2023 年度卒業

修士論文

深層学習による動画予測手法を用いた太陽全球紫外線像の時系列予測

Time-Series Prediction of Full-Disk Solar Ultraviolet Images Using Deep Learning-based Video Prediction Method

所属	新潟大学 大学院自然科学研究科 電気情報工学専攻 情報工学コース 飯田研究室
氏名	佐々木明良
学籍番号	F22C017D

概要

アブストラクト

目次

第 1 章　研究背景　3

第 2 章　動画予測　5
2.1　動画予測の定義と定式化　5
2.1.1　動画予測モデル　5
2.1.2　最適化　6
2.2　動画予測のための基礎技術　6
2.2.1　Convolutional Neural Network (CNN)　6
2.2.2　Encoder-Decoder　8
2.2.3　Long Short Term Memory (LSTM)　8
2.2.4　Attention　10
2.3　動画予測フレームワーク　10
2.3.1　ConvLSTM　11
2.3.2　PredRNN　11
2.3.3　Motion-Aware Unit(MAU)　13

第 3 章　データ　18
3.1　SDO / AIA　18
3.1.1　AIA 211 Å　18
3.1.2　AIA 193 Å　19
3.1.3　AIA 171 Å　20
3.2　前処理　21
3.2.1　不正な画像の除去　22
3.2.2　スケーリングと正規化　22
3.2.3　データセットの分割　25

第 4 章　Motion Aware Unit を用いた 1 波長を入力とした紫外線像の全球時系列予測　27
4.1　実験概要　27
4.2　実験設定　27

4.3 学習の推移 28
4.4 実験結果 28
4.4.1 全球での評価 28
4.4.2 経度依存性の評価 28
4.4.3 東側リムから出現する活動領域に対する視覚的評価 28
4.5 考察 28

第 5 章 Motion Aware Unit を用いた 3 波長を入力とした紫外線像の全球時系列予測 29
5.1 実験概要 29
5.2 学習の推移 30
5.3 実験結果 30
5.3.1 全球での評価 30
5.3.2 経度依存性の評価 30
5.3.3 東側リムから出現する活動領域に対する視覚的評価 30
5.4 考察 30

第 6 章 まとめ 31

参考文献 32

第 1 章

研究背景

宇宙天気とは、太陽活動に起因する宇宙環境の現象を指し、激しい宇宙天気の変動は地球上の電力網や衛星通信など、人間の技術システムに影響を及ぼすことがある。太陽の表面や大気における物理現象、特に太陽フレアやコロナ質量放出（CME）などの爆発的なイベントは、地球に到達する高エネルギー粒子や放射線の量を増加させ、宇宙天気の変動に大きな影響を与えることが知られている。そのため、太陽活動を観測し、宇宙天気を予測することは、人間の技術システムを宇宙天気の影響から守るための重要な課題である。

宇宙天気の予報には、太陽フレアの爆発や CME の発生を予測するものや、太陽風の地球到達時刻や強度を予測するものなどがある。これらの予測は、様々な観測機器によって得られる太陽観測データを用いて行われるが、紫外線像はその中でも重要な情報源の一つである。(DeFN、WindNet の話とか) このように、太陽の紫外線像、およびそれから得られる特徴量は、宇宙天気予報において重要な役割を果たしている。そこで、まだ観測されていない未来の紫外線画像を予測、生成することで、より早期の宇宙天気予報の実現に貢献できるのではないかと考えた。

近年、深層学習技術の発展により、「動画予測 (Video Prediction)」と呼ばれる技術が注目されている。動画予測とは、動画の一部を入力として、それに続く未来の動画を予測するタスクである。動画予測を行う深層学習モデルのアーキテクチャの多くは、Long Short-Term Memory(LSTM) などの再帰的ニューラルネットワーク (RNN) のアーキテクチャを基本とする。さらに、特徴量抽出および伝播に Convolutional Neural Network(CNN) を用いることで空間的な特徴を時系列にわたってとらえ、Decoder モデルによって動画を生成する。このような動画予測モデルは、Conv-LSTM の登場で初めて提案されて以降、様々なモデルが提案されており、自動運転や天気予報など、様々な分野での応用が期待されている。

本研究では、動画予測モデルを用いて、数日後の紫外線画像を予測、生成することを目的とする。Deep Flare Net などの多くの深層学習を用いた宇宙天気予報モデル、またそれらを用いた人間による主観的な宇宙天気予報は、現在の観測情報を用いて、数時間後から数日後までの宇宙天気を予測す

るものが多い。それらの情報源として、数日後の高精度な紫外線画像を生成することができれば、より早期の宇宙天気予報の実現に貢献できるのではないかと考えた。

そのような数日後の紫外線像の生成のために、本研究では Motion-Aware Unit(MAU) と呼ばれる動画予測モデルを用いる。MAU は、RNN や CNN を用いた基本的な動画予測モデルを基本としつつ、各時間時点における画像の処理に MAU Cell と呼ばれるモジュールを多層的に積み重ねたアーキテクチャを採用している。MAU Cell は Attention と呼ばれる機構を持ち、長期的な依存関係を適切に学習する能力を持つ。また、従来の動画予測モデルで提案されてきた Encoder-Decoder モデルや、メモリフローの改善など、様々な改良が加えられており、多くの動画予測モデルの中でも要求する計算量に対して高い予測精度を達成している。

本研究では、MAU を用いて、数日後の紫外線画像を予測、生成することを目的とするが、その評価には主に輝度強度の再現性を用いる。これは、宇宙天気予報モデルの多くではその特徴量として輝度強度を用いていること由来する。

第 2 章

動画予測

2.1　動画予測の定義と定式化

動画予測は、既知のビデオフレームの系列から未来のフレームを予測するタスクであり、教師なし学習、または自己教師あり学習の一種として位置付けられる。このタスクは、時空間的な連続性と一貫性を持つ未来のフレームシーケンスを生成することを目指す。

2.1.1　動画予測モデル

動画予測の目的は、与えられた過去のフレームシーケンスから未来のフレームを予測するモデル M を最適化することである。ほとんどの動画予測モデルにおいて、モデルの扱うシークエンスは入力長と出力長の二つに分割される。入力長は、モデルが予測を行うために必要な過去のフレームであり、一貫してモデルがその動画の時空間的ダイナミクスを学習するために使用される。出力長は、モデルが予測を行う未来のフレームであり、モデルが最終的に生成するフレームシーケンスの一部となり、この部分に対して損失が計算される。また、もっとも初めの出力フレーム以降のフレームでは、直前の出力フレームが入力フレームとして扱われる。動画予測モデル M は、出力長として位置付けられるある時間ステップ t のフレーム $\hat{X}_t$ を生成する際は、その直前までの過去のフレーム $X_{0:t-1}$ を入力として受け取り、それに基づいてそれに続く未来のフレームを生成する。このプロセスは以下のように定式化される：

$$\hat{X}_t = M(X_{0:t-1}) \tag{2.1}$$

2.1.2 最適化

最適化プロセスは、一連の学習データセットを用いて、損失関数 L を最小化するように動画予測モデル M のパラメータを調整する。このプロセスは、以下のように表現される：

$$M_{\text{optimized}} = argmin_M L(\hat{X}, X) \tag{2.2}$$

ここで、$M_{\text{optimized}}$ は最適化された予測モデルを表す。

この損失関数は、予測された未来のフレームと実際の未来のフレームとの差異を測定するために使用され、一般的には平均二乗誤差（Mean Squared Error, MSE）が用いられる。これは、予測されたフレームと実際のフレームのピクセル単位の差異を測定する。MSE は次のように定義される：

