

# マルチスペクトル cGANs による衛星画像の薄雲除去

榎本 憲二<sup>1,a)</sup> 櫻田 健<sup>1,b)</sup> 王 維民<sup>1,c)</sup> 福井 宏<sup>3,d)</sup> 松岡 昌志<sup>4,e)</sup> 中村 良介<sup>2,f)</sup> 河口 信夫<sup>1,g)</sup>

## 1. はじめに

本論文では、conditional Generative Adversarial Networks (cGANs) をマルチスペクトル画像へと拡張し、可視光の衛星画像から薄雲を除去する手法を提案する。衛星画像を用いて地上の様子を観測することで、交通サービスの改善や迅速な災害対応、環境モニタリングなどが可能となる。そのため、リモートセンシングやコンピュータービジョン、環境科学、気象学など幅広い分野で衛星画像を用いた地物の認識や、土地利用の変化検出、震災の被害把握など、多くの研究が古くから行なわれている。

一方で、衛星画像の多くが雲で覆われているため、地上の様子を安定的に観測することが難しい。雲の影響を低減するためには、波長がより長い帯域の画像を用いることが考えられ、これまでに、合成開口レーダー (SAR) 衛星の画像などに疑似カラーを施して視認性を高める手法などが提案されている [1]。しかし、波長が長くなるほど空間解像度は荒くなり、その見た目は可視光のそれと異なってくる。

そこで、本研究では、マルチスペクトル画像を入力し、雲を除去した可視光画像を生成するネットワーク Multispectral conditional Generative Adversarial Networks (McGANs) と、その学習方法を新たに提案する。具体的には、cGANs の入力をマルチスペクトル画像用に多チャンネルへと拡張し、入力には雲を合成した画像を、教師データには雲のない画像を与えることでネットワークを学習する。その際、雲のみを分離した画像を同時に教師データとして与えることで雲検出を同時に行い、雲除去の精度向上を図る。さらに、雲を合成した画像と実際に雲が写った画像をともに色補正することで、撮影条件やセンサーデータの処理方法に

よる色調の違いを吸収する。

また、衛星画像には森や海が多く写っており、実際の衛星画像から一様にサンプリングすると学習データに偏りが生じる。そのため、t-Distributed Stochastic Neighbor Embedding (t-SNE) [2] を用いてその偏りを緩和する。そして、この可視光と近赤外線の 4 バンド画像のデータセットを用いて提案手法を評価し有効性を確認する。

## 2. 関連研究

リモートセンシングの分野では、雲の影響を受けず地上を観測するためにマイクロ波を用いることが一般的であり、航空機や人工衛星の場合は、マイクロ波の低解像度を補うために合成開口レーダー (SAR) 衛星の画像を用いることがある。しかし、SAR 画像は可視光画像と比べて解像度が低く、見た目も直感的に理解しにくいものとなる。そのため、SAR 画像を直接色付けすることで視認性を向上させる手法も提案されている [1]。

コンピュータービジョンの分野では、可視光画像に写る霧を除去する手法がいくつか提案されており、RGB 画像のみを用いる手法 [3], [4] と、RGB 画像と近赤外線画像の両方を用いる手法 [5], [6] がある。前者の手法では、霧が写っていない画像に関する色情報の事前知識や仮定を用いることで、後者の手法では、可視光画像に透過性の高い近赤外線画像をガイドとして統合するなどして霧を除去している。

また、本研究に最も関連が深い研究として Generative Adversarial Networks (GANs) [7] を用いた手法が提案されている。GANs は、画像を生成する Generator と、Generator が生成した画像と教師画像を区別する Discriminator の 2 つのネットワークから構成されている。GANs の Generator は、Discriminator が教師画像と区別できないような画像を生成するよう学習するため高品質な画像を生成できる。しかし、GANs の学習は不安定であるため、Deep Convolutional Neural Network で用いられる畳み込み層と Batch Normalization を導入して学習の不安定性を解消した Deep Convolutional Generative Adversarial Networks (DCGANs) [9] が提案されている。さらに、GANs の学習に入力情報を追加して条件付ける conditional Genera-

