機械学習に基づく波浪ビデオ分析システムの構築
Construction of Wave Video Analysis System Based on Machine Learning

21N3100038H　CHEN　Jiangnan（海岸・港湾研究室）
CHEN Jiangnan / Coastal Engineering Lab.

Key Words : *Machine learning, Convolutional Neural Network, Video analysis*

1. はじめに

従来の方法で海洋や河川の波のパラメータを測定するには，ブイなどの設備を使用する必要があり，これらの設備はコストが高いだけでなく，専門家によるメンテナンスの必要もある．もし波のビデオデータを解析して波の波高などのパラメータを測定することができれば，従来の方法よりコストを大幅に削減できる．

機械学習に基づくビデオ分析システムは，多くの分野で広く使用されている．その中でも，人工ニューラルネットワークモデルの一種である3DCNN(3D畳み込みニューラルネットワーク)は，ビデオの空間的および時間的特徴を抽出でき，行動認識，ジェスチャー認識に使用されている．この研究の目的は，ビデオデータを入力，波高を出力とする 3DCNNモデルを構築し，波高測定のコストを削減することである．[1)]

2. トレーニングデータの取得

機械学習の教師あり学習モデルとして，3DCNNモデルは，事前に準備されたトレーニングデータ(この研究では，波のビデオデータと対応する波高値となる)を使用して，モデルの出力と正しい値を比較し，勾配降下法に基づく最適化手法と誤差逆伝播法で，ノードと畳み込みカーネルのパラメータを更新して，モデルの出力を正しい値に近づけることができる．

したがって，モデルをトレーニングし，その有効性を検証するには，波のビデオデータとそれに対応する波高データを取得する必要がある．

(1) ビデオと波高データの取得

本研究は，断面水槽の孤立波を撮影して，モデルトレーニングに必要な波のビデオデータと対応する波高値を取得する．実際に海面などを撮影するときの撮影角度をシミュレートするために，カメラを水面と平行に水槽の上に配置する．撮影する水面のところに波高計を設置して波高データを取得する．

(2) ビデオと波高計の同期

データをモデルに入力する前に，処理する必要がある．まず，ビデオの各フレームに対応する波高値を求めるためには，波高計の出力とビデオを同期させる必要がある．本研究では，スイッチで制御するLEDをカメラに向けて設置し，波高計の出力信号とLEDの電圧信号を同時にサンプリングする．**図-1**に示すように，造波する前と後スイッチを2回押下し，撮影したビデオのLEDの部分の明るさを分析すると，LEDが発光し始めるフレームと，LED両端の電圧が上昇した時の波高計のタイムステップ数を知ることができる．これによると，ビデオフレーム数と波高計出力タイムステップ数との線形関係を決定することができ，ビデオの各フレームに対応する波高を決定できる．

(3) ビデオの前処理

一つの孤立波のビデオに対して，波高計の出力が最大となる瞬間前0.5 sから，1 s間のビデオを切り取って，入力として3フレームごとに1フレームを選択し，10フレームの画像を取得する．トレーニングに必要な水面部分だけ残し，解像度を処理して配列化する．配列化したビデオと波高計で得た孤立波の波高を一セットとして，モデルのトレーニングに使う．

