

機械学習を用いたCygnus-X領域における大質量星に付随する分子雲の特性の調査

Investigation of the Characteristics of Molecular Clouds Associated with Massive Stars in the Cygnus-X Region Using Machine Learning

生命環境科学域 理学類 物理科学課程 電波天文学研究室 1211305117 藤本湧大 Fujimoto Yudai

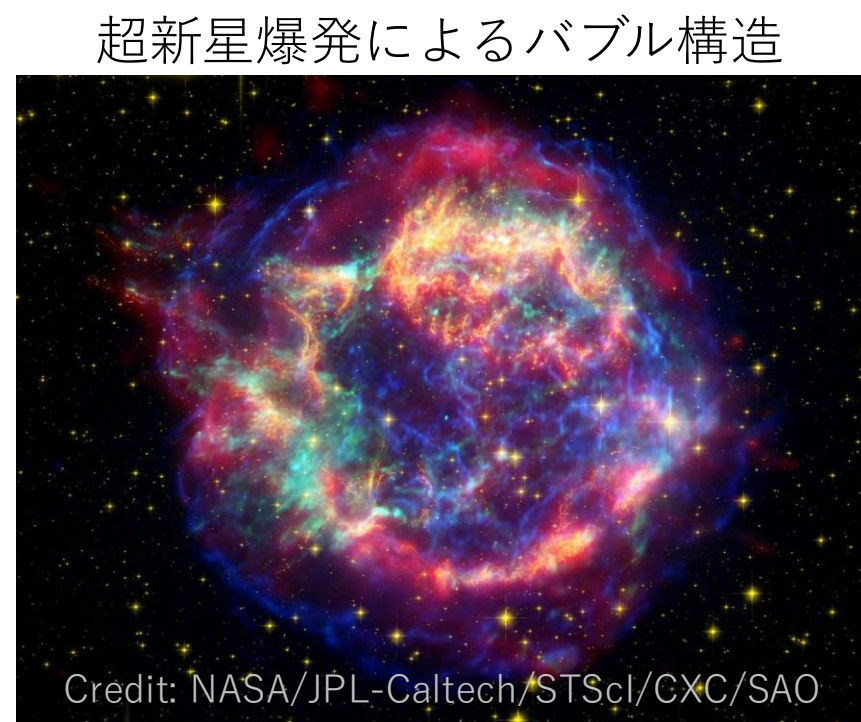
1. Abstract

近年のJWSTによる観測や深層学習を用いた観測データ解析により、銀河には超新星爆発や大質量星の輻射で形成されるバブル構造状の天体が多く存在することが分かってきた。この観測結果は、Inutsuka et al. 2015で主張されているように、バブル構造が起点となり次世代の星形成が行われていることを示唆している。本研究では、機械学習を用いた画像再構築手法の一つであるCAEを用いて、赤外線観測でバブル状の構造が見られる領域の分子雲の特性を評価した。分子雲の特性は、CAEの画像再構築の際に生じる、圧縮された入力データの特徴量を比較することによって評価した。評価の対象領域は、星形成が活発な領域であるCygnus-X領域で、CAEの学習データや評価データも同領域から作成した。学習したCAEを、バブル構造周辺の領域とCygnus-X領域全体に適用した結果、多くの特徴量が異なることが確認できた。同様の比較を、ランダムに選出した領域で行ったところ、特徴量で違いが見られるものはごく少数であった。この結果から、少なくともCygnus-X領域では、赤外線バブルを含む領域と他領域の分子雲は異なる特性を持つと考えられる。

