

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	12
5.2 ラベル設定と評価	13
5.3 先行研究との比較	16
6 考察	19
6.1 交差検証	19
6.2 評価実験について	24
7 まとめ	25
参考文献	27
謝辞	27

1 序章

1.1 研究背景

平成 30 年 7 月豪雨によって広島市で起こった土砂崩れや浸水による被害をはじめ、土砂災害は日本全国各地で年間平均 1099 件（昭和 57 年から令和 4 年まで）が発生している [1]。土砂災害の発生前には「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]。土砂災害の前兆現象を早期に検知し、住民に避難を促すことで人的被害の削減につながると期待されている。

そこで、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」に着目し、河川の監視カメラ画像から水位を推定する方法が提案されている [3]。河川の水位は水位計を用いて計測する方法が一般的ではあるが、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る。このため、安価で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 [4]。

先行研究 [3] では、河川の監視カメラ画像から多数の切抜画像を抽出し、畳込みニューラルネットワークで水面を判別し、水位を推定する方法を提案した。

1.2 研究目的

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面領域を抽出する手法（以下、SS 手法）を検討するため、その精度を評価した。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、半自動的な作成方法を検討する。

2 関連研究

河川の監視カメラ画像から水位変化の推定や検出を目的とする研究は数多くなされている。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な層である。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。

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

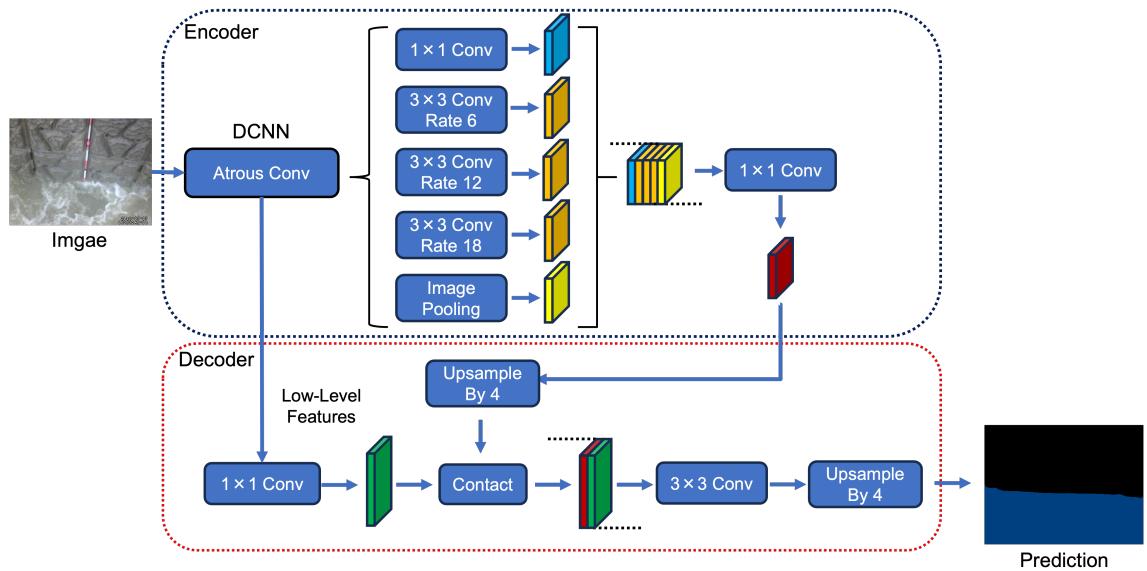


図 1: DeepLabv3+モデル

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水面クラスのソフトマックス値と元画像の縦座標値を用いて水位の変動を推定した。深層学習を用いて推定した水位変動と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

