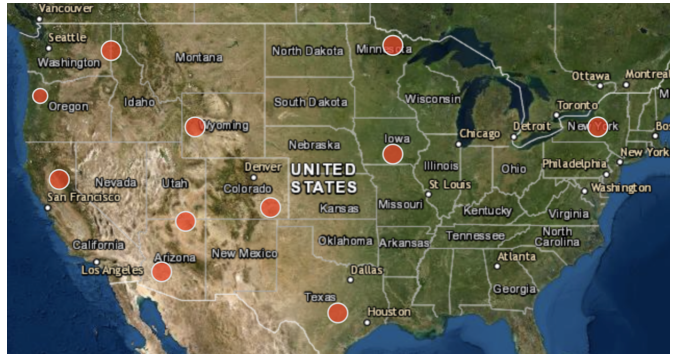


Fig. 1: Distribution of sites across CONUS with varying climatic conditions. The area of 4.9km x 4.9km is analyzed over a year to understand variations in cloud occlusions.

climate conditions. For the climatic conditions such as Dfb (Humid Continental, Mild Summers, Wet Year-Round), Dwb (Humid Continental, Mild Summers, Dry Winters), Dfa (Humid Continental, Hot Summers, Year-Round Precipitation), we observed that more than half of the satellite imagery are occluded with at least 40% cloud coverage.

Generating accurate imputation estimates of cloud-occluded observations is critical to improving the quality of geospatial applications such as precision agriculture. Locations where climate conditions result in high amounts of rainfall account for the majority of these cloud-occluded images, while the climate conditions found in deserts and humid regions account for most of the cloud-free images. Most regions are partially contaminated due to clouds for more than half the year; alleviating cloud occlusions in satellite imagery is crucial for applications that process them.

Reconstruction of information loss has been widely studied in the context of satellite imagery. Early approaches based on interpolation and propagated diffusion methods were hampered by the inability to judge recovery fidelity. More recent approaches leverage spectral bands and temporal aspects [2]–[4] to address this inability. However, these approaches face efficiency challenges stemming from large-area reconstruction and model performance sensitivity to cloud and land coverage types [5]. The deep-learning frameworks explored for this application involve image-to-image translation methods that learn how to map images with partial cloud coverage to cloud-free images. A common practice for these image-to-image translation methods is to leverage time-series relations

to increase their accuracy. However, most methods do not incorporate any trend changes over time and have poor accuracy when applied over large spatial extents. Alongside the temporal sparsity of satellite imagery, time-series modeling is further challenged by the lack of a sufficient number of cloud-free images, especially for the more rain-heavy climactic regions. Models targeting areas over large spatial extents often fail to generate realistic, cloud-free images. This is especially true when areas have large amounts of variability in their climatic conditions and topographical characteristics.

In this study, we present a novel deep learning-based model capable of predicting hyperspectral imagery occluded by clouds. Our model integrates data from different satellite constellations as well as recent historical images to capture seasonal and topographical characteristics. We also include a novel scheme to dynamically orchestrate model training over large spatial extents to improve our model’s accuracy, while also efficiently utilizing computational resources. Our dynamic training scheme supports parallel and preferential training of model instances resulting in improved accuracy and increased spatial coverage. Our CloudNet model’s training is further improved by our underlying distributed storage system to achieve data parallelism and high-throughput training over a cluster of machines.

A. Research Questions

The overarching research question guiding this study is: How can we accurately predict multi-spectral imagery over an area obscured with clouds? Within this broader context, we investigate the following research questions -

RQ-1: *How can our predictions accurately capture topographical characteristics, while accounting for short-term seasonal trends?*

RQ-2: *How can we provide high accuracy over large spatial extents that have unique ambient conditions?*

RQ-3: *How can we orchestrate computationally expensive model training over voluminous satellite datasets without compromising model accuracy?*

B. Approach Summary

The methodology we use to alleviate cloud occlusions at scale encompasses two key components: designing a deep neural network architecture and facilitating effective orchestration of model training workloads. We propose a deep neural network architecture, CloudNet, that alleviates cloud occlusions for a fixed-size reference spatial extent, S_E . The size of the spatial extent S_E is calibrated to ensure residency of the deep network — encompassing the layers and tensors that flow inside the network — within the GPU’s memory.

CloudNet falls broadly in the class of *image inpainting* efforts that reconstruct damaged or missing parts of an image. Inpainting techniques have been used in both photo editing and video compression. CloudNet extends this concept to Multi-spectral satellite imagery with visible (RGB) and non-visible bands (Near-Infrared, Thermal, etc) captured by Landsat-8 sensors. Two major impediments to training an inpainting model over multi-spectral imagery include the scarcity of

cloud-free images and the low scanning frequencies of satellite systems. These impediments adversely affect the model’s ability to account for seasonal and weather changes. CloudNet overcomes this by supplementing Landsat-8 imagery with historical data (including those from another satellite system) for the spatial extent under consideration. We complement the CloudNet’s inputted Landsat-8 imagery with historical data from both the Sentinel-2 [6] and Landsat-8 satellites. This historical data corresponds to the inputted image’s temporal and spatial location.

To preserve accuracy, while ensuring good generalization properties, we leverage regularization schemes and curb overfitting. We also train multiple model instances over a distributed cluster of machines that are tuned to different regional variations based on topological and climatic characteristics. Training an all-encompassing, global model introduces not just resource challenges (stemming from GPU memory and computational requirements), but challenges in capturing subtle variations in topology and weather across large spatial extents. Instead, our approach trains an ensemble of model instances over smaller spatial extents, which can be used together to encompass a large spatial extent. To avoid cold-starts when training model instances we have designed a dynamic hierarchical transfer learning scheme (Section IV-B) that identifies effective initializations of model parameters, incrementally refines these model parameters and preserves data locality. We leverage an imbalanced hierarchical tree structure to track our model refinements and their corresponding refinement priority. The objective function used to launch refinement tasks is based on increases in spatial coverage. Our methodology allows preferential training and refinement of model instances while reducing duplicate processes by warm-starting model instances during training.

C. Paper Contributions and Translational Impacts

Our methodology alleviates cloud occlusions in Landsat-8 imagery by leveraging other satellite collections and historical data. Our key contributions include:

- Utilizing diverse satellite collections to accurately extract structural features of topological characteristics and fast-evolving components that are subject to seasonal changes (e.g. the texture and pigmentation of crops in vegetated areas as shown in Fig. 2).
- Developing a novel dynamic hierarchical transfer learning scheme to effectively utilize computational resources and reduce model training times without compromising model accuracy.
- Creating an occlusion mitigation system that can be used to inform crucial insights into cropping systems such as soil and moisture properties, predicted crop yield, future water demands, and cross-band vegetative indices.
- Producing a model that alleviates cloud occlusions in both visible and non-visible bands by regenerating pixels occluded by clouds or cloud shadows. These bands play a crucial role in calculating evapotranspiration, measuring various vegetative indices, detecting snow cover, gauging water content, inferring

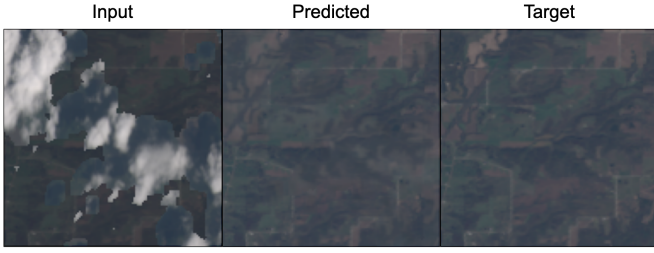


Fig. 2: An example of a Landsat-8 image generated after cloud removal by our model. 54% of the input image is covered by clouds. The predicted image has a PSNR of 36.95 dB.

plant health, quantifying soil quality, and much more.

