# 宇宙線反粒子識別を対象とした機械学習の応用と根拠の可視化

今福拓海[1]，石川博[1]，山本幸生[2]，荒木徹也[3]，
福家英之[2]，清水雄輝[4]，和田拓也[5]，中上裕輔[5]

1．東京都立大学システムデザイン研究科　2．宇宙航空研究開発機構
3．群馬大学　4．神奈川大学　5．青山学院大学大学院



This document is provited by JAXA.

# 目次

- 研究背景，目的
- 研究内容
  - Grad-CAMによる学習モデルの分類根拠部分の可視化
  - NMFによる学習データの特徴抽出
  - シミュレーションデータの非0値の値を統一したデータによる分類器の生成
- まとめ，今後の課題



This document is provited by JAXA.

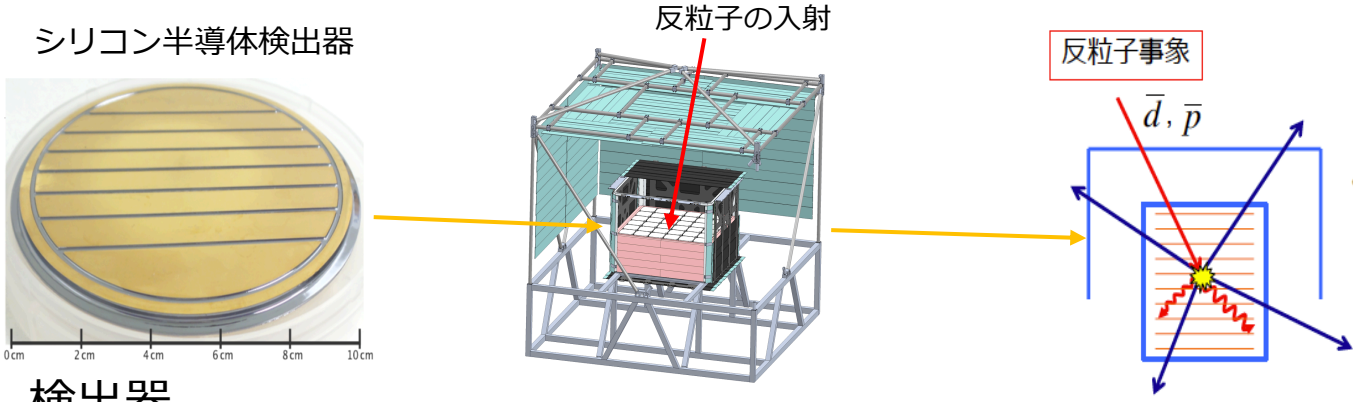
# 研究背景

- GAPS(General AntiParticle Spectrometer)

  宇宙線反粒子の高感度観測による反重陽子の検出によって暗黒物質の解明の手がかりの獲得を目指す気球実験計画

  反重陽子は暗黒物質の候補粒子の対消滅・崩壊によって生成される粒子の一つとして考えられており，予想存在量が極微小で現在まで未発見であるが，１イベントでも検出できれば暗黒物質起源の可能性が高い

- GAPS反粒子同定法




  - 検出器

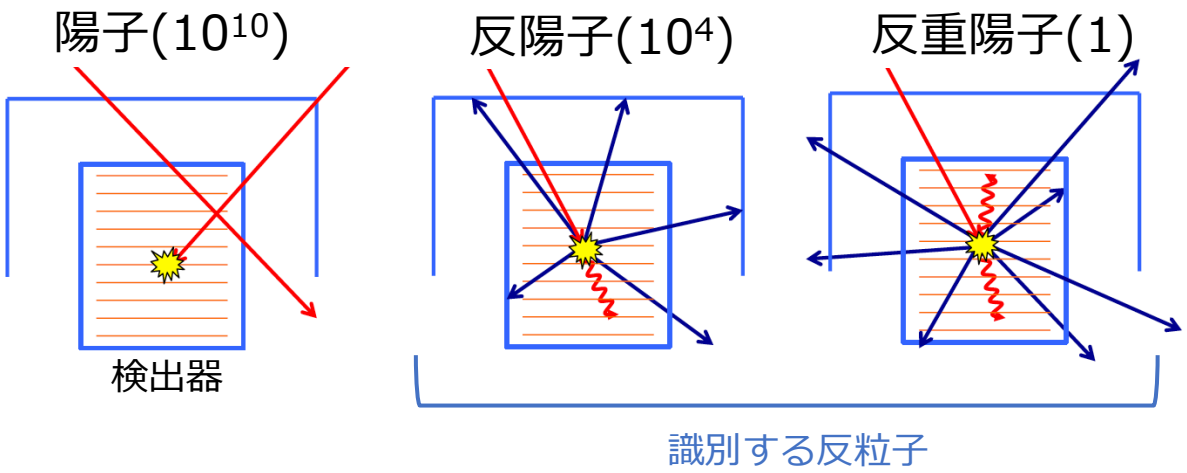
    シリコン半導体検出器群とそれを二重に取り囲むプラスティックシンチレーションカウンタ群から構成される
    シリコン検出器群の中で対消滅を起こさせ放射線状に放出される特性X線やハドロン群のエネルギー損失値を記録する



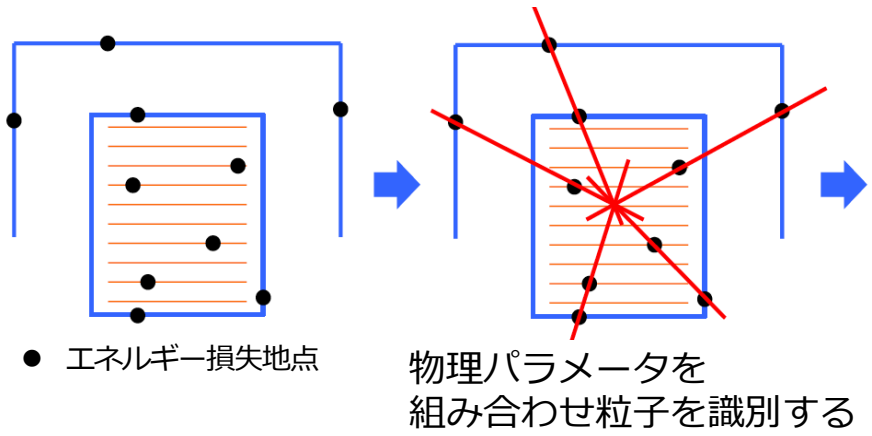
This document is provited by JAXA.