2. Introduction

◆大質量星

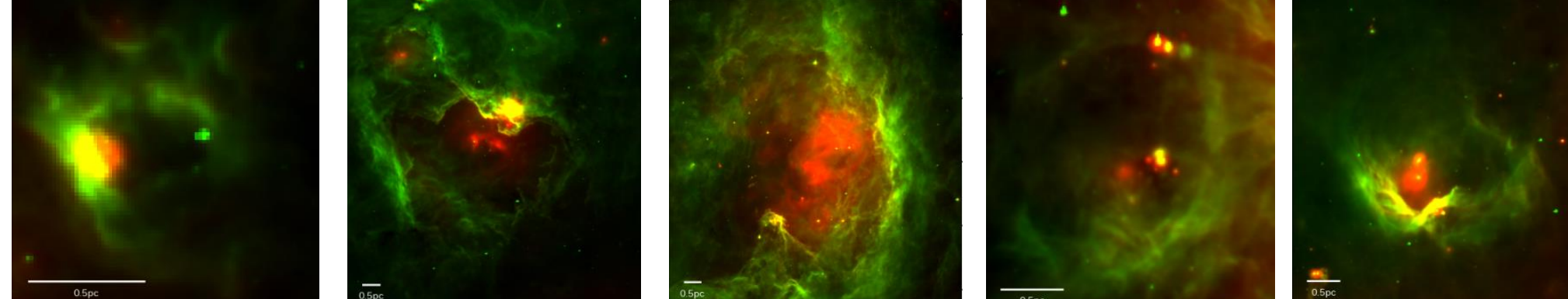
- 銀河が成長していくうえで重要な役割を果たしている
- 輻射により周囲の分子雲を押し除ける
→バブル構造が発生
- 一生の終わりに超新星爆発で星間媒質をばら撒く
→バブル構造が発生
- 太陽の8倍以上の質量を持つ星
- 通常の星形成シナリオでは誕生するとは考えにくい
→自己収縮以外の形成シナリオが必要



◆赤外線バブル構造

- 近年の観測データ解析により多数検出
- 大質量星からの紫外線放射や星風により形成

赤外線バブルの例：



Credit: NASA/spitzer space telescope

◆大質量星の形成シナリオで有力な説

- 分子雲同士の衝突によって引き起こされる
・分子雲の衝突によって生じた高密度領域で大質量星が形成
Fukui et al. 2021

- 大質量星からの輻射による新たな星形成
・大質量星からの輻射により周囲の分子雲が押し除けられる
・バブル状に圧縮された分子雲で新たな大質量星が形成
Elmegreen & Lada et al. 1977