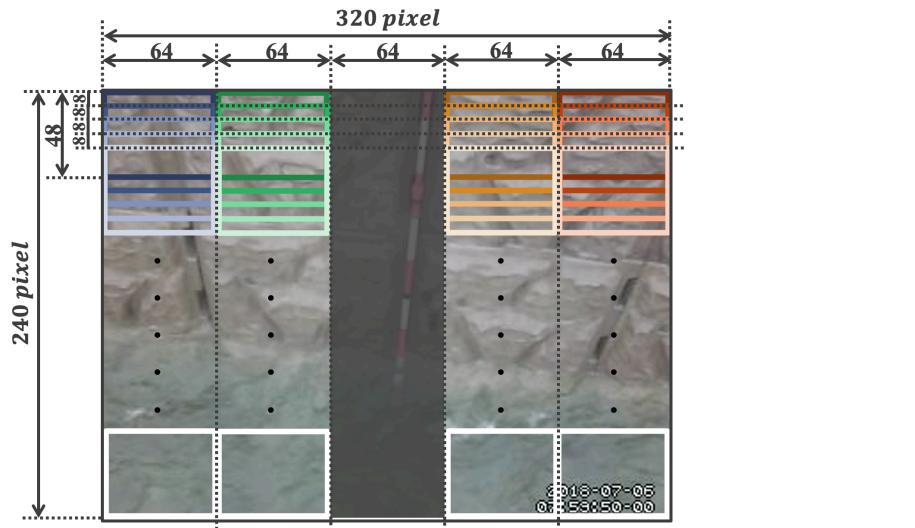


図2: 切抜画像の生成

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。

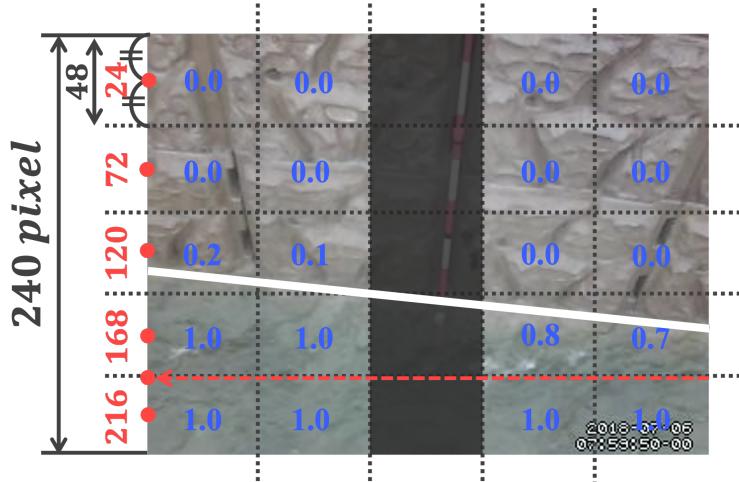


図3: 推定値算出の流れ

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、アノテーション方法が精度に影響する。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。

