

令和5年度 卒業論文

深層学習を用いた セマンティックセグメンテーションによる 水面領域抽出の精度評価

広島市立大学 情報科学部 システム工学科

氏名 中島 慶
指導教員 島 和之 准教授

2024年1月31日 提出

令和5年度 卒業論文

深層学習を用いた セマンティックセグメンテーションによる 水面領域抽出の精度評価

広島市立大学 情報科学部 システム工学科

氏名 中島 慶
指導教員 島 和之 准教授

2024年1月31日 提出

目 次

| | |
|-----------------------------|-----------|
| 1 序章 | 1 |
| 1.1 研究背景 | 1 |
| 1.2 研究目的 | 1 |
| 2 関連研究 | 2 |
| 2.1 DeepLabv3+ | 3 |
| 3 先行研究 | 4 |
| 3.1 概要 | 4 |
| 3.2 画像の切り抜きとラベル付け | 4 |
| 3.3 水位の推定方法 | 5 |
| 3.4 相関係数による評価 | 6 |
| 4 提案手法 | 7 |
| 4.1 概要 | 7 |
| 4.2 アノテーション方法 | 8 |
| 4.3 水位の推定方法 | 9 |
| 5 評価実験 | 10 |
| 5.1 学習モデルの構築と評価 | 11 |
| 5.2 先行研究との比較 | 12 |
| 5.3 ラベル設定と評価 | 15 |
| 5.4 交差検証 | 18 |
| 6 考察 | 23 |
| 7 まとめ | 25 |
| 参考文献 | 27 |
| 謝辞 | 27 |

1 序章

1.1 研究背景

広島市で起きた平成 30 年 7 月の豪雨をはじめ全国各地で年間平均 1099 件（昭和 57 年から令和 4 年までの平均）発生している [1]。土砂災害は発生前に「降雨が続くにもかかわらず川の水位が低下する」や「崖から小石がパラパラと落ちる」といった前兆現象が起こることが確認されている [2]。そういった前兆現象を早期に検知することで、住民の避難を速やかに行うことができ人的被害の削減につながると期待されている。

河川の水位変動を水位計を用いて計測する方法がある。しかし、水位計にはコストの高さや整備の難しさ、といった問題があり、限られた台数しか水位計を設置することができない。そこで、安価かつ、整備が容易な監視カメラを用いた水位推定が水位計の代替案となっている。監視カメラ画像から水位を推定する研究は数多く行われ、本研究では、河川の監視カメラから得られる画像から水面領域を抽出し、水位の推定を行う。

1.2 研究目的

先行研究で渡邊 [3] は、前兆現象の 1 つである「降雨が続くにもかかわらず川の水位が低下する」現象に着目し、河川の様子を写した監視カメラ画像を 4 列 × 25 枚の計 100 枚の切抜画像へ変換を行い、この切抜画像に畳込みニューラルネットワークを使用し、分類の際に出力されるソフトマックス値と画像の縦座標値から、水位を推定する方法を提案した。しかし、訓練データと異なる日のテストデータに対しては、推定精度が低いという問題があった。

そこで、本研究でも同じ前兆現象に焦点を当て、深層学習を用いたセマンティックセグメンテーションによる水面領域抽出する手法（以下、SS 手法）を検討するために、その精度を評価した。また、深層学習を用いたセマンティックセグメンテーションでは、一般的に、訓練データとしてアノテーション付きの画像（以下、アノテーション画像）が多数必要であり、かつ、精度に影響する。ただし、多数のアノテーション画像を手作業で作成するには過大な労力がかかるので、可能な労力で必要な精度を達成するために、半自動的な作成方法を提案する。

2 関連研究

深層学習を用いたセマンティックセグメンテーションを使用して河川の監視カメラ画像から水位推定を行った研究は存在している。Nurら[4]は、河川の監視カメラ画像から水面領域を抽出するために畳込みニューラルネットワークに基づくセマンティックセグメンテーションを提案した。DeepLabv3+[5]とSegNet[6]という二つのセマンティックセグメンテーション向け畳込みニューラルネットワークで作成したモデルの領域抽出精度の比較を行った。結果はDeepLabv3+を用いて作成したモデルの方が高精度で領域抽出できていることが確認された。また、高い抽出精度を示したDeepLabv3+モデルを用いて、実際に、5日間の水位推定と水位変動の観測を行った。推定した水位と実際の水位は強い相関関係を示し、提案手法の有効性を報告した。

2.1 DeepLabv3+

DeepLabv3+[5] は Google Research Team によって考案されたセマンティックセグメンテーション向け畳込みニューラルネットワークであり、旧バージョンのモデルから多くの改良が施された、最先端のモデルの一つである。DeepLabv3+は図1に示すようにエンコーダ・デコーダ構造を持ち、Atrous convolution と Atrous spatial pyramid pooling を備えている。

Atrous convolution は Atrous rate というストライドに関するパラメータを持っており、一般的な畳込み層より大きい受容野を持ち、効率的に畳込みすることが可能である。また、それは式(1)のように表される。式(1)の $y[i]$ は出力、 $w[k]$ は長さ k のフィルタ、 x は入力である特徴マップ、 r は「Atrous rate」を表す。

Atrous spatial pyramid pooling の特徴は、異なる Atrous rate で複数の Atrous convolution を行い、それらを重ねアップサンプリングを行い、デコーダ部分で欠損した情報の復元に使用することである。

$$y[i] = \sum_k x[i + r \cdot k] \cdot w[k] \quad (1)$$

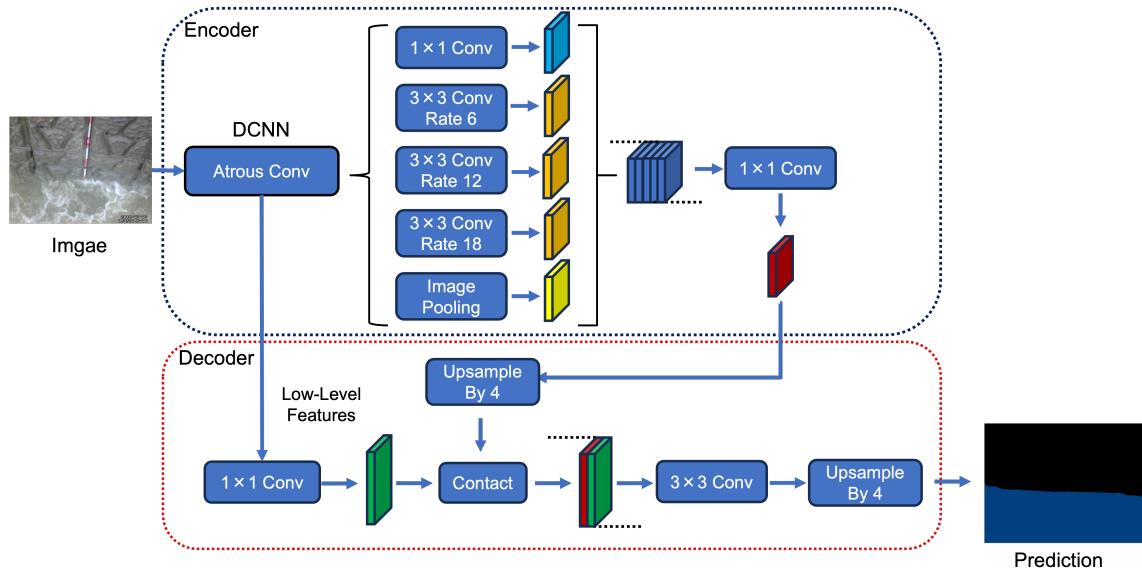


図1: DeepLabv3+モデル

3 先行研究

3.1 概要

先行研究 [3] は、土砂災害の前兆現象の一つである「降雨が続くにもかかわらず川の水位が低下する」現象に着目した。河川の監視カメラ画像を図2のように切り抜き、切抜画像が水面かどうかを畳込みニューラルネットワークによって判別を行い、判別されるときに出力されるソフトマックス値と画像の縦座標値を用いて水位の変動を推定した。深層学習を用いて水位を推定し、実際の水位変動との相関を評価をした。