Translational Impacts. Remotely sensed satellite imagery informs decision-making in several domains including land-use/land-change, agriculture, and environmental tracking. Occlusions caused by clouds adversely impact such decision-making. The core ideas in this study are translatable across other satellite collections and bands to address their occlusion issues. For example, our work could be used to alleviate occlusions in MODIS multi-spectral imagery by utilizing historical data from the Sentinel-2 and MODIS satellites. We also posit that our transfer learning scheme could be used to orchestrate other spatially explicit deep learning workloads.

D. Paper Organization

This paper is organized into six sections. Related work is outlined in Section II. In Section III we describe our experimentation to understand the characteristics of our satellite dataset. Section IV describes our methodology. Our empirical benchmarks and comparison with other models are presented in Section V. Finally, we outline conclusions and future directions in Section VI.

II. RELATED WORK

Removal of contaminated pixels in remote sensing data includes numerical equation-based models, and inpainting methods for recovering missing or cloud-contaminated pixels; more recent work includes deep learning frameworks such as the UNet model, Pix2Pix and Generative Adversarial models, etc. **Numerical Methods.** Among all these techniques [7]–[11] multi-temporal and image pixel replacement approaches were found more effective and widely implemented. In paper [7], cloudy image pixels are reconstructed based on the other cloud-free portion of the sample of the same image. On the other hand, in paper [8], the authors considered multi-temporal images to clone the information for cloudy parts and the shows boast in performance than using a single reference.

Inpainting Methods. Inpainting techniques [12]–[15] fill in pixels values covered by clouds using neighboring pixel information. Maalouf et al. [16] capture the multi-scale structure of clear regions in the single input image using Bandelet transform and multi-scale grouping. This transformation method facilitates a better understanding of the geometry of non-contaminated regions and propagating structure details to the contaminated region in the image. Inpainting methods

based on the neighborhood of the contaminated region have been explored in [17]. First, using some statistical measures cloudy pixels are detected. Next, regeneration of cloudy pixels is performed by diffusing information from the nearby clear regions using differential partial equations and nearest-neighbor interpolation. A common problem faced while using inpainting with spatial learning is the over-smoothing of reconstructed regions. Such models are highly dependent on similarity in structures and land cover types in the input cloudy image. Increasing cloud coverage and variations in land cover in the input image contribute to a drop in accuracy. Further, such models often fail to scale with increases in the area-of-interest or temporal coverage. Meng et al. [18] perform patch-by-patch recovery, via a sparse feature dictionary learning from the clear region and assigning priorities to selected patches.

Machine Learning. Another area of cloud removal models includes convolution neural networks (CNN) and spatial attention models. Such ML-based architectures are highly robust and encompass image-to-image translation methods [2]–[4], [19]–[25] that often involve fusion of clean images from other satellites for predictions. The most recent works are based on Generative Adversarial Network (GAN) architectures which involves two neural networks competing against each other. Wang et al. [26] propose the U-net generator model, where the encoder extracts high-level features using convolutional layers from the input cloudy image while downsampling the image size. The model is trained using SSIM index as the loss function, where for thick and thin clouds the discriminator generates a single loss value and loss for cloud-covered patches respectively. Gao et al. [27] further fuse the synthetic image generated by a CNN with SAR (Synthetic Aperture Radar) imagery as input and cloud-free spectral image as a target. Such models work well when input images have low cloud occlusions. With increasing occlusions, the model performance drops considerably. With thick clouds, where no partial information can be extracted from under the cloudy region, the generator model fills the region with unrealistic and blurred patches, resulting in dropped accuracy. Li et al. [28] used the GAN in a physical model for cloud distortion which considers cloud absorption while regenerating the image. Cloud-GAN [29] is adapted from the Least Square GAN for stable training. The model consists of two generator and discriminator models trained with cycle consistency loss. One generator maps cloudy to cloud-free distribution and the other one translates the generated cloud-free image back to the input cloudy image. This model suffers from over-smoothing in imputed images and highly distorted images with large cloudy regions.

Uzgent et al. [30], leverage the temporal frequency of satellite imagery and feed images captured on different timestamps as is (with clouds) to the Resnet and Encoder-Decoder model separately. This helps in capturing cloud-covered region’s spectral features from the temporally proximate images at the location that have different cloud distributions. The architecture is computationally expensive and may require considerable resource overheads when training for a large area

Band Name	Landsat-8		Sentinel-2	
	Resolution (m)	Wavelength (nm)	Resolution (m)	Wavelength (nm)
Coastal Aerosols	30	430-450	60	430-450
Blue	30	450-510	10	448-546
Red	30	530-590	10	538-583
Green	30	640-670	10	646-684
NIR	30	850-880	10	763-908
SWIR 1	30	1570-1650	20	1542-1685
SWIR 2	30	2110-2290	20	2081-2323
TIR 1	30	10600-11190	-	-
TIR 2	30	11500-12510	-	-

TABLE I: The spatial resolution and wavelength range for each Landsat-8 and Sentinel-2 band.

of interest. The approach works when images are sensed at high temporal frequencies. However, with satellites such as Landsat-8 (which is freely available), with temporal frequencies of 16 days, this poses challenges where large occlusions mean data may not be available for a month. The model outputs may not be a true representation of the crop life cycle at that point in time. CloudNet provides highly-accurate cloud-free satellite images for large spatial extents while capturing fast-evolving temporal features. CloudNet incorporates Sentinel-2 imagery to reconstruct cloud-occluded regions on unseen data for better generalization and to accurately capture temporal features.

Data management is a key consideration prior to training models. These could be based on clouds [31], data sketches [32], or distributed hash tables [33]. Our methodology leverages geohash-based partitioning of spatiotemporal data.

III. DATASET

The CloudNet model is trained on images captured by the widely used Landsat-8 satellite [34], which is equipped with an Operational Land Imager and thermal sensors. The Landsat-8's sensors are a multispectral optical system orbiting the Earth in a near-polar orbit at an elevation of 705km. Landsat-8 satellite captures the earth's land surface at 15m-100m spatial resolution every 16 days. Because of this low-temporal frequency and data corruption (caused by satellite malfunctions and cloud occlusions), getting a clear image every 16 days can be difficult.