また、抽出結果画像から変換式を用いて水位を推定する方法を提案する。

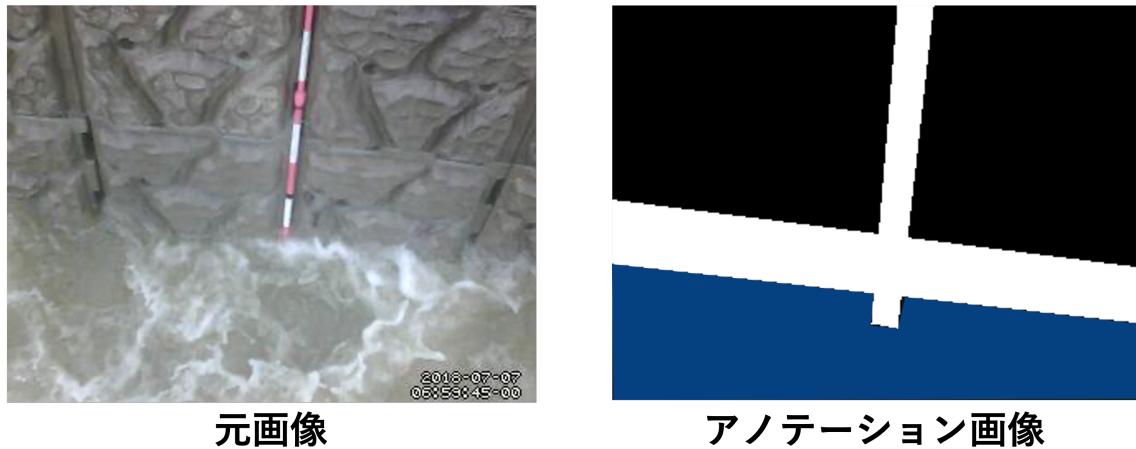


図4: 元画像とアノテーション画像の例

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。従って、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ため、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ため、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。

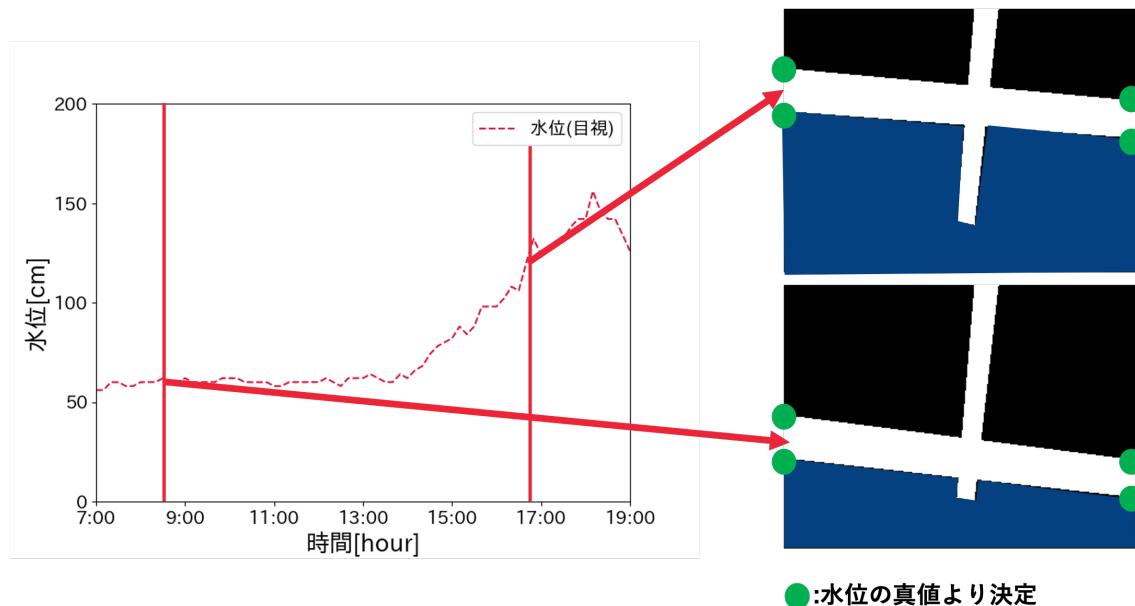


図5: 範囲決定の流れ

4.3 水位の推定方法

画像内の水面領域の面積割合と水位の散布図を図 6 に示す。また、水面領域の面積割合と水位の相関係数は 0.99 となった。従って、水面領域の面積割合から水位へ値を変換する式を作成した。この変換式を使用して、領域抽出結果画像から水位の推定を行った。