3.2 画像の切り抜きとラベル付け

河川の監視カメラ画像の幅を5等分（幅64高さ×240pixel）になるように切り分け、そのうち、水位計測ポールを含む中央の列を除外を行った。残った各4列に対して、高さを元の河川画像を5等分した大きさ（各48pixel）となるように、幅64×高さ48pixelの画像を縦に8pixelずつ移動させて、最下段まで切り抜きを行った。切り抜いた画像に対して、以下に示す3つのクラスに分類してラベル付けを行った。

- ・ 「水面」：水面が写っている画像
- ・ 「壁面」：壁面が写っている画像
- ・ 「その他」：水面と壁面の両方が含まれる画像

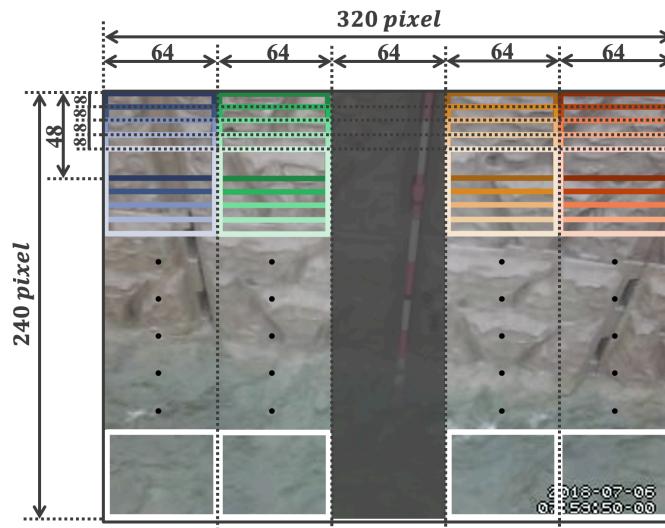


図2: 切抜画像の生成

3.3 水位の推定方法

3.2 節で生成した切り抜き画像を用いて、河川の水位を推定する。各切抜画像に対して、畳込みニューラルネットワークで学習と分類を行った際に、出力される「水面」クラスのソフトマックス値 S_{water} と、元画像における縦の座標値 y を用いて水位の推定を行った。水位の推定値 E は元画像から n 枚の切抜画像が得られたとき以下の式(2)で導出した。

$$E = \frac{\sum_n y S_{water}}{\sum_n S_{water}} \quad (2)$$

例として、水位の推定値を算出する流れを図3で説明する。なお、説明の際はわかりやすいように、簡略した5列×5枚で切り抜いた場合について考える。まず、中央の列は測定ポールを含んでいるため除外する。続いて、各切抜画像を分類し、「水面」クラスに属する確率に相当する、ソフトマックス値を算出する(切抜画像内の数値)。また、元の河川の画像における、各切抜画像の上辺と下辺の高さを二分する高さの座標値を求める。最後に、式(2)に数値を代入することで水位の推定値を算出する。例の場合、水位の推定値 $E = 190.8$ である。このとき、壁面と水面の境界線は図3の白線の部分であるが、水位の推定値は、画像中における水面の面積の半分にする縦の座標値に相当する。例で求めた水位の推定値にしたがって、縦の座標値波線の矢印を引くと、水面の面積をおよそ半分にする位置に求まっていることがわかる。

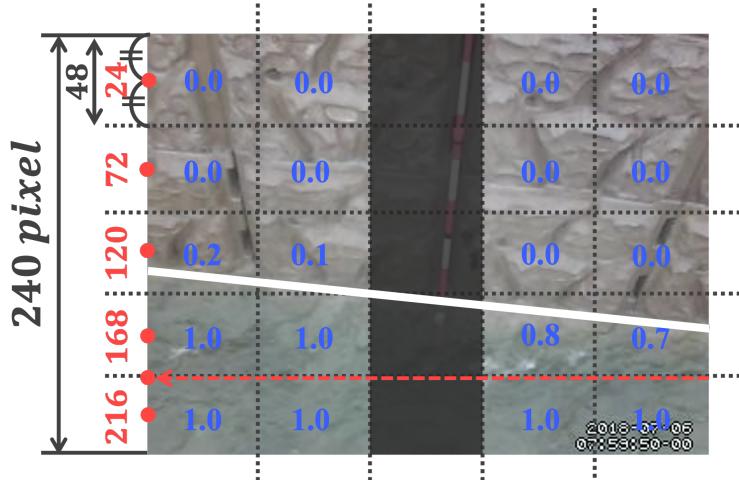


図3: 推定値算出の流れ

3.4 相関係数による評価

式 2 によって推定された推定値 E と、水位計測ポールを参考にして 10 分ごとに目視で測定した水位を線形補完した値（測定値）との相関係数を用いて評価した。

表 1: 推定値と測定値の相関

| クラス | 相関係数 |
|------|-------|
| 2 分類 | 0.956 |
| 3 分類 | 0.930 |

なお、「水面」「壁面」の 2 つのクラスで学習と分類を行った結果を 2 分類、「水面」「壁面」「その他」の 3 つのクラスで学習と分類を行った結果を 3 分類とした。

4 提案手法

4.1 概要

深層学習を用いたセマンティックセグメンテーションでは、一般的に、訓練データとしてアノテーション画像が多数必要であり、かつ、精度に影響する。また、アノテーション画像を手作業で作成する場合、過大な労力がかかる。そこで、本研究では、可能な労力で必要な精度を達成するための、半自動的なアノテーション方法を提案する。図4に使用した元画像とアノテーション画像の例を示す。

画像内の水面領域の面積割合は水位が上昇するとともに大きくなるので、水面領域の面積割合から水位へ変換する式（変換式）を作成した。SS手法を用いて領域抽出を行った画像から、変換式を用いて水位の推定を行った。

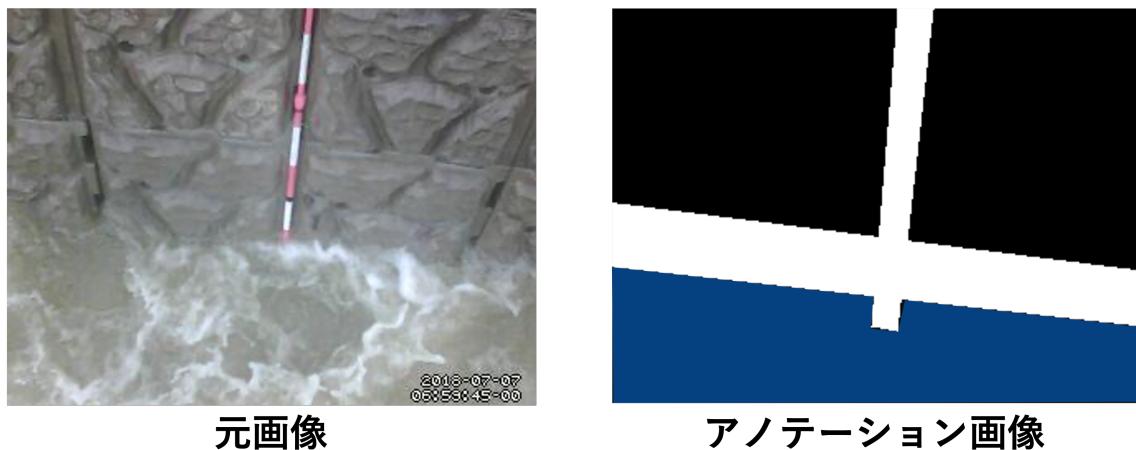


図 4: 元画像とアノテーション画像の例

4.2 アノテーション方法

本研究で扱った河川の監視カメラ画像には「水面」「壁面」「水位計測ポール」「撮影日時」の4つの範囲がある。そこで、本研究では、以下の方法によってアノテーション画像を作成した。