To overcome this lack of images, we coalesce Landsat-8 data with Sentinel-2 [6] data, which has a higher spatial resolution ($10m^2$) and orbits the Earth's surface every 5-6 days. As shown in Table I, Sentinel-2 only captures bands in the Coastal Aerosol, RGB, Near-Infrared, and Short-Wave Infrared ranges. Both satellites have spatial resolutions that are approximately the same and their spectral bands have identical wavelength ranges. For our dataset, we use Landsat-8 images captured between the years 2015 and 2019.

The Landsat-8's visible and non-visible bands are accompanied by a 16-bit QA (Quality Image) band, which describes atmospheric and surface conditions for each pixel. Clouds and cloud shadows can be identified using the QA band. The inputted cloudy images for model training are generated by overlaying pixels identified as clouds from cloudy images onto cloud-free images (Fig. 3). We do this instead of relying on satellites to capture cloudy images and a temporally proximate

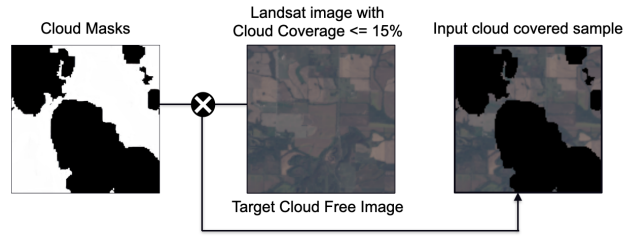


Fig. 3: Our process to generate the dataset corpus for model training and validation.

cloud-free image as the input and target data. This results in highly reliable input and target image pairs and increases the number of training samples to the cartesian product of cloudy and cloud-free images. Utilizing the cloudy image captured by satellite as is requires using the cloud-free image captured at least 15-30 days apart from the input timestamp as the target image which can result in the model not capturing fast-evolving features/trends as dynamics of a region can vary a lot in few days such as land surface temperature changes on an hourly basis.

IV. METHODOLOGY

A. CloudNet System Architecture [RQ-2]

We propose CloudNet to alleviate cloud occlusion in satellite imagery while ensuring high fidelity. Due to cloud occlusions, the temporal frequency of cloud-free Landsat-8 imagery drops considerably; resulting in approximately one cloud-free image per month. This can be quite burdensome for research efforts. However, by recovering corrupted pixel values from cloud-covered areas with inpainting, we can increase the temporal frequency of usable satellite imagery back to Landsat-8's 16-day orbiting time.

CloudNet is a deep neural network that utilizes convolutional-based computation to capture non-linear relationships between pixels from adjacent areas and different spectral bands to precisely regenerate corrupted pixels. To reduce the computational requirements of the model we train multiple model instances over small spatial extents and reduce their training by utilizing the transfer learning method of warm-starting.

Rural regions often have clearly observable variations in their satellite imagery based on agricultural and meteorological conditions. Urban settings, on the other hand, have very little observable changes (except for seasonal changes), due to roads and building structures changing little over time. Similarly, satellite imagery of heavily-forested areas have little variation due to growth in the canopies being quite slow. However, in the case of agricultural farmlands, pixel values vary considerably depending on their location, farming activities (e.g., sowing, irrigation, harvesting), crop cycle, and weather conditions. Training multiple localized models tuned to small spatial extents ensures that the model learns the patterns and trends specific to that particular region.

Fig. 4 provides a visual representation of our model. The

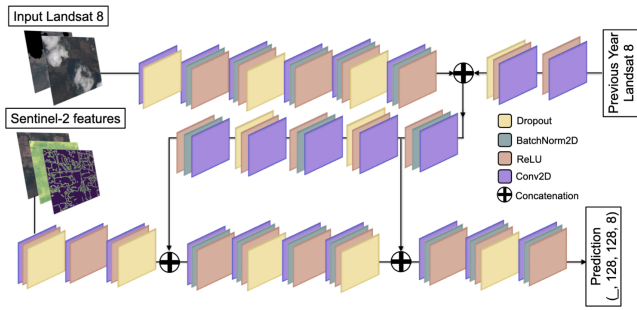


Fig. 4: The architecture of our CloudNet model. During training, the model is inputted a Sentinel-2 image (acquired within 8 days of the target image’s timestamp), edge map, and NDVI index map. The edge map and NDVI index map are derived from Sentinel-2’s spectral bands. Also inputted into the model is historical Landsat-8 data.

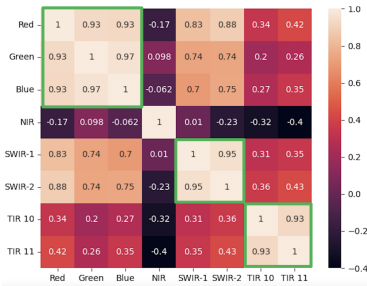


Fig. 5: A Pearson correlation matrix of our dataset’s visible and non-visible spectral bands. Highly-correlated groups can be found within the visible, thermal, and short-wave infrared bands.

inputs to the model are a Sentinel-2 image and a temporally proximate, cloud-occluded Landsat-8 image. Fast evolving features of the agricultural area such as crop condition, plant health, and density of the vegetation, can transform considerably over a few weeks. By pairing the inputted Landsat-8 data with the higher-temporal frequency Sentinel-2 data, we are able to make CloudNet’s predictions more reliable. To enhance the sharpness of edges in the satellite imagery, we first extract edges or groups of pixels that have extreme changes in the RGB bands between the Landsat-8 and Sentinel-2 imagery. We then perform edge enhancement on these extracted features to improve the visibility of roads and boundaries in the image.

The Sentinel-2’s sensors do not detect thermal bands and the thermal bands can not be directly inferred from the other Sentinel-2 bands. This is apparent in Fig. 5 where the thermal bands lack a strong correlation with any of the other non-thermal bands. However, studies have shown there is a strong inverse correlation between the thermal bands and NDVI (Normalized Difference Vegetation Index), especially during the warmer months (May through October) [35]. We leverage this inverse relationship to learn patterns in the thermal infrared bands that are observed over time. We also incorporate changes in the Landsat-8’s thermal bands captured during the same time to better learn these patterns.

Computation Parameters	Model		
	Resnet	SpaGAN	CloudNet
No. of training layers	34	61	33
GPU Memory Usage	3003 MiB	1069 MiB	996 MiB
Time per epoch	423.03 s	147.19 s	61.2 s
No. of model instances	33	96	29

TABLE II: Comparison of computational requirements for Resnet, SpaGAN, and CloudNet model during model training.

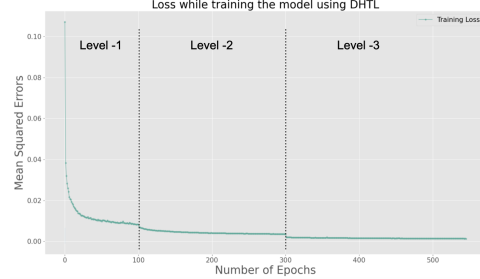


Fig. 6: Errors while training the models using our dynamic hierarchical transfer learning scheme.