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{X}_i)^2 \tag{2.3}$$

2.2 動画予測のための基礎技術

動画予測には、複数のフレーム間での時空間的な関連を捉え、未来のフレームを予測する能力が必要である。この目的を達成するためには、畳み込みニューラルネットワーク (CNN)、エンコーダ・デコーダ構造、長短期記憶 (LSTM)、そしてアテンションメカニズムといった複数の技術が組み合わされる。ここではそういった基礎技術群について説明する。

2.2.1 Convolutional Neural Network (CNN)

動画予測において、CNN は各フレームの空間的な特徴を抽出する役割を担う。動画は時間的な次元を持つ一連の画像であり、各フレームにおける空間的な特徴を理解することは、時間的な次元を解析する前の重要なステップである。

畳み込みニューラルネットワーク（CNN）は、特に画像認識や動画処理において広く用いられる深層学習の一形態である。CNN は、画像の局所的な特徴を捉えるために、畳み込み層とプーリング層を交互に繰り返すことで構成される。畳み込み層は画像から特徴を抽出するためのフィルターの役割を果たし、プーリング層は画像の特徴を圧縮する役割を持つ。CNN を用いた多くの画像処理アプリケーションにおいては、この畳み込み操作とプーリング操作が連続的に繰り返されることで、入力

データの高次元な特徴を抽出することを可能にしている。ここでは、その二つの重要な操作について説明する。

畳み込み

CNN の主な特徴は、局所的な特徴を効率的に捉えることができる畳み込み層にある。この畳み込み層は、入力された画像から特定の特徴を抽出するフィルターの役割を果たし、これにより画像の特徴を圧縮して表現することができる。

CNN おける畳み込み処理は、入力データに対する特定のカーネルの適用として理解される。カーネルは小規模な行列であり、入力データの局所的な領域に対して適用される。この畳み込み操作は、入力データを I、カーネルを K とした場合、以下の式で表される。

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n)K(m,n) \tag{2.4}$$

この操作は、入力データの全域にわたって繰り返され、最終的に特徴マップと呼ばれる新しい行列が生成される。特徴マップには、その入力データにおける特定のパターンや構造が抽出されている。

プーリング

この層の主な目的は、特徴マップの次元を減少させることである。具体的には、プーリング層は特徴マップの小さな領域を集約し、その領域内の代表的な値（最大値や平均値）を抽出する。この操作により、ネットワークは画像の局所的な変化に対してより頑健になり、より抽象的な特徴表現を学習することが可能である。

- **最大値プーリング（Max Pooling）**：この手法では、各領域の最大値が選択される。これにより、特徴マップから最も強い信号を保持し、関連性の低い信号を破棄する。最大値プーリングは、特に画像内のテクスチャや形状などの顕著な特徴を強調するのに有効である。ここで、入力となる特徴マップを F, プーリング領域のサイズを m*n, プーリング層の出力を P とすると、最大値プーリングは以下の式で表される。

$$P(i,j) = \max_m \max_n F(i+m, j+n) \tag{2.5}$$

- **平均値プーリング（Average Pooling）**：平均プーリングは、各領域の平均値を計算する。これにより、特徴マップの全体的な特性をより平滑化し、より均一な特徴表現を提供することが可能である。ここで、入力となる特徴マップを F, プーリング層の出力を P とすると、平均値プーリングは以下の式で表される。

$$P(i,j) = \frac{1}{mn} \sum_{m} \sum_{n} F(i+m, j+n) \tag{2.6}$$

2.2.2 Encoder-Decoder

エンコーダ・デコーダ構造は画像処理において広く用いられ、特に U-Net のようなアーキテクチャが代表的である。エンコーダ・デコーダ構造は、一連の入力データを処理し、それを内部表現に変換するエンコーダ部分と、この内部表現から出力を生成するデコーダ部分の二つの主要なコンポーネントから構成される。動画予測において入力データとなるのは、動画の各フレームであり、出力データはそれに続く未来のフレームである。ここでは画像処理におけるエンコーダ・デコーダ構造に焦点を当てて説明する。

エンコーダ

エンコーダは入力データを CNN によって処理し、それを高次元から低次元の表現に変換する。このプロセスは、入力データに含まれる重要な情報を抽出し、より扱いやすいサイズまたは形式に圧縮することを目的とする。動画予測アプリケーションにおいては、空間的特徴に加え、複雑な時間的特徴の依存性をモデリングするため、一般的な画像処理ディープラーニングモデルと比較して大量の計算資源を必要とする。そのため、エンコーダにより特徴を圧縮し、より低い次元で高度な特徴を抽出することは、計算コストの削減という点からも非常に有用である。

デコーダ

デコーダは本質的にエンコーダの逆処理である。動画予測においては、その直前のアーキテクチャにより生成された内部表現を受け取り、目的とする出力を生成する役割を持つ。動画予測の再帰ネットワーク内では、エンコーダにより圧縮された行列が扱われるため、そのままでは出力には不適切である。デコーダはそのような圧縮された表現を元の次元に展開し、出力に適した形式に変換する。

2.2.3 Long Short Term Memory (LSTM)

リカレントニューラルネットワーク（RNN）は、時系列データや自然言語などのシーケンシャルな情報を扱うために開発されたニューラルネットワークの一種である。RNN の特徴は、過去の情報を隠れ状態として保持し、それを利用して次の出力を生成する点にある。しかし、RNN は長期的な依存関係を捉えることに困難を抱えていた。この問題は「勾配消失問題」として知られ、ネットワークが

深くなるほど、またはシーケンスが長くなるほど顕著になる。勾配消失問題により、RNN は過去の情報を長期間保持し活用することが困難になる。

この問題を解決するために開発されたのが、長短期記憶（LSTM）である。LSTM は、RNN の基本的な枠組みを保ちつつ、特定の情報を長期間記憶する能力を強化した。LSTM の主要な特徴は、セル状態と呼ばれる内部メカニズムであり、これにより長期的な情報を保持することが可能となる。

LSTM のユニークな構造は以下の三つのゲートから成る：忘却ゲート、入力ゲート、出力ゲート。

1. **忘却ゲート（Forget Gate）**：このゲートは、セル状態に含まれる情報の一部を削除する役割を担う。LSTM ネットワークが長期的な依存関係を学習する過程で、関連性の低い古い情報を捨てることが重要である。忘却ゲートはシグモイド関数を使用して、どの情報を保持し、どの情報を忘れるかを決定する。

2. **入力ゲート（Input Gate）**：入力ゲートは、新しい情報をどの程度セル状態に追加するかを決定する。このゲートでは、シグモイド関数がどの情報を更新するかを決定し、tanh 関数が新しい候補値を生成する。そして、これら二つの値の積が新しい情報としてセル状態に追加される。

3. **出力ゲート（Output Gate）**：出力ゲートは、現在のセル状態に基づいて、ネットワークの出力を決定する。このゲートは、シグモイド関数を使用して、セル状態のどの部分が出力されるべきかを決定し、tanh 関数によって処理されたセル状態との積が最終的な出力となる。

LSTM ユニットの数学的な定式化は以下の通りである。h_t は時刻 t での隠れ状態、c_t はセル状態、x_t は入力、f_t、i_t、o_t はそれぞれ忘却ゲート、入力ゲート、出力ゲートの活性化関数を表す。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2.7}$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2.8}$$
$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{2.9}$$
$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{2.10}$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{2.11}$$
$$h_t = o_t * \tanh(c_t) \tag{2.12}$$

ここで、W と b はそれぞれ重みとバイアスを表し、σ はシグモイド活性化関数、tanh は双曲線正接活性化関数を指す。LSTM のこの構造により、長期的な依存関係を効果的にモデル化することが可能となり、特に時系列データや動画予測などの分野において有効である。