- 先行研究 – 機械学習による反粒子の識別

  - 粒子（存在比）




  - 反粒子の識別



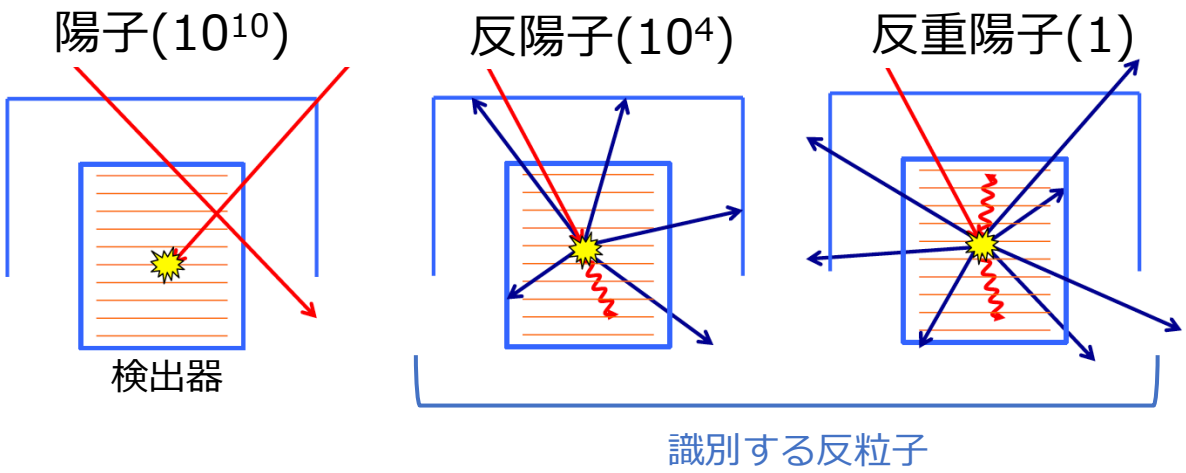

膨大な事象を
識別しなくてはならない



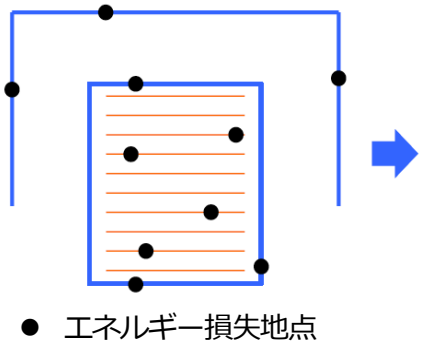
This document is provited by JAXA.

- 先行研究 – 機械学習による反粒子の識別

  - 粒子（存在比）




  - 機械学習による反粒子の識別



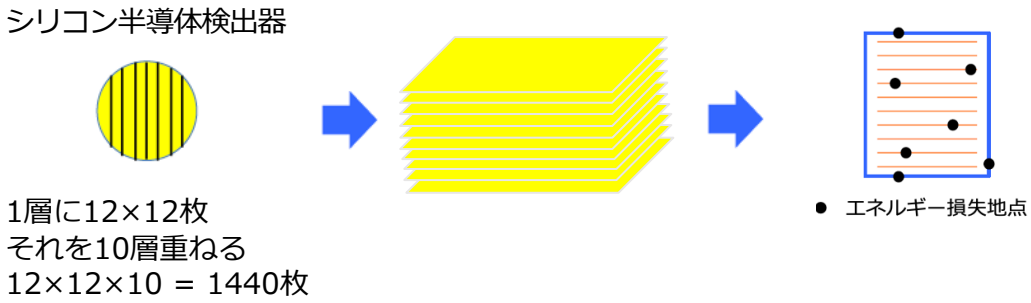

シミュレーションデータを
学習データとして識別器を生成し
粒子を識別する
高精度で人手を介さず高速で
識別できる

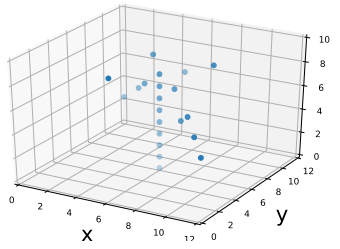


This document is provited by JAXA.

- 先行研究 – 機械学習による反粒子の識別

反陽子と反重陽子の入射と対消滅のシミュレーションによる
シリコン半導体検出器 (縦*横*高さ:12*12*10チャンネル)での
エネルギー損失値を入力データとして
CNN（畳み込みニューラルネットワーク）による識別モデルを学習させた




シミュレーションデータの例

結果として限定的な条件下では高い識別精度が確認できているが，
更なる識別精度の向上を目指している．しかし，学習データを増や
すなどの単純な方法は容量や学習時間的に厳しい

自分で作成した識別モデルの結果（使用データ数は先行研究の1/100）

| 入射角 | :0度 |
|---|---|
| Test loss | : 0.002849 |
| Test accuracy | :0.9995 |

| 入射角 | :ランダム |
|---|---|
| Test loss | : 0.1852 |
| Test accuracy | :0.9367 |

学習データ
入射粒子（ラベル）：反陽子，反重陽子
入射角　　：0度，ランダム
イベント数　　：各４万件



This document is provited by JAXA.

# 研究目的

反粒子の識別精度向上のため
学習データや識別モデルを可視化し，
反粒子の識別要因を分析することで
別の識別手法の提案を検討する



This document is provited by JAXA.