なお，波の色はほぼ単色であるため，色には波高に

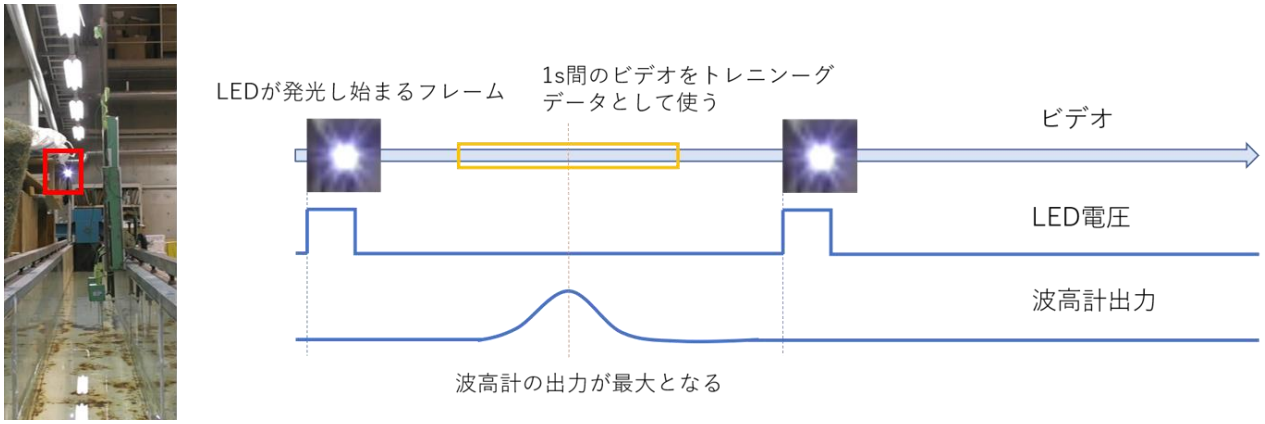

図-1ビデオと波高計出力の同期

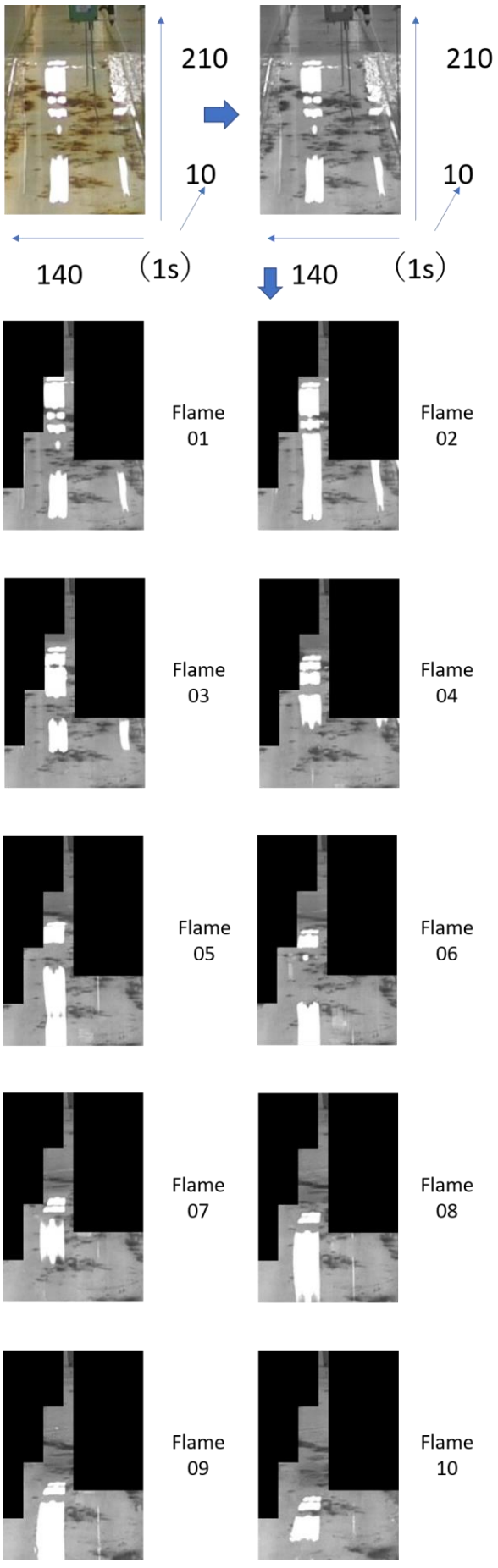

図-2　入力データ

関連する重要な情報はほぼ含まれず，カラービデオをグレースケール化して入力データのサイズを小さくすることで，モデルのパラメータ量を削減することができるので，ビデオのグレースケール化を行う．

(4)　無関係な特徴の排除

断面水槽で波を撮影すると，水面と水槽の壁にエッジが生成される．エッジのあるビデオでモデルをトレーニングすると，モデルは波高とエッジの形状の関係に適合する可能性がある．しかし，実際の波を撮影する場合，水槽の壁のように水面と接触する固定な参照物があるとは限らない．したがって，実際の波のビデオをシミュレートするには，トレーニングデータを処理する必要がある．

本研究では，**図-2**に示すように，画面中の水面と水槽の壁のエッジ，水面と波高計の接点を黒にして，モデルが実際の波のビデオから得られない情報の影響を受けないように、トレーニングデータを処理している．