2.2.4 Attention

アテンションメカニズムは、1990 年代に自然言語処理（NLP）分野で初めて提案されたが、その真価が広く認識されるようになったのは 2014 年の「Neural Machine Translation by Jointly Learning to Align and Translate」という論文での再発見以降である。この技術は、モデルが重要な情報に焦点を当て、それ以外の情報を無視する能力を提供することにより、ディープラーニングにおける重要な進歩の一つとなった。

アテンションメカニズムの中心には、クエリ（Query）、キー（Key）、バリュー（Value）の三つの概念がある。これらの要素を使用して、モデルがどの情報に注意を払うべきかを決定する。

- **クエリ（Query）**：クエリは現在注目している要素や状態を表し、モデルがどの情報に注目するかを決定する基準となる。
- **キー（Key）**：キーはデータセット内の各要素に関連付けられ、クエリとの関係を定義する。クエリとキーの間の類似性が高いほど、そのキーに関連付けられた情報に注意が向けられる。
- **バリュー（Value）**：バリューはキーに関連付けられた実際の情報を含み、アテンションメカニズムは、クエリとキーの関係に基づいて、どのバリューを重視するかを決定する。

アテンションメカニズムの基本的な操作は、クエリと各キーの間の類似性を計算し、それに基づいて各バリューの重み付き和を取ることである。この重み付き和は次のように定義される。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.13}$$

ここで、Q、K、V はそれぞれクエリ、キー、バリューの行列、d_k はキーの次元数を表す。softmax 関数は、キーとクエリ間の類似性を確率的な重みに変換する。このプロセスにより、アテンションメカニズムは、重要な情報に「注意」を集中させ、関連性の低い情報を無視することができる。

2.3 動画予測フレームワーク

動画予測のフレームワークは、一般的にエンコーダ・デコーダ構造を基本とし、その中に LSTM やアテンションメカニズムなどの機構を組み込むことで、動画の時空間的な特徴を効果的に捉えることが可能となる。ここでは、もっとも基本的な動画予測フレームワークである ConvLSTM と、その後の改良を加えた PredRNN、また本研究で用いる Motion-Aware Unit (MAU) について説明する。

2.3.1 ConvLSTM

ConvLSTM（Convolutional Long Short-Term Memory）は、Shi et al. (2015) によって提案された、動画予測とその他の時空間シーケンスデータの処理に特化したニューラルネットワークアーキテクチャである。伝統的な LSTM の枠組みを拡張し、畳み込み操作を組み込むことで、空間的な情報を効果的に処理する能力を持つ。このような特性により、ConvLSTM は、動画予測のみならず、気象予測や交通流予測など、他の時空間データ処理の応用にも適用可能であり、その応用が期待されている。

各ゲートの定義

ConvLSTM は、LSTM の各ゲート（忘却ゲート、入力ゲート、出力ゲート）とセル状態の更新に畳み込み演算を導入する。これにより、モデルは時系列データに含まれる空間的パターンを捉え、それを時間的文脈において解析することが可能となる。特に、動画や気象データなどの時空間データにおいて、局所的な空間的特徴と時間的依存関係を同時にモデル化できる。

ConvLSTM の数学的定式化は以下の通りである。ここで、$*$ は畳み込み演算、$\circ$ はアダマール積（要素ごとの積）を表す。

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \tag{2.14}$$
$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \tag{2.15}$$
$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \tag{2.16}$$
$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \tag{2.17}$$
$$H_t = o_t \circ \tanh(C_t) \tag{2.18}$$

X_t は時刻 t における入力、H_t は隠れ状態、C_t はセル状態を示し、f_t、i_t、o_t はそれぞれ忘却ゲート、入力ゲート、出力ゲートの活性化状態を表す。W と b はネットワークの重みとバイアスパラメータである。この定式化により、ConvLSTM は時空間データの空間的な特徴と時間的な特徴を統合的に処理し、高度な予測を行う能力を持つ。

ここで、N はサンプル数、Y_i は実際のフレーム、$\hat{Y}_i$ は予測フレームを表す。

2.3.2 PredRNN

最初の動画予測モデルである ConvLSTM の発表後、そのアーキテクチャを基に様々な改良が加えられてきた。本研究で用いる Motion-Aware Unit の基礎となる PredRNN は、その中でも特に代表的

なモデルである。

Yunbo Wang et al. (2017) によって提案された PredRNN は、ConvLSTM を基盤としながらも、いくつかの重要な進化と改良を経て開発された。ConvLSTM はセル状態は各層レベルに対して独立であり、時間方向でのみ更新される。このような状況では、最下層は直前の時間ステップの最上層が生成したセル状態を考慮することができない。PredRNN ではこの概念を拡張し、メモリ状態を異なる層の間で効果的に伝達することを可能にする。

時空間メモリフロー

PredRNN では、時空間メモリフローを利用して空間情報の伝達を最適化する。このメモリフローは、遠隔状態間で情報を伝達し、勾配消失問題を軽減するために設計されている。時間方向と、各時間ステップでの隠れ層方向にジグザグにメモリを流す。このように、層をまたいで情報を上方向に伝達し、時間を超えて前方向に情報を伝達することにより、空間情報の効率的な流れを実現し、動画フレーム間のより詳細な変化を捉えることができる。

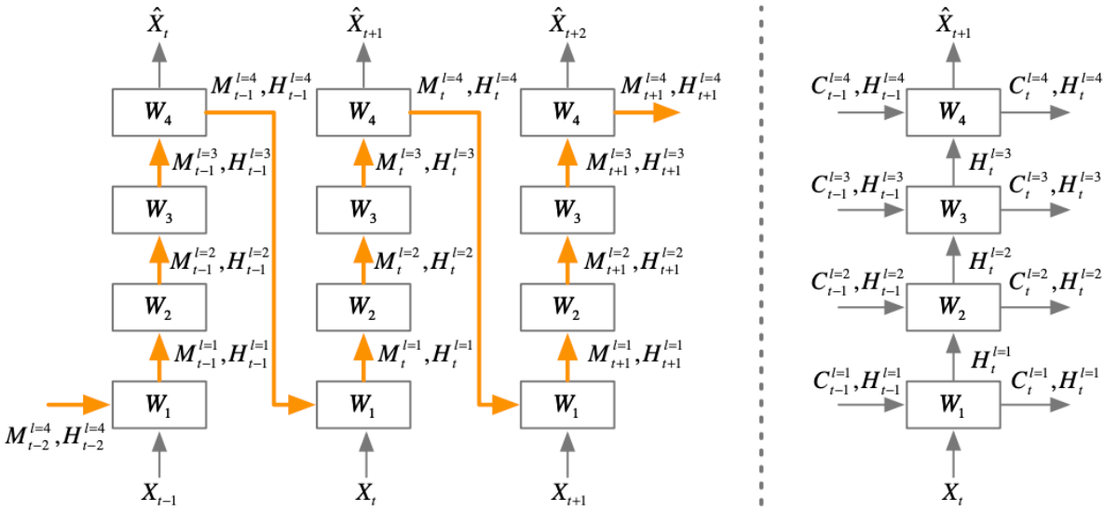


図 2.1: 左が PredRNN におけ時空間メモリフロー、右が ConvLSTM のメモリフローである。このような異なるレベルの層を通過するメモリフローにより、多様な抽象度の表現を学習することができる。

時空間 LSTM ユニット (ST-LSTM)

時空間メモリフローは空間情報の効果的な伝達を可能にするが、水平方向 (時間方向) のメモリフローを省略すると、時間的一貫性を犠牲にしてしまう。PredRNN では、標準的な LSTM ユニット

