

輪郭線からの雲画像生成と潜在表現による類似画像検索システムの構築 Cloud image retrieval system based on image generation from contours and latent representation of cloud images

中村 薫[†] 澤山 正貴[†] 土橋 宣典^{†‡}
Kaoru Nakamura Masataka Sawayama Yoshinori Dobashi

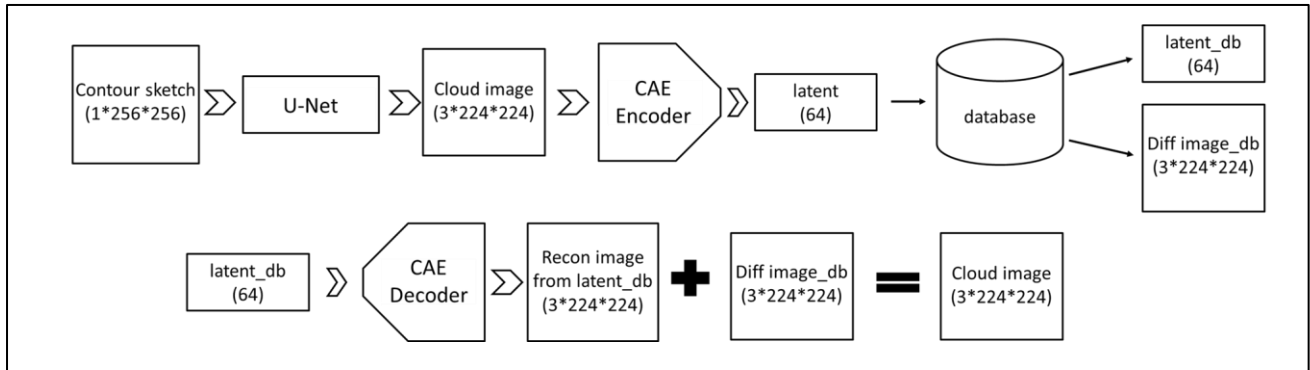


図 1 システムの概要

1. はじめに

CG(コンピュータグラフィックス)を用いて屋外シーンを生成する際、雲は写実性を高める上で非常に重要な役割を果たす。そのため、3次元で写実的な雲のシミュレーションを行う手法が数多く提案されている。しかし、これらの手法を用いて適切な雲を生成するためにはユーザが多くのパラメータを細かく調整する必要があり、初心者にとっては難しい問題となる。よって、ユーザがより直感的に求める雲の3次元モデルを得る手法が求められている。本研究はこの課題を解決する足掛かりとして、まずは2Dの雲画像に着目し、ユーザが描いた輪郭線から目的の雲画像を効率的に検索できるシステムの構築を試みる。

近年、深層学習技術の発展が進んでおり、中でも生成モデルによる画像生成の進歩は目覚ましい。そこで、本研究ではこの生成モデルを利用し、ユーザが描いた輪郭線から画像を生成するネットワークと、画像から特徴量を抽出するネットワークを用いて検索システムを構築する。従来の深層学習を用いた画像生成研究と比較して、高品質な画像生成ではなく検索システムの構築が目的であり、検索システムにとって十分な画像を生成できる範囲であれば複雑なネットワーク構造や計算量を必要とせず、簡易的なネットワークで実装できることが本研究の独自性となっている。

2. 関連研究

雲のような3Dモデルを生成する手法の研究はいくつか存在する。例えば、Miyazakiらは流体シミュレーションにより雲の3D形状を生成する手法を提案している[1]。Wietherらはユーザが描いたスケッチを入力とした雲の3Dモデリング手法を提案している[2]。一方でこれらの手法で写実的な雲を生成するためにはシミュレーションのパラメータを適切に設定したり、適切な輪郭を描いたりする必要があり、初心者にとっては難しい問題となる。

深層学習技術を用いた雲の3Dモデル生成に関連する研究としてはLinらのものが挙げられる[3]。この研究では入

力として少数のステレオ画像のペアを受け取り、それらから雲の密度分布を推定するネットワークが構成されている。この手法の課題として、複数の画像ペアが必要である点や密度分布の詳細成分(高周波成分)が再現できず、生成された雲が写実的とはいえないことがある。

雲画像の検索システムに関する研究としてTaoらによるSkyFinderが挙げられる[4]。このシステムは空の画像を検索するシステムを構築しており、青空や曇り空などのカテゴリ、レイアウト、太陽の位置などの空の属性を抽出し、これらの組み合わせにより画像検索を可能にしている。また、空画像のグラフを構築することで対応する属性空間内での画像探索をスムーズに行える。

ユーザが描いたスケッチから3Dモデルを検索する研究としてはSuzukiらのものが挙げられる[5]。このシステムはユーザが描いた2次元の輪郭線からクエリとなる3次元形状を生成し、その形状からデータベースに保管されている3次元雲モデルの中から適切なものを検索する。本研究はこのシステムをベースとし、クエリ形状生成や既存の雲3Dモデルの特徴抽出部分に深層学習技術を用いた検索システムの構築を目指し、その足掛かりとして深層学習技術を用いた雲画像検索システムを構築する。

3. 提案手法

本システムの概要を図1に示す。データベースには雲画像データに対して畳み込みオートエンコーダ(CAE)で抽出された潜在表現と(元画像 - 潜在表現から再構成された画像)により計算される差分画像が保存されている。