<sup>1</sup> 名古屋大学

<sup>2</sup> 産業技術総合研究所

<sup>3</sup> 中部大学

<sup>4</sup> 東京工業大学

a) enoken@ucl.nuee.nagoya-u.ac.jp

b) sakurada@nagoya-u.jp

c) weimin@ucl.nuee.nagoya-u.ac.jp

d) fhiro@vision.cs.chubu.ac.jp

e) matsuoka.m.ab@m.titech.ac.jp

f) r.nakamura@aist.go.jp

g) kawaguti@nagoya-u.jp

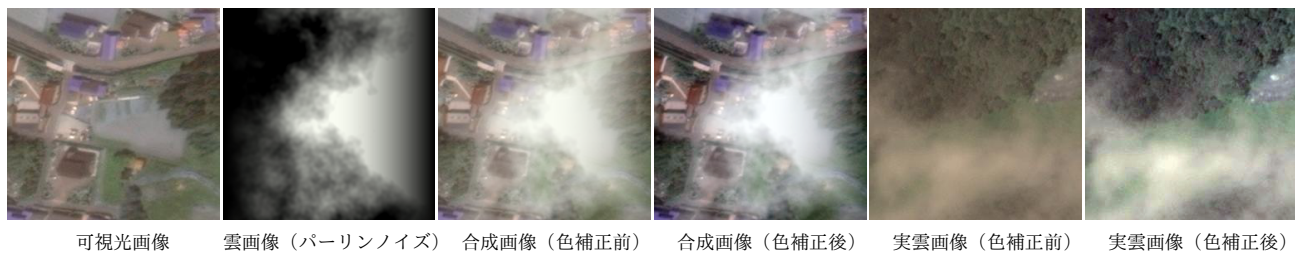


図 1 雲の合成と色補正の例. 左から可視光画像, パーリンノイズで生成した雲画像, 可視光画像と雲画像をアルファブレンドで合成した画像, 合成画像の色補正後. 右二つは異なるエリアを撮影した実際の雲画像の色補正前後. 色補正により合成画像と実際の雲画像の色調を揃える.

tive Adversarial Networks (cGANs) も提案されている [8].

cGANs と DCGANs を用いた画像生成の研究は幅広く行われており, 傷の修復や, 雨や雪のような特定物体の除去 [10], [11] などに応用されている. 特に [12] では, U-Net [13] をベースとした Generator と PatchGAN [14] の Discriminator で構成された cGANs を提案しており汎用的に高品質な画像を生成できる. U-Net の Generator は Encoder と Decoder の層間にスキップ結合を導入することで, Encoder の畳み込み層で欠落する空間的特徴を Decoder の各層に伝える. また, PatchGAN の Discriminator は, 教師画像と生成画像のパッチを複数区別することで, 高周波な構造をモデル化することができる. 一方で, これらの cGANs を用いた手法は可視光画像のみを生成モデルに入力して画像を生成しており, 遮蔽等で見えない領域を周辺情報から尤もらしく予測しているだけとも考えることができる.

これらの研究を踏まえて, 本研究では, 可視光画像とより長い波長領域の画像を組み合わせ, 可視光画像の雲を除去するネットワーク (McGANs) を構築する. 雲で見えない領域をその周辺情報から予測するだけでなく, 雲を透過する波長帯域の画像と統合することで, 異なる波長帯域の画像間で整合性が取れた可視光画像を生成する. SAR 画像のような雲を完全に透過するマイクロ波の画像も含めたネットワークを実現することを最終目標とし, 本稿では, まず, 可視光画像と, その次に波長が長く空間解像度が高い近赤外線画像を用いて衛星画像の薄雲除去を行う.

### 3. 雲除去用衛星画像データセットの構築

本研究では, 地球観測衛星 WorldView-2 から撮影された衛星画像を使用する. 可視光画像, 近赤外線画像ともに解像度は  $20,000 \times 20,000$ , 画像分解能は  $0.5 \text{ [m/pixel]}$  である. 実際に学習で用いる衛星画像は比較的雲がかかっておらず, 主に都市部が写っている 8 枚を選択した. 各衛星画像は未計測領域を含んでおり, それらを除いた  $256 \times 256$  の画像を切り出し, 約 37,000 枚の学習用画像を取得した.

#### 3.1 雲画像の合成

衛星画像から雲を除去するネットワークを学習するためには, 入力として雲が写っている画像を, 教師データとして同じ場所の雲が写っていない画像が必要である. しかし, 異なる時刻に同一エリアを撮影しても, 照明などの撮影条件や地表の変化によりその見えは大きく異なったものとなる. そこで, 雲がかかっていない画像に雲画像を合成することで学習データを作成する. また, 雲を合成した画像と実際の雲画像の色調の違いを吸収するため両画像ともに色補正を行う.

