

土地利用分類における拡散モデルを用いたデータ拡張による精度改善

田 銘池¹ 重井 徳貴²

概要: 航空リモートセンシングによる土地利用分類は重要であるが、学習データの確保が難しく、特に少数クラスでは外観類似により分類精度が不安定になりやすい。本研究では、テクスチャ解析によるスケール推定と画像の統計的属性の離散化を行い、その結合分布をモデル化するデータ拡張手法を提案する。分布に基づく LoRA 微調整と制約付き生成を拡散モデルに適用し、ドメイン乖離を抑えた合成画像を生成する。実験の結果、GAN 等の既存手法と比較して高品質な合成画像を安定して生成でき、データ拡張手法として用いた場合、幾何・色変換や GAN 系手法よりも分類精度が一貫して向上し、特に少数クラスの識別性能改善に有効であることを示す。

キーワード: 機械学習, データ拡張, 拡散モデル, LoRA, 土地利用分類

Data Augmentation Using Diffusion Models for Improved Land-Use Classification

Abstract: Securing training data for aerial remote-sensing land-use classification is difficult, often causing instability in classifying visually similar minority classes. To address this, we propose a data augmentation framework that models the joint distribution of texture-based scale estimates and discretized image statistics. By applying LoRA fine-tuning and constrained generation to a diffusion model, we produce synthetic images with reduced domain shift. Experiments show that our approach stably generates higher-quality images than GAN-based methods and consistently improves classification accuracy over geometric, color-based, and GAN-based augmentations, particularly for minority classes.

Keywords: Machine Learning, Data Augmentation, Diffusion Models, LoRA, Land-Use

1. はじめに

航空リモートセンシング画像に基づく土地利用分類は、農業、都市計画、環境モニタリングなどの分野において、地域の状況を定量的に把握するための基盤技術である。一方で、季節変動や撮像条件のばらつき、ならびに観測可能な地理的範囲の制約により、十分な量と多様性を備えたラベル付き学習データを継続的に確保することは容易ではない [1]。また、土地利用クラス間で外観が類似する場合には識別が困難となり、特に少数クラスにおいては学習の不安定化や再現率の低下が生じやすいという問題がある。本研

究では、国土地理院の航空写真 [2] を用い、竹林・針葉樹林・水田・茶畑の 4 クラス土地利用分類を対象とする。本研究で扱うデータセットにおいても、クラスごとの画像枚数に偏りが存在し、竹林や茶畑といった少数クラスの存在が、分類性能向上のボトルネックとなっている。

データ不足に対する代表的な対策として、幾何変換や色変換に基づく従来のデータ拡張 [3]、ならびに Generative Adversarial Network (GAN) などの既存手法を用いたデータ拡張が広く検討されてきた [4]。しかし、前者は画像の幾何的特性を操作するに留まり、画像内のオブジェクトやシーンの意味的な多様性を実質的に増やすことはできないという限界がある。一方、後者は新規パターンの生成が可能であるものの、学習の不安定さやモード崩壊に加え、意図した属性を持つ画像を生成するための制御性の確保が困

¹ 鹿児島大学工学部 先進工学科 情報・生体工学プログラム
1 Chome-21-24 Korimoto, Kagoshima 890-0065, Japan

² 鹿児島大学大学院理工学研究科 情報科学専攻 情報科学プログラム 協創情報コース

難であるという課題が指摘されている。

近年、拡散モデルは高品質な画像生成能力をもとに、合成データ拡張の有力な選択肢として注目を集めている [5]。しかし、合成データを土地利用分類の学習に利用する際には、合成データは見た目が自然であるだけでは不十分であり、実データ分布と整合した位置に配置されている必要がある。特に、分布整合性の確保とドメインシフトの抑制が、分類性能を左右する重要な要因となる。

そこで本研究では、拡散モデルを用いて意味的に豊かで、かつ実データ分布に整合した合成画像を生成し、土地利用分類の性能向上に資する生成的データ拡張フレームワークを提案する。本手法では、訓練データの特性を統計的にモデル化し、条件付き生成と分布整合制約を組み合わせることで、合成データに起因するドメインシフトを抑制する。特に、観測スケール (Zoom) は、テクスチャの粒度や境界の見え方を通じて分類性能に強く影響する要因であるため、生成過程において制御可能な条件として導入する。

本研究の主な貢献は、以下の三点にまとめられる。

- (1) テクスチャ解析と統計的離散化による生成条件の明示化: 航空写真に対し、画像のテクスチャ特徴量 (ラプリアン分散やエッジ密度) に基づく弱教師あり学習を用いて観測スケールを推定する手法を構築した。さらに、色調・輝度統計に基づく環境属性の離散化 (トークン化) を組み合わせることで、曖昧な画像特徴を物理的特性に基づいて分離 (disentangle) し、生成過程における詳細な制御を可能にした。
- (2) 結合確率分布に基づくドメイン整合性と多様性の確保: 訓練データにおける観測スケールと環境属性の条件付き結合確率分布 $P(\text{attribute} | \text{zoom})$ をモデル化による生成の多様化と、分布統計に基づく厳格なフィルタリングを統合した。これにより、各属性の共起関係を維持しつつ、実データの統計的性質と整合した合成データを生成し、ドメインシフトを体系的に抑制した。
- (3) 生成品質の定量的評価と分類精度への寄与: 生成画像の品質評価 (FID) において、提案手法は GAN や分布整合を行わない拡散モデル (LoRA) と比較して最も実データ分布に近い数値を達成し、モード崩壊や分布乖離を抑制できることを確認した。この高い分布適合性が分類器の学習に有効に作用し、結果として土地利用分類タスクにおいて、macro-F1 および少数クラスの識別性能を一貫して改善することを確認した。