The input cloud-occluded image is fed to six blocks of convolutional layers with the ReLU activation followed by a batch normalization layer and a dropout layer present in only alternate blocks. At each block level, we double the number of the output features to capture low-level features from the input images. These extracted features are then concatenated with the historical Landsat-8 image’s low-level features. The merged features are then passed through six more blocks of stacked Conv2D layer, BatchNorm, and dropout layers. At each block, the high-level features of concatenated features are extracted to capture the topological feature and texture of cloud-occluded regions.

The ancillary information from Sentinel-2 sensors are then passed to three such blocks and merged with input Landsat-8 feature maps. The convolutional layers also assist to bring both Sentinel-2 and Landsat-8 images to the same spatial resolution. The dropout layer is a regularization method applied to prevent overfitting of the model on training samples via sparse activation of output neurons. Using dropouts has shown to allow the model in generalizing well over unseen testing samples [36]. As the model becomes denser, we add a skip connection between the 9th and 21st convolution block, to mitigate the issue of vanishing gradients. We maintain the image’s spatial extent at 128pixels² and limit the number of skip connections in the model for faster learning and low GPU memory consumption. We train multiple model instances of CloudNet each specializing over smaller spatial extents. We have designed a dynamic hierarchical transfer learning scheme (see Section IV-B), that lowers the computational footprint involved in training CloudNet instances simultaneously across multiple cores and machines.

CloudNet is trained and tuned across spatial extents based on our hierarchical transfer learning scheme (Section IV-B) to avoid cold starts during training while refining multiple model instances for different spatial extents. In Fig. 6, we report the training errors while specializing from a global model to

the regional model for a refinement depth of 3 as the model converges at each level. Storing multiple training weights, each focused on predicting spectral values in a certain area, helps in generating high-accuracy, cloud-free images even when clouds are present in 80-100% of the input image. Table II summarizes the overall training computation requirements to achieve target accuracy for CloudNet and other state-of-the-art models.

Multiple CloudNet and SpAGAN model instances can be fitted into the GPU memory simultaneously. We train the model instances at each hierarchical level until they coverage and show no further accuracy improvements. The next level of training begins at a reduced learning rate (by a factor of 0.1) while specializing in the smaller spatial extent. CloudNet requires 29 model instances for training approximately 60K images to achieve the desired accuracy. Each SpAGAN model instance is trained on 700 images, as suggested by the authors, however, resulting in a high model instance count. On other hand, we locally train model instances on each machine for Resnet resulting in 33 models in total.

Loss Method. The cloud removal training is focused towards regions that are occluded by clouds and cloud shadows. Currently, the majority of optical satellite sensors are coupled with additional quality assessments (QA) bands, or cloud masks to identify the pixels as cloud or cloud-free with the confidence percentage. Therefore, identifying clouds and generating cloud masks, allows us to leave the clear spatial region as is and focus on cloud-occluded parts in an image. This is achieved through the loss method based on the Cloud-Adaptive Regularized Loss [21]. The training loss function is applied to the model-generated image and target image coupled with a 2D cloud mask map. The loss function is a weighted score of MAE between the cloudy pixels and MSE between clean pixels across all bands identified using the cloud mask. The following equation sums up our training loss function:

$$Loss = \sum (|Predicted - Target| * scale_factor * cloud_mask) + \sum |Predicted - Target|^2$$

Here, the cloud mask is represented by the value of 1 and 0 for cloudy and non-cloudy pixels respectively. We penalize the absolute errors between the actual occluded pixels by scaling using a fixed scaling factor. The MAE (Mean Absolute error) further helps in dealing with outliers and is widely used because of its robustness. We combine this loss with MSE between the complete predicted and target image, where larger errors are penalized more. We let the gradient weights be trained by learning patterns in the clean neighboring region on the given input image while focusing on corrupted areas.

B. Dynamic Hierarchical Transfer Learning [RQ-3]

Our methodology targets reducing the computational requirements for model training. Training a model from the ground-up for every spatial extent, S_E , introduces computation tractability challenges. Once a model is trained (and meets

performance thresholds), that instance can be used to alleviate cloud occlusion for all spatial extents, of size S_E , within the larger encompassing spatial extent. Our methodology identifies when a model is applicable for larger spatial extents and when they must be refined and targeted for smaller spatial extents. Further, when creating and training new model instances, our scheduling mechanism targets spatial coverage as the objective function that should be maximized. Our transfer learning algorithm can be leveraged for large-scale distributed training of any spatially explicit machine learning model.

We start by partitioning the spatial extent under consideration, the continental U.S. or CONUS, into a set of contiguous and non-overlapping spatial extents, S_i . The CloudNet model is then trained for each spatial extent S_i . We use the notation, $M(S_i)$, to represent a model trained with data from region S_i ; each data item used during model training and validation is an image representing spatial extents of size S_E . Note that the input to our CloudNet model includes Landsat-8 and Sentinel-2 images for spatial extent S_E . If the model $M(S_i)$ does not meet our performance thresholds, we hierarchically partition the region into subregions and train models for each subregion. A larger spatial extent S_i is partitioned into N non-overlapping, equal-sized spatial extents each of which is represented as S_{ij} . In particular, $S_i = \sum_{j=0}^{N-1} S_{ij}$ as shown in Fig. 7a and 7b.

Our hierarchical partitioning scheme preserves spatial contiguousness. For each subregion S_{ij} , we assess the suitability of the parent model. We assess the performance of the model using a validation set with data for S_{ij} if the model $M(S_i)$ satisfies the performance threshold; then model training is skipped for those subregions. We assert that $M(S_{ij}) = M(S_i)$ if the model, $M(S_i)$, is deemed performant for the smaller spatial extent S_{ij} . This accounts for cases, where the model may not generalize for the entire encompassing spatial extent, but would be applicable for other smaller, spatial extents within the larger spatial extent. Model refinements over the smaller spatial extent are performed only when the model for the larger, encompassing spatial extent $M(S_i)$ is deemed non-performant for S_{ij} . Further, these model refinements (retraining while avoiding cold starts) are targeted and performed only for spatial extents where the model is underperforming.

Rather than perform model training using cold starts, subregions where the parent model is not performant use the parent model's weight vectors as the starting condition. This has two advantages. First, because the parent model was trained with data from a much larger spatial extent it has seen data with larger, more diverse variations. Second, the data that the parent model was trained on also includes data from the subregion under consideration. Ultimately, models refined using the parent model's weight vectors as the starting condition benefit from generalization while accounting for subtle regional variations that impact performance. Initial layers of the model are held fixed, while the latter layers are deemed trainable. The models for the subregions are trained using data from that subregion, and the weight vectors in the final layers of the model instance are trained to account for subtle regional variations.