◆星間空間に広がる分子雲

星形成は、大質量星の性質や形成シナリオを考慮すると、大質量星が形成するバブル構造による分子雲の衝突や圧縮が繰り返されることによって進んでいると考えられる。



赤外線バブルが見られる領域の分子雲は大質量星から影響を受けている可能性が高い

目的

赤外線バブルが見られる領域と他領域の分子雲では空間分布と速度分布の特徴が異なるのか統計的に判断する

◆問題点

- x, y, vの3次元データで空間的解析をする必要がある
- 膨大な天文データの三次元的解析は非常に困難
→機械学習を用いて分子雲の特性を評価

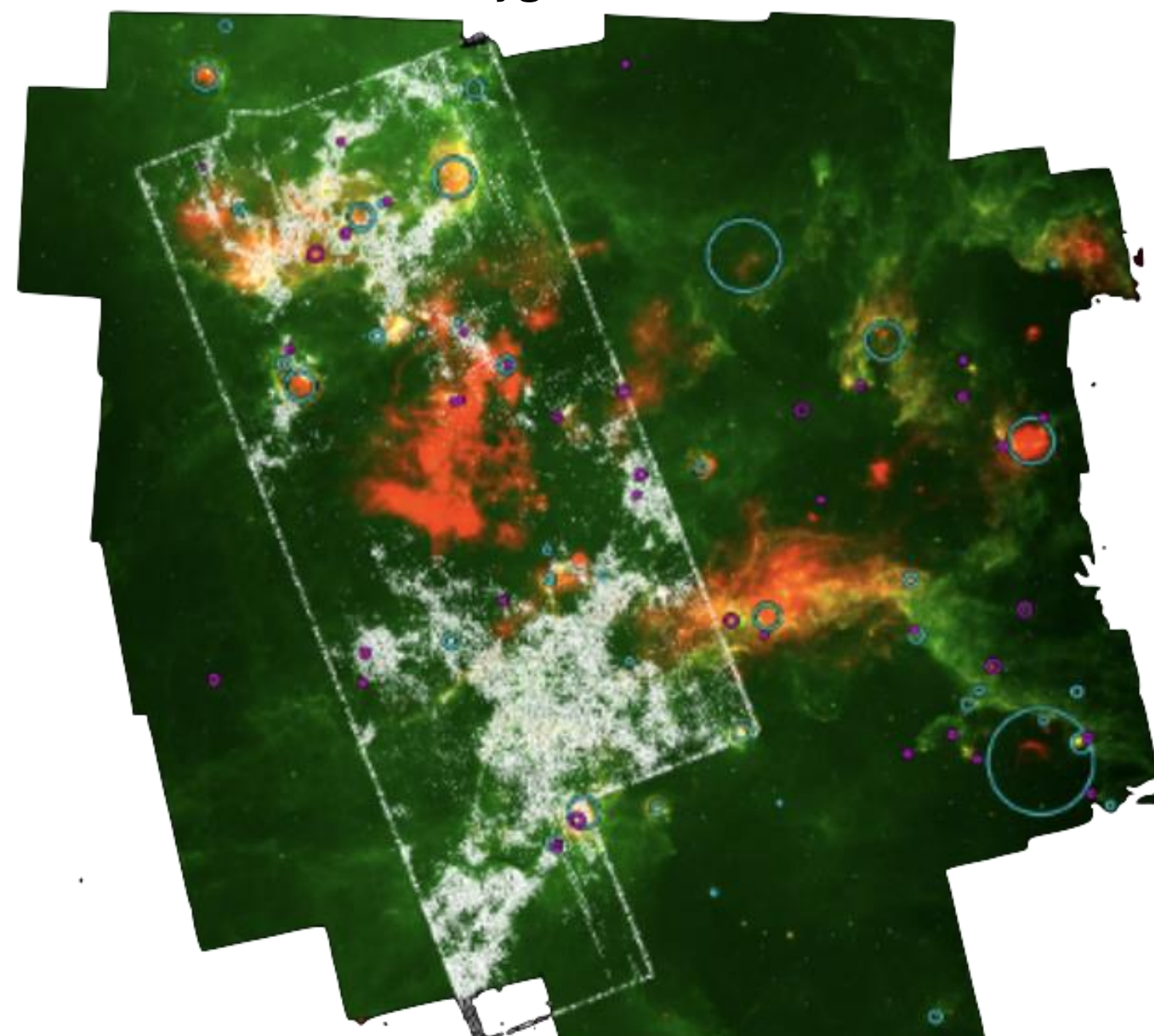
3. Verification Methods

◆検証領域

大質量星から影響を受けている分子雲(以降バブル領域)とその他の領域で特徴が異なるのかどうかの検証をCygnus-X領域にて行った。この領域には以下の特徴がある。

- 星形成が活発に行われている場として有名
- 分子雲のトレーサーである $^{12}\text{CO}(J=1-0)$ 輝線の、空間x, yの2次元、速度1次元を含む、3次元データが存在
- 先行研究によりバブル構造が多数検出済

