

GrabCutとLAMAを用いた高環境適応型 インテリアシミュレータの実装

山村 慧心¹ 山森 一人^{2,a)}

概要: 本研究では、ARを用いたインテリアシミュレーションにおいて、配置したい3D家具オブジェクトと背景映像との不整合を軽減する手法を提案する。GrabCutにより不要な既存の家具領域を精密に消去し、LAMAによって消去部分の背景を復元、Stable Diffusionを用いて3D家具オブジェクトを周囲に適応させる。LAMAにおいて、新たな3D家具オブジェクトを配置してなお残る欠損部をマスクとすることで復元領域の指定を自動化し、部分畳み込み演算により欠損部の影響を排除しつつ背景復元を行う。提案手法により補正した室内画像についてBRISQUEによる評価およびアンケートを行い、良好な結果を得た。

キーワード: GrabCut, Large mask inpainting, Augment Reality, Stable Diffusion

Implementation of a highly adaptive interior simulator using GrabCut and LAMA

Abstract: In this study, we propose a method to reduce the mismatch between 3D furniture objects and background images in an AR-based interior simulation system. The proposed method automatically restores the background that remains after the new 3D furniture objects are placed using the partial convolution operation. Finally, the created AR image is tuned by Stable Diffusion. BRISQUE evaluated the room images corrected by the proposed method, and a questionnaire survey was conducted, and the results were favorable.

Keywords: GrabCut, Large mask inpainting, Augment Reality, Stable Diffusion

1. はじめに

現実空間を映した映像内に仮想オブジェクトを重ね合わせる拡張現実 (Augment Reality) 技術を応用したアプリケーションが増加している。図1に示したMarket.us[1]による調査では、2033年までにAR市場規模は5,917億米ドルに達すると予想されている。また、2024年から2033年までの予測期間に39.5%の安定した年平均成長率を記録するとも推定されている。AR技術が発展・普及していく根拠として、多岐にわたる分野でARを用いたアプリケーションの成功が挙げられる。例を挙げると、

- ゲーム分野: Pokemon Go [2]

- 災害シミュレーション: 「災害体験AR」アプリ [3]
- プロモーション: Jack Daniel'sのARプロモーション [4]

など、枚挙に暇がない。様々な分野に応用されているARアプリケーションだが、中でも特に取り上げるべきなのが、インテリアコーディネート为目的としたARアプリケーションである。AR技術を用いたインテリアコーディネートアプリケーションでは、現実世界を映した映像内に仮想の3D家具オブジェクトをユーザー自身が部屋に配置することで、配置後の雰囲気や空間の調和を事前に視覚化することができる。Chakrabortyら[5]によって提案されたARIDでは、ユーザーが仮想的に家具を部屋に配置し、その配置後の見た目や雰囲気を事前に確認できる。同様に、Moralesら[6]が開発したInteriorARでは、カメラで撮影した現実空間の映像上に3D家具オブジェクトを配置し、家具を設置した場合の空間全体の様子を視覚的にブ

¹ 宮崎大学工学研究科
Graduate School of Engineering, University of Miyazaki,
Japan

² 宮崎大学工学教育研究部
Faculty of Engineering, University of Miyazaki, Japan

a) yamamori@cs.miyazaki-u.ac.jp

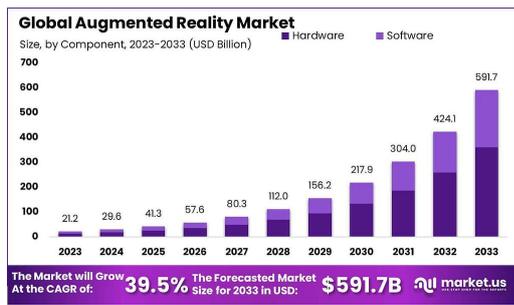


図 1: 世界の AR における市場規模予測 [1]
Fig. 1 Global Augmented Reality Market[1]

レビューする機能を提供する。こうした特徴から、AR 技術はオンラインショッピングのような場面において、ユーザーの意思決定を大きく支援することができる。インテリアシミュレーションという用途に着目すれば、空間の寸法やレイアウトの制約をリアルタイムで考慮したシミュレーションを提供可能であり、物理的な試行錯誤を減らすという利点もある。

一方、現実的かつ自由度の高い体験を提供するという観点では、これらのツールは依然として技術的な課題が多い。たとえば、3D 家具オブジェクトの陰影や光沢感を既存の空間に合わせることや、既存家具と仮想家具との重なり調整などはまだ不十分である。先述の ARID は仮想家具の配置を視覚化することに重点を置いているが、対象としている室内に既存家具がある場合、仮想家具を画像に配置した際に、既存家具との重なり処理に難があり、現実の状況に即した利用には限界がある。同様に Inter AR も既存家具や壁を適切に考慮する仕組みを備えていない。そのため、仮想家具と既存家具の間に不自然な重なりや視覚的な違和感が生じることがある。