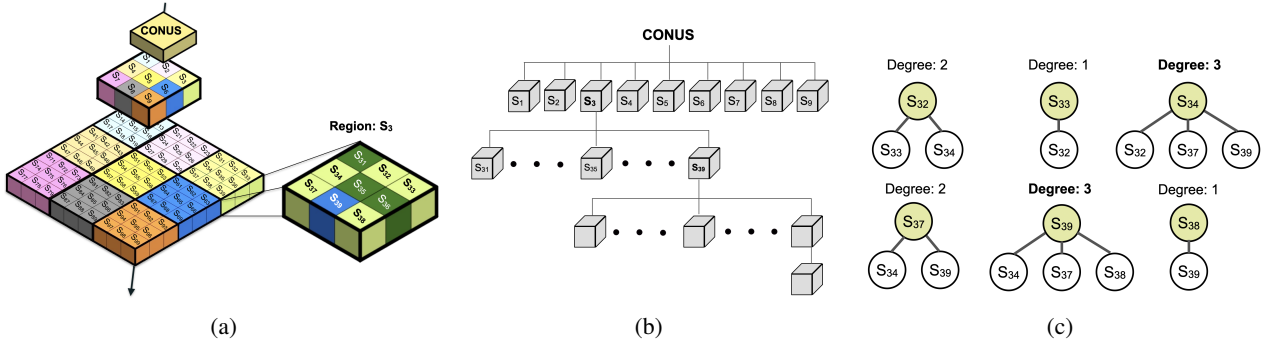


Fig. 7: An overview of the model refinement scheme is CloudNet. Models are incrementally refined to meet the performance objectives. (a) The spatial extent is partitioned into a set of contiguous, non-overlapping spatial extents. Model for S_{39} represents the case where the regional model is performant for the subregion (breaking tie randomly between S_{39} and S_{34}) (b) Depth of a node in the tree represents the degree of refinement for that spatial extent. (c) Within a hierarchically partitioned region, the subregions are organized as a graph and the extent with the highest degree is selected for the model training.

To inform scheduling and refinement of model training we organize spatial extents in an imbalanced, hierarchical tree structure – the refinement tree. We consult the tree to inform the launching of computational tasks to ensure spatial coverage. The tree structure allows us to accurately estimate the spatial extents for which CloudNet model instances are available. Each node maintains the PSNR (Peak Signal-To-Noise Ratio) for the trained models allowing us to identify models that have met their performance threshold and those that have not. The objective function for selecting a spatial extent to construct/refine a model is to maximize the spatial extent for which the model is being built. We consult the refinement tree to perform model refinement for spatial extents where the model is underperforming. These refinements are targeted, hierarchical, and incremental. Our methodology preferentially trains model instances over spatial extents and then refines these models to create new instances for smaller spatial extents. The process is performed recursively for increasingly finer-grained spatial extents till the model is deemed performant. When the model for a spatial extent underperforms, we recursively split the region and repeat the model refinement process. The recursive spatial splits stop once the desired accuracy is reached. Further, the depth of the tree represents the degree of refinements that a model has gone through for a given spatial extent.

To identify the order of modeling tasks within a region, we organize the subregions within a spatial extent as a graph. Each subregion is represented as a vertex in this graph; an edge between two vertices exists only if they have a shared boundary. Once such a graph is constructed, the spatial extent chosen for model refinement is the node with the highest degree (i.e., the number of edges) as shown in Fig. 7c. When a model is being trained for the vertex, its adjacent vertices (representing neighboring subregions with shared boundaries) are not considered for training. Once the model is trained, we assess the performance of this sibling model at neighboring subregions (adjacent vertices in the graph). If the model is deemed performant, then that model is used. If not, a check is made to see which of the two

models (parent or sibling) had better performance for the particular subregion. We use the model that performed better as the initial starting condition for subregions for refining the model for that subregion. We improve our training cost by isolating independent model instances to train simultaneously on multiple cores and machines. All the models at the same depth in the tree with different parent models or all the underperforming siblings models with the same parent model are trained simultaneously on different machines or separate cores in the same machine while preserving data locality. The model instances that meet our specified PSNR accuracy are utilized to perform predictions on unseen data.

Our methodology allows us to: (1) preferentially train models that maximize spatial coverage while identifying spatial extents where models underperform and must be refined. (2) Leveraging parameters from trained models (either parents or siblings) allows us to minimize duplicate processing to converge faster with fewer epochs. Cumulatively, these allow us to ensure spatial coverage for CloudNet while ensuring frugal utilization of resources.

V. EVALUATION

We perform several experiments to assess the performance of our proposed methodology. The experiments were performed in a cluster of 33 machines (Xeon E5-2620, 64 GB Memory); machines also had a Quadro P2200 GPU (5GB of memory with 1280 cores). CloudNet was designed and implemented using Tensorflow-GPU 2.4.1. The distributed training of the model was performed using Horovod, which is based upon the ring-all reduce strategy [37]. Horovod uses NVIDIA’s NCCL library to perform optimized weights averaging across machines. The strategy ensures that each node in the cluster performs pairwise communications with only two other nodes significantly reducing network communications in the cluster.

A. Study Region

The focus of this work is recovering the corrupted spectral values over agricultural areas, by reconciling complex crop patterns, plant lifecycle, and irrigation requirements over time. We perform our experiments over a spatial extent encompass-

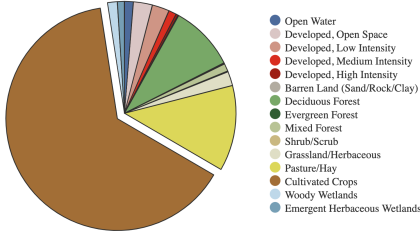


Fig. 8: Breakdown of the study region IOWA for our model training based on Landcover types. We target agricultural farmlands for our study which covers roughly 60% of Iowa.

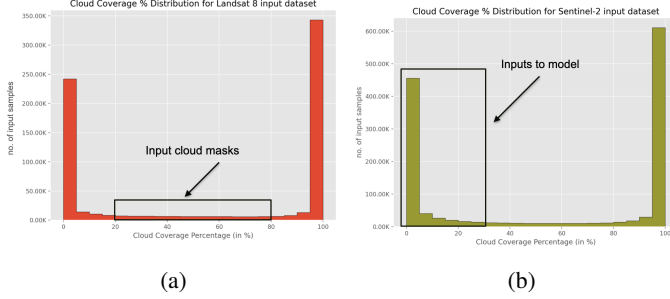


Fig. 9: Distribution of input training dataset for Landsat-8 and Sentinel-2 based on cloud coverage.

ing the state of Iowa with total spatial coverage of approximately $451.7\text{km} \times 332.8\text{km}$. In Fig. 8, we break down our study area based upon the topological features based on National Land Cover Database (NLCD) which provides data on the land cover at a 30m resolution [5]. The model is also exposed to other varieties of land covers such as developed urban regions and water bodies. Our model trains using Landsat-8 images captured in the months when crops are grown and harvested (May to November). The input dataset, comprising cloudy Landsat-8 images, is approximately 50 GB in size. This is supplemented with auxiliary datasets consisting of the historical Landsat-8 imagery which is also 50 GB and Sentinel-2 imagery with a total size of 125 GB. The model instances are trained using 225 GB of the dataset with a total of 62,788 input samples. Fig. 9a and 9b shows the distribution of our input training dataset based on their cloud-occlusion percentage. More than 40% of our input data is occluded by clouds; the degree of occlusion is at least 20% of the spatial region. We consider images with the degree of cloud occlusions between 20-80% when training the CloudNet model. Similarly, for fusing Sentinel-2 inputs, images with a degree of cloud occlusions less than 30% are considered.