- ・「水面」と「壁面」の範囲は水位によって、変化するので、まず目視によって10分毎の水位を推定し、その測定値から線形補間によって自動的に範囲を決定した。
- ・確実なラベル付けを行うため、「水面」と「壁面」の境界が曖昧な範囲は、学習対象から除外した。
- ・「水位計測ポール」と「撮影日時」の範囲は時間によらず変化しないので、目視で決定した。

「水面」と「壁面」の範囲決定の様子を図5に示す。

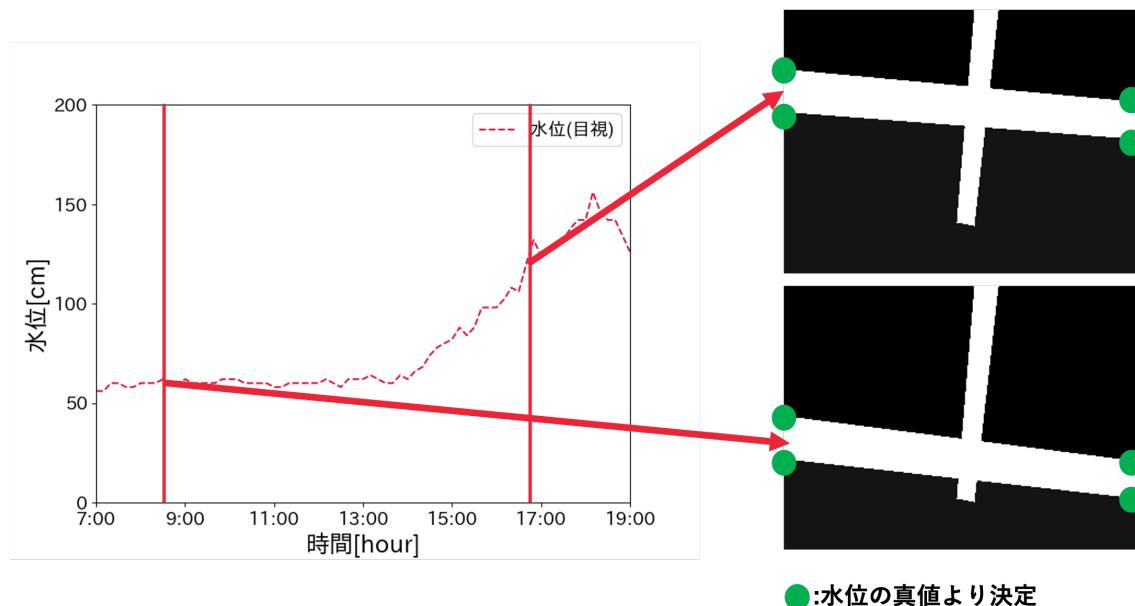


図5: 範囲決定の流れ

4.3 水位の推定方法

SS 手法によって領域抽出された画像内の水面領域と推定されたピクセル数を全体のピクセル数 (all) で割り , 画像内の水面領域の面積割合を取得し , 変換式を用いて水位の推定を行った . 水位の推定値 E は抽出された画像内に , 水面領域だと推定されたピクセルが n 個の時以下の式 (3) で算出した .

$$E = 243.78 \times \frac{n}{all} - 23.23 \quad (3)$$

5 評価実験

目視で計測した水位と SS 手法を用いて推定した水位の RMSE(Root Mean Square Error) を先行研究 [3] で比較することで , SS 手法の評価を行う . また , 提案したアノテーション方法で作成したモデルの領域抽出精度を評価することで提案手法の有効性を示す .

本研究では河川画像として , 本学情報工学科モニタリングネットワーク研究室にて運用中の広島市安佐北区三入地区桐原川の監視システムにおいて撮影された図 6 の監視カメラ画像を使用した . また , 置込みニューラルネットワークによるセマンティックセグメンテーションのシステムとして , DeepLabv3+[5] を用いた .



図 6: 桐原川の監視カメラ画像

5.1 学習モデルの構築と評価

セマンティックセグメンテーションモデルの学習回数を変化させて領域抽出を行い、IoU(Interaction over Union)[7]を評価することでモデルの適切な学習回数の調査を行った。

領域抽出結果についての評価は、Mean IoUを用いて行われるのが一般的である。IoUは式(4)で表せられる予測領域と正解領域のオーバーラップ率を意味する値である。式(4)を構成するTP,FP,FNの定義は図7に示す。すべての画像に対するIoUをクラス毎に平均したAverage IoUを、さらに全クラスで平均した値がMean IoUとなる。ただし、本研究では、クラス数が少なくまた、水面クラスの抽出結果に対してのみ評価を行うことを目的としているので、水面クラスのAverage IoU(以下,IoU)に着目して評価を行うこととした。

$$IoU = \frac{TP}{(TP + FP + FN)} \quad (4)$$

| | | 予測 | |
|----|-----------------------------|-------------------------------|-------------------------------|
| | | Positive (水面 と予測) | Negative (背景 と予測) |
| 正解 | Positive (水面 領域) | 真陽性 TP (True Positive) | 偽陰性 FN (False Negative) |
| | Negative (背景 領域) | 偽陽性 FP (False Positive) | 真陰性 TN (True Negative) |

図7: 領域抽出における分類問題の混合行列の定義

2018年7月6日の7:00から19:00までの画像を訓練データ(8147枚)に、テストデータには同年の7月7日同時間帯の画像データ(8086枚)を用いた。また、モデル作成に使用したアノテーション画像は以下に示す設定(設定A)でアノテーションした。

- ・ 「水面」 : 水面ラベル
- ・ 「壁面」 : 背景ラベル
- ・ 「水位計測ポール」 : 学習除外ラベル
- ・ 「撮影日時」 : 水面ラベル

学習回数を 50,100,200,300...1000 と変化させてセマンティックセグメンテーションモデルセマンティックセグメンテーションモデルを作成した。作成したモデルを用いて領域抽出を行い、IoU を求めた。図 8 はモデルの学習回数と IoU の関係を表したグラフである。