AR を用いたインテリアシミュレーションにおいて、仮想家具配置の際の物理的制約を適切に反映する仕組みが欠けている点は、重要な問題である。このような欠点を克服するためには、既存家具を正確に認識し、仮想家具との相対的位置関係を考慮した AR 画像生成法が必要である。

インテリアコーディネートにおける上記の欠点を克服するには、画面中の既存家具を削除し、削除された空間の背景を自然に再構築し背景を再現、仮想家具を最適な配置し、陰影や光沢感を考えた技術が必要である。

本研究では、Rother ら [7] によって提案された GrabCut を用いた精密な画像セグメンテーションと、Suvorov ら [8] が開発した大規模マスクインペインティングアルゴリズム (LAMA) による深層学習ベースの背景補完を組み合わせたフレームワークを提案する。GrabCut は、ユーザーが指定した初期マスクや領域に基づき物体を抽出する手法である。グラフカットアルゴリズムに基づく最適化プロセスとガウス混合モデル (GMM)[9] を活用することで、対象物の輪郭を正確に識別し、不要な領域を除去する。LAMA は

畳み込みニューラルネットワーク (Convolutional Neural Networks, CNN) [10] を基盤としたインペインティング手法 [11] であり、欠損領域を学習済みのデータに基づいて予測し補完する。LAMA は、周辺ピクセルの特徴やパターンを分析し、視覚的に一貫性のある背景を再現する点で優れている。さらに、Runway ら [12] が開発した Stable Diffusion モデルを利用し、配置された 3D 家具オブジェクトの陰影や光沢感を調整する。これにより、3D 家具オブジェクトが周囲の環境に自然に溶け込み、実際の写真のような品質の画像を得ることができる。

GrabCut, LAMA, StableDiffusion の技術を統合することで、妨げとなる既存家具を画面上から除去し、背景をシームレスに復元しつつ、3D 家具オブジェクトをユーザーに違和感を与えることなく配置することが可能とする。

本論の構成は以下の通りである。第 2 章では、GrabCut, LAMA, Stable Diffusion について説明する。第 3 章では、提案するフレームワークを紹介し、その実装の詳細について述べる。第 4 章では、BRISQUE による定量評価とユーザーアンケートによる定性評価の 2 つの実験結果を示し、システムの性能を評価する。第 5 章で本結論を統括し、今後の課題について述べる。

2. 関連技術

2.1 GrabCut

GrabCut は、2004 年に Rother ら [7] の研究チームによって、デジタル画像内の前景と背景を効果的に分離するために提案された手法である。GrabCut は従来のグラフカットアルゴリズムを改良し、ユーザー操作を活用した効率的かつ高精度な前景と背景の分離を実現している。

GrabCut は、エネルギー最小化問題に基づく画像セグメンテーションアルゴリズムを利用しており、以下の 2 つの主要要素を組み合わせて動作する。

- **ガウス混合モデル (Gaussian Mixture Model:GMM)**[9]: 前景と背景の色分布をモデル化するために使用される。各ピクセルの色値は、前景および背景それぞれに属する複数のガウス分布に基づいて分類される。
- **マルコフ確率場 (Markov Random Field:MRF)**[13]: 隣接ピクセル間の関係をモデル化する要素であり、空間的一貫性を保つ役割を持つ。これにより、前景と背景の境界が滑らかになるようにセグメンテーション結果を調整する。

GrabCut は、式 (1) のエネルギー関数 $E(L, \theta, z)$ を最小化することで、前景と背景を分離する。

$$E(L, \theta, z) = U(L, \theta, z) + V(L, z), \quad (1)$$

ここで、 L は各画素が前景または背景のどちらに属するか

を示すラベル情報, θ は前景・背景の色分布を統計的にモデル化するための GMM のパラメータ, z は画素の集合で, 画像データそのものである.

- **データ項** $U(L, \theta, z)$: ガウス混合モデルを用いて, 各ピクセルの色情報を基に前景または背景のいずれかに属する確率をモデル化する項である:

$$U(L, \theta, z) = \sum_{i \in I} -\log P(z_i | \theta_{L_i}), \quad (2)$$

ここで, I は画像内のすべての画素の集合, z_i は画素 i の色情報 (ベクトル), θ_{L_i} は画素 i のラベル L_i に対するガウス混合モデルのパラメータ (ベクトル) である.

- **スムーズネス項** $V(L, z)$: 画素 i に隣接するピクセル j 間のラベル一貫性を確保する項である:

$$V(L, z) = \sum_{(i,j) \in N} \beta \cdot [L_i \neq L_j] \cdot \exp\left(-\gamma \cdot \|z_i - z_j\|^2\right), \quad (3)$$

ここで, β は隣接ピクセル間の不一致を抑制する重み, γ は色差の影響を調整するパラメータ, N は画像内で隣接する画素ペアの集合, L_j は画素 j のラベル, z_j は画素 i に隣接する画素 j の色情報 (ベクトル) を表す.

GrabCut のプロセスは, 以下の通りである.

