Structural Representation-Guided GAN for Cloud Removal

Implementation in PyTorch based on IEEE GRSL 2025

Tools Used: PyTorch • CUDA • VS Code • SEN12MS-CR Dataset

The Imperative of Cloud Removal

Satellite imagery is a cornerstone for numerous applications, yet its utility is frequently hampered by pervasive cloud cover. Clouds obscure crucial ground information, severely impacting the reliability and completeness of remote sensing data.

This challenge affects critical domains:

- **Disaster Monitoring:** Rapid assessment of flood, fire, or earthquake damage.
- **Agriculture & Crop Evaluation:** Accurate yield prediction and disease detection.
- **Urban Mapping:** Up-to-date infrastructure planning and change detection.
- **Climate & Environmental Studies:** Monitoring deforestation, ice caps, and ocean health.

The goal is to accurately recover the underlying terrain, providing a clear, unobstructed view of the Earth's surface.

Challenges in Cloud Removal:

- Thick clouds completely hide ground truth, making reconstruction difficult.
- Traditional interpolation methods often produce blurring or unrealistic artifacts.
- Maintaining structural consistency (e.g., roads, rivers) under reconstruction is complex.

Our objective is to develop a robust method to remove clouds and reconstruct high-quality, structurally correct images.

SEN12MS-CR: The Foundation Dataset

To train and validate our cloud removal model, we utilize the **SEN12MS-CR dataset**, a specialized collection designed for this exact purpose.

Sentinel-2 Based

Derived from Sentinel-2 satellite imagery, ensuring consistency with real-world remote sensing data.

Paired Imagery

Consists of synchronized cloudy and cloud-free image pairs, providing direct ground truth for supervised learning.

Multi-Spectral Bands

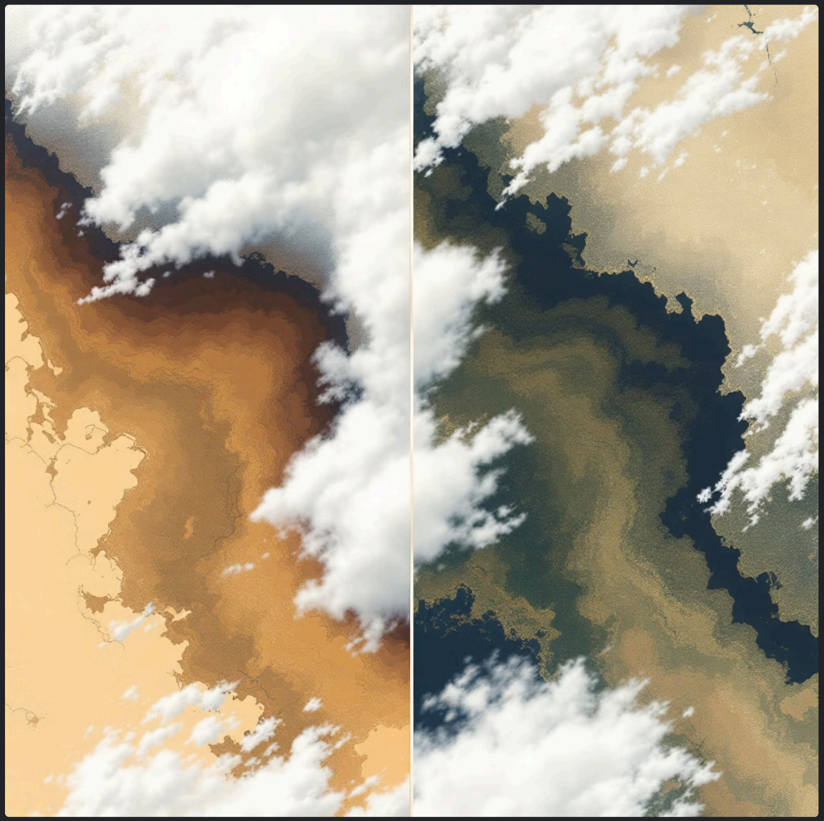
Includes various spectral bands (beyond just RGB), crucial for comprehensive land cover analysis and robust reconstruction.

Temporally Aligned

Scenes are temporally aligned, meaning cloudy and clear images capture the same location at slightly different, yet close, times to ensure minimal land cover change.