を、時空間メモリセルとゲート構造を導入した時空間 LSTM(ST-LSTM) ユニットに置き換える。ST-LSTM ユニットは、時間メモリセルと時空間メモリセルの両方を維持し、それらを縦方向および横方向に流すことにより、一定の時間一貫性を担保しながら、異なる抽象度レベルでの特徴を捉える。また、ST-LSTM は 1x1 の畳み込み層を使用して次元削減を行い、隠れ状態の次元をメモリセルと同じにする。この構造により、PredRNN は時空間データの複雑なダイナミクスを捉え、より正確な予測を生成することができる。

時空間 LSTM の数学的定式化は以下の通りである。ここで、$*$ は畳み込み演算を、$\circ$ はアダマール積（要素ごとの積）を示す。W と b は重みとバイアスパラメータ、σ はシグモイド活性化関数、tanh は双曲線正接活性化関数を指す。X_t は時刻 t の入力、H_t^l は隠れ状態、C_t^l は標準的な LSTM セル、M_t^l は時空間メモリセルを表す。

$$g_t = \tanh(W_{xg} * X_t + W_{hg} * H_{t-1}^l + b_g) \tag{2.19}$$
$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1}^l + W_{mi} \circ M_{t-1}^l + b_i) \tag{2.20}$$
$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1}^l + W_{mf} \circ M_{t-1}^l + b_f) \tag{2.21}$$
$$C_t^l = f_t \circ C_{t-1}^l + i_t \circ g_t \tag{2.22}$$
$$g'_t = \tanh(W'_{xg} * X_t + W_{mg} * M_{t-1}^l + b'_g) \tag{2.23}$$
$$i'_t = \sigma(W'_{xi} * X_t + W_{mi} * M_{t-1}^l + b'_i) \tag{2.24}$$
$$f'_t = \sigma(W'_{xf} * X_t + W_{mf} * M_{t-1}^l + b'_f) \tag{2.25}$$
$$M_t^l = f'_t \circ M_{t-1}^l + i'_t \circ g'_t \tag{2.26}$$
$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1}^l + W_{co} * C_t^l + W_{mo} * M_t^l + b_o) \tag{2.27}$$
$$H_t^l = o_t \circ \tanh(W_{1\times1} * [C_t^l, M_t^l]) \tag{2.28}$$

この構造により、PredRNN は時空間データの複雑なダイナミクスを捉え、より正確な未来予測を生成する。

2.3.3 Motion-Aware Unit(MAU)

Motion-Aware Unit（MAU）は、Zheng et al.(2021) によって発表された、フレーム間のダイナミクスをより効率的に捉えるために提案された新しい動画予測アーキテクチャである。MAU は、PredRNN と似た積層 LSTM の構造を持ち、その LSTM ユニットを MAU セルによって置き換えている。MAU セルは、注意（Attention）モジュールと融合（Fusion）モジュールの 2 つの部分から構成されており、PredRNN における時空間 LSTM ユニットをさらに拡張したものである。このような変更により、時間的受容野

本研究では、この MAU を用いて太陽全球紫外線画像の予測を行う。

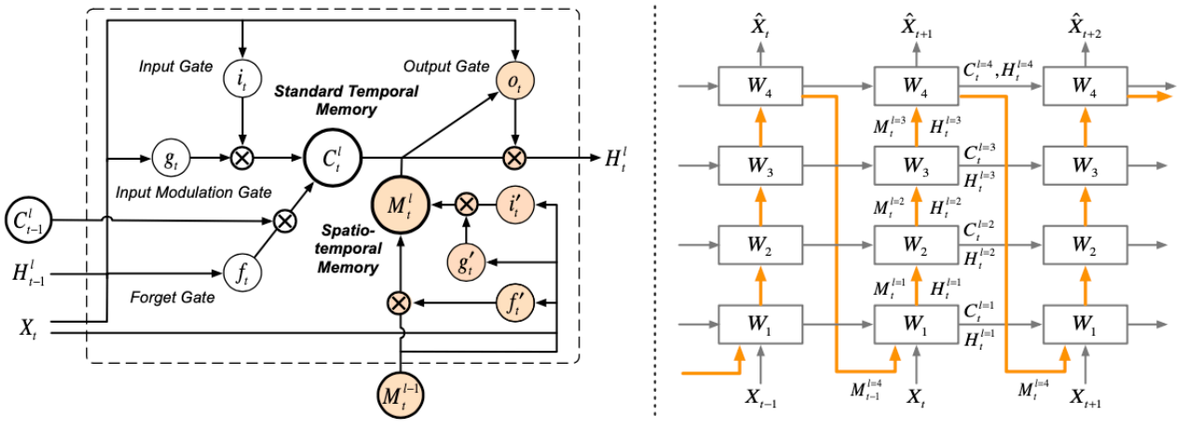


図 2.2: 左が ST-LSTM ユニット、右が PredRNN のメモリフロー。オレンジ色のノードは、従来の ConvLSTM と異なる PredRNN 独自の構造を示す。PredRNN 内のオレンジ色の矢印は、時空間メモリ M_t^l の移行経路を示す。

アーキテクチャ

動画予測モデルでは、出力中の時間ステップが進むほど、時間情報の不確実性が増加するため、予測誤差が劇的に加速してしまう。この問題を解決するため、動画予測モデルはより幅広い時間ステップから有用な特徴を保存し活用する必要がある。すなわち、時間的受容野を拡張する必要がある。このような問題に対するアプローチとして、Yunbo et al. (2019) によって提案された、三次元畳み込みを導入した E3D-LSTM があったが、非常に高い計算コストを必要とし、性能の改善は限定的であった。Motion-Aware Unit (MAU) は、このような課題を解決するため、Attention 機構を導入した新しいアーキテクチャを提案している。ここでは、Attention 機構を用いたアーキテクチャと、その統合や効率化のために導入されたいくつかの特徴について説明する。

- **エンコーダ・デコーダ構造**: MAU は、エンコーダ・デコーダ構造を基本としている。ConvLSTM や Pred-RNN などの他の動画予測モデルではエンコーダ・デコーダ構造は採用されていないが、MAU ではエンコーダ・デコーダ構造を採用することで、より効率的な特徴抽出を可能にしている。一定程度抽象化された情報を再帰的ネットワークに入力することで、より多くの MAU セルの積層を行っても、計算コストを抑えることができる。

$$\hat{X}_t = \mathrm{Dec}[\mathrm{MAU}(\mathrm{Enc}(X_{t-\tau:t-1}))] \tag{2.29}$$