2. データセット

2.1 使用データセット

本研究では、国土地理院が公開する「CNN による地物抽出用教師データセット (以下、本データセット)」[2] および地理院地図を基に、分類学習用のデータセットを構成する。

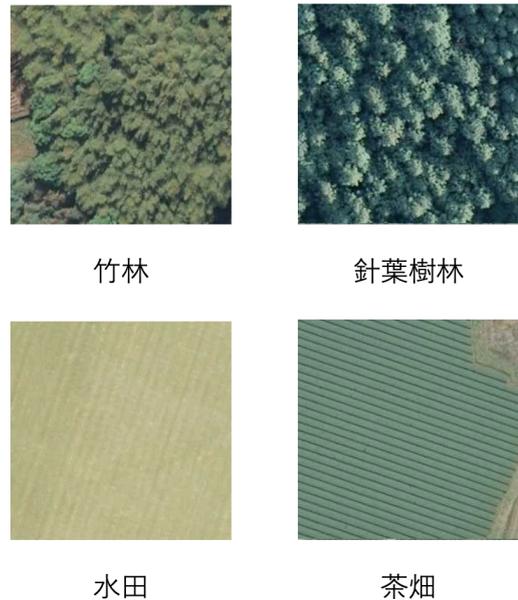


図 1 4 種類の土地利用データセットの例
Fig. 1 Examples of four land use categories



図 2 作成手法による地物スケールの差異 (左: 分割, 右: 縮小)
Fig. 2 Scale difference caused by dataset construction methods (Left: Split, Right: Shrunk)

本データセットは、2007 年度から 2018 年度に撮影された地上画素寸法 20 cm 級のデジタル航空カメラ画像を対象とし、セマンティックセグメンテーション用の教師データとして多様な地物領域がラベリングされたものである [2]。

本研究ではこの中から、竹林 (bamboo)、針葉樹林 (pine)、水田 (rice)、茶畑 (tea) の 4 カテゴリに着目する。各クラスの代表例を図 1 に示す。先行研究 [3] と同様の手順で切り出した航空写真パッチを入力し、クラスラベルを出力とする土地利用分類を行う。

2.2 パッチ作成と観測スケールの定義

本データセット [2] の仕様では、まず地上画素寸法 20 cm の空中写真 (572 × 572 ピクセル) から、(i) 縦横それぞれ 2 分割したパッチ 4 枚、および (ii) 全体を 286 × 286 に縮小したパッチ 1 枚が作成されており、1 枚の親画像から合計 5 枚の 286 × 286 パッチ (画像とラベルのペア) が提供されている [6]。

この作成過程により、データセット内には実質的な空間

表 1 データセットのクラス別内訳

クラス	訓練	検証	テスト	合計
竹林 (bamboo)	376	47	47	470
針葉樹林 (pine)	816	102	102	1020
水田 (rice)	1768	221	221	2210
茶畑 (tea)	136	17	17	170
合計	3096	387	387	3870

解像度 (Spatial Resolution) が異なる二種類の画像が混在することになる。すなわち、分割由来のパッチ (高空間解像度・拡大視) と、縮小由来のパッチ (低空間解像度・広域視、地上画素寸法約 40 cm 相当) である (図 2 参照)。

この観測スケール (地物の見かけの大きさ) の差異は、テクスチャ粒度や境界の見え方に影響を与え、学習の不安定化要因となり得る。そこで本研究では、このスケール差を離散的な条件 (Zoom トークン) として定義し、これを用いて後述する生成プロセスを明示的に制御する。ただし、提供される学習用データセットには、各画像がどちらの作成過程 (分割または縮小) を経たかを示すメタデータは付与されていない。したがって、後述する提案手法 (第 4 節) において、画像特徴からこの観測スケールを自動推定する必要がある。

2.3 学習・検証・テスト分割

データはクラス内でランダムに分割し、訓練・検証・テストの比率を 8:1:1 とした。各データのクラス別内訳を表 1 に示す。表から分かるようにクラス間の不均衡が顕著であり、特に竹林および茶畑が少数となっている。そこで本研究では、このデータ不均衡による学習への悪影響を緩和するため、後述するデータ拡張を用い、精度の改善を図る。

3. 関連研究

本章では、本研究に関連する先行研究を、(1) 土地利用分類における従来のデータ拡張、および (2) 生成モデル (GAN・拡散モデル) によるデータ合成の観点から整理し、既存手法の課題と本研究の位置付けを明確にする。