Cygnus-X領域



白：今回使用した領域

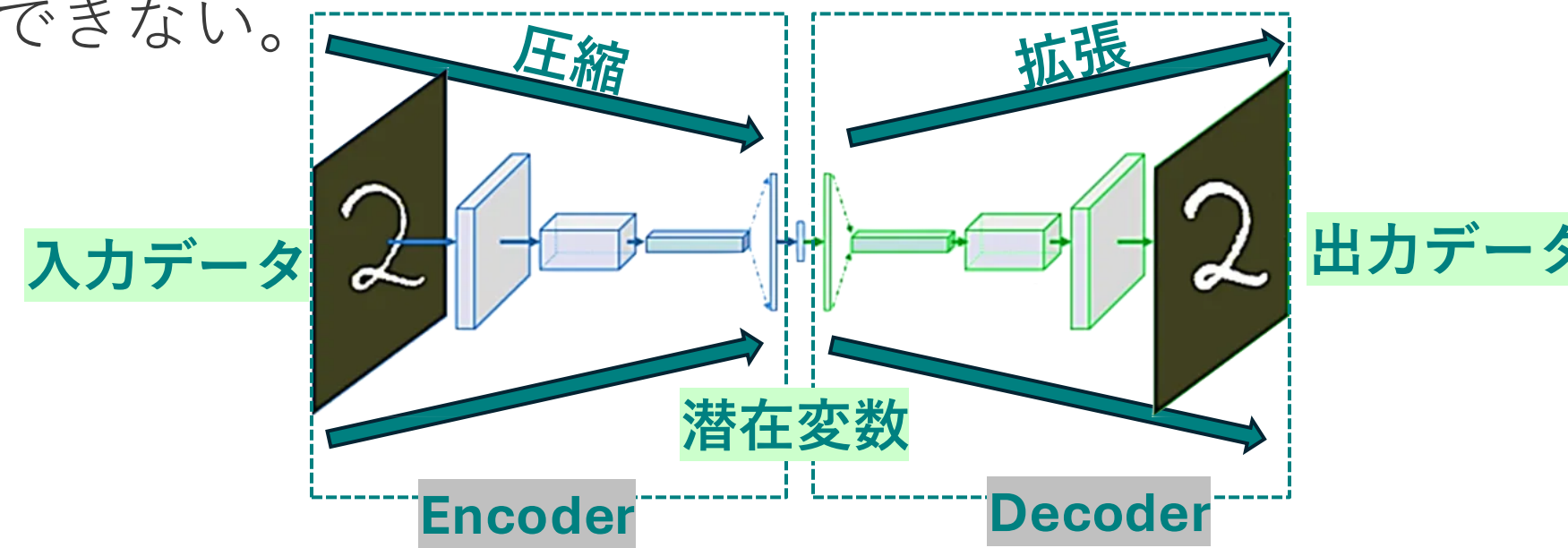
シアン：Tharindu et al. 2019

マゼンダ：Nishimoto et al. 2025

Credit: NASA/spitzer space telescope

◆機械学習モデル

今回使用した機械学習モデルはCAE(Convolutional Autoencoder)というものである。このモデルは入力データの特徴量を、潜在変数と呼ばれる任意の数の変数に圧縮することができる。今回は潜在変数の数を300、200、100個でそれぞれ実験を行った。潜在変数の数は多すぎると不要な特徴まで抽出してしまい、少なすぎると十分に特徴を抽出できない。