# 本研究で行ったこと

- Grad-CAMによる学習モデルの識別根拠部分の可視化
- NMFによる学習データの特徴抽出
- シミュレーションデータの非0値の値を統一したデータによる識別器の生成



This document is provited by JAXA.

# 本研究で行ったこと

- **Grad-CAMによる学習モデルの識別根拠部分の可視化**
- NMFによる学習データの特徴抽出
- シミュレーションデータの非0値の値を統一したデータによる識別器の生成



This document is provited by JAXA.

# Grad-CAMによる学習モデルの識別根拠部分の可視化

- Grad-CAM
  CNNの判断根拠の可視化技術
  CNNが識別のために注視している範囲をヒートマップで表示

以下は学習モデルが入力画像を”犬”と識別している場合



入力画像



Grad-CAM

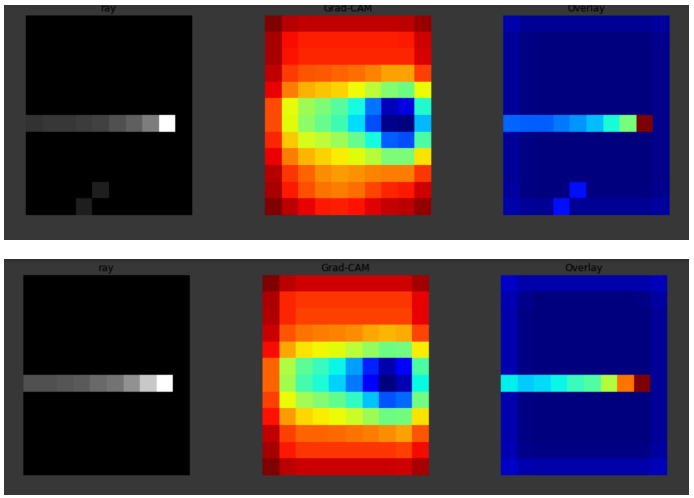


This document is provited by JAXA.

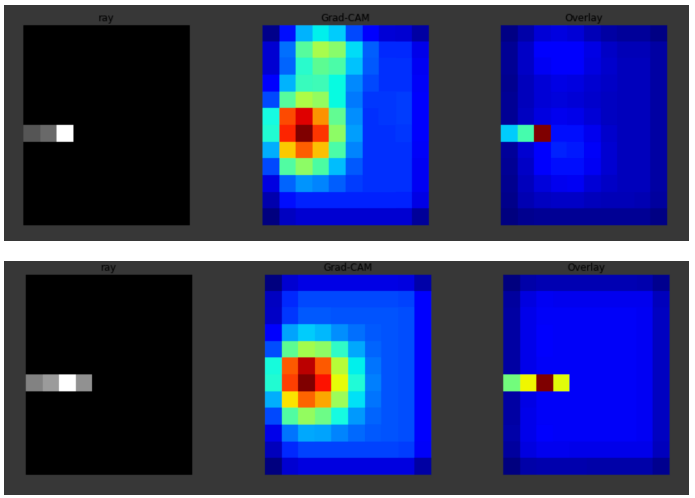
# Grad-CAMによる学習モデルの識別根拠部分の可視化

- シミュレーションデータ（入射角固定）の断面図

反重陽子




反陽子




入射方向→

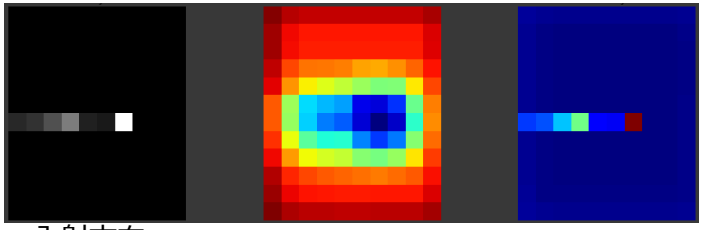
- 対消滅を起こしていると考えられる最もエネルギー損失値の高い場所から広がるような範囲を注視している
- 反陽子に比べて反重陽子はかなり奥まで入射している
- 人目でも判断できそうなくらい違いがあるため機械学習でもかなり高い精度の識別ができたのではないか



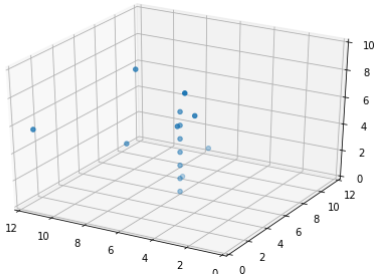
This document is provited by JAXA.