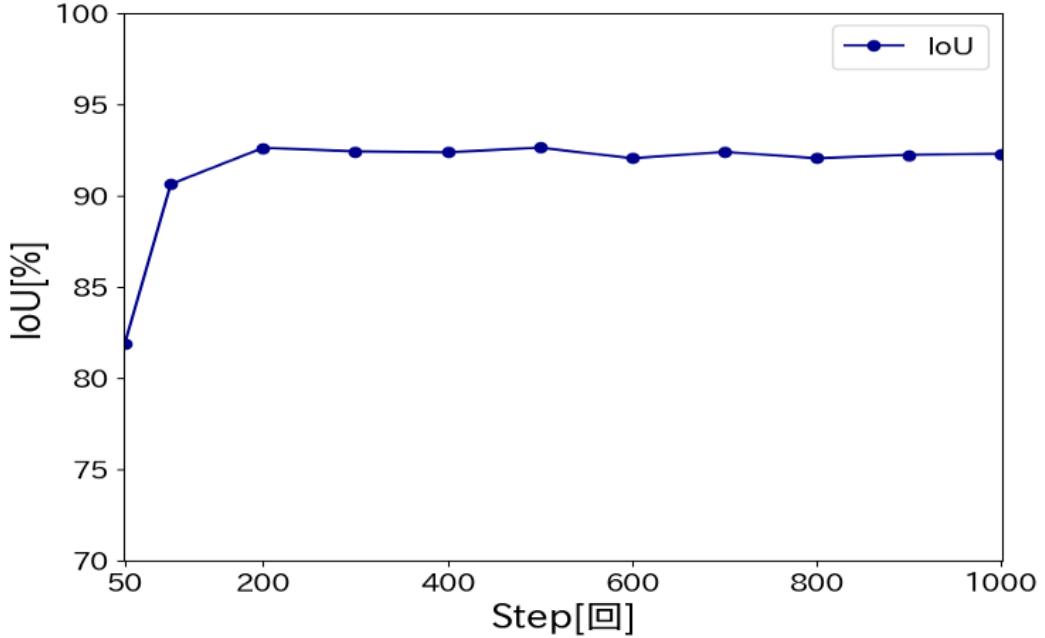


図 8: 学習回数と IoU の関係

5.2 先行研究との比較

SS 手法の領域抽出精度を評価するため、時系列の河川の監視カメラ画像に先行研究 [3] の手法と SS 手法を適応し、推定した水位と目視で計測した水位との RMSE を比較した。先行研究のラベル設定と統一するため、設定 A でアノテーションを行い、訓練データには 7 月 6 日の画像を、テストデータには 7 月 6 日と 7 月 7 日の画像を使用した。

SS 手法と先行研究の手法に対し、水位を推定したとき、測定値との RMSE は表 2 の通りとなった。また、測定値と比較した結果を図 9 から図 12 に示す。

表 2: 推定値と測定値の RMSE

| | 7月6日 | 7月7日 |
|-------|-------|--------|
| SS 手法 | 8.05 | 15.02 |
| 先行研究 | 26.00 | 163.44 |

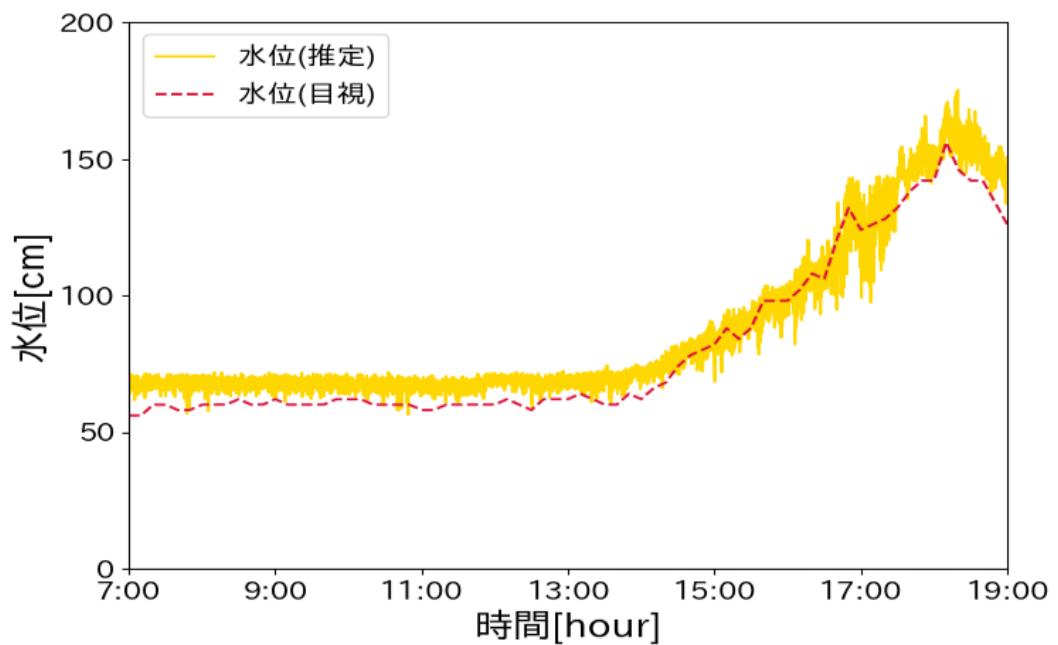


図 9: SS 手法の推定結果 (テスト : 7 月 6 日)

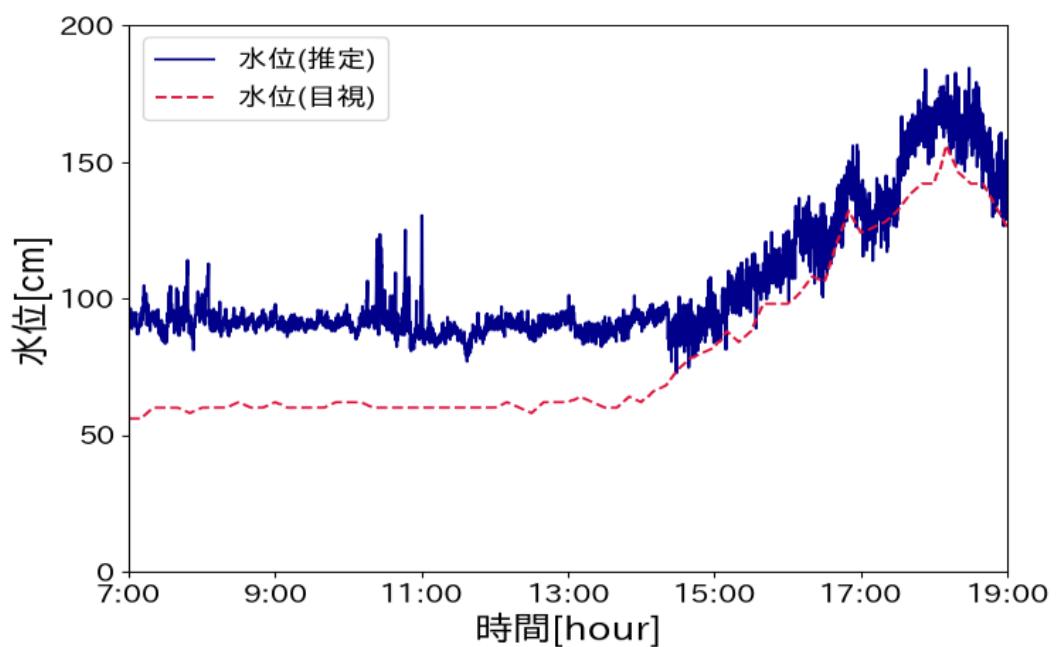


図 10: 先行研究の手法の推定結果 (テスト : 7 月 6 日)