- (1) **初期化**: ユーザーが画像上に矩形を描き, この矩形内を前景候補, 外側を背景候補とする.
- (2) **ガウス混合モデル**: 矩形内外のピクセルの色分布に基づき, 前景および背景の GMM パラメータを求める.
- (3) **グラフの作成**: 画像をグラフ構造に変換し, エッジの重みをデータ項とスムーズネス項で定義する. データ項 $E(L, \theta, z)$ の重みが大きいほど前景もしくは背景とのつながりが大きい. スムーズネス項 $V(L, z)$ の重みが大きいと隣接する画素間の色差が小さいことを意味する.
- (4) **グラフカットによる最適化**: 初期 GMM パラメータを基に, エッジの重みを利用して, エネルギー E を最小化し, グラフカットアルゴリズム [14] により初期セグメンテーションを取得する.
- (5) **反復処理**: セグメンテーション結果を用いて GMM パラメータを再計算し, ラベルを更新する処理を繰り返す. 以上の反復により, セグメンテーション結果が徐々に改善される.

2.2 LAMA

LAMA は, 2021 年に Suvorov ら [8] により提案された手法である. LAMA は, 従来の画像インペインティング手法と比較して特に大規模な欠損領域を精巧に復元する能力で注目を集め, 生成的手法と従来の畳み込み技術を融合し

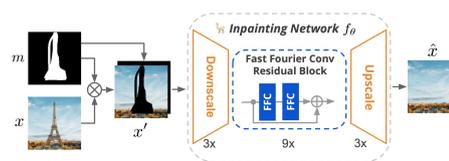


図 2: LAMA のプロセス [8]

Fig. 2 LAMA's process

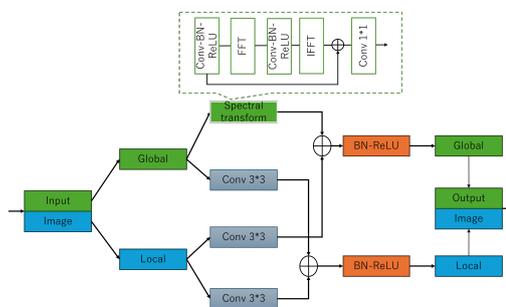


図 3: LAMA における FFC

Fig. 3 FFC in LAMA

た新しいアプローチが注目された.

LAMA は, ローカルおよびグローバルの両視点で画像の欠損領域を補完する設計を特徴とする. ローカルなテクスチャ情報とグローバルな構造情報を並行して処理することで, 復元領域が周囲と自然に融合することを目指している.

図 2 に, LAMA のインペインティングプロセスを示す. LAMA のプロセスは以下の 2 ステップから構成される.

- (1) **入力画像とマスク**: 欠損領域を示すマスク m と入力画像 x が与えられる. このとき, マスク m を適用した画像 x' を用いて処理を行う. LAMA の処理では, このマスクされた画像 x' を入力として使用する.
- (2) **ローカルおよびグローバル特徴の抽出**: 図 2 中の “Downscale” で画像を低次元表現に変換する. 次に, Fast Fourier Convolution (FFC) [15] を用いて画像全体の情報と局所的なテクスチャ情報を同時に抽出する. 最後に “Upscale” で低次元表現を元の解像度に戻す.

LAMA の補完能力を支える中心的な技術が, Fast Fourier Convolution (FFC) である. FFC は, 局所畳み込みとグローバル変換を組み合わせた手法であり, 広範囲の情報を活用して補完の質を向上させる.

図 3 に示すように, FFC は以下の 2 つの経路で構成される:

- **局所畳み込み (ConvLocal)**: 入力画像の局所的なテクスチャ情報を抽出する. 畳み込みカーネルを用いて, 隣接する画素間の関係を捉える.
- **グローバル変換 (FFT)**: Fast Fourier Transform (FFT) を用いて, 画像全体の構造情報を抽出する. LAMA のグローバル変換では, FFT により画像を周波数空間に変換し, 大域的な特徴を解析することで広範囲の情報を捉える.

FFC は式 (4) で定義される。

$$\mathbf{y} = \text{ConvLocal}(\mathbf{x}) + \text{FFT}(\mathbf{x}), \quad (4)$$

ここで、 \mathbf{x} は入力特徴マップ、 \mathbf{y} は FFC の出力を表す。

青色の Local Branch では、空間領域における 3×3 の畳み込み演算を通じて、近傍の詳細な特徴を抽出する。Local Branch での演算を式 (5) に示す。

$$\mathbf{y}_l = \text{ReLU}(\text{BN}(\mathbf{W}_{\text{local}} * \mathbf{x} + \mathbf{b}_{\text{local}})), \quad (5)$$

ここで、 $\mathbf{W}_{\text{local}}$ は 3×3 の畳み込みカーネル、 $\mathbf{b}_{\text{local}}$ はバイアス項、BN はバッチ正規化処理 (Batch Normalization)、 $\text{ReLU}(\cdot)$ は活性化関数である。