3.1 土地利用分類における従来のデータ拡張

航空写真を用いた土地利用分類では、学習データの不足やクラス不均衡に起因する過学習を抑制するため、幾何変換や色変換に基づくパラメトリックなデータ拡張が広く用いられてきた。代表的な手法として、画像の反転、回転、クロップ、輝度・コントラスト調整などが挙げられる。これらはモデルの不変性を高め、汎化性能の改善に寄与する。

一方で、これらの手法が付与できる多様性は、あくまで既存画像の「補間変換」の域を出ない。そのため、土地利用の識別に本質的な意味の変化 (例: 地物境界の形状変化や、異なる観測環境下のテクスチャ変動) を伴う新規サンプルを生成することは困難である。結果として、特に少

数クラスや外観が類似するクラスにおいて識別境界が十分に形成されず、分類精度の向上が頭打ちになるという課題が残る。

3.2 生成モデルによるデータ拡張

上述した従来手法の限界を補うため、データ分布そのものを学習し、新規画像を合成する生成モデルの活用が検討されている。

3.2.1 GAN による合成拡張と課題

Generative Adversarial Network (GAN) [8] は、生成器と識別器の敵対的学習により写実的な画像を生成する既存手法であり、リモートセンシング分野においても少数クラスの補填に応用されてきた。特に StyleGAN2-ADA[10] のように、少量データ下での学習安定化を図る手法も提案されている。

しかし、GAN によるデータ拡張には依然として本質的な課題が残されている。第一に学習の不安定性である。本研究のように「クラス不均衡」かつ「観測スケールが混在」するデータセットにおいては、識別器が特定の支配的な特徴に過剰適合しやすく、その結果としてモード崩壊が発生し、生成画像のテクスチャや背景が単一化する傾向がある。第二に、生成内容の制御の困難さである。GAN は通常、潜在空間上のランダムノイズから画像を生成するため、観測スケールや撮影条件といった具体的な物理属性を意図通りに操作・指定して生成することは容易ではない。土地利用分類においては、不足している特定の条件をピンポイントで補完することが求められるが、GAN ではこのようなきめ細かな条件付け学習の収束が難しく、結果として分類タスクに有効な多様性を持つデータを十分に供給できないという限界がある。

3.2.2 拡散モデルと LoRA による拡張の可能性

近年、拡散モデル (Diffusion Models) [11] は、段階的なノイズ除去過程を通じて高品質かつ多様な画像生成を実現し、地球観測 (EO) 分野でのデータ拡張においても注目されている [5]。特に、事前学習済みの基盤モデル (例: Stable Diffusion) に対し [12]、Low-Rank Adaptation (LoRA) [13] を用いて軽量に領域適応を行う手法は、計算資源の制約下でも対象ドメインの特徴を反映した画像を得られる点で有望である。

一方で、通常の LoRA による微調整は、テキスト条件への追従性に主眼が置かれており、リモートセンシング画像に固有の「物理的な観測条件 (空間解像度・スケール)」や「地理的整合性」を明示的に制御する機構を持たない。とりわけ、本研究の対象データのように高空間解像度と低空間解像度が混在する場合、条件付けが曖昧なまま学習を行うと、高周波成分 (細粒度テクスチャ) と低周波成分が混同され、実データ分布から乖離した画像が生成される恐れがある。

したがって、拡散モデルを用いた拡張を土地利用分類へ有効に接続するためには、生成過程において観測スケールを明示的に制御し、かつ合成サンプルを実データの分布特性に整合させる新たな手法が必要となる。

4. 提案手法：制約付き LoRA による生成的データ拡張

4.1 概要

本研究では、竹林などの少数クラスのデータ不足と観測スケールの混在という課題に対し、実データ分布との整合性を保ちながらデータ拡張を行う手法を提案する。提案手法の全体像を図 3 に示す。本手法は、画像統計量に基づく「自動化」と「分布制御」を特徴とする以下のステップで構成される。

- (1) **Distribution Estimation:** 訓練データにアノテーションが付与されていない点を補うため、擬似的な参照セットを構築して Zoom 分類器を学習する。これを用いて全画像に Zoom トークンを付与し、さらに統計的手法で属性トークンを抽出して、結合確率分布