# Grad-CAMによる学習モデルの識別根拠部分の可視化

- 反陽子を反重陽子と誤分類したもの

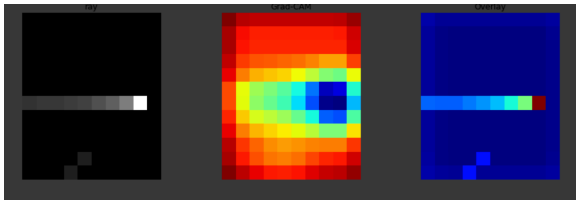


入射方向→

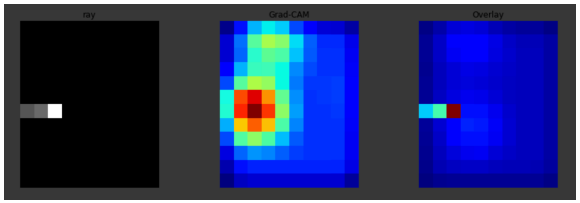



入射方向↑

反重陽子




反陽子



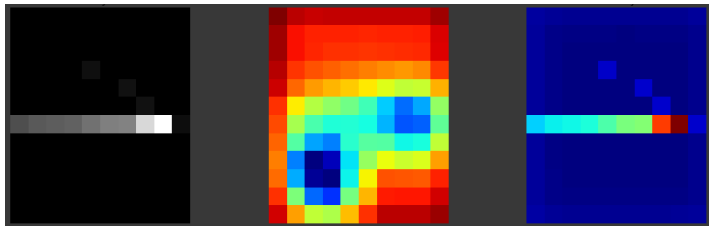

他の反陽子に比べて対消滅・崩壊までの入射経路が比較的長く
反重陽子と識別してしまっている



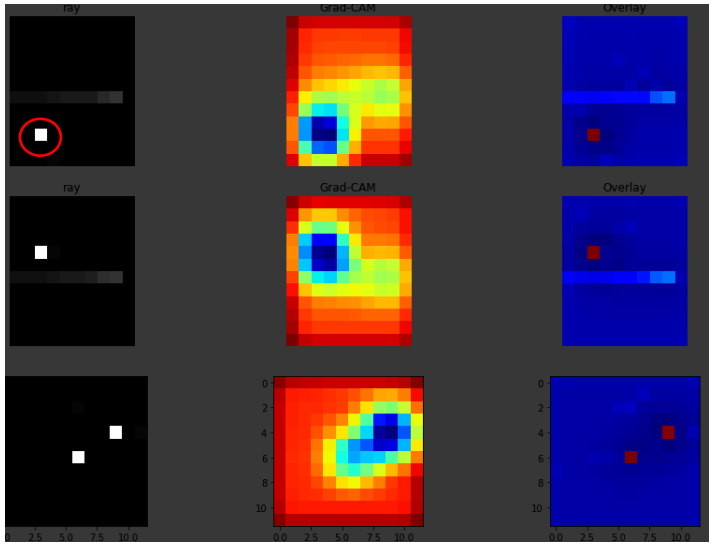
This document is provited by JAXA.

# Grad-CAMによる学習モデルの識別根拠部分の可視化

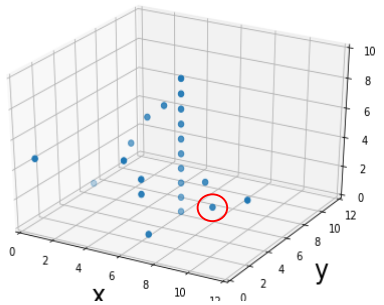
- 反重陽子を反陽子と誤分類したもの



入射方向→




3次元の各方向に断面図を足し合わせたもの




入射方向↑

図のデータのエネルギー損失最大値：87.6

反重陽子のエネルギー損失最大値の平均値：37.2
反陽子のエネルギー損失最大値の平均値：29.6
反重陽子のエネルギー損失最大値の分散：363.2
反陽子のエネルギー損失最大値の分散：249.1

学習モデルはシミュレーションごとの
エネルギー損失値の最大値周辺を見ている

意図しない認識でCNNが識別して
しまっている

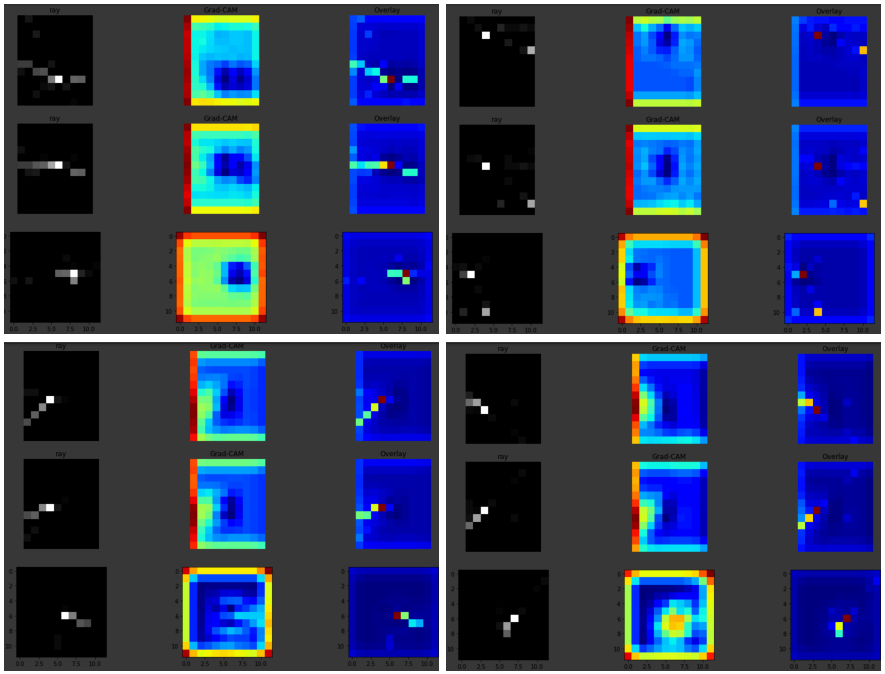


This document is provited by JAXA.