緑色の Global Branch は、周波数領域での特徴抽出を通じて広域的な構造情報を捉える。Global Branch のプロセスでは、初めに入力特徴マップ \mathbf{x} に 1×1 の畳み込みを適用し、バッチ正規化処理と ReLU 関数により式 (6) を用いて \mathbf{x}'_g を得る。

$$\mathbf{x}'_g = \text{ReLU}(\text{BN}(\mathbf{W}_{1 \times 1} * \mathbf{x} + \mathbf{b}_{1 \times 1})). \quad (6)$$

ここに、高速フーリエ変換 (FFT) を適用して \mathbf{x}'_g を周波数領域 \mathbf{X}_g に変換する。

$$\mathbf{X}_g = \text{FFT}(\mathbf{x}'_g). \quad (7)$$

続いて周波数領域での畳み込みを実行し、再びバッチ正規化処理と ReLU 関数を適用し、式 (8) により \mathbf{X}'_g を得る。

$$\mathbf{X}'_g = \text{ReLU}(\text{BN}(\mathbf{W}_{\text{fft}} \cdot \mathbf{X}_g + \mathbf{b}_{\text{fft}})), \quad (8)$$

ここで、 \mathbf{W}_{fft} は周波数領域の畳み込みカーネルであり、“ \cdot ” は要素ごとの積を表す。

\mathbf{X}'_g を逆フーリエ変換 (IFFT) で空間領域に戻して \mathbf{x}''_g を得る。

$$\mathbf{x}''_g = \text{IFFT}(\mathbf{X}'_g). \quad (9)$$

最後に、空間領域での 1×1 畳み込みを適用する。

$$\mathbf{x}'''_g = \mathbf{W}_{\text{global}} * \mathbf{x}''_g + \mathbf{b}_{\text{global}}. \quad (10)$$

Local Branch と Global Branch の出力は式 (11) により統合される。

$$\mathbf{y} = \text{ReLU}(\mathbf{y}_l + \mathbf{x}'''_g). \quad (11)$$

以上の処理により、局所的な詳細情報と大域的な情報を融合し、より自然で一貫性のある特徴マップが生成される。

2.3 StableDiffusion

Stable Diffusion は、Runway ら [12] が画像生成や編集のために開発したモデルであり、Latent Diffusion Models (LDMs)[16] を基盤としている。LDMs は、画像を直接ピクセル単位で処理するのではなく、事前学習済みのエンコー

ダを使って「潜在空間」と呼ばれる圧縮されたデータに変換する。これにより、データ量が大幅に減り、計算負荷が軽減される。画像を圧縮することでデータの次元を下げ、重要な特徴や構造を保持しながら効率的に処理を行う。その後、潜在空間上での操作結果を、事前学習済みのデコーダを使って高解像度の画像空間に復元する。この仕組みにより、計算負荷を抑えながらも高品質な画像生成を可能にしている。

Stable Diffusion の中心的な技術は、ノイズ除去を用いた拡散プロセスにある。まずランダムなノイズを入力として生成を開始し、ノイズを徐々に除去しながら、対象の構造や特徴を再現していく。この手法は、画像補完 (インペインティング) や解像度の向上、スタイルの変換など、幅広い応用が可能である。

本研究では、Stable Diffusion を活用して、画像内の照明や影のディテールを調整し、3D 家具オブジェクトをより現実的に表現する。この工程により、仮想オブジェクトと元の画像が自然に統合され、視覚的一貫性を保つ。

3. 提案手法

3.1 フレームワークの概要

提案するシステムの全体的なワークフローを図 4 に示す。この図 4 は、既存家具 (以下、オブジェクトと呼称する) のセグメンテーションから最終的な出力画像の生成までの処理の流れを表している。本フレームワークのプロセスは以下のステップで構成される。

- (1) **入力画像とオブジェクトの選択**: ユーザーが画像内のオブジェクトを選択し、オブジェクトを囲むバウンディングボックスを描画する。バウンディングボックスの手動設定により、システムは前景オブジェクトと背景を区別する。区別の後にユーザーによる入力をもとに前景と背景をより細かく分類することで、システムは対象オブジェクトの境界を正確に把握し、セグメンテーションの精度を向上させる。
- (2) **モルフォロジー変換によるオブジェクトセグメンテーション**: GrabCut を用いて、選択されたオブジェクトを画像から切り離し、初期のバイナリマスクを生成する。ただし、初期マスクには、オブジェクトと背景の類似性やセグメンテーション精度の限界によって、小さな欠損領域や境界の途切れが生じる場合がある。これらを補うために、膨張および収縮のモルフォロジー変換 [17] を適用する。膨張処理はマスクのエッジを拡張して欠損領域を埋め、収縮処理はノイズや不要なアーティファクトを除去する。これらの処理を組み合わせることで、不要なオブジェクトの除去と背景復元に適したクリーンなマスクを生成する。
- (3) **3D 家具オブジェクトの配置とマスクの改良**: 不要なオブジェクトの削除後、ユーザーは新しい家具の 3D