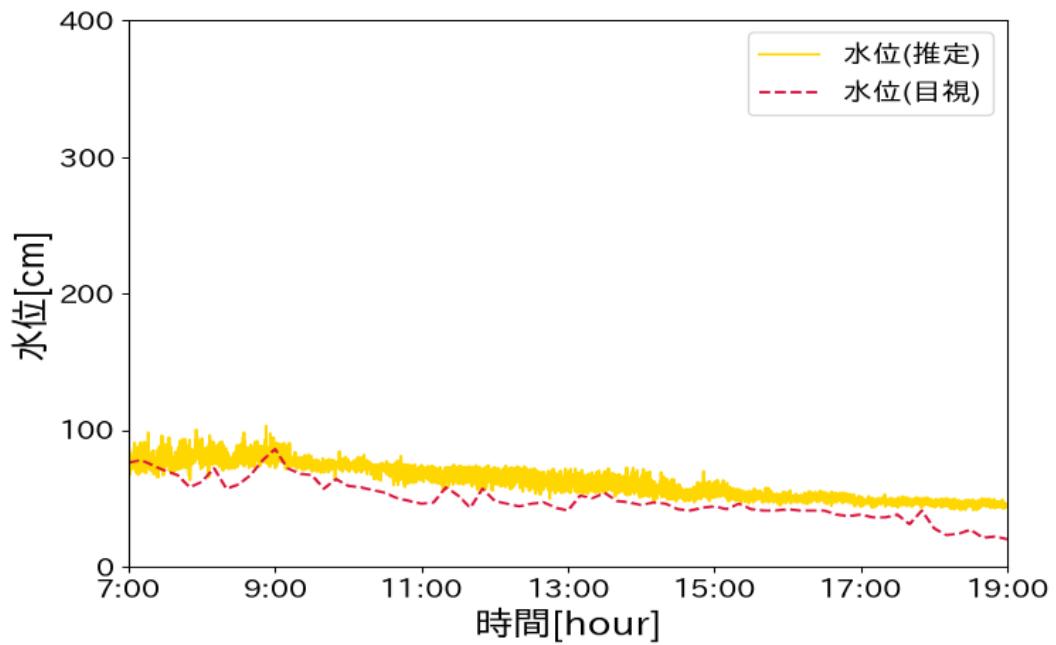


図 11: SS 手法の推定結果（テスト：7月 7日）

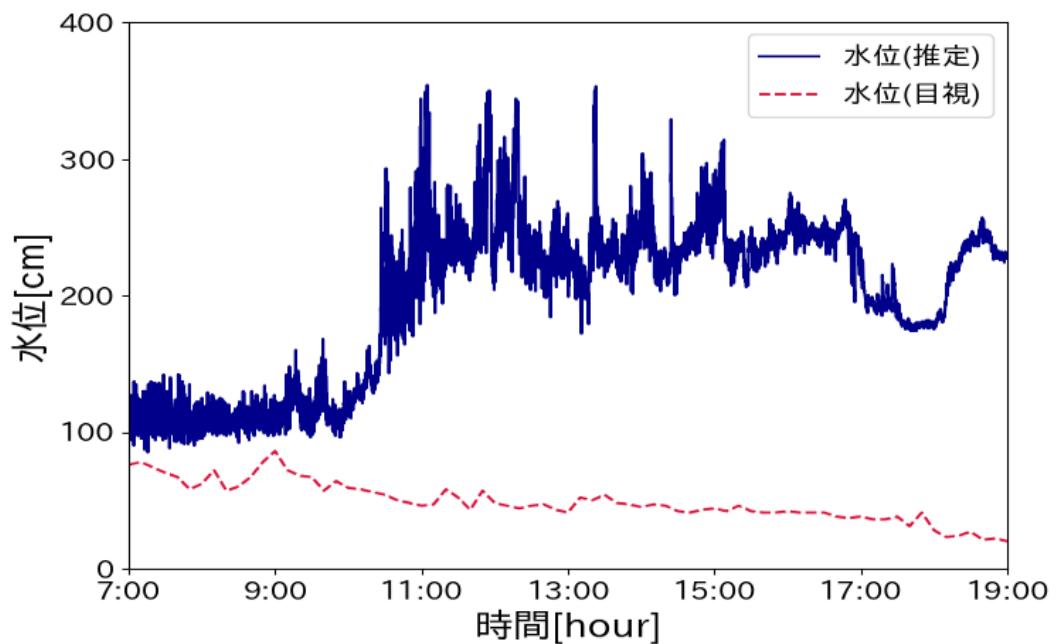


図 12: 先行研究の手法の推定結果（テスト：7月 7日）

5.3 ラベル設定と評価

アノテーション画像の水位計測ポールに対するラベル設定を変化させ，モデルを作成し，推定した水位と目視で計測した水位との RMSE と IoU を評価することで，適切な水位計測ポールに対するラベル設定の調査を行った．

設定 A と設定 B で作成したアノテーション画像を使用して作成したモデルで，水位の推定を行った．訓練データとテストデータには 5.1 節と同じデータを使用した．以下に設定 B の概要を示す．

- ・水位計測ポールに水位計測ポールラベルを与える．
- ・水位計測ポール以外のラベル設定は設定 A と統一する．

設定 A と設定 B に対し，水位を推定したとき，測定値との RMSE と IoU は表 3 の通りとなった．また，それぞれの水位推定結果を図 11 と図 13 に，時間別の IoU を図 14 に，領域抽出例を 15 に示す．

表 3: RMSE と IoU

| | RMSE | IoU |
|------|-------|-------|
| 設定 A | 15.02 | 92.30 |
| 設定 B | 15.66 | 88.57 |