  ここで、Enc はエンコーダ、Dec はデコーダを表し、τ は入力シーケンスの長さを表す。

- **注意 (Attention) モジュール**: 注意モジュールは、先述の時間的受容野の拡張のために導入

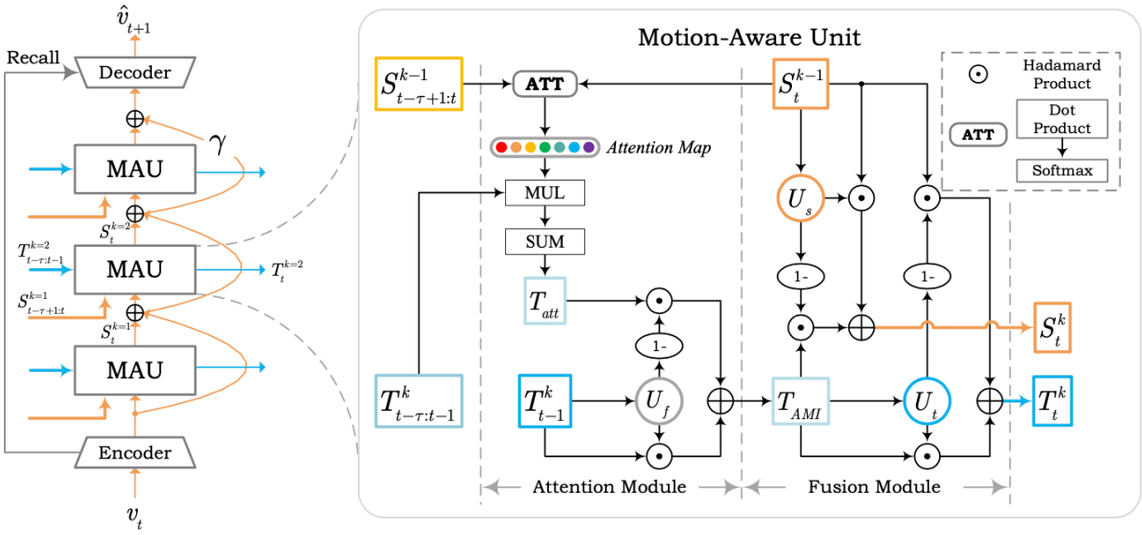


図 2.3: 左: MAU セルを積み重ねた MAU モデルの構造。オレンジ色の矢印は空間方向のメモリフローを示し、青色の矢印は時間方向のメモリフローを示す。エンコーダ・デコーダ構造は各時間ステップにおいて一回ずつ適用され、情報を効果的に圧縮している。右:MAU セルは、注意（Attention）モジュールと融合（Fusion）モジュールの 2 つの部分から構成されている。

される。このモジュールの導入により、異なる時間状態に対して異なるレベルの注意を払い、予測に対してもっとも相関の高い状態に注意を集中させることが期待されている。長期的な運動情報としての T_{att} は以下のように計算される。

$$T_{\text{att}} = \sum_{j=1}^{\tau} \alpha_j \cdot T_{t-j}^{k} \tag{2.30}$$

ここで、α_j は時間状態 T_{t-j}^{k} に対するアテンションスコアを表す。T_{att} は、アテンションスコアによって、予測結果に対して長期的な相関性を持つ時間状態を考慮することができるが、短期的な時間状態 T_{t-1}^{k} も考慮する必要がある。そこで、それらを融合するためのゲート U_f を導入し、長期的な時間状態 T_{att} と短期的な時間状態 T_{t-1}^{k} を融合した T_{AMI} を以下のように計算する。

$$U_f = \sigma(W_f * T_{t-k-1}) \tag{2.31}$$

$$T_{AMI} = U_f \odot T_{t-k-1} + (1 - U_f) \odot T_{\text{att}} \tag{2.32}$$

- **融合 (Fusion) モジュール**: 融合モジュールは、拡張された運動情報 T_{AMI} と現在の空間的状態を適切に統合する役割を持つ。融合プロセスでは、時間的更新ゲート U_t と空間的更新ゲート U_s を導入する。

$$U_t = \sigma(W_{\text{tu}} * T_{\text{AMI}}) \tag{2.33}$$
$$U_s = \sigma(W_{\text{su}} * X_t) \tag{2.34}$$

ここで、W_{tu} と W_{su} はそれぞれ時間的更新ゲートと空間的更新ゲートの重みを表す。このゲートを利用して、時間状態 T_t^k と空間状態 S_t^k を以下のように計算する。

$$T_t^k = U_t \odot (W_{tt} * T_{\text{AMI}}) + (1 - U_t) \odot (W_{st} * S_t^{k-1}) \tag{2.35}$$
$$S_t^k = U_s \odot (W_{ss} * S_t^{k-1}) + (1 - U_s) \odot (W_{ts} * T_{\text{AMI}}) + \gamma \cdot S_t^{k-1} \tag{2.36}$$

ここで、W_{tt}、W_{st}、W_{ss}、W_{ts} はそれぞれ時間状態と空間状態の重みを表し、γ は学習の安定化を図るための残差項の係数である。

- **情報リコール**: MAU では、エンコーダとデコーダ間で情報損失を防ぐために、情報リコールスキームが採用されている。これは U-Net などでで用いられるスキップ接続に似た構造を持っている。これにより、デコーダは多レベルのエンコードされた情報を考慮することができ、予測の視覚的品質を向上させることができる。

主な実験結果

MAU は、複数のデータセットで評価され、その中には Moving MNIST、KITTI、Caltech Pedestrian、TownCentreXVID、Something-Something V2 が含まれる。ここでは、Moving MNIST データセットに関する既存の動画予測モデルによる性能との比較の結果を示す。表 2.1 は、異なる手法による Moving MNIST データセットにおける定量的な結果を示している。MAU は、構造的類似性 (Structual Similarity, SSIM) スコアと平均二乗誤差 (Mean Squared Error, MSE) スコアの両方において、最も優れた結果を示している。2.2 は、LSTM ユニットを積層する動画予測モデルのバックボーンを統一し、それぞれの場合でのパラメータと推論時間に焦点を当てた比較を行なっている。これによれば、MAU は他の手法と比較して、より少ないパラメータ数とより短い推論時間で動作することができ、さらに MSE と SSIM の両方において最も優れた結果を示している。

表 2.1: Moving MNIST データセットにおけるビデオ予測方法の定量的結果（10 フレーム→ 10 フレーム）。低い MSE と高い SSIM スコアはより良い視覚的品質を示す。

方法	**SSIM/frame↑**	**MSE/frame↓**
ConvLSTM (NeurIPS2015)	0.707	103.3
FRNN (ECCV2018)	0.819	68.4
VPN (ICML2017)	0.870	70.0
PredRNN (NeurIPS2017)	0.869	56.8
PredRNN++ (ICML2018)	0.898	46.5
MIM (CVPR2019)	0.910	44.2
E3D-LSTM (ICLR2019)	0.910	41.3
CrevNet (ICLR2020)	0.928	38.5
MAU (w/o recalling)	0.931	29.5
MAU	**0.937**	**27.6**

表 2.2: Moving MNIST データセット（10 フレーム→ 10 フレーム）に対する比較。公平な比較のため、すべてのモデルのエンコーダとデコーダは同じ構造をしており、すべてのモデルは MSE 損失に基づいて Adam オプティマイザーを用いて訓練されている。

方法	バックボーン	**MSE↓**	**SSIM↑**	パラメータ	推論時間
ConvLSTM (NeurIPS2015)	4 × ConvLSTMs	102.1	0.747	0.98M	16.47s
ST-LSTM (NeurIPS2017)	4 × ST-LSTMs	54.5	0.839	1.57M	17.74s
Casual-LSTM (ICML2018)	4 × Casual-LSTMs	46.3	0.899	1.80M	21.25s
MIM (CVPR2019)	4 × MIMs	44.1	0.910	3.03M	45.13s
E3D-LSTM (ICLR2019)	4 × E3D-LSTMs	40.1	0.912	4.70M	57.21s
RPM (ICLR2020)	4 × RPMs	42.0	0.922	1.77M	18.01s
MotionGRU (CVPR2021)	4 × MotionGRUs	34.3	0.928	1.16M	17.58s
MAU	4 × MAUs	**29.5**	**0.931**	**0.78M**	**17.34s**

第 3 章

データ

3.1 SDO / AIA

モデルの学習及び評価データとして、NASA の Solar Dynamic Observatory(SDO)[1] の Atmospheric Imaging Assembly(AIA)[2] で撮影された紫外線観測データを用いた。

SDO は NASA の Living With a Star（LWS）プログラムの一つとして 2010 年 2 月に打ち上げられた太陽観測衛星である。AIA、Helioseismic and Magnetic Imager(HMI)、Extreme Ultraviolet Variability Experiment(EVE) などの高い空間解像度、時間分解能を持つ観測機器を搭載し、地上では不可能な多くの波長でのデータを提供する。その観測データを用いることにより、太陽物理学、宇宙天気、また地球環境に関する理解や洞察を深めることが期待されている。