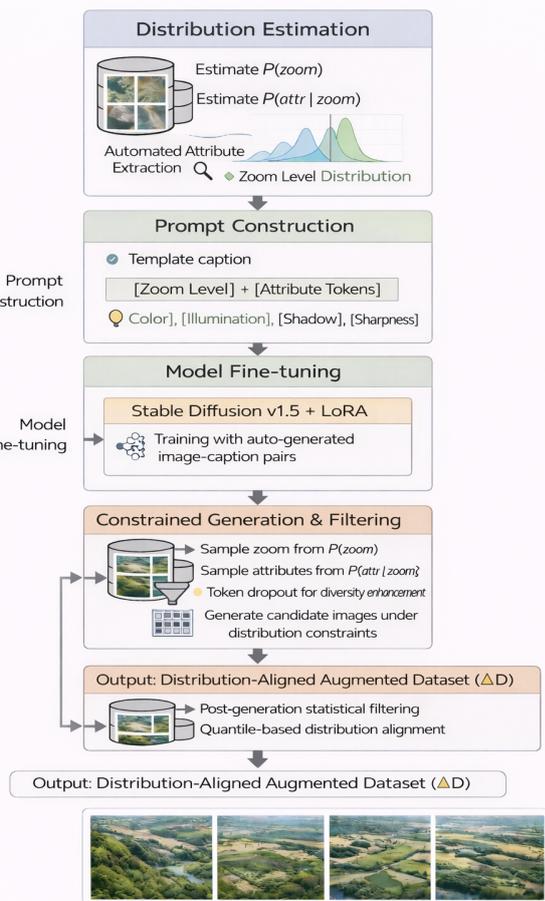


図 3 提案手法（制約付き LoRA による生成的データ拡張）の全体概要

Fig. 3 Overview of the proposed method (generative data augmentation via constrained LoRA).

$\hat{P}(\text{attr} | \text{zoom})$ を推定する。

- (2) **Prompt Construction:** 推定されたトークンをテンプレートに埋め込むことで、学習用キャプションを自動生成する。VLM（視覚言語モデル）に依存せず、画像統計量のみから物理的特性を記述するプロンプトを構築する点が特徴である。
- (3) **Model Fine-tuning:** 自動生成されたキャプションと画像のペアを用い、Stable Diffusion を LoRA により微調整する。モデルはトークンと具体的な画像特徴の対応関係を学習する。
- (4) **Constrained Generation & Filtering:** 推定分布からのサンプリングに加え、トークン・ドロップアウトによる多様化を適用して画像を生成する。その後、統計的フィルタリングによって分布外のノイズを排除し、実データ分布に適合した高品質な拡張データセット (ΔD) を作成する。

4.2 観測条件のトークン化と分布推定

生成画像を実データの物理的特性に適合させるため、以下の手順で画像特徴を離散トークン化し、その分布をモデル化する。

4.2.1 擬似マルチスケール参照セットの構築

テキスト解析によって観測スケールを推定するための教師データとして、Web 地図タイルをベースに、画像処理によって意図的に空間解像度を変化させた「擬似参照セット (Pseudo-Reference Set)」を構築した。具体的には、中心切り出し (Crop) とリサイズ (Resize) の比率を変えることで、以下の 2 つの有効ズームレベル (Effective Zoom) を持つアンカー画像群を作成した。

- ref_18.52: 200px を切り出して拡大 (高空間解像度)
- ref_18.16: 256px を 286px へリサイズ (低空間解像度)

これらは物理的な空間解像度と厳密に一致するものではないが、テキストの周波数特性における「相対的な高低関係」を学習するための代理ラベル (Proxy Label) として機能する。

4.2.2 テキスタ解析による Zoom 分類とトークン定義

次に、構築した参照セットを用いて、ターゲットデータセットの Zoom レベルを推定する。各画像から、解像感に関連深いテキスト特徴量 $\phi(x)$ (ラプラシアン分散, エッジ密度, FFT 高周波成分比, 平均輝度, コントラストの 5 次元ベクトル) を抽出する。これらを入力とするロジスティック回帰分類器 (Logistic Regression) を参照セットで学習させ、訓練データ全画像の分類を行った。その結果、データセット内の画像は主に ref_18.52 に近いグループと、ref_18.16 に近いグループの二つに明確に分離された。この分類結果に基づき、各画像の観測スケールトークンを以下のように定義する。

表 2 属性トークンの定義と離散化基準

属性種別	算出指標	トークン (離散化基準)
Color	HSV 色空間における色相ヒストグラム ($S > 0.18, V > 0.15$)	green ($70^\circ \leq H \leq 160^\circ$)
		yellow ($35^\circ \leq H < 70^\circ$)
		mixed ($G > 0.25 \ \& \ Y > 0.18$)
		low_sat (mask < 0.05)
Light	明度 V の平均値	bright ($V > Q_{0.66}$)
		normal ($Q_{0.33} \leq V \leq Q_{0.66}$)
		dark ($V < Q_{0.33}$)
Shadow	暗部画素の割合 (gray < 0.12)	high ($\rho_{\text{shadow}} \geq Q_{0.80}$)
		low ($\rho_{\text{shadow}} < Q_{0.80}$)
Sharp	Laplacian 分散 ($\text{Var}(\nabla^2 x)$)	high ($\text{Var} > Q_{0.66}$)
		normal ($Q_{0.33} \leq \text{Var} \leq Q_{0.66}$)
		low ($\text{Var} < Q_{0.33}$)

- **High-Res Token (z_{high})**: ref_18.52 に分類された画像. データセット内の高空間解像度データに対応する. 実装上の識別子は zoom_18.52 である.
- **Low-Res Token (z_{low})**: ref_18.16 に分類された画像. データセット内の低空間解像度データに対応する. 実装上の識別子は zoom_18.16 である.

以降, 本稿では物理的意味に基づき, これらを抽象化されたトークン $z_{\text{high}}, z_{\text{low}}$ として記述する.