3.　本研究のモデル設計

(1)　3DCNN（3D畳み込みニューラルネットワーク）

本研究で使用される3DCNNモデルは，主に全結合レイヤー，3D畳み込みレイヤー，およびプーリングレイヤーで構成されている．

全結合レイヤーは最も基本的な人工ニューラルネットワークのレイヤーであり，複数のノードで構成されている．ノードは前のレイヤーの出力ベクトルにノードの重みベクトルを掛けドット積を計算する．得た値をノードのバイアスと加算し，非線形の活性化関数を介して，このノードの出力値が取得される．一つのレイヤーのすべてのノードの出力値がこのレイヤーの出力ベクトルを形成する．

ビデオデータの特性に応じて，3D畳み込みレイヤーはローカル結合とパラメータ共有，二つの特性を持っている．3D畳み込みレイヤーでは，ノードの代わりに3D畳み込みカーネルを使用し，入力された3次元行列から出力を計算し，入力の空間および時間的特徴を抽出できる．[2]プーリングレイヤーでは，プーリングカーネルによって入力データのサイズを減らすことができる．畳み込みレイヤーとプーリングレイヤーを組み合わせることで，時間と空間の情報を含む入力データを次元削減し，特徴ベクトルに変換できる．

(2)　本研究のモデル設計

既存の汎用大型3DCNNモデルは，複雑な環境で目標特徴を抽出することができる．しかし，そのようなモデルは多くのパラメータを持ち，モデルのトレーニングには大量のデータが必要であり，結果としてコストが高くなり，本研究の目的と矛盾する．低コストで波ビデオの解析を実現するには，タスクに適した小型モ

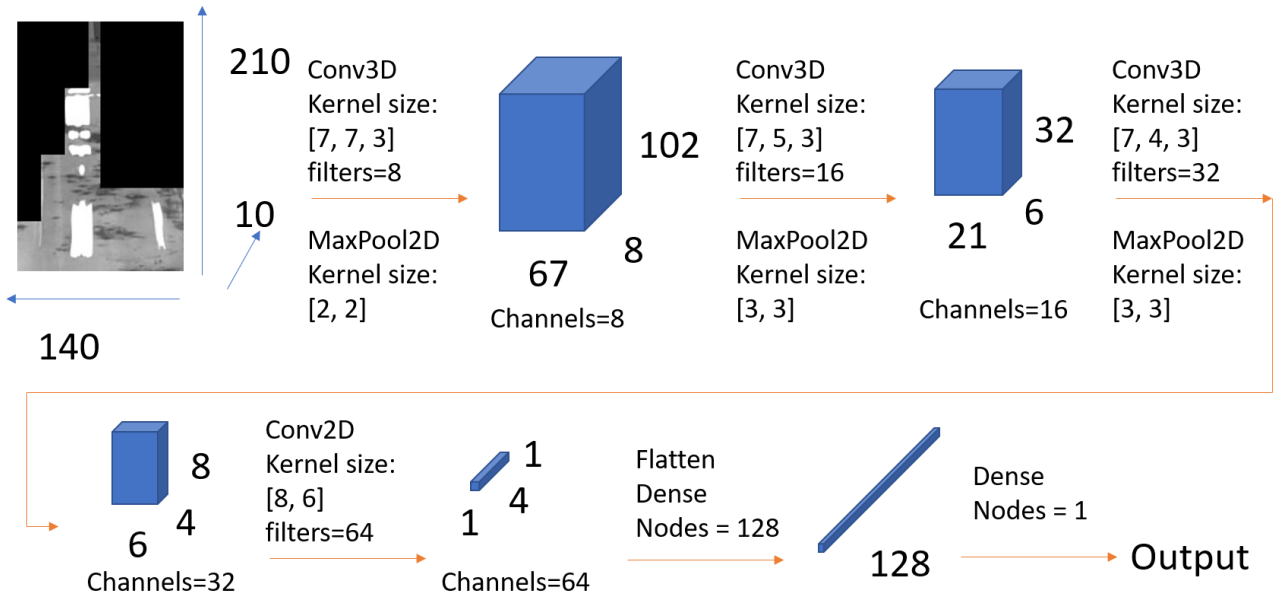

図-3　モデルの構造

デルが望ましいと考えられる．

本研究では，3D畳み込みと2D最大値プーリングを使用して，ビデオデータをビデオの空間と時間の特徴を含む特徴テンソルの集合に変換する．次に2次元畳み込みを使用してデータの次元を削減する．[3]その結果の特徴ベクトルを全結合レイヤーに入力し，出力の波高データを取得する．モデルの具体的な構造，畳み込みカーネルのサイズ，およびモデル計算中特徴テンソルの形状変化を図-3に示す．