AIA は主に太陽大気を観測する観測機器であり、4 つの望遠鏡で構成されている。また、7 つの極紫外線フィルターと、2 つの紫外線フィルター、および 1 つの可視光フィルターを持ち、広範な温度帯で太陽大気を観察することを可能にしている。本研究で用いられる 171 Å、193 Å、211 Åの 3 つのフィルターは、36 秒間隔で撮影され、4096 × 4096、約 1.5 秒角の空間解像度を持つ

これらのデータは Joint Science Operations Center(JSOC) によって提供されており、Python の太陽物理学を支援するライブラリである Sunpy を用いてダウンロードすることができる。

3.1.1 AIA 211 Å

表 3.1 のように、211 Å (21.1 nm) のフィルターは、約 200 万 K の 14 価鉄 (Fe XIV) イオンが放射するスペクトルを捉えるために特化している。この波長での観測は、活動領域のコロナを観測するのに最適である。図 3.1 は、これらの特性を示している。

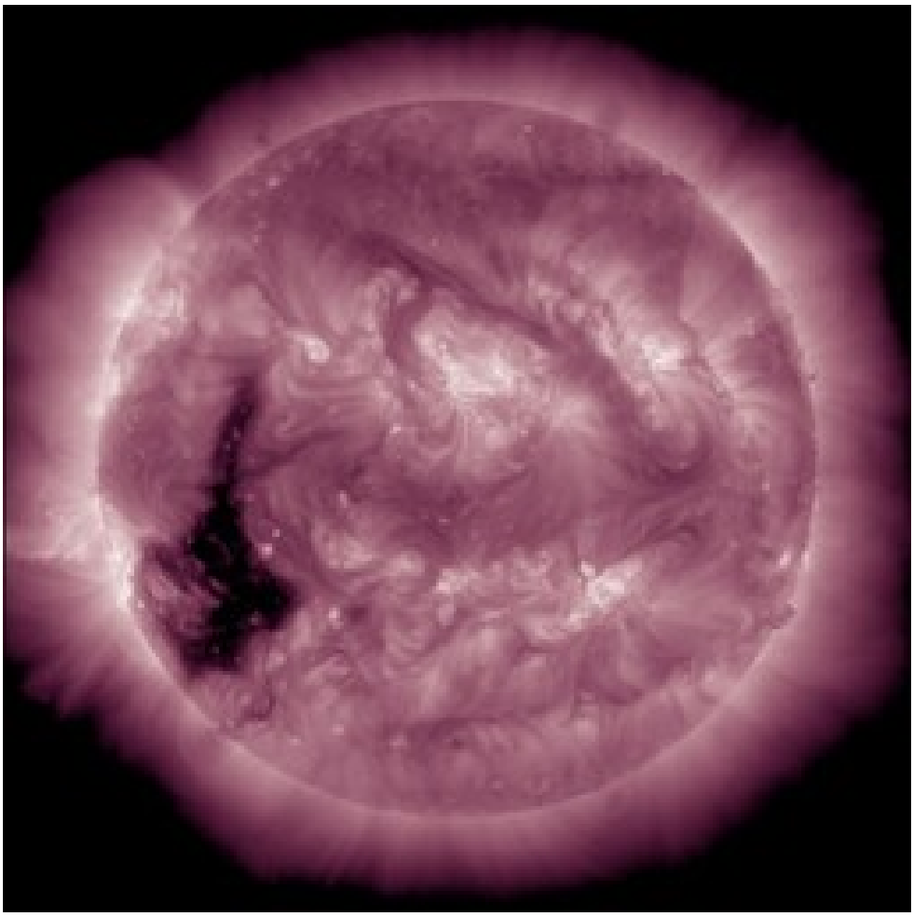


図 3.1: SDO/AIA の 211 Åフィルターで撮影された太陽全球紫外線像。強調のために紫色に色付けされている。球面の中上部から中下部には明るく輝く活動領域が見られ、左下部に暗くコロナホールが観測できる。

3.1.2 AIA 193 Å

表 3.1 のように、193 Å (19.3 nm) のフィルターは、約 150 万 K の 12 価鉄 (Fe XII) イオンが放射するスペクトル、または約 2000 万 K の 24 価鉄 (Fe XXIV) イオンが放射するスペクトルを捉えるために特化している。前者は主にコロナの中程度の高温領域を観測するために用いられ、後者は、主にコロナの高温フレアプラズマを観測するために用いられる。さらに、コロナホールも強調して観測することができる。図 3.2 は 193 Åフィルターで捉えた太陽像を示している。

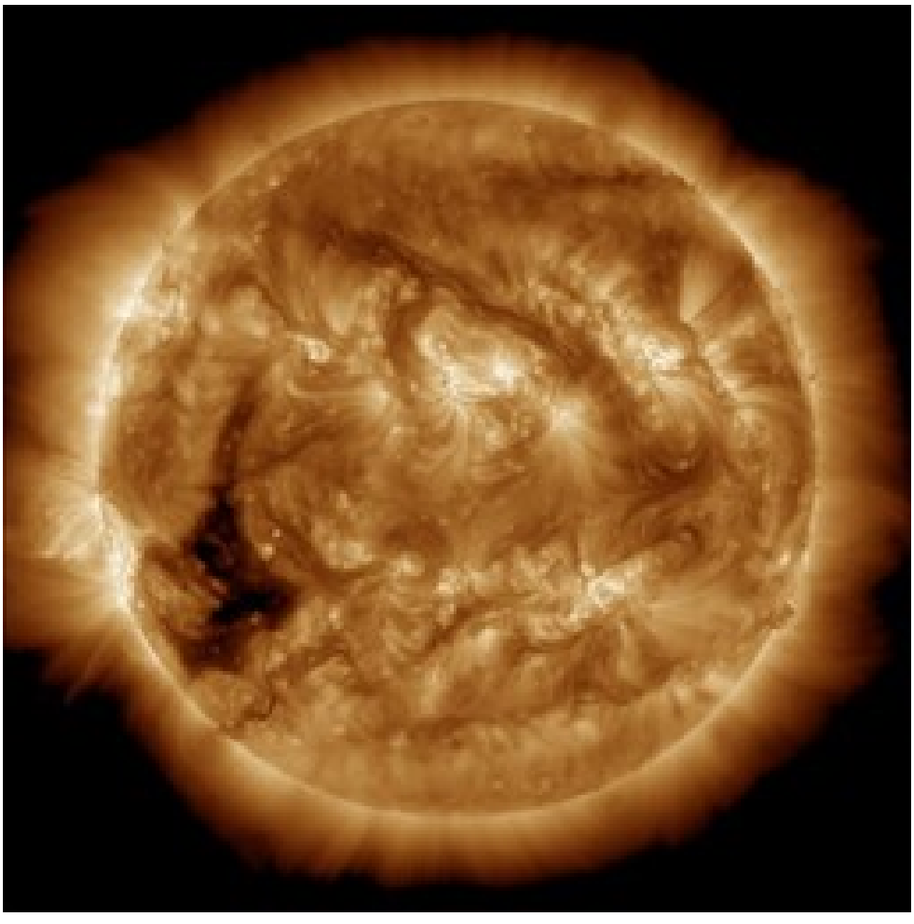


図 3.2: SDO/AIA の 193 Åフィルターで撮影された太陽全球紫外線像。強調のために橙色に色付けされている。

3.1.3 AIA 171 Å

表 3.1 のように、171 Å (17.1 nm) のフィルターは、約 60 万 K の 9 価鉄 (Fe IX) イオンが放射するスペクトルを捉えるために特化している。この波長での観測は、太陽のコロナループ、静穏領域コロナ、コロナホールなどの磁気構造を詳細に観察することができる。図 3.3 に示される 171 Åフィルターによる観測は、これらの特徴を捉えている。