4.2.3 属性トークン (外観特徴)

Zoom トークンに加え, 生成画像の多様性と整合性を制御するため, 4 種類の外観属性 Color, Light, Shadow, Sharpness を定義する. 各属性は画像処理により算出された統計値に基づき, 訓練データの分位点を用いて離散化される. 定義を表 2 に示す.

4.2.4 結合分布のモデル化

全訓練画像 (竹林クラス) に対して付与されたトークンを集計し, Zoom トークン $z \in \{z_{\text{high}}, z_{\text{low}}\}$ と属性トークン集合 a の結合頻度分布 $\hat{P}(\text{attr} | \text{zoom})$ を算出し, 実データの物理的相関をモデル化する.

4.3 LoRA による領域適応

推定されたトークンを用いてプロンプトを構築し, Stable Diffusion v1.5 を LoRA により微調整する. プロンプトは以下のテンプレートをを用いて自動生成される.

aerial orthophoto, remote sensing top-down
view of dense bamboo forest canopy texture,
<zoom>, <color>, <light>, <shadow>, <sharp>

ここで zoom_18.52 (z_{high}) または zoom_18.16 (z_{low}) の文字列が挿入される. なお, 学習段階ではトークンのドロップアウトは適用せず, 全ての属性トークンを常にプロンプトに含める.

4.4 制約付き生成とフィルタリング

4.4.1 目標分布に基づくサンプリング

データ拡張を行う際は, ランダム生成ではなく, 実データの統計分布に基づいたサンプリングを行う. まず, Zoom トークンを目標分布 $P_{\text{target}}(\text{zoom})$ に従って決定し, 次に属性トークンを条件付き分布 $\hat{P}(\text{attr} | \text{zoom})$ からサンプリングする.

4.4.2 推論用プロンプトの構築

決定された属性列 t_{attr} に基づき, 最終的な入力プロンプトを作成する. ここで, 学習データの過学習を防ぎ, 生成画像の多様性を向上させるため, トークン・ドロップアウト (Token Dropout) を適用する. 具体的には, Zoom トークンを除く属性トークンを確率 p_{drop} でランダムに除外した新たなトークン列 t'_{attr} を生成に使用する.

推論プロンプトは, 生成対象を記述するポジティブプロンプト P_{pos} と, 回避すべき要素を記述するネガティブプロンプト P_{neg} のペアで構成される. まず, P_{pos} は, ドロップアウト適用後の属性トークン列 t'_{attr} にドメイン固定テキスト T_{fixed} を結合して構築する.

$$P_{\text{pos}} = T_{\text{fixed}} \oplus t'_{\text{attr}} \quad (1)$$

ここで, T_{fixed} は学習時と同様の "aerial orthophoto..." 等のドメイン記述であり, テンプレートの一貫性を保つ. 一方, P_{neg} には生成品質を阻害する要素 (道路, 建物, 低画質など) を抑制するための固定テキストを使用する (詳細は 5.3 節参照).

最終的な画像生成は, この $(P_{\text{pos}}, P_{\text{neg}})$ を用いた Classifier-Free Guidance [14] により行われる. このドロップアウトによる制約の緩和は, 多様な画像を生成する一方で分布外のサンプルを生む可能性があるため, 後段の統計的フィルタリングによる品質保証が不可欠となる.

4.4.3 分布整合フィルタリング

生成された画像集合 ΔD に対し, 最終的な品質保証として統計的フィルタリングを適用する. 生成画像から再度 4.2.2 節および 4.2.3 節と同様の手順で特徴量 (エッジ密度や属性値) を算出し, それらが実データ (竹林クラス) の 5%–95% 分位点範囲内に収まる画像のみを採用することで, LoRA が生成しがちな分布外 (Out-of-Distribution) のサンプルを自動的に排除する.

最終的に選別された画像を訓練データに統合し, 分類器の学習に用いる.

5. 実験

5.1 実験設定: 土地利用分類器 (Small CNN+)

土地利用分類器には, 軽量かつ高い特徴表現能力を持つ Small CNN+ を採用した. これは畳み込み層と Max Pooling を基本とし, Batch Normalization と Dropout などを導入して学習安定性と汎化性能を高めたアーキテク

チャである [15]. 訓練時は, データ拡張として水平反転 (確率 20%) および輝度・コントラスト変動 ($\pm 5\%$) を適用した. 最適化には Adam (学習率 1×10^{-4} , バッチサイズ 32) を用い, 損失関数にはクロスエントロピー損失を使用した.

5.2 比較手法の構築

提案手法 (LoRA-align) の有効性を検証するため, 以下の 5 つのベースライン手法と比較を行った. Real-only 以外の手法では, 竹林クラスの実データ (376 枚) に対して 224 枚の拡張データを追加し, 合計 600 枚に拡張する設定した.

5.2.1 Real-only

拡張を行わず, 実データのみ (竹林クラス 376 枚) で学習させたベースラインである. これを性能改善の基準とする.

5.2.2 Oversampling

少数クラスの実データをランダムに復元抽出 (Oversampling) し, 単純に枚数を増やす手法である.