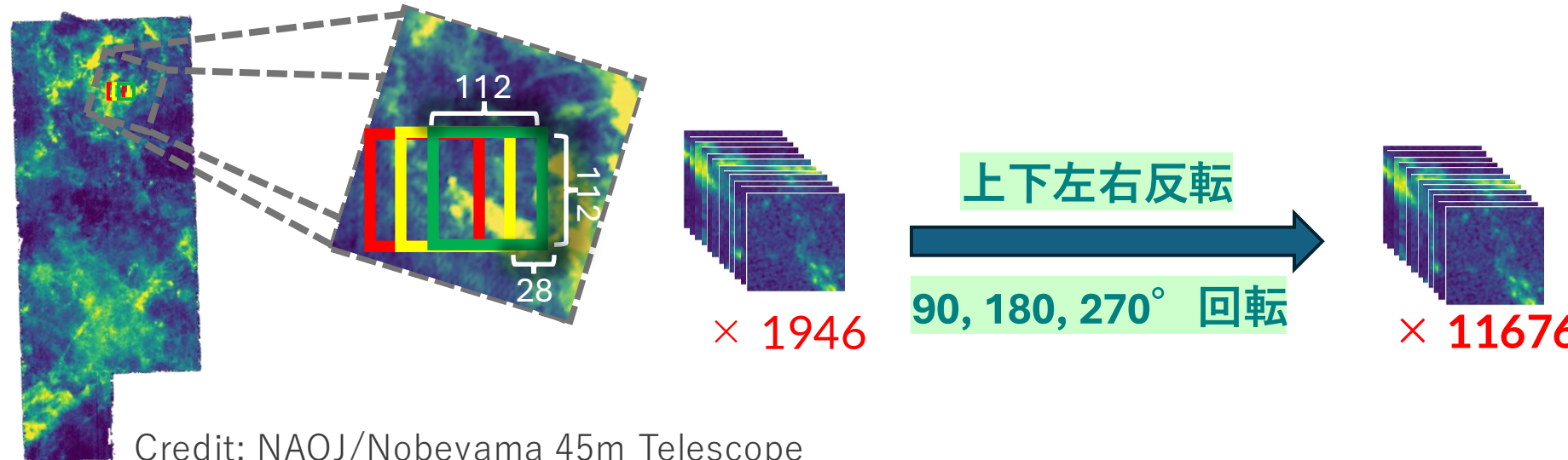


◆学習データ

学習データは以下のように作成した。

- 112pix×112pixを28pixずつ移動し切り取り(計1946個)
- エラーデータの処理
- 強度を0~1で正規化
- 回転、反転したデータを作成

最終的なデータの形状は112×112×12pix
データ数は11676個となった。



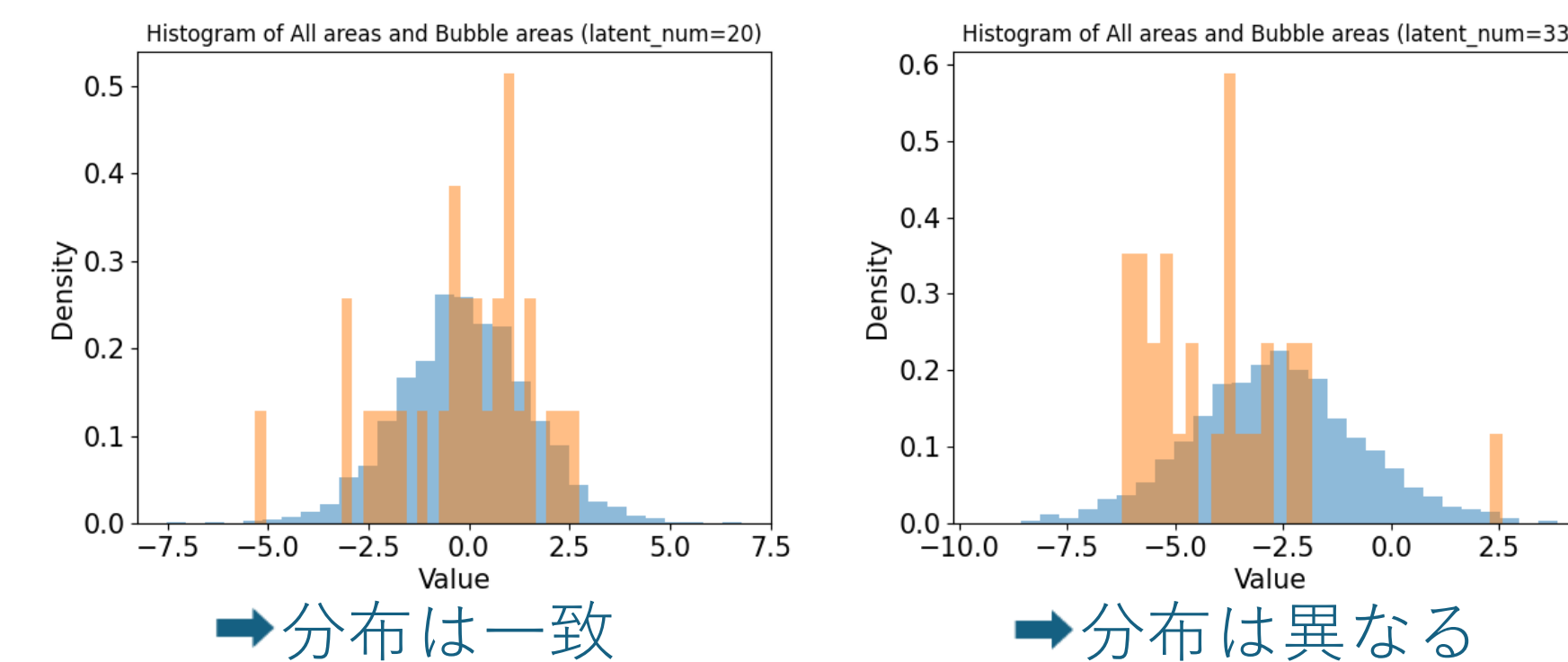
Credit: NAOJ/Nobeyama 45m Telescope

◆学習目標

このモデルの学習目標は入力データを出来るだけ精度良く再現することである。特徴を圧縮した潜在変数から入力データを再現することができれば、その潜在変数は入力データの特徴量を圧縮したものであるといえる。今回の場合、入力データの総ピクセル数112×112×12=150528個の情報量を、たった数100個の潜在変数に圧縮することができる。

◆K-S検定(コルモゴロフ-スミルノフ検定)