Dataset Split & Properties:

- **Training Pairs:** 58 high-quality paired images.
- **Validation Pairs:** 12 pairs for model tuning and performance monitoring.
- **Test Pairs:** 14 unseen pairs for final evaluation.
- **Image Resolution:** All images are uniformly **256 × 256** pixels, suitable for efficient deep learning training.
- **Cloud Variations:** Contains a diverse range of cloud types, from thin wisps to thick, dense formations, making the model robust.



SRG-GAN: Structural Representation-Guided GAN

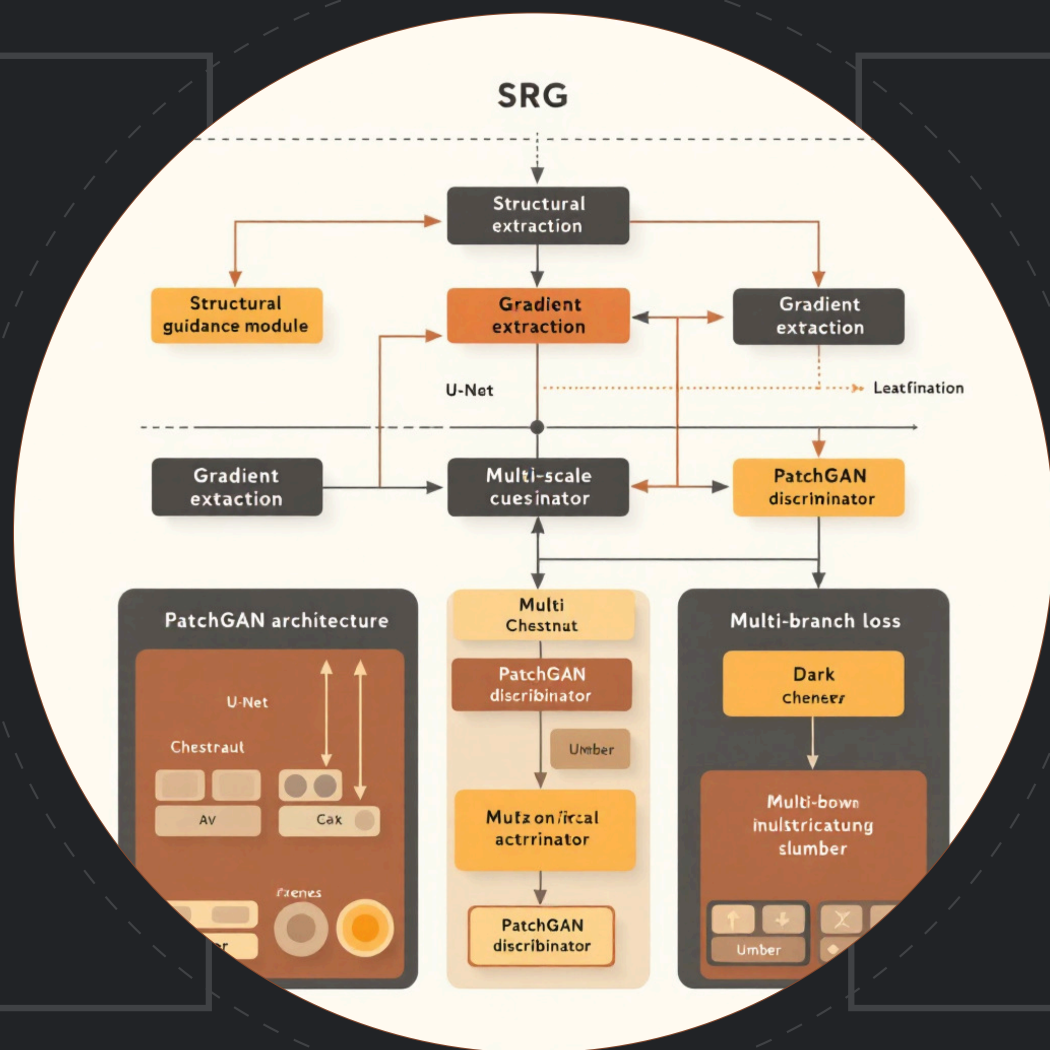
Our approach leverages a Generative Adversarial Network (GAN) enhanced with structural guidance to achieve superior cloud removal. The **Structural Representation-Guided GAN (SRG-GAN)** is designed to not only remove clouds but also preserve and reconstruct fine ground details.