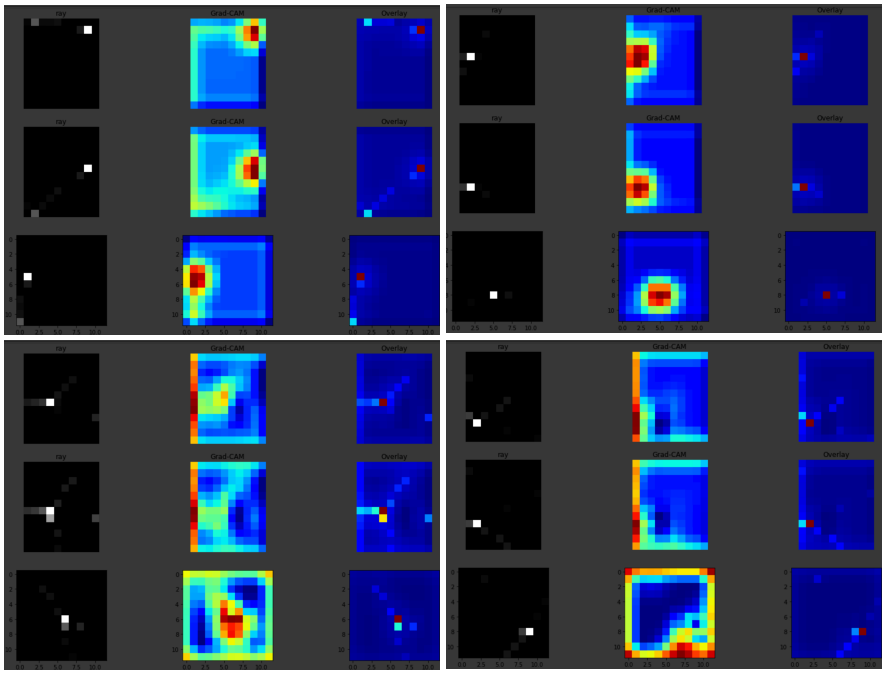
# Grad-CAMによる学習モデルの識別根拠部分の可視化

- シミュレーションデータの断面図，入射角ランダム

反重陽子



反陽子



- 入射角が固定の時と同じく，対消滅を起こしていると考えられる最もエネルギー損失値の高い場所から広がるような範囲を注視している
- 入射角がランダムだと入射の形跡が対消滅後に飛散した形跡と混じり，識別が難しい
- 人目で分類することは極めて困難



This document is provited by JAXA.

# 本研究で行ったこと

- Grad-CAMによる学習モデルの識別根拠部分の可視化
- **NMFによる学習データの特徴抽出**
- シミュレーションデータの非0値の値を統一したデータによる識別器の生成

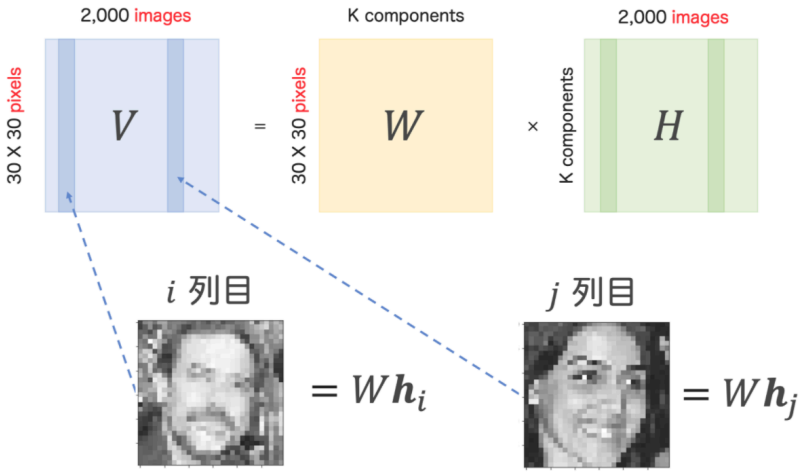


This document is provited by JAXA.

# NMFによる学習データの特徴抽出

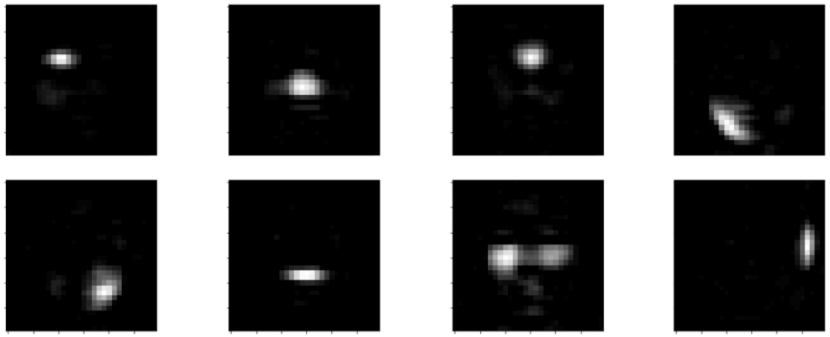
- *Non-negative Matrix Factorization*(非負値行列因子分解)

  負の値が含まれていない行列の掛け算形式で、元の行列を分解する




←NMFの分解イメージ

W： 基底行列
K：基底の次元数
H：係数行列



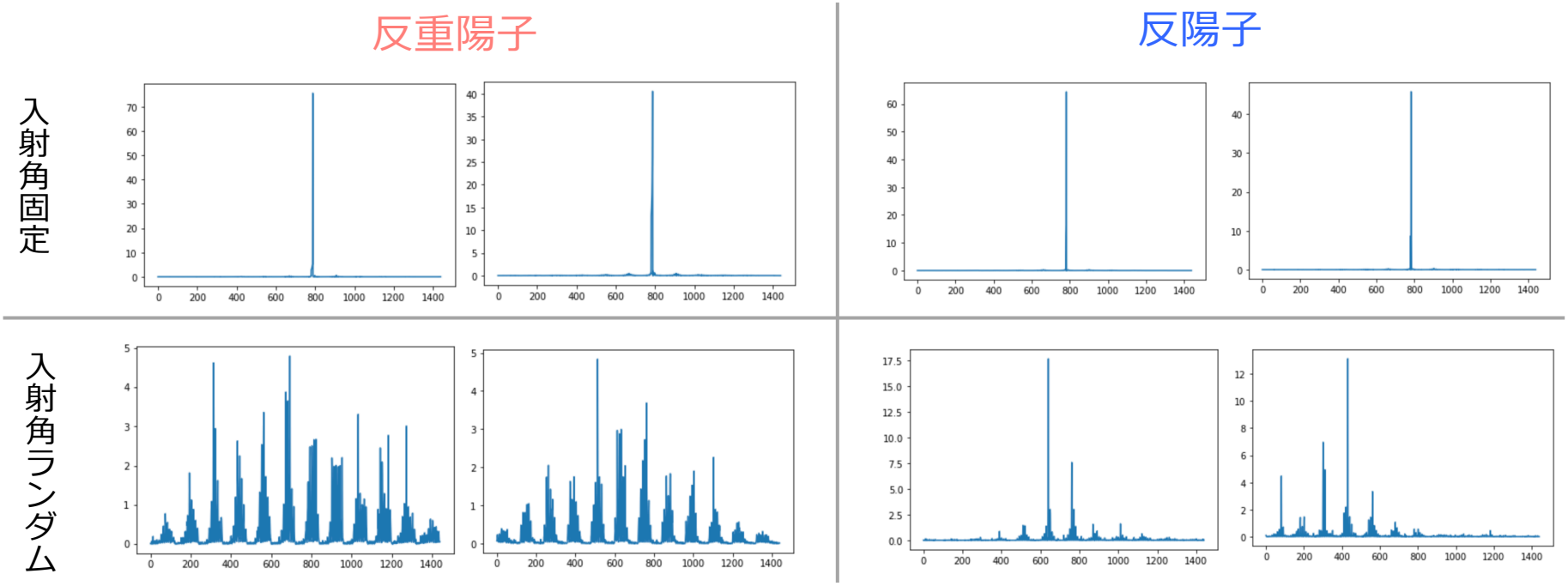
←
K = 90 とした時の
Wの各列を画像形式に
したもの



This document is provited by JAXA.