Landsat-8’s sensors capture the entire Earth in $185\text{km} \times 180\text{km}$ scenes. However, training machine learning models using these tiles is infeasible due to the accompanying memory size and residency requirements. To efficiently manage this voluminous dataset, we utilize a data partitioning scheme based on geohashes. Leveraging the distributed hash table (DHT), the framework achieves a load-balanced distribution for fast retrieval of the input samples. The proposed data-storage system partitions satellite imagery by breaking down input images along geohash boundaries [38].

The geohash algorithm is a hierarchical spatial indexing scheme where regions are defined by character strings, and the region is partitioned into subregions by reducing the spatial extent of the tile and increasing the length of the geohash string. Each spatial extent is distributed based on their respective geohash key. This controls the number of geohash characters to consider while distributing tiles. Data distributing in this fashion ensures distribution based on the spatial proximity of the image, i.e., geocodes spatially closer to each other are co-located on the same machine. If we were to use a 3-character geocode to inform distribution in the DHT; for the area associated with geocode “9xp”, all the encapsulated spatial regions such “9xpdw” or “9xpmq” will reside on the same machine in the cluster. Because the partitioning scheme is deterministic and independent of the satellite system, the input files are co-located as well, regardless of the satellite system, avoiding data movements when co-processing imagery from Sentinel-2 and Landsat-8 sources.

B. Vegetative Indices

Vegetative indices are functions, expressed over combinations of spectral bands that are formulated to accentuate the vegetative properties of a spatial region. Indices that are highly sensitive to plant biomass are commonly utilized in scheduling crop irrigation, predicting crop evapotranspiration rates, and monitoring droughts among others. These vegetative indices primarily depend on high-resolution reflectance values captured by sensors in the visible and infrared wavelengths. However, due to occlusion by clouds, the vegetation index value is significantly hampered. For example, the NDVI over high-density vegetation areas generally ranges between 0.6 to 1, and small positive values close to 0 (typically indicating dead plants or bare soils). However, in the presence of clouds or water, the NDVI value of a pixel drops below 0.

$$GCI = \left(\frac{NIR}{Green} \right) - 1 \quad (1)$$

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} \quad (2)$$

$$ARVI = \frac{NIR - (2 * Red) + Blue}{NIR + (2 * Red) + Blue} \quad (3)$$

$$NDSI = \frac{Green - SWIR}{Green + SWIR} \quad (4)$$

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (5)$$

The scatter plot (in Fig. 9) depicts the different vegetative indices calculated using model-generated reflectance values and actual cloud-free pixels over 8,000+ different spatial locations. We calculate (1) GCI - Green Chlorophyll Index (eq.1) used to monitor plant health by calculating the chlorophyll content in leaves, (2) NDMI - Normalized Difference Vegetation Index (eq.2) that is used to estimate moisture level to monitor droughts and fuel-level in fire areas, (3) ARVI -

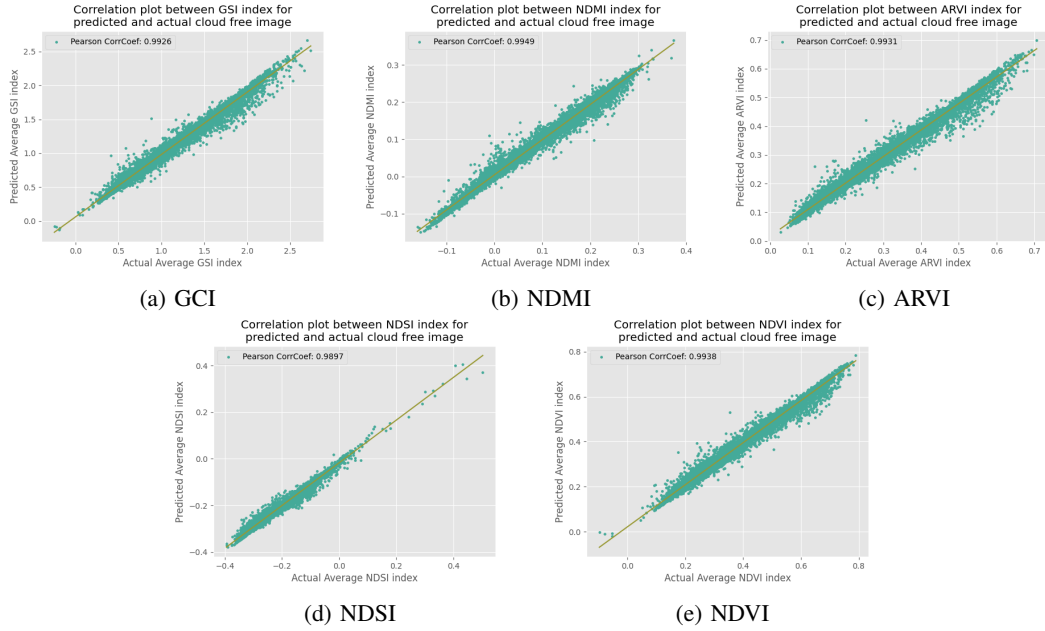


Fig. 10: Scatter plot contrasting the model imputed cloud-free vegetative index value (average over input image) and the actual cloud-free average vegetative index value. We utilize different spectral band combinations to capture the correlation between the model generated and the remotely sensed reflectance values.

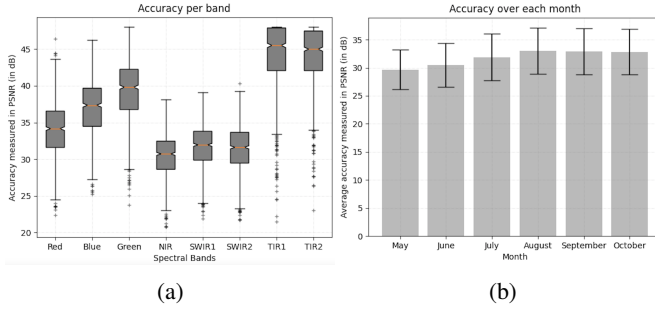


Fig. 11: (a) CloudNet testing errors for each band. NIR, SWIR1, and SWIR2 bands have the lowest PSNR of 31 dB. Thermal band predictions are highly accurate with a PSNR of 45 dB. (b) Monthly accuracy of CloudNet.

Atmospherically resistant vegetation index (eq.3) for robust monitoring of vegetation area which are prone to aerosols, (4) NDSI - Normalized Difference Snow Index (eq.4) for snow detection and, (5) NDVI - Normalized Difference Vegetation Index (eq.5) vegetative index. The Pearson correlation coefficient higher than 0.98 for all five indices shows a strong linear relationship between the predicted and target values and demonstrates that model imputed pixels (using CloudNet) are highly reliable for real-life applications and can precisely mitigate the clouds.