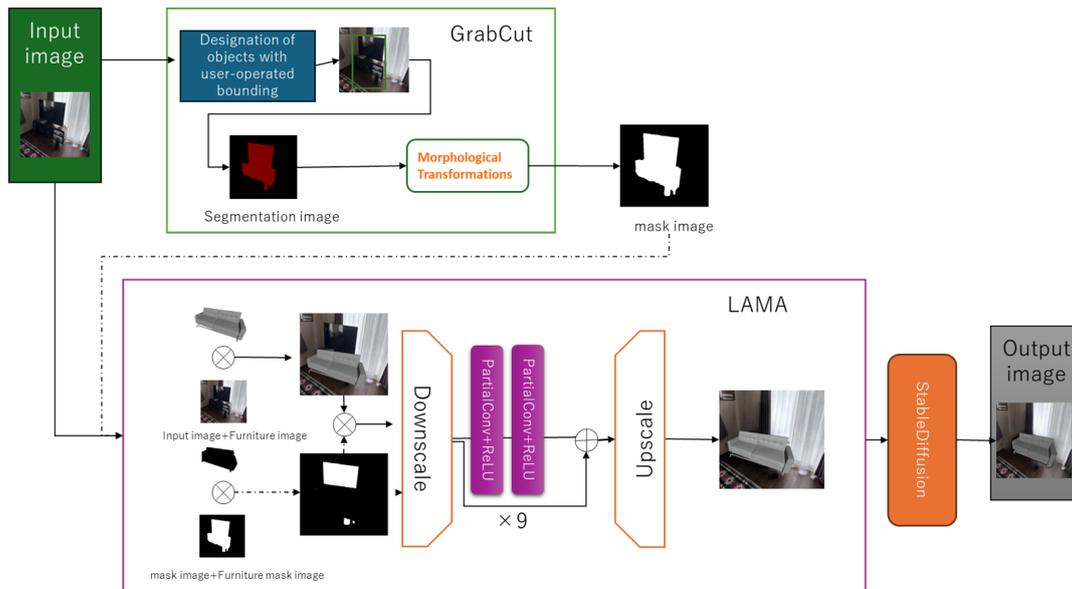


図 4: 提案システムの全体的なワークフロー
Fig. 4 Overall workflow of the proposed system

家具オブジェクトを配置する。この時、元の家具と重なる領域は LAMA によるインペインティング処理の障害物として扱われる。そのため、3D 家具オブジェクトの配置位置を反映した新しいマスクを生成し、既存のマスクに重ね合わせて重複領域を作成し、更新されたマスクと家具が配置された画像を使用して、LAMA によるインペインティングを実行する。

- (4) **LAMA による背景インペインティング**: 3D 家具オブジェクトを配置した新たな入力画像と、新しいマスクを LAMA で処理を行い、背景を補完する。この詳細は 3.2 節で説明する。
- (5) **品質評価と出力画像**: 背景復元と 3D 家具オブジェクトの配置が完了した後、Stable Diffusion を使用して画像を洗練する。改良するステップでは、陰影、光沢感、および 3D 家具オブジェクト周辺の細部を調整し、視覚的な一貫性を強化する。

3.2 オリジナル LAMA を用いた背景復元

3D 家具オブジェクトの配置とマスクの調整が完了した後、LAMA (Large-mask inpainting) を使用して背景を再構築する。ベースとする LAMA のバッチ正規化 (Batch Normalization, BN) と ReLU[18] を、部分畳み込み (Partial Convolution) と ReLU に置き換えることで、インテリアコーディネートに特化した処理を実現する。

3.2.1 マスク画像の更新

背景復元にオリジナル LAMA フレームワークを適用する前に、マスク画像を更新するための前処理ステップを導入する。従来の LAMA のようなインペインティングフレームワークは、入力画像とそのマスクだけを使い欠損領域を特定する。しかし、オリジナルの LAMA では、新た

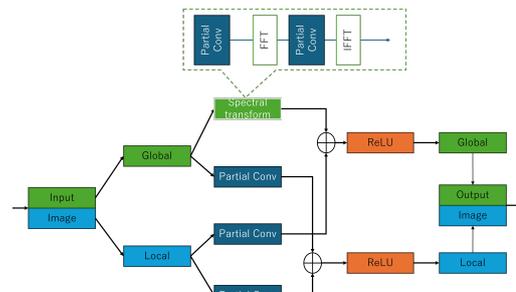


図 5: Partial Convolution+ReLU
Fig. 5 Partial Convolution+ReLU

に追加されるオブジェクトを考慮していないため、背景復元中に領域の重複や不整合が発生する可能性がある。

本研究で提案する手法では、3D 家具オブジェクトの配置をマスク生成プロセスに組み込むことでこの制約を克服する。まず、ユーザーが 3D 家具オブジェクトを画像内の所望の位置に配置し、新しいオブジェクトを設置する場所を指定する。このとき、3D 家具オブジェクトのマスクを元のセグメンテーションマスクに重ね合わせてマスク画像を更新する。更新されたマスクからは、新たに配置された 3D 家具オブジェクトに対応する領域がインペインティング処理から除外される。