# NMFによる学習データの特徴抽出

基底の次元数 K = 2




入射角固定の場合ではエネルギー損失値を確認できる位置が限定されているため特徴の違いは見られなかった

入射角ランダムの場合では反陽子に比べて反重陽子はエネルギー損失値の広がりが強いことがわかる

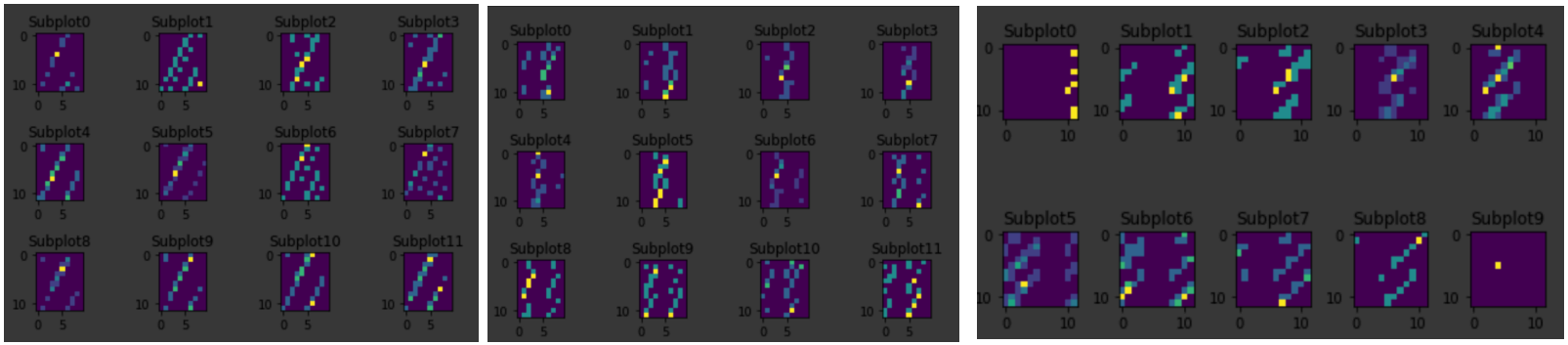


This document is provited by JAXA.

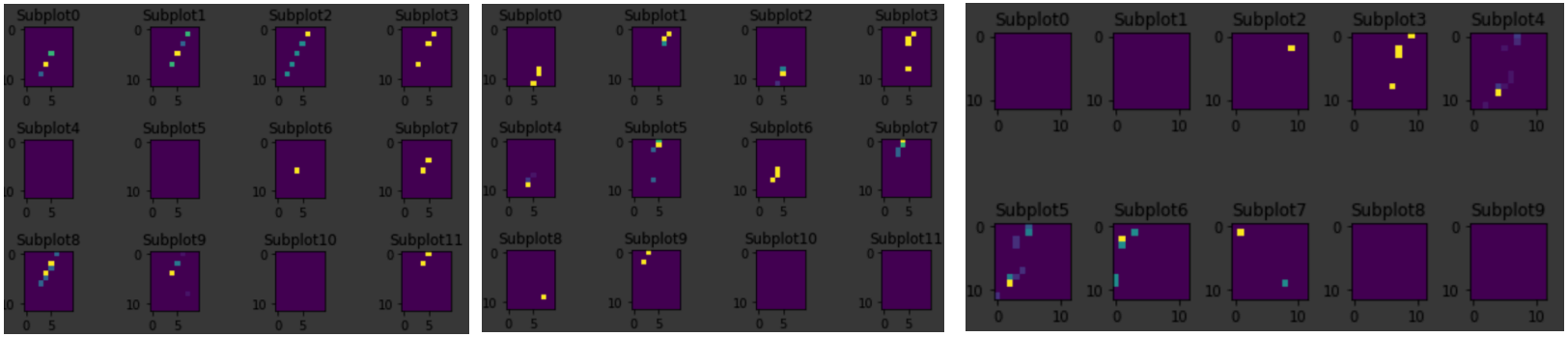
# NMFによる学習データの特徴抽出

NMF分解したシミュレーションデータの基底ベクトル(12*12*10)の断面図

- 反重陽子の基底ベクトル 1/2




- 反陽子の基底ベクトル 1/2




同じ方向に特徴が現れているが反重陽子の方が広がりがあることが確認でき，
入射角ランダムの場合でも，対消滅が起きるまでの入射が反重陽子の方が長い
ことがわかる



This document is provited by JAXA.

# 本研究で行ったこと

- Grad-CAMによる学習モデルの識別根拠部分の可視化
- NMFによる学習データの特徴抽出
- **シミュレーションデータの非0値の値を統一したデータによる識別器の生成**



This document is provited by JAXA.