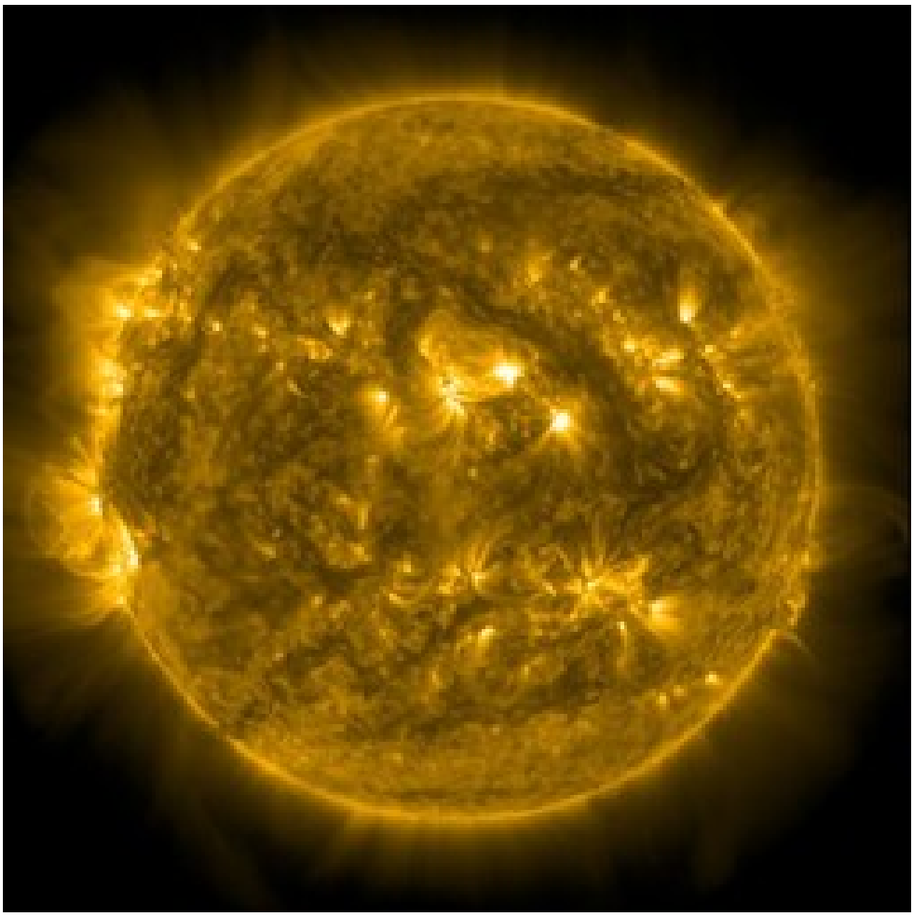


図 3.3: SDO/AIA の 171 Åフィルターで撮影された太陽全球紫外線像。強調のために黄色に色付けされている。193 Å、211 Åでは観察できない、コロナホールなどの静穏領域のスペクトルも観測できる。

3.2 前処理

本研究で用いるデータセットには、SDO/AIA のデータが提供されている 2010 年 5 月から、2022 年 10 月までのデータが含まれている。この期間に存在するデータから、4 時間ごとにデータを抽出し、各波長ごとに約 22000 枚をデータセットに含んでいる。データは JSOC により提供されているものをダウンロードし、その後、不正な画像を除去し、正規化やスケーリングを行った後、学習用、検証用、テスト用に分割した。この手順をアルゴリズム**??**に示す。これらのデータを、24 枚の画像を 1 セットとして分割する。各セットは 24 枚の時系列に並んだ画像で構成され、太陽全球の空間的情報の時間的変化を捉えている。24 枚のうち、前半の 12 枚、すなわち 48 時間までを入力シークエンス 、

表 3.1: SDO AIA の 171, 193, 211(Å) の特性

フィルター (Å)	主要イオン	大気の領域	温度帯 (K)
171	Fe IX	静穏領域コロナ, 上層遷移領域	6.3×10^5
193	Fe XII, XXIV	コロナと高温フレアプラズマ	$1.5 \times 10^6, 2.0 \times 10^7$
211	Fe XIV	活動領域コロナ	2.0×10^6

後半の 12 枚、すなわち 52 時間から 96 時間までを出力シークエンスとして扱う。学習の際は、入力シークエンスに対して出力シークエンスを教師データとして扱い、テストの際は入力シークエンスに続くモデルにとって未知の出力シークエンスを再現できるか検証する。

このデータセットは第 24 太陽活動周期の初期から、第 25 周期の初期までの観測データを網羅している。この時間範囲には、太陽活動の活発性が高いフェーズと低いフェーズの両方が含まれている。従って、このデータセットは太陽活動の活発性に依存しない可能性が高く、その汎化能力に対する期待が一定程度裏付けられる。

3.2.1 不正な画像の除去

SDO/AIA 望遠鏡で撮影された全球画像には、露光時間が他の画像より極端に低い、画像内に太陽全体を捉えていない、などの不正な画像が含まれている。確認することができた主な不正な画像を図 3.4 に示す。

これらの画像は、モデルの学習に悪影響を及ぼす可能性があるため、データセットから除去した。機械学習のタスクによっては、十分なデータセットがあれば、モデルが不正な画像に対する頑健性を獲得し、不正な画像がデータセットに含まれていても、学習結果にあまり大きな影響を与えない場合がある。しかし、本研究で行う動画予測は、データセットに含まれる画像がそのまま教師データとなるため、不正な画像は損失関数の計算、またはモデルの評価に大きな影響を与えるため、慎重に除去する必要がある。データの除去には、FITS ファイルのヘッダーに記録された各キーワードの値に対して閾値を設定して判定したのち、numpy による輪郭検出を用いた月蝕判定関数により不正な画像を排除した。

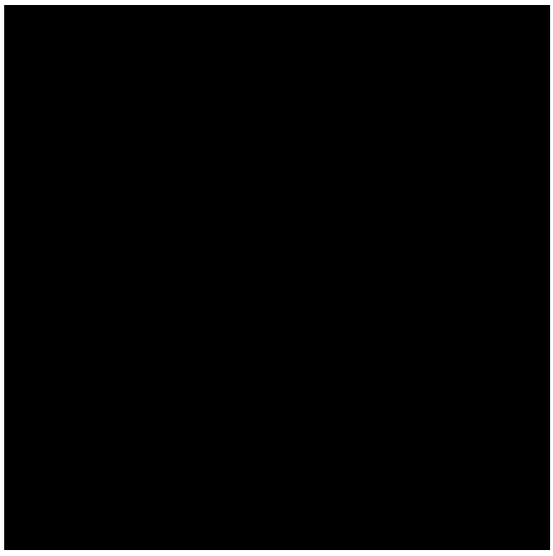
3.2.2 スケーリングと正規化

太陽動画データセットの前処理として、以下のステップを実施する。

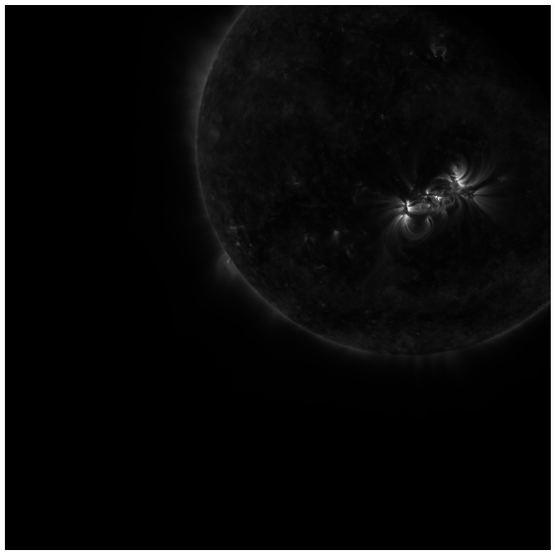
1. **正規化**: クリッピング処理されたデータを 0 から 1 の範囲に正規化する。ここで、ノイズによる負の値を削除し、極端に大きい外れ値の影響を削減するために、画像内の全ピクセル値に対して最小値