新しい 3D 家具オブジェクトが存在する領域を明示的に除外することで、インペインティングはオブジェクト周囲の背景復元に集中できる。不必要な重複や矛盾する復元が排除されるため、最終画像の視覚的一貫性が向上する。更新されたマスクは、オリジナル LAMA フレームワークが新しい家具と調和した背景を生成する上で重要な役割を果たす。

3.2.2 部分畳み込み+ ReLU 処理

LAMA のアーキテクチャをインテリアコーディネータタスク向けに特化させるため、提案手法では図 5 のように局所のおよび、大域的な特徴抽出経路の両方に部分畳み込みを統合する。欠損領域を画像情報を持つピクセル（以下、有効ピクセルとする）に干渉させることなく処理できるようになり、シームレスな背景復元が可能になる。

局所特徴抽出経路では、従来のバッチ正規化を伴う畳み込みを部分畳み込みに置き換える。部分畳み込みは、入力特徴マップ x_l にバイナリマスク m を適用し、有効ピクセルのみを畳み込み結果に寄与させる。局所の特徴抽出経路での演算を式(12)に示す。

$$x'_l = \text{ReLU} \left(\frac{W_l * (x_l \cdot m)}{\text{Sum}(m) + \epsilon} \right), \quad (12)$$

ここで、

- W_l : 畳み込みカーネルの重み。
- $*$: 畳み込み演算。
- \cdot : 要素ごとの積。
- m : バイナリマスク（有効ピクセルは 1、欠損ピクセルは 0）。
- ϵ : ゼロ割りを防ぐための小さな定数。
- $\text{Sum}(m)$: バイナリマスク m 内の有効ピクセル数。

大域的特徴抽出経路では、部分畳み込みを空間領域と周波数領域の両方で適用する。まず、入力特徴マップ x_g に部分畳み込みを適用して空間的特徴を抽出する。

$$x''_g = \text{PartialConv}(x_g, m). \quad (13)$$

次に、得られた特徴を高速フーリエ変換 (FFT) によって周波数領域に変換する。

$$X_g = \text{FFT}(x''_g). \quad (14)$$

その後、周波数領域でも部分畳み込みを適用する。

$$X'_g = \text{PartialConv}(X_g, m). \quad (15)$$

最終的に、逆フーリエ変換 (IFFT) を用いて周波数領域から元に戻す。

$$x'_g = \text{IFFT}(X'_g). \quad (16)$$

局所的経路と大域的経路の出力 x'_l と x'_g を結合し、ReLU 活性化関数を適用することで最終的な復元結果が得られる。

$$y = \text{ReLU}(x'_l + x'_g). \quad (17)$$

以上処理により、局所的な細かいテクスチャとグローバルな構造を統合し、インテリアコーディネータタスクにおける背景復元を効率的かつ自然に行う。

3.3 Dreambooth を用いた Stable Diffusion の適用

Dreambooth[19] は、Stable Diffusion のような事前学習済みの画像生成モデルを特定のオブジェクトやスタイルに適応させるためのファインチューニング技術である。少量のデータセット（5~20 枚の画像）を用いてモデルを微調整し、形状、テクスチャ、素材特性といった対象物の特徴を学習させる。

本研究では、3D 家具オブジェクトを様々な視点からレンダリングし、その特徴を Dreambooth を通じて Stable Diffusion に学習させる。Dreambooth の手法により、3D 家具オブジェクトがリビングルームや屋外スペースなど異なる環境に自然に溶け込む画像を生成する。これにより、3D 家具オブジェクトを直接配置した際に生じる陰影や光沢感の不整合、遠近感の歪みといった視覚的な問題を補正する。

4. 評価実験

本章では、提案フレームワークを定量的、および定性的な手法で評価する。定量的評価では、BRISQUE [20] を用いて、元の LAMA と提案手法による出力画像の品質を比較する。一方、定性的評価では、ユーザー調査を通じて幾何学的整合性、光学的整合性、全体的な満足度などの観点から評価する [21]。評価には、20 枚の画像（オリジナル画像 10 枚、Kaggle の Living Dataset [22] から 10 枚）と、CGTrader [23] から取得した 20 個の 3D 家具オブジェクトを使用する。また、今回は作成された 20 枚の中から評価の高いもの、低いもの、中間のもの二つずつの紹介、評価する。

4.1 定量的評価のための BRISQUE

BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) は、Mittal ら [20] が提案した手法である。BRISQUE は参照画像を必要とせずに画像の知覚品質を評価する無参照型の評価指標であり、空間領域での自然シーン統計 (NSS) からの偏差を分析することで自然な画像として評価する。

BRISQUE は、平均差分コントラスト正規化係数などの特徴を抽出し、それらをガウス分布でモデル化することで知覚品質を定量化する。スコアが低いほど画像品質が高いことを示し、これは人間の視覚的感覚と一致する。