# シミュレーションデータの非0値の値を統一したデータによる識別器の生成

シミュレーションデータのエネルギー損失値（非0値）を
全て1.0に統一し，学習データとして学習モデルを生成

学習データ
入射粒子（ラベル）：反陽子，反重陽子
入射角　　：0度，ランダム
イベント数　　：各４万件（入射角:0度，ランダム）

| | 入射角固定 | 入射角ランダム |
|---|---|---|
| 値統一前 | Test loss　　：0.00285<br>Test accuracy：0.9995 | Test loss　　：0.1852<br>Test accuracy：0.9367 |
| 値統一後 | Test loss　　：0.0294<br>Test accuracy：0.99425 | Test loss　　：0.837<br>Test accuracy：0.812 |

識別精度は入射角固定，ランダムのどちらも低下したが
位置情報だけでもある程度の識別ができることがわかる

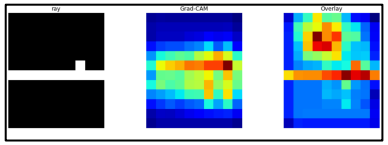


This document is provited by JAXA.

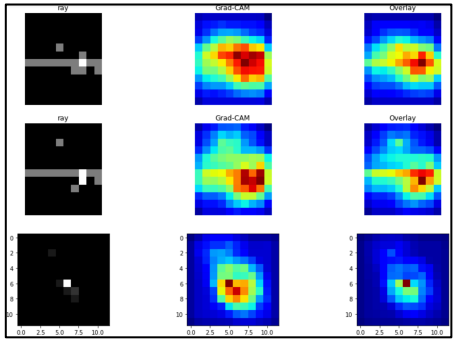
# シミュレーションデータの非0値の値を統一したデータによる識別器の生成

## Grad-CAMによる学習モデルの識別根拠部分の可視化（入射角固定）

反重陽子

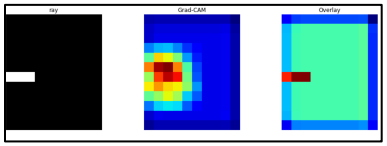


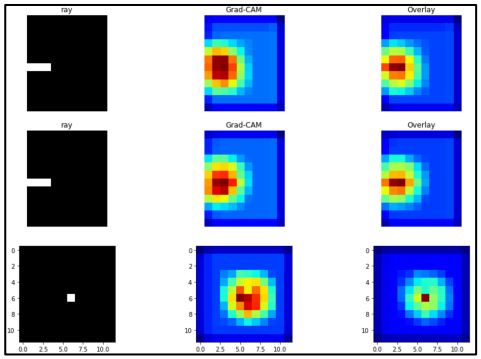

断面図



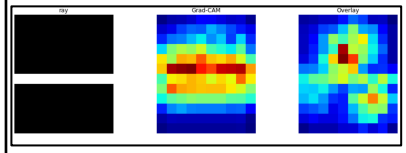

各次元方向に足し合わせたもの

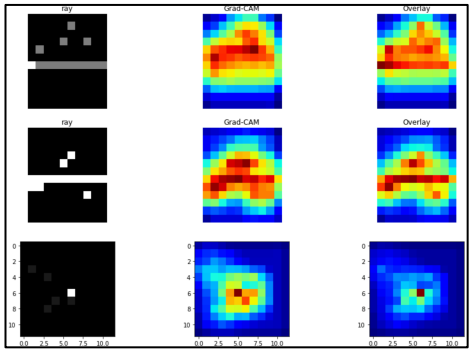
反陽子



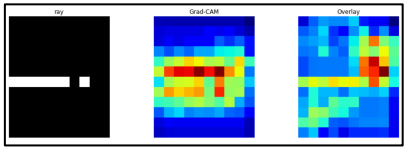



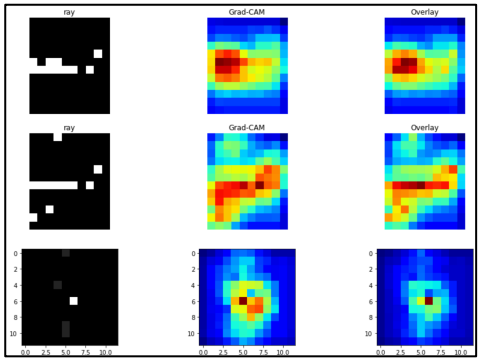

反陽子を
反重陽子と誤分類




反重陽子
を反陽子と誤分類




対消滅・崩壊後のエネルギー広がりよりも
それまでの入射の長さをもとに識別していると考えられる

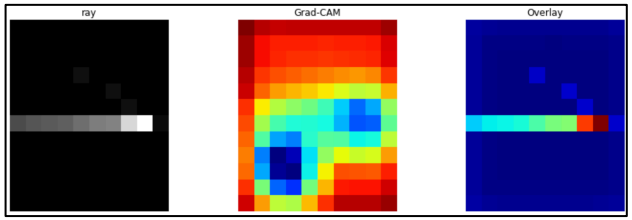


This document is provited by JAXA.