具体的には, パーリンノイズ [15] を用いて雲画像を生成し, 生成した雲画像と可視光画像をアルファブレンドにより合成する. 図 1 にパーリンノイズを用いた雲の合成と色補正の例を示す. 左から可視光画像, パーリンノイズで生成した雲画像, 可視光画像と雲画像をアルファブレンドで合成した画像, 合成画像の色補正後である. さらに, 右二つの画像は異なるエリアを撮影した実際の雲画像の色補正前後である. 色補正により合成画像と実際の雲画像の色調の違いを緩和していることが分かる.

#### 3.2 t-SNE を用いたデータセットの均一化

地球上の土地被覆には面積的に偏りがあり, 本研究で利用する衛星画像にも森や海が多く写っている. そのため, 実際の衛星画像からランダムに画像をサンプリングすると学習データ内で画像の種類に偏りが生じる. そこで, 特定カテゴリーへのオーバーフィッティングを避けるため t-SNE を用いて学習データの偏りを緩和する.

まず, ImageNet で学習された AlexNet を用いて, 各画像から 4,096 次元の特徴ベクトルを抽出する. 抽出した特徴ベクトルに t-SNE を適用することで, 各画像を 2 次元空間にマッピングする. そして, 2 次元特徴空間から一様に画像を 2,000 枚サンプリングし, 学習データを作成することで, データの偏りを緩和する.

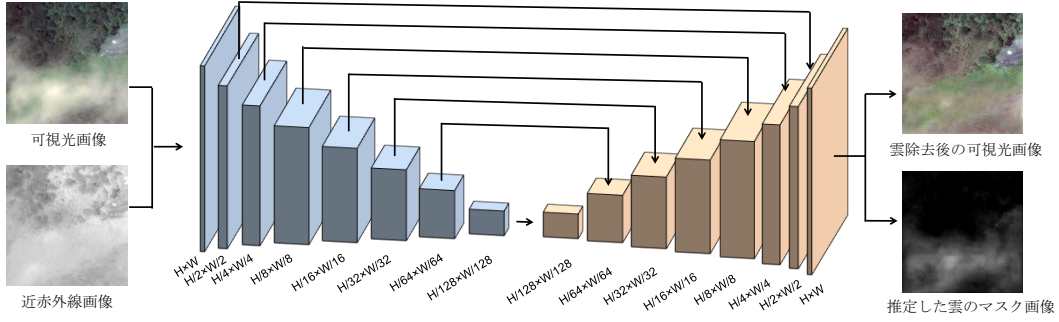


図 2 McGANs の Generator のネットワーク構造

表 1 McGANs の構成

Encoder	Decoder	Discriminator
CR (64, 3, 1)	CBRD (512, 4, 2)	CBR (64, 4, 2)
CBR (128, 4, 2)	CBRD (512, 4, 2)	CBR (128, 4, 2)
CBR (256, 4, 2)	CBRD (512, 4, 2)	CBR (256, 4, 2)
CBR (512, 4, 2)	CBR (512, 4, 2)	CBR (512, 4, 2)
CBR (512, 4, 2)	CBR (256, 4, 2)	C (1, 3, 1)
CBR (512, 4, 2)	CBR (128, 4, 2)	
CBR (512, 4, 2)	CBR (64, 4, 2)	
CBR (512, 4, 2)	C (4, 3, 1)	

#### 4. Multispectral conditional Generative Adversarial Networks (McGANs)

本論文では、可視光画像とより長い波長領域の画像を統合し、可視光画像上の雲を除去するために、cGANs の入力をマルチスペクトル画像へと拡張した Multispectral conditional Generative Adversarial Networks (McGANs) を提案する。ネットワーク構成を図 2 と表 1 に示す。[12] で提案された cGANs をベースに、入力を RGB-NIR の 4 チャンネルに拡張する\*1。さらに、出力を雲除去後の RGB 画像と雲のマスク画像（グレースケール）の計 4 チャンネルへと拡張し、雲のマスク画像を同時に推定することで精度向上を図る。入力の RGB-NIR 画像、出力の RGB 画像と雲のマスク画像は各チャンネルを  $[-1, 1]$  に正規化しネットワークに与える。