5.2.3 Traditional DA

伝統的なデータ拡張手法である. 本実験では航空画像の特性を考え, 回転やせん断に伴う画像欠損 (黒領域) の発生を抑制するため, 前処理として画像の周囲にリフレクションパディング (画像サイズの 30%) を付与した. その上で, 各画像に対して以下の幾何学的変換を適用する.

- **反転:** ランダムな水平・垂直反転 (各 $p = 0.5$).
- **アフィン変換:** 一様分布に基づくランダム変換 (平行移動 $\pm 20\%$, 回転 $\pm 60^\circ$, せん断 $\pm 60^\circ$).
- **スケーリング:** スケール 0.9 ~ 1.0 のランダムクロップ (アスペクト比固定).

5.2.4 LSGAN

生成データ拡張の比較として, 学習安定性の高い **Least Squares GAN (LSGAN)** [16] を用いた. 潜在変数 $z \in \mathbb{R}^{400}$ から 256×256 の画像を生成するように Generator と Discriminator を 500 エポック学習させた (Optimizer: Adam, 学習率: $G = 3.0 \times 10^{-4}$, $D = 1.0 \times 10^{-4}$). 学習後, ランダムなノイズから 224 枚の合成画像を生成した.

5.2.5 LoRA

提案手法の核心である「Zoom トークン化」と「属性トークン化」の効果を分離して検証するためのベースラインである. 画像の物理的特徴 (Zoom や Color 等) を区別せず, ドメイン全体を代表する一般的な固定プロンプトのみを用いて LoRA を学習させたモデルである.

- (1) **学習:** 訓練データ全画像に対し, 属性トークンを含まない固定的なキャプション ("aerial bamboo forest, dense bamboo canopy, natural texture") を一律に付与し, 提案手法と同等の設定で LoRA を学習した.
- (2) **推論:** 学習時と同じプロンプトを入力として画像を生

表 3 LoRA 学習および生成のハイパーパラメータ設定

Stage	Parameter	Value
<i>Training (Stage II-III)</i>		
	Base Model	SD v1.5
	Resolution	512×512
	LoRA Rank / Alpha	16 / 16
	Learning Rate	5×10^{-5}
	Batch Size	2
	Total Steps	4,000
<i>Generation (Stage IV)</i>		
	Target Zoom Dist.	$z_{high} : 0.7, z_{low} : 0.3$
	Token Dropout	0.25
	Smoothing (α)	1.0
	Inference Steps	{20, 25, 30, 35} (Random)
	CFG Scale	{4.0, 5.0, 6.0} (Random)
	Sampler	DDIM
	Filter Range	5th - 95th percentile
	Max Keep Images	400(Select 224)

成した. 制約付き生成や, 生成後の統計的フィルタリングは適用せず, 出力画像をそのまま採用した.

5.3 LoRA-align (提案手法)

5.3.1 開発環境

本研究の実装および評価は, Ubuntu 20.04 LTS 上で NVIDIA GeForce RTX 4090 (24GB VRAM) を用いて行った. 深層学習フレームワークには PyTorch 2.3.1+cu121 (CUDA 12.1) を使用し, LoRA の学習には **sd-scripts**^{*1} ライブラリを用いた. 基盤モデルには Stable Diffusion v1.5^{*2} を使用した.

5.3.2 LoRA 学習と画像生成 (Stage I-IV)

提案手法によるデータ拡張は, トークン推定, LoRA 学習, 制約付き生成, フィルタリングの順に行われた. 各部分の主要なハイパーパラメータを表 3 にまとめた.

5.3.2.1 Stage I: トークン推定

4.2 節で述べた手順に従い, 訓練データ (竹林クラス) に対して Zoom および属性トークンを付与した. Zoom 分類器の特徴量 $\phi(x)$ には, ラプラシアン分散, エッジ密度 (閾値 0.12), FFT 高周波成分比, 輝度平均・分散の 5 次元を用いた. 推定にはロジスティック回帰を用い, 参照データ (z_{high}, z_{low}) により学習させた.

5.3.2.2 Stage II-III: LoRA 学習

推定されたトークンとキャプションを用い, **sd-scripts** ライブラリを用いて LoRA の学習を行った. 計算資源の制約と学習効率を考慮し, 解像度は 512×512 , 混合精度 (fp16) を利用した.

*1 入手先: <https://github.com/kohya-ss/sd-scripts> (commit: 206adb6)

*2 入手先: <https://huggingface.co/stable-diffusion-v1-5/stable-diffusion-v1-5>

5.3.2.3 Stage IV: 制約付き生成とフィルタリング

学習済み LoRA を用いた画像生成において、生成画像の多様性を確保するため、Zoom トークンを除く属性トークンに対しドロップアウト（確率は表 3 参照）を適用した上でプロンプトを構築した。入力プロンプトは、ポジティブ P_{pos} とネガティブ P_{neg} のペアにより定義される。

$$P_{\text{input}} = P_{\text{pos}} \text{ with } P_{\text{neg}} \quad (2)$$

ここで、 P_{pos} はトークン列と固定テキストを結合した "`<zoom>`, `<color>`, `<light>`, `<shadow>`, `<sharp>`, aerial orthophoto, nadir top-down, bamboo canopy, landcover" とし、 P_{neg} には "low quality, blurry, buildings, roads, cars" 等の阻害要因を指定した。