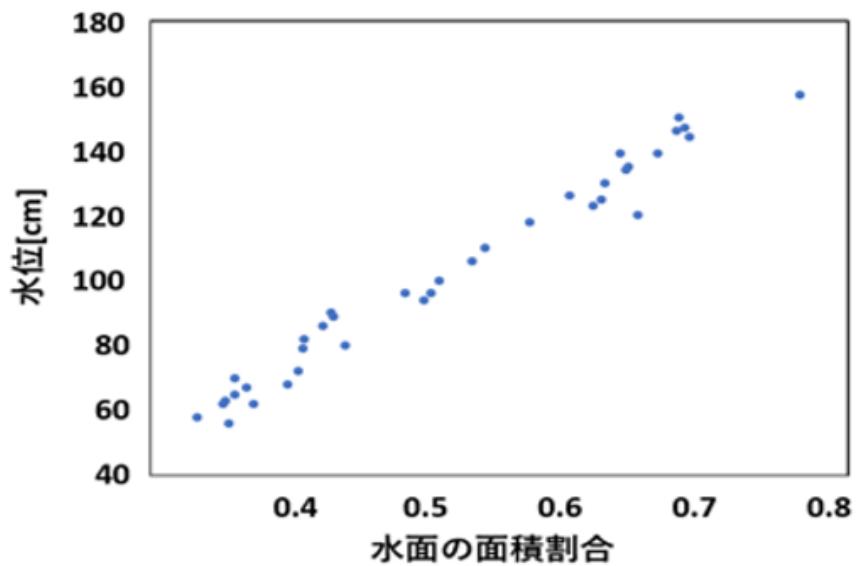


図 6: 水面領域の面積割合と水位の散布図

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 7 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 7: 桐原川の監視カメラ画像

領域抽出結果についての評価は，IoU(Interaction over Union)[7] と F 値 [8] を用いて行う。IoU は図 8 と式 (3) で定義さ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。

F 値は図 8 と式 (4),(5),(6), で表される値である。感度は正解画像の中で水面領域である画素数に対して，予測結果においても水面領域と判定した割合である。また，適合率は予測結果の中で水面領域と判定した画素数について，正解画像においても水面領域である割合である。そして，F 値は一般にトレードオフの関係になる感度と適合率の両方を評価する。今回はすべての画像に対する水面クラスの平均 IoU と平均 F 値に着目して評価を行う。

		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 8: 領域抽出における分類問題の混合行列の定義

$$\text{IoU} = \frac{\text{TP}}{(\text{TP} + \text{FP} + \text{FN})} \quad (3)$$

$$\text{感度} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (4)$$

$$\text{適合率} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (5)$$

$$\text{F 値} = 2 \times \frac{\text{適合率} \times \text{感度}}{(\text{適合率} + \text{感度})} \quad (6)$$

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，その結果を IoU と F 値を利用して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．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を作成した．また，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．