Generator

U-Net backbone producing clear images

Structural Guidance

Gradient extraction and multi-scale cues



Discriminator

PatchGAN assessing local realism

Loss Network

Multi-branch losses on outputs

Key Components:

- **Generator (U-Net Backbone):** A powerful convolutional network responsible for generating the cloud-free image. Its U-Net architecture is ideal for image-to-image translation tasks due to its ability to capture both local and global context.
- **Structural Guidance Module:** This innovative component extracts critical structural information, such as gradients and edge maps, from the input image. This guidance is then fed into the Generator, helping it to reconstruct realistic textures and preserve geometric details.
- **PatchGAN Discriminator:** A discriminator that operates on image patches rather than the entire image. This encourages the generator to produce high-frequency details and coherent local structures, making the output images more realistic.
- **Multi-branch Loss Network:** A sophisticated loss function combining various components to guide the training process effectively, ensuring both pixel-level accuracy and structural integrity.

Why Structural Guidance is Crucial:

Integrating structural cues directly into the generation process significantly improves detail retention, preserves terrain edges, and enhances reconstruction accuracy, especially under conditions of thick cloud cover where traditional methods falter.

Multi-faceted Loss Functions for Robust Reconstruction

The success of SRG-GAN hinges on a comprehensive suite of loss functions, each targeting a specific aspect of image quality and structural integrity. By combining these losses, we ensure the model learns to produce highly accurate, visually pleasing, and structurally consistent cloud-free images.

L1 Reconstruction Loss

Measures pixel-wise absolute difference between generated and ground truth images, ensuring basic pixel accuracy.

Perceptual Loss

Utilizes features extracted from a pre-trained VGG-16 network to compare high-level content similarity, focusing on human-perceived quality.

Style Loss

Based on Gram matrices from VGG-16 features, this loss encourages the generated image to match the texture and style of the ground truth.

Structural Loss

Specifically designed to maintain the integrity of geographical features and patterns.

Gradient Consistency Loss

Enforces similarity in image gradients between the generated and target images, crucial for sharp edges and fine details.

Adversarial GAN Loss

The core GAN loss drives the generator to produce images that are indistinguishable from real cloud-free images to the discriminator.

Benefits: This multi-loss approach allows for superior texture recovery, robust structure preservation, and significantly reduced artifacts, leading to a more realistic and actionable terrain appearance in the reconstructed images.

Training Pipeline: Precision and Monitoring

Our training process is meticulously configured to ensure optimal model performance and reproducibility, leveraging powerful tools and methodologies to achieve robust cloud removal capabilities.

Training Setup

- **Epochs:** 100 cycles to allow for thorough learning.
- **Batch Size:** 4 images per batch, balancing memory usage and training stability.
- **Image Size:** 256 × 256 pixels, consistent with the dataset.
- **Optimizer:** Adam, chosen for its adaptive learning rate capabilities.
- **Learning Rate:** `1e-4`, carefully tuned for convergence.
- **Hardware:** Dedicated CUDA GPU acceleration for efficient computation.

Key Training Features

- **Cloud Synthesis:** Incorporates both real cloudy samples and synthetically generated cloud patterns to enhance model generalization.
- **Checkpoint Saving:** Regularly saves model weights, allowing for recovery from interruptions and selection of the best performing model.
- **TensorBoard Visualizations:** Comprehensive logging of loss curves, image outputs, and other metrics for real-time performance monitoring and debugging.
- **Validation Every Epoch:** Evaluates model performance on a separate validation set after each epoch, preventing overfitting and guiding hyperparameter tuning.

This rigorous pipeline ensures the model is trained effectively, producing consistent and high-quality results for cloud removal.

Quantitative Performance: Demonstrating Accuracy

Objective metrics provide a clear measure of the SRG-GAN's effectiveness on the test set, affirming its capability to reconstruct cloud-free images with high fidelity and structural integrity.

30.14dB

PSNR (Peak Signal-to-Noise Ratio)

Indicates accurate pixel-level reconstruction with minimal noise.

0.809

SSIM (Structural Similarity Index)

Confirms excellent preservation of structural information and perceptual quality.

0.799

Correlation Coefficient (CC)

Demonstrates a strong linear relationship between the generated and ground truth images.

0.064

RMSE (Root Mean Square Error)

Quantifies the average magnitude of the error, indicating low prediction variance.