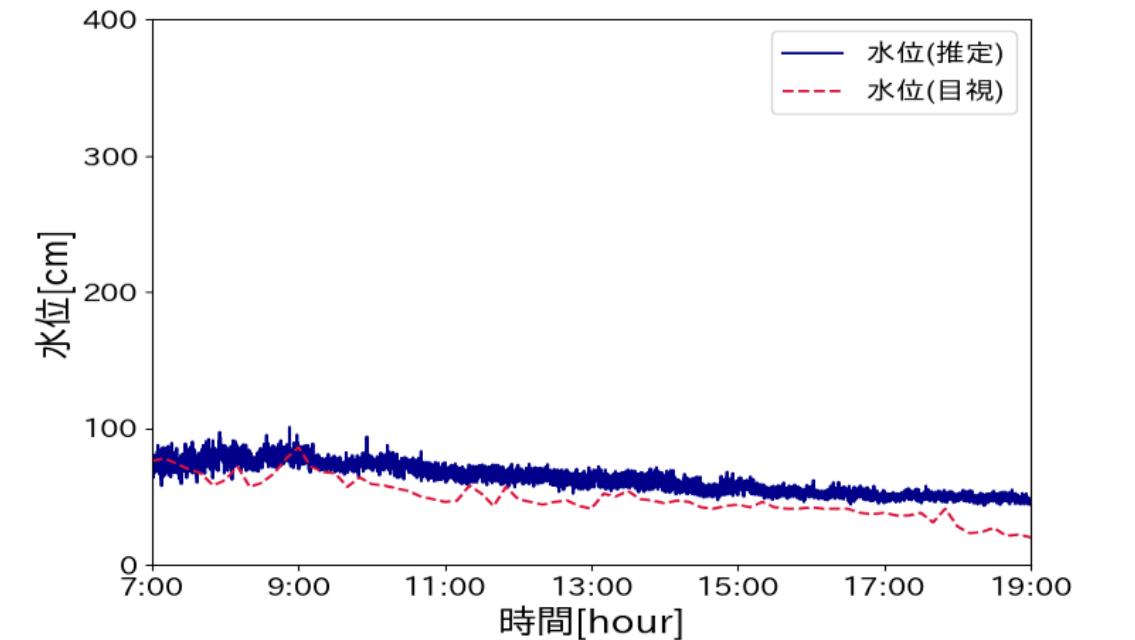


図 13: 設定 B の推定結果

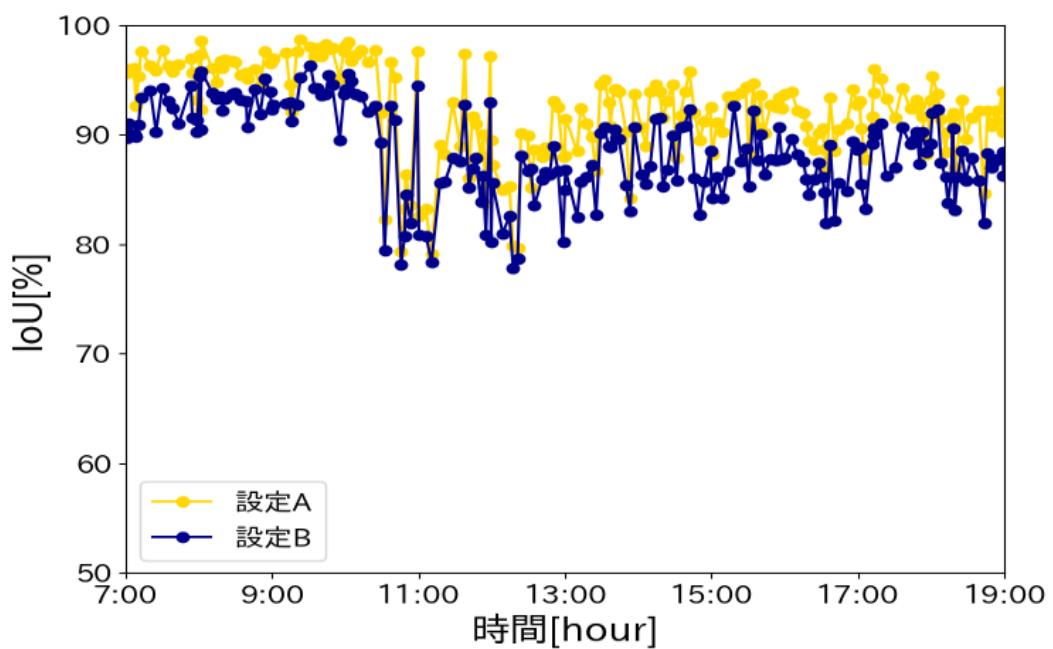
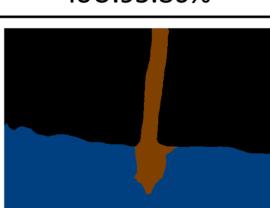
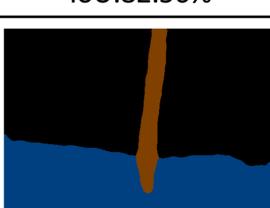
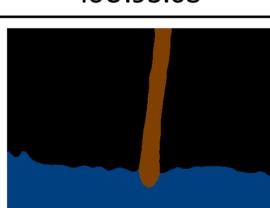


図 14: 時間別 IoU

| | | 画像1 | 画像2 | 画像3 |
|----------|-----|---|--|---|
| 入力画像 | |  |  |  |
| 正解画像 | |  |  |  |
| 領域抽出結果画像 | 設定A |  |  |  |
| | 設定B |  |  |  |

■ : 背景
 ■ ■ : 水面
 ■ ■ ■ : 水位計測ポール

図 15: 領域抽出結果例

5.4 交差検証

2018年,7月5,6,7日の3日分の画像データを使用して交差検証（例：7月5日と6日を訓練データ，7月7日をテストデータとして使用）を行い，提案手法の評価を行う。

推定した水位と測定値とのRMSEとIoUは表4の通りとなった。また，それぞれの水位推定結果を図16から図18に，時間別のIoUを図19から図21に，領域抽出結果例を22に示す。

表4: RMSEとIoU

| | 7月5日 | 7月6日 | 7月8日 |
|----------|-------|-------|-------|
| IoU(%) | 88.33 | 91.05 | 90.36 |
| RMSE(cm) | 17.32 | 7.73 | 8.32 |