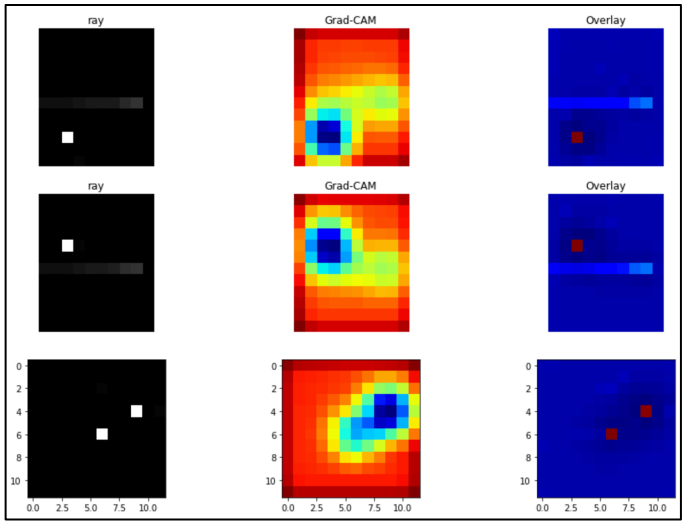
# シミュレーションデータの非0値の値を統一したデータによる識別器の生成

以前の学習モデルで誤分類したもの

- 反重陽子を反陽子と誤分類

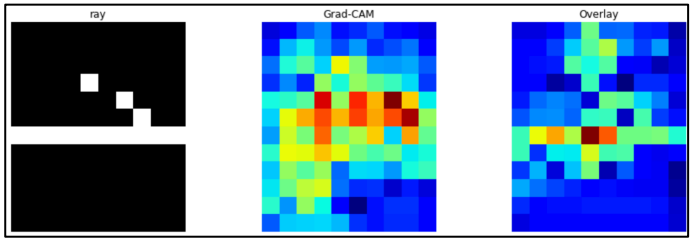


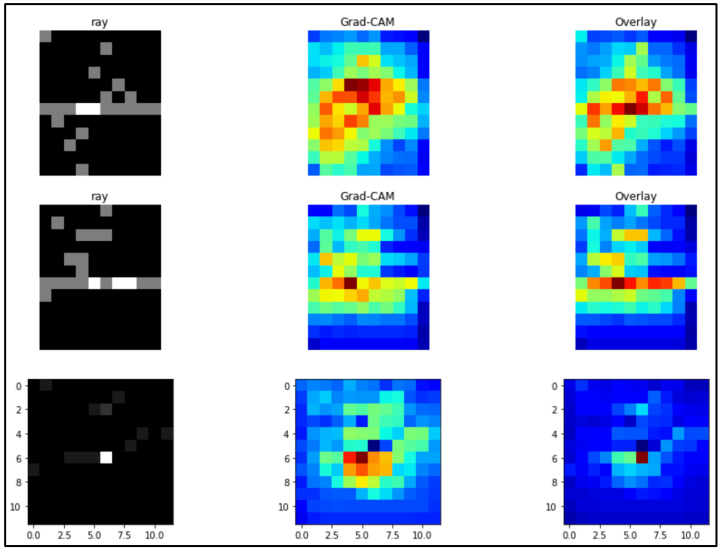

断面図




3次元の各方向に断面図を足し合わせたもの

- 反重陽子を反重陽子と正しく分類



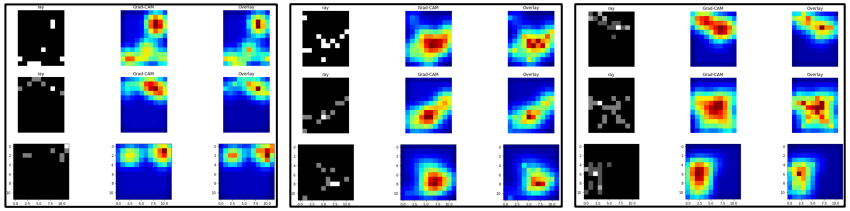



This document is provited by JAXA.

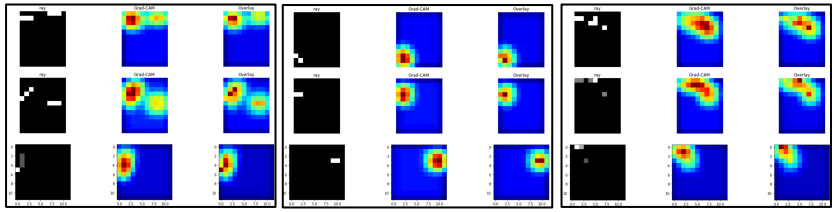
# シミュレーションデータの非0値の値を統一したデータによる識別器の生成

## Grad-CAMによる学習モデルの識別根拠部分の可視化（入射角ランダム）

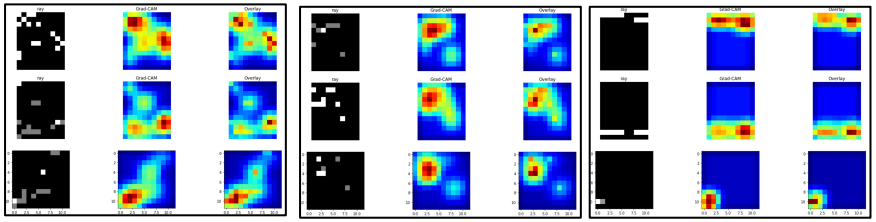
反重陽子



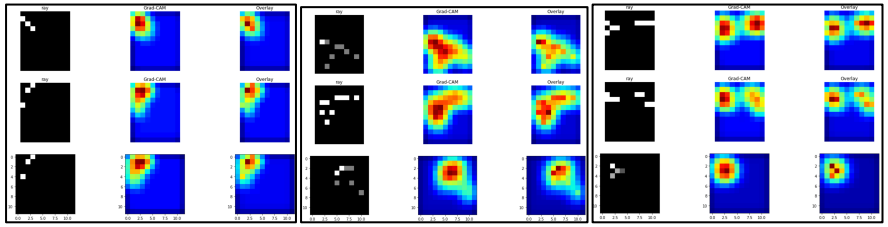
反陽子



反陽子を反重陽子と誤分類



反重陽子を反陽子と誤分類



入射角がランダムになると識別根拠部分を可視化しても
識別根拠そのものを分析することは難しい



This document is provited by JAXA.

# まとめ

- Grad-CAMによる学習モデルの分類根拠部分の可視化

  入射角固定におけるCNNの学習モデルにおいて，意図しない根拠で識別し，高精度の識別を行なっていた

- NMFによる学習データの特徴抽出

  入射角ランダムのシミュレーションデータのNMF分解によって
  反重陽子が反陽子よりエネルギーの拡散が強いことが分かった

- シミュレーションデータの非0値の値を統一した
  データによる分類器の生成

  非０値の位置情報だけでもある程度の精度を保ったまま
  識別器が生成できた
  ただし，対消滅・崩壊後のエネルギー広がりよりも
  それまでの入射の長さをもとに識別している



This document is provited by JAXA.

# 今後の展望

- 入射形跡だけでなく対消滅・崩壊後のエネルギーの広がりも根拠とした分類器生成方法の考案
- 時間情報などの，現在学習に使用していない情報を取り入れた学習データによる識別器の生成
- 入射角がランダムの場合のシミュレーションデータや学習モデルの詳しい分析



This document is provited by JAXA.