生成後、実データ分布との整合性を保証するため、以下の統計量集合 S を用いてフィルタリングを行った。

$$S = \{\mu_{\text{gray}}, \sigma_{\text{gray}}, \text{Var}(\nabla^2), \rho_{\text{edge}}, r_{\text{hf}}, \rho_{\text{shadow}}\} \quad (3)$$

生成画像 x_{gen} は、全ての統計量 $s \in S$ が実データの 5%–95% 分位点範囲内にある場合のみ採用され、最終的に 286×286 にリサイズされてデータセットに追加された。

5.4 評価指標

分類性能の評価には、不均衡データセットにおいて標準的に用いられる **Macro-F1** および **Average Precision (AP)** を採用する。また、モデルの識別能力を総合的に評価するため **Macro-ROC-AUC** も併せて報告する。

5.5 実験結果と考察

5.5.1 定量評価：ベースラインとの比較

提案手法 (LoRA-align) と 5 つの比較手法の分類性能を表 4 に示す。ここでは、竹林クラスに対して 224 枚の拡張データを追加し、データ数を他クラスと均衡させている (Real-only を除く)。

全体傾向：表 4 に示す通り、提案手法は全ての指標において最高値を記録した (Macro-F1: 0.9486, AP: 0.9870)。「Real-only」と比較して大幅な性能向上が見られるだけでなく、「Oversampling」や「Traditional DA」に対しても明確な優位性を示している。

生成手法間の比較：同じ生成的アプローチである「LSGAN」や「LoRA」と比較しても、提案手法は優れている。この要因は、生成画像の分布整合性 (FID) の差に現れている。竹林クラスにおける FID を比較すると、「LSGAN (230.89)」はモード崩壊により著しく高く、「LoRA (145.87)」も分布の乖離が見られた。一方、提案手法は 141.95 と最も低い値 (実データとの高い類似性) を達成しており、この「質の高い分布近似」が、分類精度の向上と整合している。すなわち、提案手法は分布外のノイズ生成

表 4 分類性能の比較 (竹林クラス追加数=224 枚)

Method	Macro-F1	Macro-AP	ROC-AUC
Real-only	0.9244 ± 0.0048	0.9772 ± 0.0024	0.9961 ± 0.0009
Oversampling	0.9347 ± 0.0131	0.9807 ± 0.0058	0.9967 ± 0.0010
Traditional DA	0.9343 ± 0.0127	0.9803 ± 0.0045	0.9966 ± 0.0011
LSGAN	0.9397 ± 0.0077	0.9843 ± 0.0046	0.9976 ± 0.0006
LoRA	0.9363 ± 0.0048	0.9829 ± 0.0045	0.9971 ± 0.0006
LoRA-align (ours)	0.9486 ± 0.0042	0.9870 ± 0.0022	0.9977 ± 0.0004

表 5 Zoom トークンによる精度の比較

Method	Macro-F1	Macro-AP	ROC-AUC
LoRA-zoom	0.9391 ± 0.0121	0.9850 ± 0.0015	0.9973 ± 0.0003
LoRA-align (ours)	0.9486 ± 0.0042	0.9870 ± 0.0022	0.9977 ± 0.0004

を抑制し、分類器にとって有効な学習サンプルを提供できたことが、FID と精度の双方から裏付けられた。

5.5.2 アブレーション実験：Zoom トークンの有効性と拡張性

提案手法における「観測スケールの制御」の有効性を検証するため、比較実験を行った。比較対象とした「LoRA-zoom」は、学習画像が持つ本来の空間解像度差を無視し、全ての画像に対して**単一の共通 Zoom トークン**のみを付与して学習させたモデルである (属性トークンやフィルタリング条件は提案手法と同一)。

結果を表 5 に示す。Zoom トークンを区別しなかった「LoRA-zoom」に対し、提案手法 (LoRA-align) は Macro-F1 において約 0.9 ポイントの精度向上を達成した。

本実験では 2 種類のスケール ($z_{\text{high}}, z_{\text{low}}$) のみでの検証であったが、本手法の本質的な価値は「任意の空間解像度レベルへの制御可能性」にある。今後、ドローン撮影高度が可変的なデータセットなど、より多様な空間解像度が混在する環境に適用した場合、この Zoom 制御機能は、生成画像の空間解像度感を自在に操るための強力なインタフェースとして、その真価をさらに発揮すると期待される。



図 4 竹林クラスの生成画像の比較

Fig. 4 Qualitative comparison of generated bamboo images

5.5.3 定性評価

各手法によって生成された竹林クラスの画像例を図4に示す。「LSGAN」はテキストチャが平滑化されやすく、竹林特有の高周波成分の再現に課題が見られた。「LoRA」は高空間解像度なテキストチャを生成できるものの、Zoomトークンと属性トークンによる制御がないため、実データには存在しない極端な遠景や不自然なアングルが混入している。一方、提案手法「LoRA-align(ours)」は、Zoomトークン (z_{high}, z_{low}) によって空間解像度を適切に制御し、推定した属性トークンに基づいて外観特性も整合させる。さらに統計的フィルタリングを経ているため、実データと見分けがつかない自然な画像が生成されている。この「質の高い」データ拡張が、前述の定量評価における高い分類精度に寄与したと結論付けられる。