Interpretation of Results:

- High PSNR (30.14 dB): Points to a very accurate pixel-by-pixel reconstruction, signifying low distortion in the output.
- High SSIM (0.809): Suggests that the model effectively maintains the structural similarity and perceptual quality of the original scene, making the reconstructed images visually compelling.
- Low RMSE (0.064): Reflects minimal error between the generated image and the ground truth, validating the model's predictive accuracy.
- CC close to 1 (0.799): A strong correlation coefficient indicates that the model's output closely matches the ground truth in terms of overall patterns and distributions.

These metrics collectively affirm that the SRG-GAN provides both numerically accurate and perceptually pleasing cloud removal.

Qualitative Results: Exemplary Performance

Visual inspection of the model's output demonstrates its remarkable capability to restore obscured terrain, revealing high detail and structural integrity that aligns closely with ground truth.

Cloudy Input

The original satellite image, partially or heavily obscured by cloud cover.

SRG-GAN Prediction

The model's reconstructed image, revealing the underlying terrain.

Ground Truth

The actual cloud-free image of the region for comparison.

These examples highlight the SRG-GAN's strengths:

- **High Detail Recovery:** Fine features like river networks, roads, and land divisions are effectively reconstructed.
- **Clean Cloud Removal:** Clouds are removed without leaving noticeable artifacts or unnatural blurs.
- **Terrain Structure Preservation:** The overall landscape morphology and object contours are maintained, leading to visually coherent results.
- **Quantitative Alignment:** The good cases typically correspond to high SSIM and PSNR values (>30 dB), confirming their accuracy.

The visual resemblance between the prediction and ground truth is compelling, underscoring the model's efficacy.

Qualitative Results: Addressing Challenging Cases

While SRG-GAN demonstrates strong performance, extremely dense cloud cover or complex atmospheric conditions present inherent limitations. These challenging cases highlight areas for future improvement and underscore the complexity of cloud removal.

Cloudy Input (Extreme)

An input image with very thick and noisy clouds, providing minimal ground information.

SRG-GAN Prediction

The model's attempt at reconstruction, showing a plausible but smoothed terrain.

Ground Truth

The clear ground truth image, revealing intricate details not fully recovered.

Observations from challenging scenarios:

- When ground information is nearly absent due to **extremely thick or noisy clouds**, the model struggles to accurately recover fine details.
- The model tends to produce a **smooth, plausible reconstruction** in these areas, rather than introducing artifacts. However, some fine structures, such as small buildings or narrow pathways, may be generalized or completely missed.
- This limitation is common across even the most advanced cloud-removal models, as they cannot infer information that is fundamentally missing from the input.

Potential Areas for Improvement:

- **More Diverse Training Data:** Expanding the dataset to include a wider variety of extreme cloud conditions and geographical features.
- **Enhanced Structural Priors:** Integrating more sophisticated prior knowledge about ground structures into the model.
- **Cloud-Shadow Modeling:** Explicitly accounting for cloud shadows, which can further complicate reconstruction.
- **Higher-Capacity Networks:** Exploring architectures with greater representational power to capture more complex patterns.

Implementation Details: Code Structure and Ecosystem

The entire SRG-GAN framework is implemented in PyTorch, ensuring flexibility, scalability, and integration with modern deep learning practices. Our structured codebase facilitates reproducibility and future extensions.

`models/`

Contains definitions for the Generator (U-Net) and Discriminator (PatchGAN) architectures.

`losses/`

Houses implementations of all custom loss functions: Perceptual, Style, Structural, and Gradient Consistency losses.

`utils/`

Utility functions for image operations, performance metrics (PSNR, SSIM), and data transformations.

`train.py`

The main script orchestrating the training loop, including data loading, model updates, and validation.

`inference.py`

Script for running the trained model on new cloudy images to generate cloud-free predictions.

`checkpoints/`

Directory for saving trained model weights and optimizer states, allowing for model loading and continuous training.

`evaluate.py`

A dedicated script for calculating and reporting quantitative metrics (PSNR, SSIM, CC, RMSE) on test datasets.

Additional Features for Enhanced Workflow:

- **TorchScript Export Support:** Enables model optimization and deployment to production environments with ease.
- **TensorBoard Monitoring:** Provides interactive dashboards for visualizing training progress, loss curves, and output images.
- **Synthetic Cloud Generation:** Includes scripts to augment the dataset with synthetically generated clouds, improving model robustness to

Thankyou