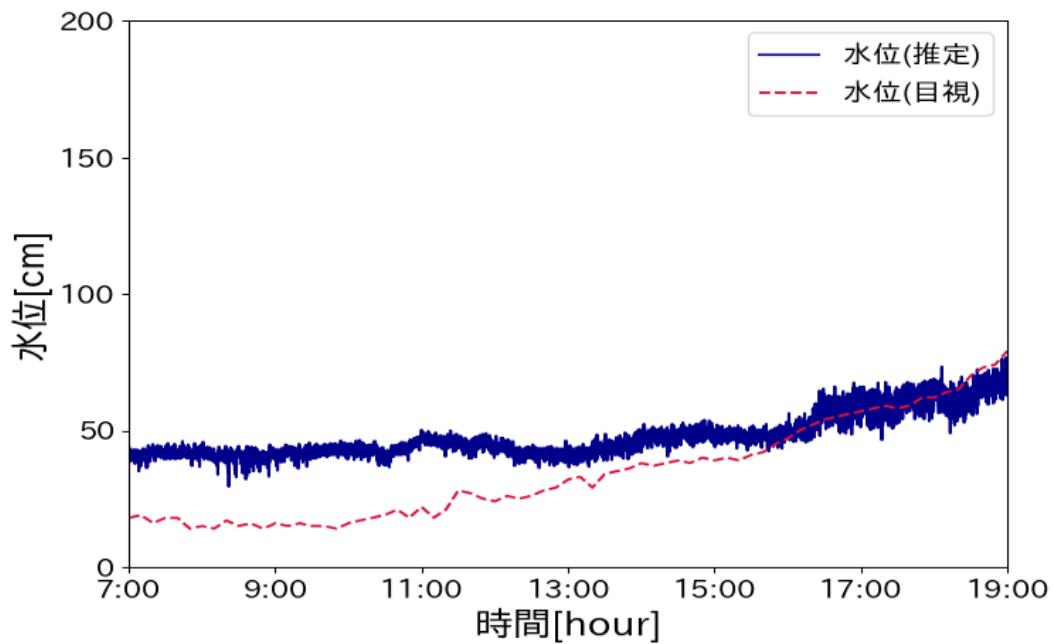


図16: 水位の推定結果（テスト：7月5日）

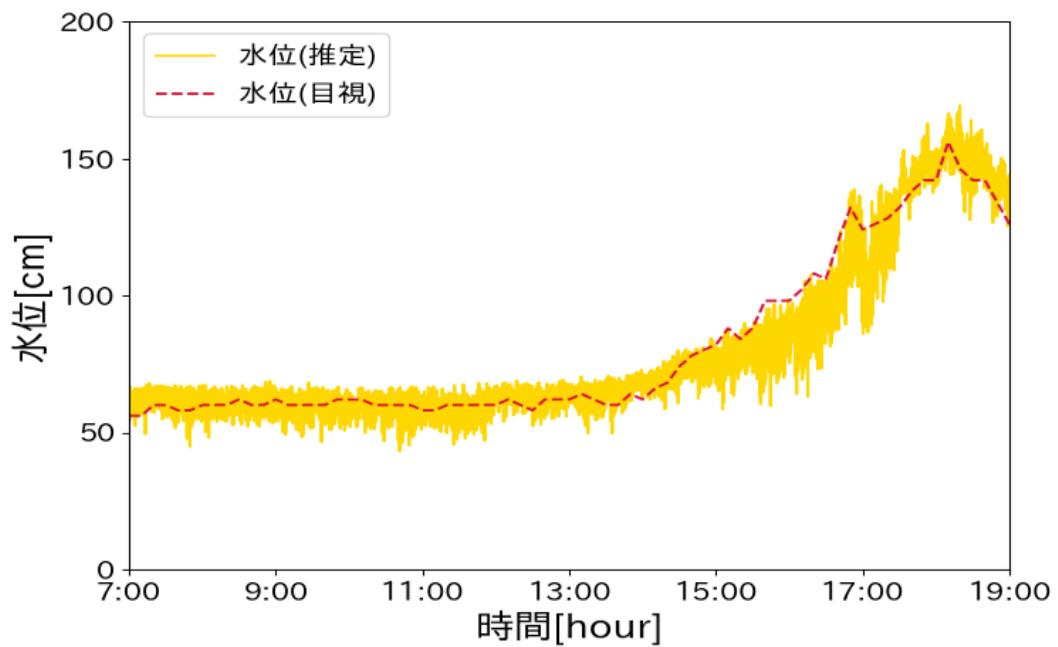


図 17: 水位の推定結果（テスト：7月6日）

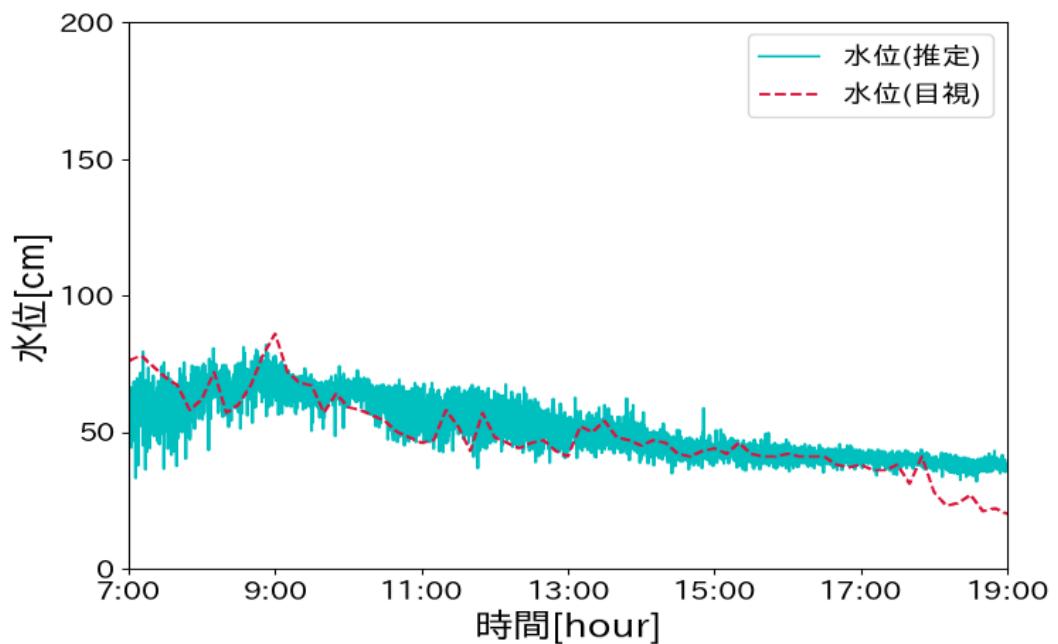


図 18: 水位の推定結果（テスト：7月7日）

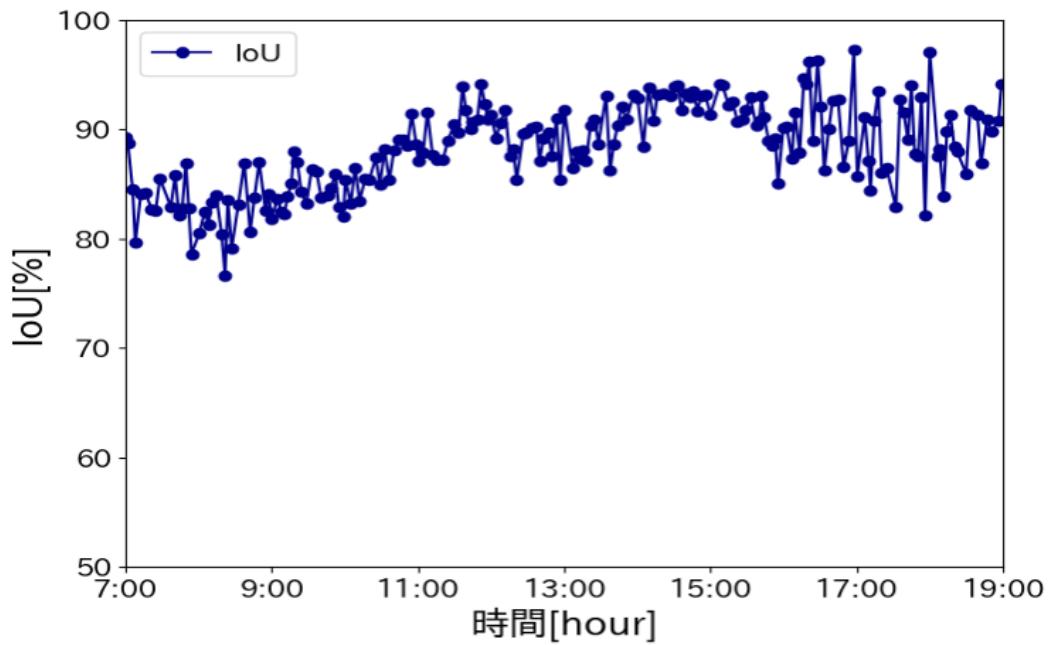


図 19: 時間別 IoU (テスト : 7月 5 日)

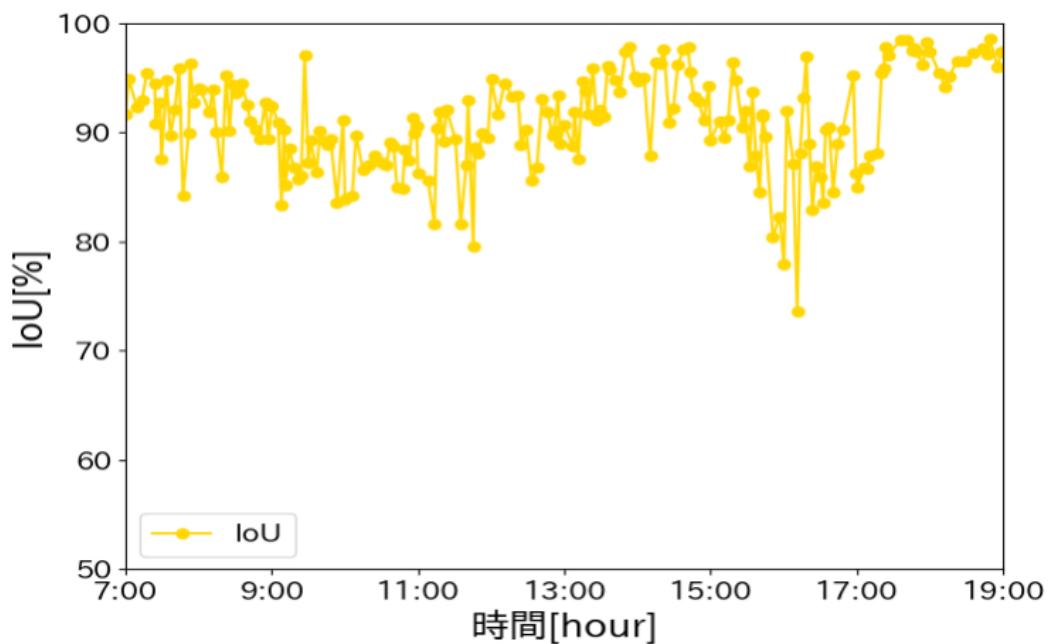


図 20: 時間別 IoU (テスト : 7月 6 日)

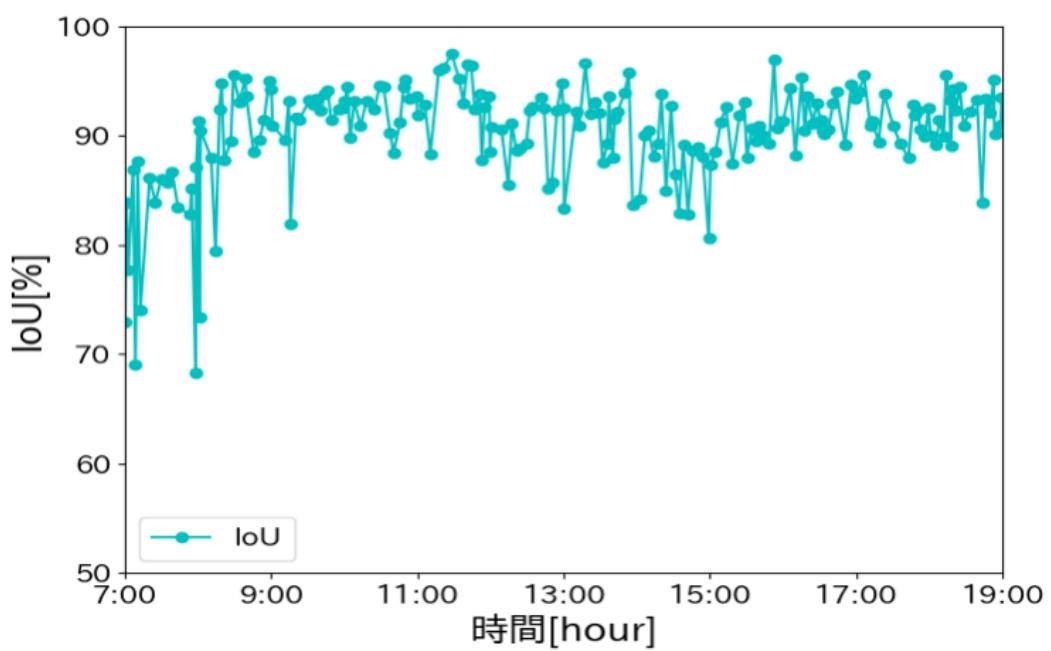


図 21: 時間別 IoU (テスト : 7月 7日)