##### ネットワーク構造

本実験で使用した McGANs のネットワーク構造の詳細を表 1 に示す。ここで、Convolution, Batch Normalization, ReLU の各レイヤーを C, B, R で、ドロップアウトを含む場合は D で表す。括弧内の数字は、左から畳み込みフィルタの枚数、サイズ、ストライドを表している。また、Encoder と Discriminator で使用する ReLU は全て Leaky ReLU を用いる。

入力のマルチスペクトル画像を  $I_M$ 、教師データである出力の RGB 画像と雲のマスク画像を計 4 チャンネルの画

像として  $I_T$  とすると、Generator の  $L1$  損失関数  $\mathcal{L}_{L1}$  は以下の式で定義される。

$$\mathcal{L}_{L1} = \frac{1}{4HW} \sum_{c=1}^4 \sum_{v=1}^H \sum_{u=1}^W \lambda_c |I_T^{(u,v,c)} - \phi(I_M)^{(u,v,c)}|_1 \quad (1)$$

$\lambda_c$  は出力の各チャンネルの重み\*2を、 $\phi(I_M)$  は入力画像  $I_M$  に対する学習されたネットワークによる推定値を表す。

#### 5. 評価実験

提案手法の有効性を確認するために評価実験を行った。先述の通り、衛星から異なる時刻に同一エリアを撮影しても、照明などの撮影条件や地表の変化によりその見えは大きく異なったものとなるため、雲で遮蔽された領域の正解データを取得することは難しい。そのため、本実験では、可視光と近赤外線の衛星画像に提案手法である McGANs を適用することで、可視光画像の雲を除去し視認性が向上しているかを確認する。

学習には 3 節で作成した雲のない可視光画像、パーリンノイズで生成した雲画像、雲を合成した可視光画像、近赤外線画像を 2,000 組用いた。雲のない可視光画像と雲を合成した画像、さらに、テスト時に入力する実際に雲が写った可視光画像に対して色補正を行う。ミニバッチ数を 1、エポック数 500 に設定した。

さらに、近赤外線画像のみから可視光画像を推定するネットワーク (NIRcGANs)、可視光画像のみから可視光画像を推定するネットワーク (RGBcGANs) と比較することで、マルチスペクトル画像の有効性を確認する。NIRcGANs では、学習データとして McGANs のトレーニングデータの中から、雲のない可視光画像と近赤外線画像のみを利用し、ネットワークの入力として近赤外線画像を、教師データとして雲のない可視光画像を与える。RGBcGANs では、学習データとして McGANs のトレーニングデータの中から、雲を合成した可視光画像と雲のない可視光画像のみを利用し、ネットワークの入力として雲を合成した可視光画像を、教師データとして雲のない可視光画像を与える。McGANs と同様に、ミニバッチ数を 1、エポック数を 500