潜在変数の比較を行う際に、K-S検定と呼ばれる方法を用いた。この検定は、設定した閾値を基準に二つのヒストグラムの分布が一致するか否かを判断できる。

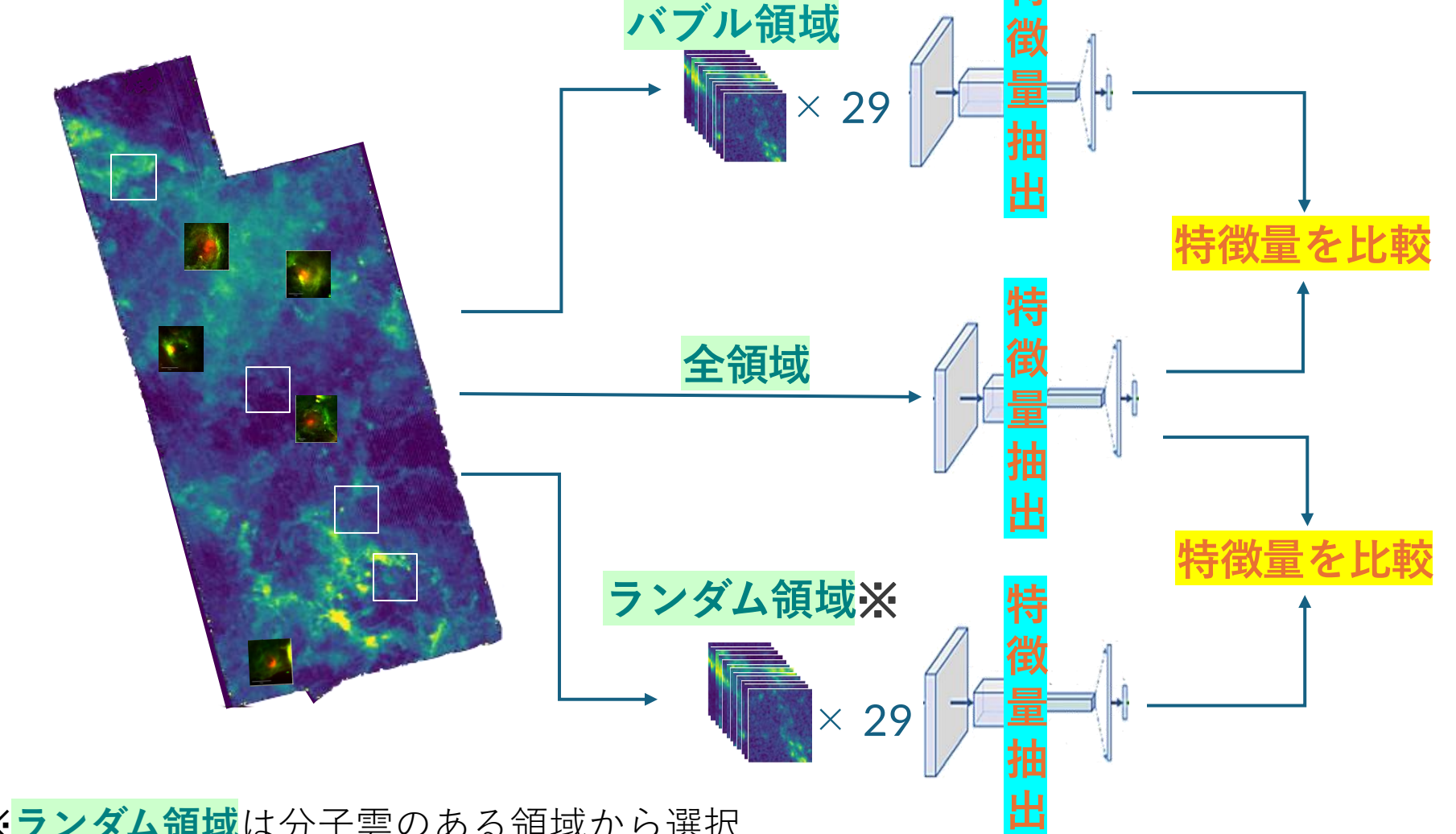


→分布は一致

→分布は異なる

◆検証方法

まず切り取った29個のバブル領域をCAEに入力し、特徴量を抽出、潜在変数に変換。その後、全領域から抽出した潜在変数と比較し、K-S検定によって異なると判断される潜在変数の数を数えた。対照実験として分子雲のあるランダムな領域を29個切り出し同様にK-S検定を用いて全領域と異なる潜在変数の数を数え、二つの結果を比較した。この際、バブル領域と他で特徴が同じなのであれば、二つの比較結果は似たようなものになり、特徴が異なるのなら、比較結果も同様に異なるはずである。