6. おわりに

本研究では、土地利用分類におけるデータ不足という課題に対し、拡散モデルを用いたデータ拡張手法を提案し、その有効性を検証した。具体的には、「従来手法」や「LSGAN」に加え、近年注目される「拡散モデル」と「LoRA」を用いた微調整によるデータ生成を比較・検討した。

実験の結果、分類器の精度改善においては、単に高空間解像度で写実的な画像を生成することよりも、生成される画像の分布を学習データの分布にいかにかに適合させるかが重要であることを明らかにした。特に、本研究で対象としたデータセットは、作成過程に起因して、提案手法は各画像のズームレベルを明示的に条件付けることで、「高空間解像度パッチの鋭いエッジ」と「縮小パッチの平滑化されたテキストチャ」を区別して学習させた。加えて、トークン・ドロップアウトによる「多様性の確保」と、統計的フィルタリングによる「分布外ノイズの排除」を統合したことで、実データと整合しつつも多様性に富んだ拡張データが得られ、結果として既存手法を上回る分類精度を達成した。

今後の課題として、データセットの精製と適用範囲の拡大が挙げられる。本研究で使用した国土地理院のデータセットには依然としてノイズが含まれるため、より厳密なアノテーションが施されたデータセットに対して本手法を適用し、検証することが望まれる。また、今回は竹林クラスに焦点を当てたが、同様にテキストチャ特徴が複雑な他の植生クラスへの展開も検討する必要がある。

さらに、本実験では離散的な空間解像度のみでの検証であったが、本手法を用いればその中間的な空間解像度や、さらに広域的な視点を持つ画像を連続的に生成・補間できる可能性がある。これは、異なる高度で撮影された複数のデータセット間のドメインギャップを埋めるための「空間解像度正規化」技術としても応用が期待され、将来的に航空写真認識・リモートセンシング解析の精度向上に寄与すると考えられる。

参考文献

- [1] 国土地理院: AI を活用した地物自動抽出に関する研究(第5年次), 入手先 (<https://www.gsi.go.jp/common/000254456.pdf>) (2026/01/29 閲覧).
- [2] 国土地理院: CNN による地物抽出用教師データセット, 入手先 (<https://gisstar.gsi.go.jp/gsi-dataset/index-ja.html>) (2026/01/29 閲覧).
- [3] 平島ほか: 航空写真からの CNN による土地利用分類の半教師あり学習を用いる精度改善, 知能と情報, Vol.33, No.1, pp.520–524 (2021).
- [4] Moharram, M. A. and Sundaram, M.: Data Augmentation for Land Cover Classification Using Generative Adversarial Networks, 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4432–4435 (2021).
- [5] Sousa, T., Ries, B. and Guelfi, N.: Data Augmentation in Earth Observation: A Diffusion Model Approach, *arXiv preprint arXiv:2406.06218* (2024).
- [6] 国土地理院: G S I データセット-19 (竹林), 入手先 (<https://gisstar.gsi.go.jp/gsi-dataset/19/index.html>) (2026/01/29 閲覧).
- [7] Shorten, C. and Khoshgoftaar, T. M.: A survey on image data augmentation for deep learning, *Journal of Big Data*, Vol.6, No.1, Article 60 (2019).
- [8] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative adversarial nets, *Advances in Neural Information Processing Systems (NeurIPS)*, Vol.27, pp.2672–2680 (2014).
- [9] Jozdani, S., Mahdianpari, M., Salehi, B. and Homayouni, M.: A review and meta-analysis of generative adversarial networks and their applications in remote sensing, *Journal of Applied Geodesy*, Vol.16, No.4, pp.287–311 (2022).
- [10] Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J. and Aila, T.: Training generative adversarial networks with limited data, *Advances in Neural Information Processing Systems (NeurIPS)*, Vol.33, pp.12104–12114 (2020).
- [11] Ho, J., Jain, A. and Abbeel, P.: Denoising diffusion probabilistic models, *Advances in Neural Information Processing Systems (NeurIPS)*, Vol.33, pp.6840–6851 (2020).
- [12] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B.: High-resolution image synthesis with latent diffusion models, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.10684–10695 (2022).
- [13] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. and Chen, W.: LoRA: Low-rank adaptation of large language models, *International Conference on Learning Representations (ICLR)* (2022).
- [14] Ho, J. and Salimans, T.: Classifier-free diffusion guidance, *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications* (2021).
- [15] Song, J., Gao, S., Zhu, Y. and Ma, C.: A survey of remote sensing image classification based on CNNs, *Big Earth Data*, Vol.3, No.3, pp.232–254 (2019).
- [16] Mao, X., Li, Q., Xie, H., Lau, R. Y. K., Wang, Z. and Smolley, S. P.: Least squares generative adversarial networks, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp.2794–2802 (2017).