C. Analyzing CloudNet across spectral bands

Next, we benchmark model performance across different spectral bands (Fig. 11a). Among the visible spectra, the red band has the lowest PSNR (a measure of image quality) accuracy of 34 dB, whereas the green and blue bands perform well with a PSNR accuracy of more than 35 dB. To improve

model performance in thermal bands, we incorporate the NDVI index map generated from Sentinel-2 images due to its inverse relationship. CloudNet precisely captures this relation and shows the highest PSNR of 45 dB for both thermal bands. The drop in performance is accounted for primarily by low accuracy in infrared bands, where accuracy drops by 22.5% from other bands. The quality band in Landsat-8 images occasionally fails to identify the cloud-occluded pixels, resulting in the misclassification of cloudy pixels as clean pixels. Such pixels results in significantly low accuracy as the model predicts cloud-free pixels and in the target image same pixel is cloud-covered.

In Fig. 11b, we break down model performance on a monthly basis. Across months, the crop life cycle evolves from seeding to harvesting. In Iowa, the planting of corn and soybean crops occurs during the late weeks of May and crops start blooming by the end of August. This affects the change in density of vegetation across months. We observe an improving performance of the model in the early months after which the accuracy becomes consistent from August to October.

D. Comparison with SpAGAN and Resnet architecture

We compare CloudNet's performance with state-of-the-art cloud removal models such as Resnet [21] and SpAGAN [39]. The aforementioned models are trained on the same input samples with 20-80% cloud coverage.

Resnet, a deep residual neural network [21], was developed to remove cloud occlusions from Sentinel-2 satellite images. The network consists of skip connections between double or triple layers. The Resnet model comprises residual blocks containing a 2D convolution layer, followed by a ReLU activation function, a 2D convolution layer, and a residual scaling

Image Quality Index	Model		
	Resnet	SpAGAN	CloudNet (Ours)
PSNR (in dB)	31.885	39.8	34.28
PSNR-B (in dB)	31.803	16.219	34.59
RASE	183.59	390.62	144.83

TABLE III: Comparison of performance by Resnet, SpAGAN, and CloudNet model using different image quality indices.

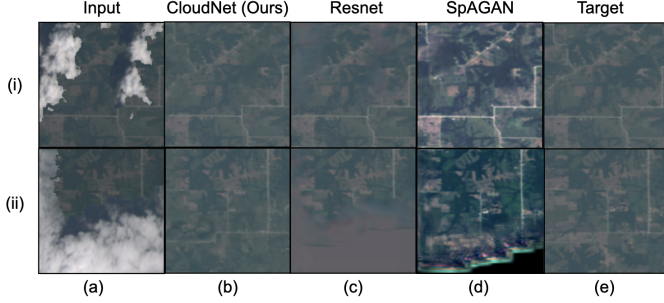


Fig. 12: Representative examples of a cloud-free image generated by CloudNet, Resnet, and SpAGAN respectively. Here, PSNR-B is calculated over all bands (i) Cloud Coverage:38.44% PSNR-B: 36.42 dB, 29.35 dB, 28.78 dB (ii) Cloud Coverage:72% PSNR: 28.63 dB, 24.98 dB, 22.275 dB.

layer. The authors propose a custom loss function called Cloud-Adaptive Regularized Loss to optimize prediction while preserving the cloud-free region of the input image and making the model learn only the cloud-covered mappings.

The Spatial Attention GAN (SpAGAN) [39] is an image-to-image translation neural network composed of generator and discriminator networks. The generator uses a Resnet architecture with skip connections between sequential layers with 17 Resnet blocks and four spatial attention mappers. These spatial attention mappers convolutional layers and weighted matrices to create an attention matrix that focuses on cloud-occluded areas of the image. To focus on these cloudy areas the SpAGAN loss function takes the MSE of the attention matrix and a matrix where cloudy areas have a value of one while leaving the cloud-free areas unchanged.

Table III, summarizes the average performance of the aforementioned models on the Red, Blue, Green, NIR, SWIR1, SWIR2, and thermal bands. We measure PSNR, PSNR-B (Block Sensitive Peak Signal-to-Noise Ratio), and RASE (Relative Average Squared Error). The SpAGAN model fails to generate a high-quality cloud-free image on multiple bands resulting in a PSNR-B accuracy of 16.21 dB. Using RASE and PSNR-B, we can measure image similarity on the block level, considering the blurriness and over-smoothness artifacts. These artifacts have the most adverse impact on SpAGAN, while CloudNet and Resnet show reasonable accuracy and reduced errors. Our model, CloudNet, achieves an overall PSNR-B accuracy of 34.59 dB.

E. Variation in the percentage of cloud-occluded region

To measure the sensitivity of the aforementioned models to the degree of cloud-occlusion, we measure the performance on input samples in RGB bands with 20-40%, 40-60%, and 60-80% clouds (Fig. 12). To measure performance, we calculate

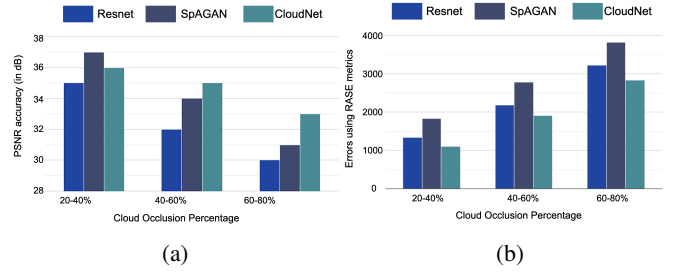


Fig. 13: Comparing CloudNet performance with Resnet and SpAGAN models with increasing cloud-occluded regions (a) Accuracy is measured in terms of PSNR (b) Testing errors are measured in terms of RASE.

PSNR which is commonly used image quality measure. However, PSNR metrics do not translate well with block artifacts and blurriness in the image, therefore as a fair comparison, we report RASE, which is sensitive towards artifacts.

Clouds can be classified based on their altitude [40]. Some can be classified as cirrus clouds which are thin strands at high altitudes, dense and cauliflower-shaped cumulus clouds that are found in warmer regions at low altitudes, and altocumulus clouds present in middle altitudes and characterized by globular masses. We report model imputed images after cloud removal in Fig. 12 on various cloud types and cloud coverage. SpAGAN outperforms CloudNet on RGB bands when clouds are scattered and sparsely spread out on RGB bands. The cloud-free regions are well preserved by all the models. However, we observe an over-smoothing in the cloud-occluded area generated by Resnet. This is in Fig. 13a as well, for 20-40% cloud coverage; Resnet has the lowest PSNR of 35.2 dB. With increasing cloud coverage the accuracy of SpAGAN, Resnet, and CloudNet drops by 14.5%, 16.981% and, 11.66% respectively. This is expected due to the majority of the area being covered by clouds; however, CloudNet is able to preserve the underlying topological features in areas with dense clouds with comparably lower blurriness in the image. The accuracy is maintained at a PSNR of 33.2 dB, while RASE errors in CloudNet are then the lowest compared to other models.

VI. CONCLUSION

We described our methodology to mitigate cloud occlusions in remotely sensed satellite imagery by leveraging historical variations in Landsat-8 and Sentinel-2 satellite systems that have different spatial and temporal resolutions. Our methodology encompasses the deep network, hierarchical transfer learning, and scalable orchestration of model training workloads.