Algorithm 1 太陽動画データセット作成アルゴリズム

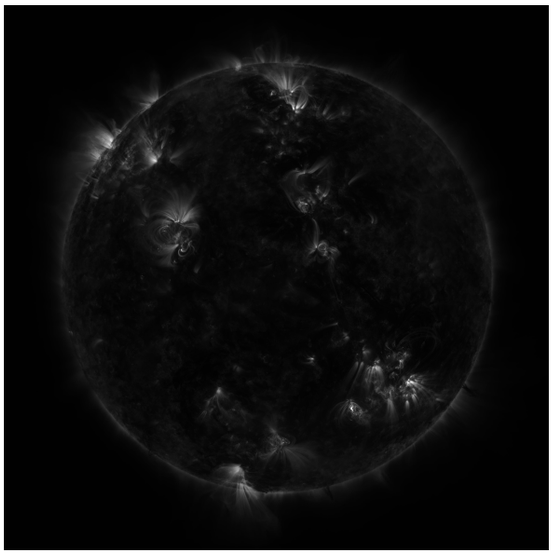
```
procedure CreateSolarDataset
    for each wavelength in wavelengths do
        images ← DownloadData(wavelength)
        images ← ValidateAndReplaceImages(images)
        all_images.extend(images)
    end for
    dataset ← CreateDataset(all_images)
    processed_dataset ← PreprocessDataset(dataset)
    train, val, test ← SplitDataset(processed_dataset)
    return train, val, test
end procedure
function DownloadData(wavelength)
    Download data for given wavelength
    return images
end function
function ValidateAndReplaceImages(images)
    for each image in images do
        if not ValidateImage(image) then
            alternative ← FindAlternativeImage(image.timestamp)
            images.replace(image, alternative)
            ValidateAndReplaceImages(images)                ▷ Recursive call
        end if
    end for
    return images
end function
function CreateDataset(images)
    Create dataset by grouping 24 images
    return dataset
end function
function PreprocessDataset(dataset)
    Apply preprocessing to each image in dataset
    return processed_dataset
end function
function SplitDataset(dataset)
    Split the dataset into train, validation, and test sets
    return train, val, test
end function
```



(a) 短い露光時間により、極端に暗い画像。



(b) 太陽が画像の中心にない画像。



(c) 衛星が回転しており、正しい角度で太陽が撮影されていない画像。活動領域の少ない左下部と右上部が極である。

図 3.4: SDO/AIA により観測された不正な画像の例

を 0、最大値を 10000 に設定した:

$$I_{normalized}(x,y) = \frac{min(max(I(x,y),0),10000)}{10000} \tag{3.1}$$

2. **平方根スケーリング**: ダイナミックレンジの広さに対応するために、正規化されたデータに平方根スケーリングを適用する。この過程は以下の式で示される。図のヒストグラムに示すように、スケーリングが適用されていないデータは、下位 5% 程度の範囲に極端に輝度が集中している。平方根スケーリングを適用することで、輝度間の相対的特徴を犠牲にせずに、極端な差を緩和することができる。

$$I_{scaled}(x,y) = \sqrt{I_{normalized}(x,y)} \tag{3.2}$$

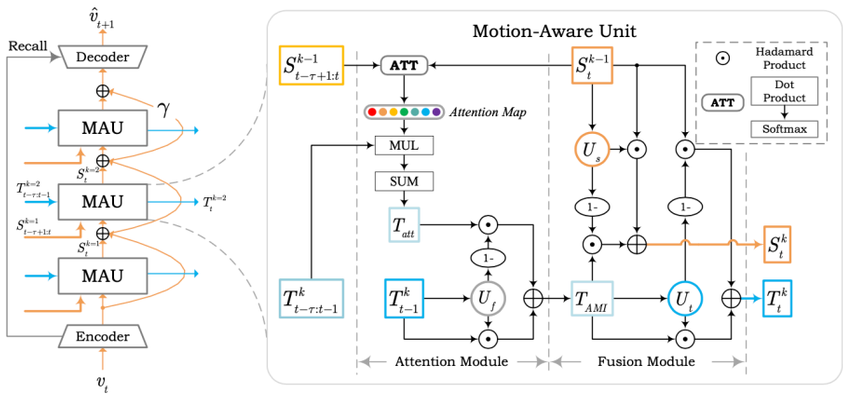


図 3.5: 正規化されたデータに対する平方根スケーリングの効果。左: 正規化されたデータのヒストグラム。右: 平方根スケーリングを適用したデータのヒストグラム。

3. **リサイズ**: 効率的な処理のために、4192px x 4192px の画像を 512px x 512px の解像度にリサイズした。この処理は、画像の空間解像度を低下させるが、太陽全球の大規模な構造を捉えるには十分である。

3.2.3　データセットの分割

このようにして作成されたデータセットは、約 1000 セットになり、これを学習用データセットに約 800 セット、検証データセットに 50 セット, テストデータセットに 50 セットというように分割した。

実験	実験 1	実験 2
入力波長	211 Å	171 Å,193 Å,211 Å
出力波長	211 Å	
総枚数	22000	66000
セット数	232	
セットごとの枚数	入力 12 → 出力 12	
解像度	512 * 512	

表 3.2: 各実験でのデータセット

第 4 章

Motion Aware Unit を用いた 1 波長を入力とした紫外線像の全球時系列予測

4.1　実験概要

実験 1 では入力、出力ともに 211 Å フィルターで得られたデータを利用した。これは 211 Å フィルターで撮影された紫外線像が、コロナホールと活動領域といった、二つの太陽円盤上の大規模構造をバランスよく明瞭に表現し、本研究のモデルの効果検証に適していると考えたためである。

4.2　実験設定

各ハイパーパラメータの設定を表 4.1 に示す。

ハイパーパラメータ	値
バッチサイズ	4
エポック数	100
学習率	0.0005
損失関数 a	MSE
カーネルサイズ	(5, 5)
MAU Cell 数	16

表 4.1: 実験 1 でのハイパーパラメータ

4.3　学習の推移

4.4　実験結果

4.4.1　全球での評価

平均輝度とその誤差

画像類似度

単純差動回転モデルとの比較

4.4.2　経度依存性の評価

平均輝度とその誤差

単純差動回転モデルとの比較

4.4.3　東側リムから出現する活動領域に対する視覚的評価

4.5　考察

第 5 章

Motion Aware Unit を用いた 3 波長を入力とした紫外線像の全球時系列予測

5.1　実験概要

実験 2 では入力に 171 Å、193 Å フィルターで得られたデータを追加で利用した。これらの波長を追加することで、より広範な温度帯に渡る太陽活動をモデルが学習することを期待している。

5.2　学習の推移

5.3　実験結果

5.3.1　全球での評価

平均輝度とその誤差

画像類似度

単純差動回転モデルとの比較

5.3.2　経度依存性の評価

平均輝度とその誤差

単純差動回転モデルとの比較

5.3.3　東側リムから出現する活動領域に対する視覚的評価

5.4　考察

第 6 章

まとめ

謝辞

参考文献

[1] W Dean Pesnell, B J Thompson, and PC Chamberlin. *The solar dynamics observatory (SDO)*. Springer, 2012.

[2] James R Lemen, Alan M Title, David J Akin, Paul F Boerner, Catherine Chou, Jerry F Drake, Dexter W Duncan, Christopher G Edwards, Frank M Friedlaender, Gary F Heyman, et al. The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo). *Solar Physics*, Vol. 275, pp. 17–40, 2012.