\*1 遠赤外線やマイクロ波など、他の波長領域の画像を追加する場合は、このチャンネル数を増やして拡張する。

\*2 本論文の実験では全て  $\lambda_c = 1$  に固定した。



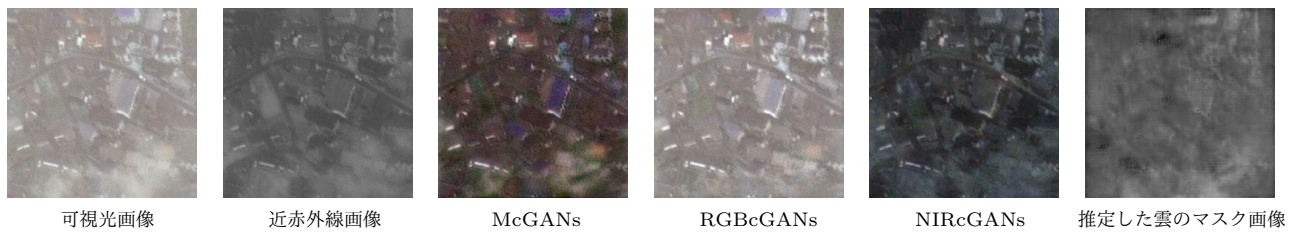


図 3 雲除去の例

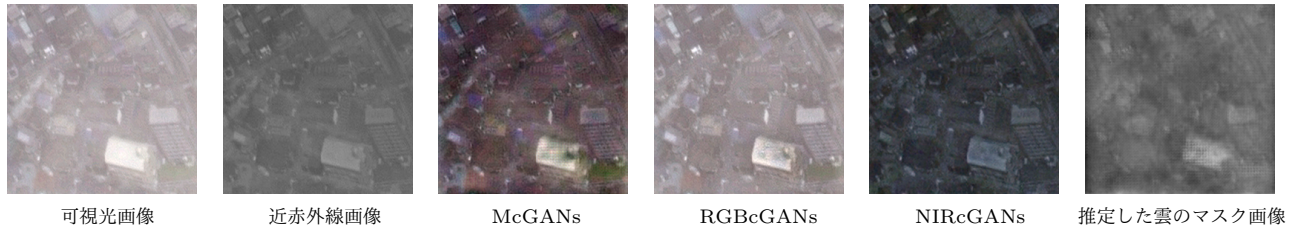


図 4 白い物体による誤認識の例

に設定した。

実験結果の一部を図 3, 4 に示す。図 3 では、近赤外線画像に雲が写っている領域以外は雲が除去できていることが分かる。マルチスペクトル画像を用いた場合 (McGANs) と、近赤外線画像のみを用いた場合 (NIRcGANs) を比較すると、雲が写っている可視光画像で薄く見えている物体の色が、雲を除去した画像にも反映されていることから、単純に近赤外線画像に対して尤もらしい色付けをただけではないことが分かる。

一方で、図 4 では推定した雲のマスク画像から、白い物体を雲と誤認識していることが分かる。これは白い物体と雲が重なると、可視光画像と近赤外線画像だけではそれらを分離することが難しいことを表している。

## 6. まとめ

本論文では cGANs をマルチスペクトル画像へと拡張し、可視光の衛星画像から薄雲を除去する手法を提案した。さらに評価実験を行い、近赤外線画像に雲が写っていない場合は雲を除去できることを確認した。本研究では、地上の様子を観測することが目的であり、近赤外線画像で雲が写っている領域を [10] のように見た目の尤もらしさだけで補完することは行っていない<sup>\*3</sup>。今後は、より波長の長い遠赤外線画像や SAR 画像を追加して、全ての雲を除去するネットワークを構築する。

## 参考文献

- [1] R. Furuta, “Synthetic Aperture Radar (SAR) Utilization for Disaster Management”, Technological Seminar on Environmental Monitoring, 2014.
- [2] L. Maaten, G. Hinton, “Visualizing Data using t-SNE”, JMLR, vol.9, no. Nov, pp. 2579–2605, 2008.
- [3] K. He, J. Sun, X. Tang, “Single Image Haze Removal Using Dark Channel Prior”, TPAMI, vol.33, no.12, pp. 2341–2353, 2011.
- [4] D. Berman, S. Avidan, “Non-Local Image Dehazing”, CVPR, pp. 1674–1682, 2016.
- [5] L. Schaul, C. Fredembach, S. Süsstrunk, “Color Image Dehazing using the Near-Infrared”, ICIP, pp. 1629–1632, 2009.
- [6] C. Feng, S. Zhuo, X. Zhang, L. Shen, “NEAR-INFRARED GUIDED COLOR IMAGE DEHAZING”, ICIP, pp. 2363–2367, 2013.
- [7] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, “Generative Adversarial Nets”, NIPS, 2014.
- [8] M. Mirza, S. Osindero, “Conditional Generative Adversarial Nets”, arXiv, 2014.
- [9] A. Radford, L. Metz, S. Chintala, “UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS”, arXiv, 2015.
- [10] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, A. A. Efros, “Context Encoders: Feature Learning by Inpainting/”, CVPR, pp. 2536–2544, 2016.
- [11] H. Zhang, S. Vishwanath, V. M. Patel, “Image De-raining Using a Conditional Generative Adversarial Network”, arXiv, 2017.
- [12] P. Isola, J. Zhu, T. Zhou, A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks”, arXiv, 2016.
- [13] O. Ronneberger, P. Fischer, T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, MICCAI, pp. 234–241, Springer, 2015.
- [14] C. Li, M. Wand, “Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks”, ECCV, pp. 702–716, 2016.
- [15] K. Perlin, “Improving Noise”, ACM Transactions on Graphics (TOG), vol.21, no.3, pp. 681–682, 2002.

<sup>\*3</sup> 可視光・近赤外線ともに雲に遮蔽されている領域の予測を行う場合は、可視光画像に対する近赤外線画像の雲の透過率をモデル化し、近赤外線画像にも雲を合成する必要がある。