| 入力画像 | 正解画像 | 領域抽出結果 画像 |
|---|------|--------------|
| 7月5日(上：7:00,下：12:00) | | |
| | | |
| | | |
| 7月6日(上：8:00,下：17:00) | | |
| | | |
| | | |
| 7月7日(上：7:00,下：18:00) | | |
| | | |
| | | |
| ■ : 背景 ■ : 水面 | | |

図 22: 領域抽出結果例

6 考察

IoU 以外にも F 値 [8] という領域抽出結果の評価指標が存在する。また、それは式 (5), (6) と図 7 で表される。F 値は、トレードオフの関係である Precision と Recall の両方の評価を行う。

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \quad (5)$$

$$BF = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

しかし、F 値と IoU は F 値 = IoU という関係が常に成り立つため、F 値を評価に使用しなかった。

以下に、 $Fscore = IoU$ が常に成り立つこと証明する。

$$\begin{aligned} Fscore - IoU &= Fscore - IoU - 0 \\ &= 2 \times \frac{Precision \times Recall}{Precision + Recall} - \frac{TP}{(TP + FP + FN)} \\ &= \frac{2 \times TP}{(2TP + FP + FN)} - \frac{TP}{(TP + FP + FN)} \\ &= \frac{TP(FN + FP)}{(2TP + FN + FP)(TP + FP + FN)} \\ &\quad TP, FP, FN \text{ は正の整数なので,} \\ &= TP(FN + FP) - 0 \\ &\quad \text{同様に} \\ &= TP - 0 \end{aligned}$$

TP は正の整数なので、常に $Fscore = IoU$ となる。また、IoU が 0 % もしくは 100 % の時 $Fscore = IoU$ となる。

5.1 節では、セマンティックセグメンテーションモデルを構築する最適な学習回数をモデルの領域抽出精度から調査した。図4より、200回からIoUは収束していることが確認できた。したがって、モデルの学習回数は最低でも200回以上必要であるといえる。

5.2 節では、SS手法を用いて推定した水位と目視で計測した水位とのRMSEを先行研究の手法の結果と比較した。表4より、SS手法を用いた場合の方がRMSEは小さくなつた。先行研究の手法には、訓練に使用したデータと別のデータに対してテストを行うと精度が低くなるという問題があった。しかし、SS手法のRMSEからSS手法は新規のデータに対しても高精度で領域抽出ができていることが分かる。

5.3 節では、水位計測ポールに対する適切なラベル設定の調査を行つた。水位計測ポールの範囲に学習除外ラベルと水位計測ポールラベルを与え、作成したモデルのIoUとRMSEを比較した。表4、図14より、学習除外ラベルを与えた場合の方がIoUもRMSEも良い結果になることが確認できた。図4の領域抽出結果例から、水位計測ポールに水位計測ポールラベルを与えた場合、水面領域に覆いかぶさる形で水位計測ポールの推定を行つてゐることわかる。このような推定により、IoUが低くなつたと考えられる。また、今回、水面領域面積の割合から水位の推定を行つたので、水位計測ポール分水面領域面積が減少することにより、水位が少し低く推定されたと考えられる。

5.4 節では交差検証を行い、提案したアノテーション方法の評価を行つた。提案手法を用いて作成したアノテーション画像を使用したモデルのIoUはすべての日で、85%以上という結果になり、高精度の領域抽出が行えていることがわかる。よつて、提案したアノテーション方法の有効性を示すことができた。しかし、他の日に比べて7月5日はIoUもRMSEも悪くなつた。特に、7月5日の7:00から12:00の範囲でIoUが低いことが、図19より確認できる。図22より、7月5日の午前中は水位が低く、水面が澄んでいて、肉眼でも水面と壁面の境界面を発見しにくくなつてゐることが影響したと考えられる。また、7月6日は図20より、16:00から17:00の間で、7月7日は図21より7:00から9:00の間で領域抽出精度が低い瞬間がある。原因として図22のように、河川の濁りが濃くなつてゐることが抽出に影響したと考えられる。

7 まとめ

本研究では、深層学習を用いたセマンティックセグメンテーションによって河川の監視カメラ画像から水位を推定することを目的とし、訓練データとして必要なアノテーション画像を半自動的に作成する方法を提案した。交差検証を行った結果、提案手法を用いて作成したモデルは全ての日で IoU が 85 %以上となり、本手法の有効性が確認できた。しかし、水面の状態によって精度が低くなるといった問題が見受けられた。また、水位の推定方法として SS 手法を用いた手法を検討した。SS 手法と先行研究の手法を適用して推定した水位の誤差を比較した。その結果、SS 手法の RMSE が 15.02cm、先行手法が 163.44cm となり、SS 手法の精度を確認したこと、SS 手法の有効性を示せた。今後の課題としては、他の河川に対する適用実験やアノテーション画像の自動生成などが挙げられる。

参考文献

- [1] 国土交通省：令和4年の土砂災害発生件数は795件，砂防NEWS Press Release , 2023年3月
- [2] 国土交通省水管理・国土保全局砂防部：「土砂災害警戒避難に関わる前兆現象情報の活用のあり方について」, <http://www.mlit.go.jp/common/001021004.pdf>, 2006年3月.
- [3] 渡邊 康平：土砂災害の前兆現象検知を目的とした画面分割と深層学習を用いた水位変動の推定，広島市立大学大学院情報科学研究科システム工学専攻修士論文，2023年1月。
- [4] Nur Atirah Muhadi ,Ana Mijic , Ahmad Fikri Abdullah , Siti Khairunniza Bejo , Muhammad Razif Mahadi: Deep Learning Semantic Segmentation for Water Level Estimation Using Surveillance Camera, , 2021年10月.
- [5] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, Hartwig Adam: Encoder-Decoder with Atrous SeparableConvolution for Semantic Image Segmentation, , 2018年9月.
- [6] Vijay Badrinarayanan, Alex Kendall , and Roberto Cipolla, Senior Member: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017年12月.
- [7] 青島 亘佐, 山本 拓海, 中野 聰, 中村 秀明: 深層学習によるセグメンテーション手法を用いたコンクリート表面の変状領域の検出, , 2020年7月.
- [8] 山根達郎, 全邦釤,Tatsuro Yamane, Pang-jo Chun : Deep learning による Semantic Segmentation を用いたコンクリート表面ひび割れの検出，構造工学論文集 Vol.65A, 2019年3月.

謝辞

本研究に関して多大なるご指導を頂きました広島市立大学大学院情報科学研究科システム工学専攻の島和之准教授に心より御礼申し上げます。平素より島准教授には、研究の進め方について丁寧にご指導とご鞭撻を賜りました。重ねて感謝いたします。また、本実験で使用する河川の画像をご提供くださいました同学情報工学専攻モニタリングネットワーク研究室西正博教授に厚く感謝申し上げます。