本研究では、元の LAMA と提案フレームワークで処理された画像の品質を比較するために BRISQUE を使用する。

4.2 定量的評価結果

image1 に関しては、提案手法が 1.09、LAMA が 1.71 であり、約 0.61 ポイントの改善がみられる。一方で、image2 では提案手法が 26.31、LAMA が 26.41 であり、スコアの差は非常に小さい。これは、画像の特性によっては LAMA



図 6: 元画像 (左から image1,image2,image3,image4)

Fig. 6 Input Images(From left: image1, image2, image3, image4)



図 7: 処理後画像 (左から output1, output2, output3, output4)

Fig. 7 Output Images(From left: output1, output2, output3, output4)

表 1: BRISQUE による自然画像スコア

Table 1 BRISQUE Scores for Image Quality Evaluation

Image	Proposal Process	LAMA Process	Discrepancy
image1	1.09	1.71	-0.61
image2	26.31	26.41	-0.10
image3	26.61	29.93	-3.31
image4	56.39	53.94	2.45

のみでも十分な品質が確保される場合があることを示唆している。

image3 では、提案手法が 26.61, LAMA が 29.93 であり、約 3.31 ポイントの大幅な改善が見られた。これは、LAMA よりも提案手法の方が、より自然なインペインティングが行われた可能性を示している。

一方で、image4 では提案手法が 56.39, LAMA が 53.94 となっており、BRISQUE 値は約 2.45 ポイント低下した。この結果は、シンプルかつ家具が少ない部屋においては、提案手法が LAMA よりも画質を低下させる可能性があることを示唆している。

4.3 ユーザーによる定性的評価

ユーザー評価の結果は、幾何学的整合性、光学的整合性、満足度の 3 つの観点で示す。幾何学的整合性は家具モデルの形に違和感があるかどうか、光学的整合性は家具モデルの色や光に違和感があるかどうか、満足度は雰囲気マッチしているかについてアンケートを取る。評価スケールは 1 から 5 までで、5 が最も高い品質である。図 6(a), 図 6(b), 図 6(c) にユーザアンケートの結果をグラフに示す。

4.4 定性的評価結果

幾何学的整合性では、output1 と output3 の評価は高かったが、output2 と output4 の評価は低い。特に out-

put4 では、オブジェクトの遠近感やサイズが背景と適切に調整されていないことが低評価の要因と考えられる。光学的整合性では、output1 や output3 が高評価を得た一方で、output4 では評価が低い。これは、光の当たり方や陰影がユーザー評価によって違和感があると評価された可能性を示唆する。満足度評価では、output1 は高い評価を得ており、提案手法により背景復元やオブジェクト統合の精度が向上していることが示された。一方、output4 では低い評価が多く、3D 家具オブジェクトが部屋とマッチしていない可能性がある。

5. まとめ

本研究では、インテリアコーディネートシステムにおいて、既存家具と仮想家具との重なり調整問題と、照明や影の不整合による視覚的違和感の問題に着目した。課題を解決するため、GrabCut によるオブジェクトセグメンテーションを用いて既存家具を除去し、オリジナル LAMA による背景復元で一貫性を確保しながら 3D 家具オブジェクトを配置するフレームワークを提案した。さらに、Stable Diffusion を活用し、3D 家具オブジェクトの陰影や光沢感を調整することで、背景との視覚的調和を実現した。

提案手法の有効性は、BRISQUE スコアによる定量評価とユーザー調査による定性評価で確認された。特に複雑な背景や細かいテクスチャを持つシナリオでは、従来手法を上回る品質向上が見られた。一方で、一部のシナリオでは形状保持や照明調整に課題が残ることが明らかとなった。

本研究は、インテリアコーディネートの自由度とリアリティを向上させる新たなアプローチを提示した。今後は、提案手法の精度向上や Stable Diffusion の改良を進める。

謝辞 先生のおかげで無事終わることが出来ました。

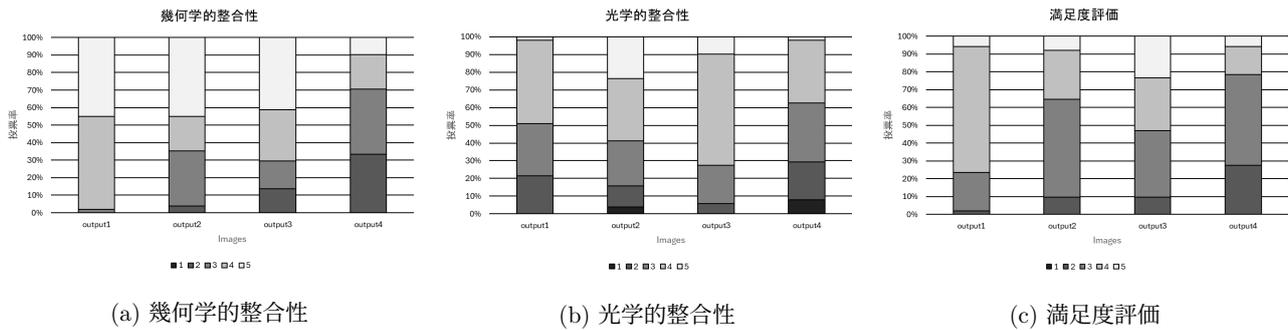


図 8: ユーザーアンケートによる定性評価

Fig. 8 Qualitative evaluation based on user questionnaires

参考文献

[1] Market.us: “Augmented Reality Statistics: Trends and Insights,” 2025. Available: 入手先 (<https://scoop.market.us>).