4.　トレーニングの結果

本研究では，断面水槽で撮影された波のビデオと測定された波高データを使用して，120セットのデータを含むトレーニングセットと16セットのデータを含む検証セットを作成し，モデルをトレーニングした．

モデルのトレーニングにランダム性が含まれるため，100回のトレーニング回数で，モデルを三回トレーニングした．その結果，トレーニングセットでの平均誤差が0.0424 cmで，検証セットの平均誤差が0.1525 cmとなる．（参考として，撮影時計測した最小波高は3.83 cm，最大波高は5.50 cmである．）

三回のトレーニングの中，一回を例としてトレーニング回数に応じて変化するモデル誤差の曲線を図-4に示す．最初の数回のトレーニングで，モデルの誤差は大幅に減少する．モデルはバッチ単位でパラメータを更新するため，一時的に検証セットの平均誤差はトレーニングセットの平均誤差よりも小さくなる．5回目から60回目のトレーニングではモデルの誤差は徐々に減少し，60回目のトレーニング以降はモデル誤差の減少速度が段々に小さくなり，最後に，トレーニングセットのモデルの平均誤差は約0.04 cmで安定した．検証セットの平均誤差は約0.16 cmで安定している．

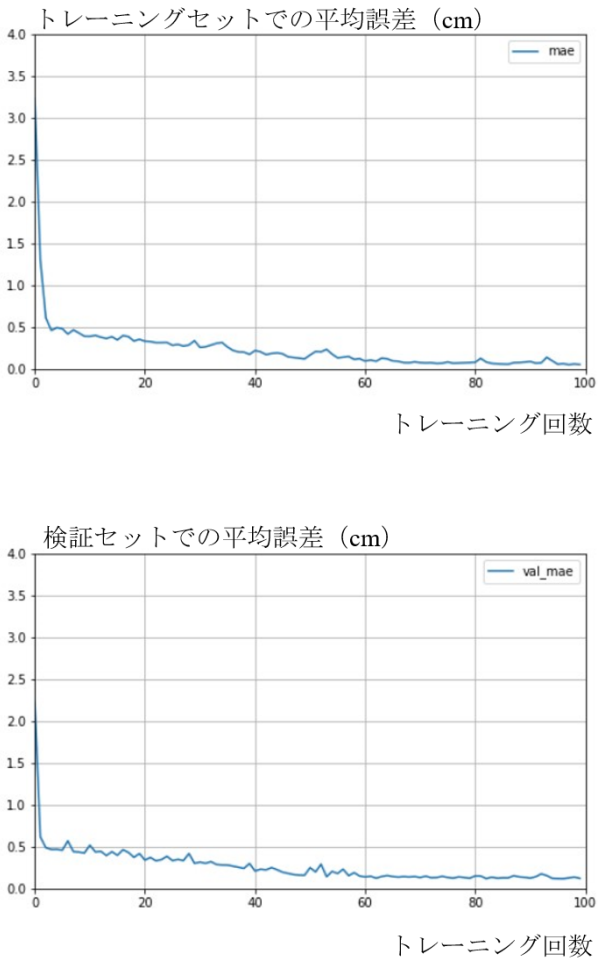

図-4　10フレームのデータセットでモデルのトレニーング結果

5. 時間的情報による精度への影響

既存のウェーブビデオ分析方法は，多くの場合，ビデオを独立したフレームに分割するか，いくつかのフレームで平均画像を作成してから，[4]エッジ検出またはCNNを使用して分析を行うが，この場合，ビデオに含まれる時間的情報が破棄されてしまう．