RQ-1 CloudNet captures the spatial characteristics from historical imagery and also utilizes high-frequency scans from Sentinel-2 for incorporating short-term seasonal trends. By leveraging data from different satellite systems for the same spatial extent, the model-generated images can capture the topological features of the cloud-occluded spatial region.

RQ-2 Rather than build an all-encompassing model, CloudNet is designed for smaller spatial extents. Having multiple

model instances, each calibrated to smaller spatial extents, allows the framework to account for subtle regional variations and preserve accuracy.

RQ-3 Training deep learning models is computationally intensive involving the tuning of a large number of parameters. Our dynamic hierarchical transfer learning scheme prevents cold-starts (and the associated computational costs) during model training. Additionally, the scheme also reduces the number of model instances that need to be trained. Our DHT-based scheme facilitates data locality during model training.

VII. ACKNOWLEDGEMENT

This research was supported by the National Science Foundation [OAC-1931363, ACI-1553685], the National Institute of Food and Agriculture [COL0-FACT-2019], and a Jack Cochran Family Professorship.

REFERENCES

- [1] M. D. King et al. Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites. *IEEE transactions on geoscience and remote sensing*, 51(7):3826–3852, 2013.
- [2] Y. Chen et al. Thick clouds removing from multitemporal landsat images using spatiotemporal neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2020.
- [3] Q. Zhang et al. Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:148–160, 2020.
- [4] N.-B. Chang et al. Smart information reconstruction via time-space-spectrum continuum for cloud removal in satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(5):1898–1912, 2015.
- [5] C. G. Homer et al. The national land cover database. Technical report, Reston, VA, 2012.
- [6] M. Drusch et al. Sentinel-2: Esa’s optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, 120:25–36, 2012.
- [7] Q. Cheng et al. Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal mrf model. *ISPRS journal of photogrammetry and remote sensing*, 92:54–68, 2014.
- [8] C.-H. Lin et al. Cloud removal from multitemporal satellite images using information cloning. *IEEE transactions on geoscience and remote sensing*, 51(1):232–241, 2012.
- [9] D.-C. Tseng et al. Automatic cloud removal from multi-temporal spot images. *Applied Mathematics and Computation*, 205(2):584–600, 2008.
- [10] E. H. Helmer and B. Ruefenacht. Cloud-free satellite image mosaics with regression trees and histogram matching. *Photogrammetric Engineering & Remote Sensing*, 71(9):1079–1089, 2005.
- [11] M. Li et al. Producing cloud free and cloud-shadow free mosaic from cloudy ikonos images. In *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477)*, volume 6, pp. 3946–3948. Ieee, 2003.
- [12] L. Lorenzi et al. Inpainting strategies for reconstruction of missing data in vhr images. *IEEE Geoscience and remote sensing letters*, 8(5):914–918, 2011.
- [13] Q. Cheng et al. Inpainting for remotely sensed images with a multichannel nonlocal total variation model. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1):175–187, 2013.
- [14] Z. Xu and J. Sun. Image inpainting by patch propagation using patch sparsity. *IEEE transactions on image processing*, 19(5):1153–1165, 2010.
- [15] C. Zhang et al. Gaps-fill of slc-off landsat etm+ satellite image using a geostatistical approach. *International Journal of Remote Sensing*, 28(22):5103–5122, 2007.
- [16] A. Maalouf et al. A bandelet-based inpainting technique for clouds removal from remotely sensed images. *IEEE transactions on geoscience and remote sensing*, 47(7):2363–2371, 2009.
- [17] A. C. Siravenha et al. Evaluating inpainting methods to the satellite images clouds and shadows removing. In *International Conference on Signal Processing, Image Processing, and Pattern Recognition*, pp. 56–65. Springer, 2011.
- [18] F. Meng et al. A sparse dictionary learning-based adaptive patch inpainting method for thick clouds removal from high-spatial resolution remote sensing imagery. *Sensors*, 17(9):2130, 2017.
- [19] Q. Zhang et al. Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8):4274–4288, 2018.
- [20] W. Li et al. Thin cloud removal with residual symmetrical concatenation network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 153:137–150, 2019.
- [21] A. Meraner et al. Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:333–346, 2020.
- [22] S. Ji et al. Simultaneous cloud detection and removal from bitemporal remote sensing images using cascade convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1):732–748, 2020.
- [23] Y. Zi et al. Thin cloud removal for multispectral remote sensing images using convolutional neural networks combined with an imaging model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:3811–3823, 2021.
- [24] K. Bruhwyler et al. Lightweight, embeddings based storage and model construction over satellite data collections. In *2020 IEEE International Conference on Big Data (Big Data)*, pp. 246–255. IEEE, 2020.
- [25] P. Khandelwal et al. Mind the gap: Generating imputations for satellite data collections at myriad spatiotemporal scopes. In *2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, pp. 92–102. IEEE, 2021.
- [26] X. Wang et al. Thin and thick cloud removal on remote sensing image by conditional generative adversarial network. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1426–1429. IEEE, 2019.
- [27] J. Gao et al. Cloud removal with fusion of high resolution optical and sar images using generative adversarial networks. *Remote Sensing*, 12(1):191, 2020.
- [28] J. Li et al. Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:373–389, 2020.
- [29] P. Singh and N. Komodakis. Cloud-gan: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1772–1775. IEEE, 2018.
- [30] B. U. UzKent et al. Cloud removal in satellite images using spatiotemporal generative networks. 2019.
- [31] S. L. Pallickara et al. Scientific data management in the cloud: A survey of technologies, approaches and challenges. *Handbook of Cloud Computing*, pp. 517–533, 2010.
- [32] T. Buddhika et al. Synopsis: A distributed sketch over voluminous spatiotemporal observational streams. *IEEE Transactions on Knowledge and Data Engineering*, 29(11):2552–2566, 2017.
- [33] M. Malensek et al. Evaluating geospatial geometry and proximity queries using distributed hash tables. *Computing in Science & Engineering*, 16(4):53–61, 2014.
- [34] B. L. Markham et al. Landsat 8: status and on-orbit performance. In *Sensors, Systems, and Next-Generation Satellites XIX*, volume 9639, pp. 60–67. SPIE, 2015.
- [35] D. Kumar and S. Shekhar. Statistical analysis of land surface temperature-vegetation indexes relationship through thermal remote sensing. *Ecotoxicology and environmental safety*, 121:39–44, 2015.
- [36] N. Srivastava et al. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [37] A. Sergeev and M. Del Balso. Horovod: fast and easy distributed deep learning in tensorflow. *arXiv preprint arXiv:1802.05799*, 2018.
- [38] G. Niemyer. Geohash. <http://www.geohash.org/>, 1999.
- [39] H. Pan. Cloud removal for remote sensing imagery via spatial attention generative adversarial network. *arXiv preprint arXiv:2009.13015*, 2020.
- [40] W. Menzel et al. Improved cloud motion wind vector and altitude assignment using vas. *Journal of Applied Meteorology and Climatology*, 22(3):377–384, 1983.