[2] Niantic, Inc.: “Pokémon GO: Augmented Reality Game by Niantic, Inc.,” 2016. Available: 入手先 (<https://nianticlabs.com>).

[3] 東京海上日動: “自治体のオープンデータを活用した「災害体験 AR」の高度化,” 2021. Available: 入手先 (<https://www.tokiomarine-nichido.co.jp>).

[4] Jack Daniel’s: “Interactive AR App: Showcasing Brand History through Bottle Label Scanning,” 2025. Available: 入手先 (<https://www.xr-lifedig.com>).

[5] Chen, X. Y. and Kanaparan, G.: “ARID—An Augmented Reality Mobile Application for Interior Design,” in *Resilient and Responsible Smart Cities*, Springer International Publishing, 2023, pp. 3–17.

[6] Moares, R., Jadhav, V., Bagul, R., Jacob, R., Rajguru, S., and Rakhi, K.: “Inter AR: Interior Decor App Using Augmented Reality Technology,” *SSRN Electronic Journal*, 2020. Available: 入手先 (<https://api.semanticscholar.org/CorpusID:213232720>).

[7] Rother, C., Kolmogorov, V., and Blake, A.: “GrabCut: Interactive Foreground Extraction Using Iterated Graph Cuts,” in *ACM SIGGRAPH 2004 Papers*, ACM, New York, 2004, pp. 309–314.

[8] Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., and Lempitsky, V.: “Resolution-robust Large Mask Inpainting with Fourier Convolutions,” arXiv preprint, 2021. Available: 入手先 (<https://arxiv.org/abs/2109.07161>).

[9] Reynolds, D. A.: “Gaussian Mixture Models,” in *Encyclopedia of Biometrics*, Springer, Berlin, 2009, pp. 741–663.

[10] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D.: “Back-propagation Applied to Handwritten Zip Code Recognition,” *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[11] Bertalmio, M., Bertozzi, A. L., and Sapiro, G.: “Navier-Stokes, Fluid Dynamics, and Image and Video Inpainting,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, vol. 1, 2001, pp. I-I.

[12] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B.: “High-Resolution Image Synthesis With Latent Diffusion Models,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 10684–10695.

[13] 片岡 駿: “マルコフ確率場における統計的機械学習,” *電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review*, vol. 11, no. 4, pp. 256–, 2018.

[14] Boykov, Y., Veksler, O., and Zabih, R.: “Fast approximate energy minimization via graph cuts,” in *Proceedings Eighth IEEE International Conference on Computer Vision (ICCV 2001)*, vol. 1, pp. 377–384, 2001.

[15] Fuoli, D., Van Gool, L., and Timofte, R.: “Fourier Space Losses for Efficient Perceptual Image Super-Resolution,” in *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 2340–2349.

[16] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B.: “High-Resolution Image Synthesis with Latent Diffusion Models,” *CoRR*, vol. abs/2112.10752, 2021. Available: 入手先 (<https://arxiv.org/abs/2112.10752>).

[17] Fisher, R., Perkins, S., Walker, A., and Wolfart, E.: “Morphological Operations,” 1997. Available: 入手先 (<https://homepages.inf.ed.ac.uk/rbf/HIPR2/morops.htm>).

[18] Nair, V. and Hinton, G. E.: “Rectified Linear Units Improve Restricted Boltzmann Machines,” in *Proc. 27th Int. Conf. Machine Learning (ICML)*, Haifa, Israel, 2010, pp. 807–814.

[19] Ruiz, R., Li, S., and Belkacem, A. A.: “DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation,” arXiv preprint, 2022. Available: 入手先 (<https://arxiv.org/abs/2208.12242>).

[20] Mittal, A., Moorthy, A. K., and Bovik, A. C.: “No-Reference Image Quality Assessment in the Spatial Domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[21] 田中 智己, 中林 拓馬, 加戸 啓太, and 平沢 岳人: “Web アプリケーションへ展開した AR インテリアシミュレータへの評価の分析,” *日本建築学会計画系論文集*, vol. 79, no. 698, pp. 1063–1069, 2014.

[22] Kaggle: “Room Dataset: High-Quality Interior Images for Vision Tasks,” 2020. Available: 入手先 (<https://www.kaggle.com/datasets>).

[23] CGTrader: “3D Models Marketplace for Designers and Artists,” 2023. Available: 入手先 (<https://www.cgtrader.com>).