システムが動作する際、まずユーザが描いた輪郭スケッチを輪郭線-画像生成ネットワークに入力することで雲画

[†]北海道大学 Hokkaido University

[‡]プロメテック CGリサーチ Prometech CG research

像が生成される。この画像を CAE のエンコーダに入力することで潜在表現を得る。この潜在表現をデータベースにある潜在表現と比較し、最も近い潜在表現とそれに対応する差分画像を取り出す。この際コサイン類似度で潜在表現同士の類似度を測っている。次に取り出した潜在表現を CAE のデコーダ部分に入力して再構成画像を生成して、最後に再構成画像と差分画像を加算することで検索結果となる画像を推定する。

以下では、CAE と輪郭線-画像生成ネットワークについて詳しく説明する。

3.1 CAE

本手法において、CAE は畳み込みブロック 5 個と全結合層から構成されたエンコーダと同様に構成されたデコーダからなる簡易的なネットワークである。各畳み込みブロックは畳み込み層と活性化関数からなり、エンコーダにおいてはそれらの間にバッチノーマライゼーションが行われる。活性化関数としては最終出力には Sigmoid を、それ以外には LeakyReLU を用いている。また、エンコーダの全結合層の後にドロップアウトを確率 $p = 0.2$ で行い、汎化性能を高めている。入力は 224×224 のカラー画像であり、潜在表現は 64 次元とした。

3.2 輪郭線-画像生成ネットワーク

本手法では輪郭線から画像を生成するネットワークとして U-Net[6]を基盤として用いた。U-Net とはセグメンテーション用に構築された畳み込みニューラルネットワーク(CNN)である。エンコーダとデコーダからなるモデルであり、特徴としてエンコーダから得られた中間特徴をデコーダに渡す Skip connection を採用することで細かな情報を失いにくくなっている。本手法においては畳み込み層とバッチノーマライゼーション、及び活性化関数からなる畳み込みブロックを 7 個積み重ねたものをエンコーダ及びデコーダとして用いた。活性化関数は最終出力には Tanh を、それ以外には LeakyReLU を用いている。入力は 256×256 の輪郭画像であり、出力は 256×256 のカラー画像である。

また、学習方法として pix2pix[7]をベースとしたアルゴリズムを用いた。pix2pix は conditional GANS(cGAN)を活用したアルゴリズムであり、生成器に U-Net を用い、識別器として畳み込み層により構成された Patch GAN を用いる。生成器に与えられる条件として、本システムにおいてはユーザが描いた輪郭が与えられる。学習時には既存の雲画像のデータセットから抽出した輪郭を条件として生成器に与え、そこから再構成された画像を識別器に入力した際の敵対的損失と、元の画像と再構成画像との L1 損失を加算することで全体の損失とした。また、PatchGAN は U-Net と同様の畳み込みブロックを 3 個積み重ねたものを用いた。

4. 結果

この節では提案手法の有効性を検証するために実施した実験とその結果について述べる。実験は CPU が Intel Xeon w7-2495X、GPU が NVIDIA RTX A5500 の計算機にて行った。学習に用いるデータセットとしては高画質な画像を提供するプラットフォームである Unsplash が提供している Unsplash API により雲の画像を 4800 枚取得し用いた。また、汎化性能を高めるためにランダムで学習画像に平行移動や



水平反転、輪郭に対するブラーを行うことでデータ拡張を

図 2 各輪郭と再構成画像

行い、全体で 14400 枚とした。

学習を行った後に、学習に使用した輪郭並びに汎化性能を測るためにその輪郭にノイズを付加したものをを用いて正しい画像が出力されるか検索精度を計測した。ノイズとしては輪郭を一定確率でマスクするもの、輪郭にブラーをかけて弱めの閾値で 2 値化するもの、輪郭の解像度を下げよう一度元の大きさにリサイズするものの 3 つを利用した。図 2 に元画像及び各輪郭とそれらにおける生成画像を示す。検索精度は操作がない輪郭及びマスクした輪郭については 98.8%、ブラーをかけた輪郭については 85.6%、解像度を削減したものについては 91.0%となった。

学習に用いた輪郭についても精度が 100%にならなかった理由として、雲が薄く広がっている画像など輪郭が適切に抽出されていなかった画像が複数個あったため、その影響が考えられる。また、用いるデータセットに背景などの統一性を持たせることでさらなる精度向上が期待できる。

5. おわりに

本研究では輪郭線を入力として適切な雲画像を検索するシステムを深層学習技術により実装した。今後はこのシステムを基盤として 3D モデル検索システムを構築する。

参考文献

- [1] Miyazaki Ryo, et al. "Simulation of cumuliform clouds based on computational fluid dynamics", Eurographics 2002 - Short Presentations (2002).
- [2] Wither Jamie, et al. "Rapid Sketch Modeling of Clouds", Proceedings of the Fifth Eurographics Conference on Sketch-Based Interfaces and Modeling, pp113-118 (2008)
- [3] Lin Jacob, et al. "Volumetric Cloud Field Reconstruction", arXiv preprint arXiv:2311.17657 (2023)
- [4] Tao Litian, et al. "SkyFinder: Attribute-based Sky Image Search", ACM Transactions on Graphics, Vol.28, No.3 (2009).
- [5] Suzuki Kei, et al. "A sketch-based system for cloud volume retrieval from simulated dataset for realistic image synthesis", Proceedings of the 14th ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry, pp51-54 (2015).
- [6] Olaf Ronneberger, et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation", Medical Image Computing and Computer-Assisted Intervention 2015, pp234-241 (2015).
- [7] Philip Isola, et al. "Image-to-Image Translation with Conditional Adversarial Networks", 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp5967-5976 (2017).