ビデオに含めている時間的情報の量がモデルの波高の判断に影響するかどうかを検証するため，この研究では同じビデオデータを使用して，10フレームの時系列画像を含むデータセットに対して，3フレームの画像のみを含むデータセットを作成し，モデルをトレーニングした(入力データのサイズにより，モデル3D畳み込みレイヤーを1つのみにして，残りは2D畳み込みを使用する)．モデルの平均誤差は，**図-5**と**表-1**のように示す．モデルのパラメータが少ないため，最初の数回のトレーニングで，モデルの誤差は10フレームのデータセットを使用したモデルよりも速く減小するが，最終的な誤差は10フレームのデータセットを使用したモデルより大きい傾向が見られる．

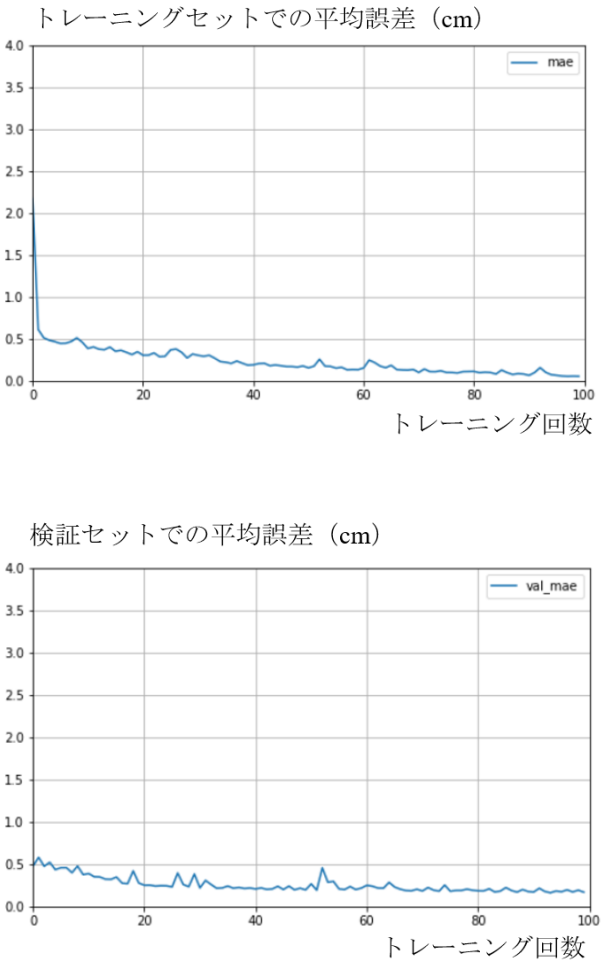

図-5 3フレームのデータセットでモデルのトレニンーグ結果

表-1 トレーニング結果の比較

データセット	tra_mae(cm)	val_mae (cm)
10フレーム	0.0424	0.1525
3フレーム	0.0529	0.1752

6. 結論

3DCNNモデルは，波のビデオデータから空間的および時間的特徴を抽出し，波高などの情報を得ることができる．

なお，モデルのトレーニングに使うデータに含む時間的情報の量は，モデルが波高を判断する精度に影響する．

7. 今後の課題

本研究では断面水槽で撮影したビデオデータを用いて3DCNNモデルをトレーニングし，一定の成果を得られたが，現実に波を撮影する場合，照明条件や環境の変化は複雑になり，取得したデータにはより多くの無関係な情報が含まれていることが想定される．3DCNNが実際の波のビデオ分析に応用する際に，どのような精度を達成できるか，どのくらいのデータを必要とするか，ビデオデータにどのような前処理が適切かはまだ検証する必要がある．

実際の波のビデオデータとそれに対応する波高データが入手できれば，異なる照明条件と環境におけるモデルの汎化能力と転移学習を使用する可能性も検討できる．

参考文献

1) Ji, S., Xu, W., Yang, M. and Yu, K： 3D convolutional neural networks for human action recognition.， IEEE transactions on pattern analysis and machine intelli-gence 35.1: 221-231,2012.
2) Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, Manohar Paluri; : Learning Spatiotemporal Features With 3D Convolutional Networks，Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 4489-4497.
3) Simonyan K, Zisserman A. : Two-stream convolutional networks for action recognition in videos. Advances in neural information processing systems. 2014;27.
4) 宮下 侑莉華，中村 友昭，菊 雅美，趙 容桓，水谷 法美：深層学習による海岸画像を用いた波浪推定に関する検討，土木学会論文集 B2(海岸工学)，2022，78巻，2号，p. I_127-I_132.