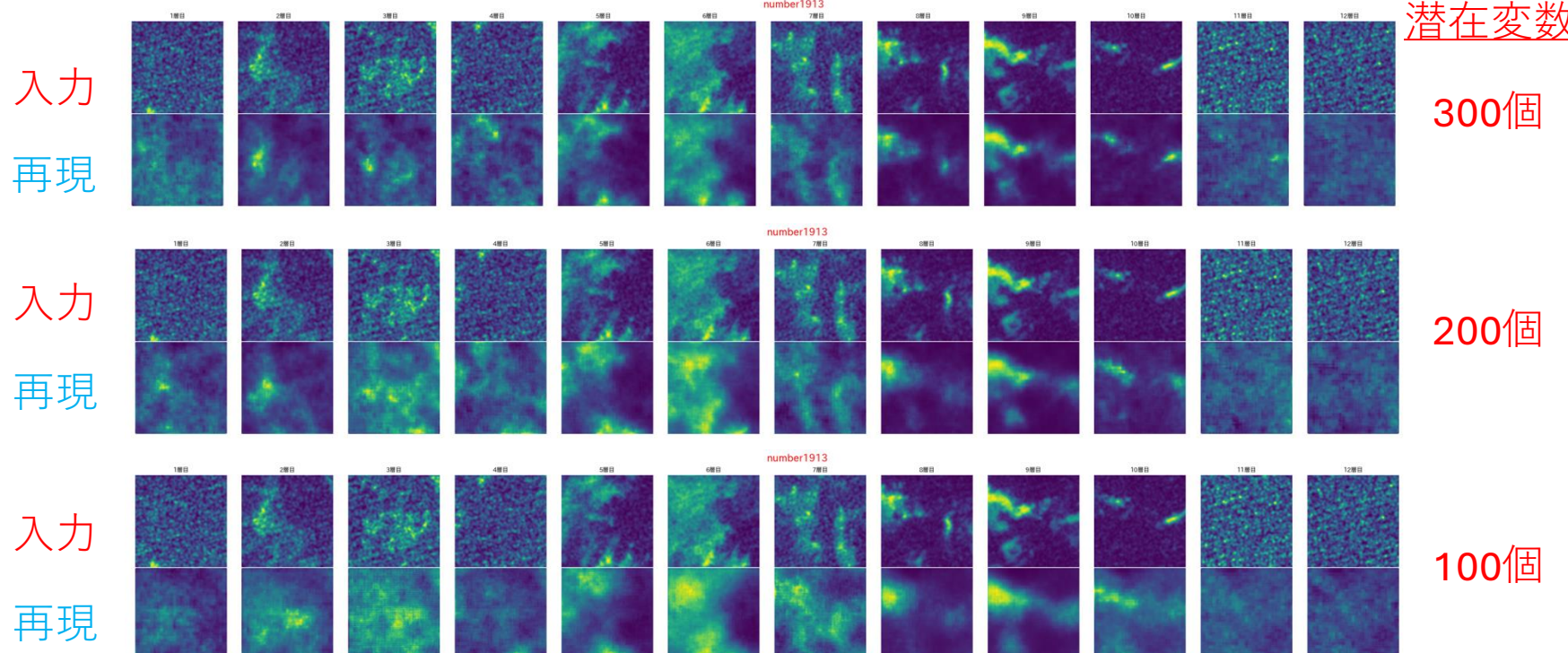


※ランダム領域は分子雲のある領域から選択

4. Results

◆学習結果

学習済みモデルの入力データの再現結果は以下の通り。



潜在変数が多いほどよく特徴を抽出できていることが、再現画像から分かる。

◆バブルとランダム領域の比較結果

潜在変数の数	300	200	100
異なる潜在変数の数(バブル領域)	25	22	12
異なる潜在変数の数(ランダム領域)	2.8	1.5	0.5

※ランダム領域は10回の平均値を記載

- ランダムに選出した領域
→全領域と異なると判断された潜在変数はほとんどなかった
- バブル領域
→全領域と異なると判断された潜在変数は、ランダムに選出した領域と比べ明らかに多かった

結論

赤外線バブルが見られる領域の分子雲は他の領域とは異なる空間と速度分布の特徴を持っている

5. Summary

We evaluated the features of molecular clouds in regions with infrared bubbles in the Cygnus X region using machine learning. As a result, we found different features between regions with infrared bubbles and randomly selected regions. This result suggests that, at least in the Cygnus-X region, the molecular clouds in the regions with infrared bubbles have peculiar characteristics.

6. Future works

- Identify the physical meaning of the extracted features.
- Verify whether the same results can be obtained in regions other than Cygnus-X.
- Use the model trained in this experiment to discover regions with unique characteristics.