作成したモデルを用いて領域抽出を行い，学習回数毎に IoU と F 値を求めた．図 9 はモデルの学習回数と抽出精度の関係を表したグラフである．

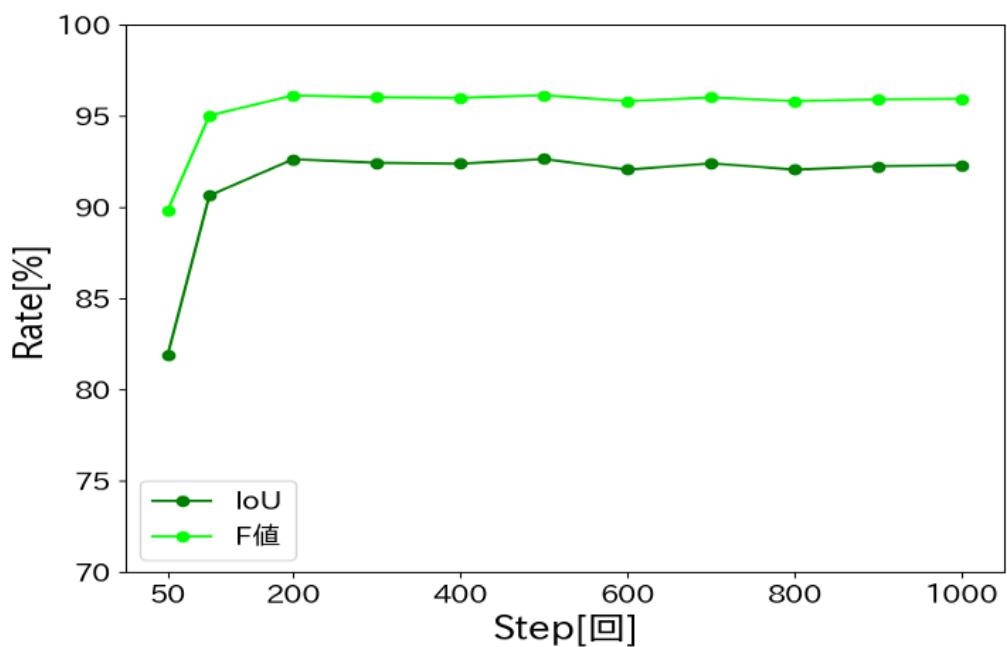


図 9: 学習回数と IoU,F 値の関係

5.2 ラベル設定と評価

2種類のアノテーション設定で構築したモデルの領域抽出精度をIoUとF値で比較することで、適切な水位計測ポールラベルの調査を行う。設定Aと設定Bの二つの設定でアノテーションを行った。図10で示すように、設定Aは水面と撮影日時に水面ラベル、壁面に背景ラベル、水位計測ポールを学習除外領域に設定した。また、設定Bは、水位計測ポール以外は設定Aと統一し、水位計測ポールに水位計測ポールラベルを与えた。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。

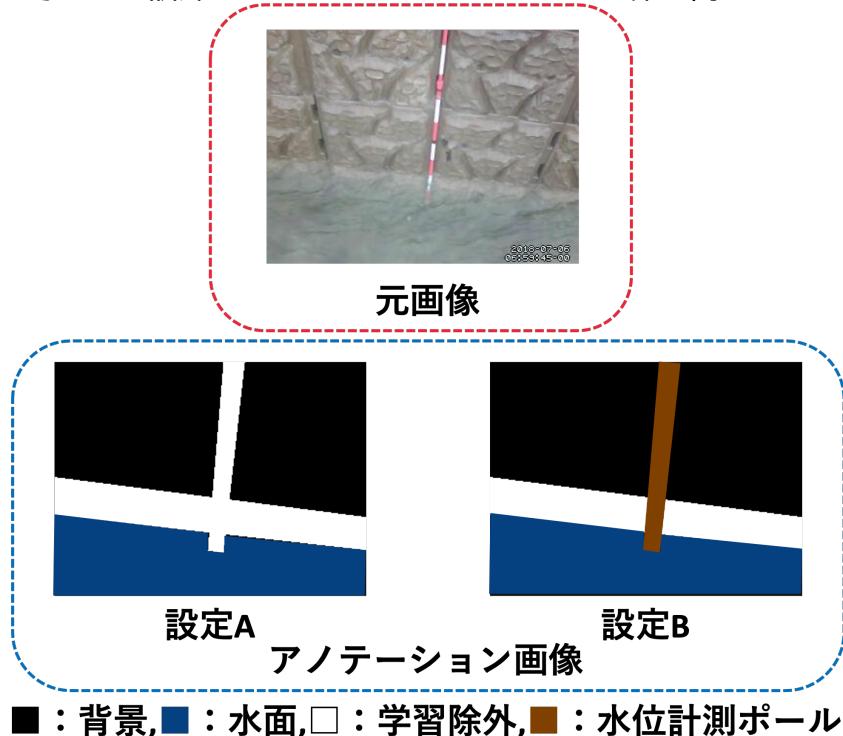


図10: アノテーション画像例

それぞれの設定での、IoUとF値は表2の通りとなった。また、時間別のIoUを図11に、時間別のF値の値を図12に示す。それぞれのモデルの抽出結果例を図13に示す。

表2: IoU(%)とF値(%)

	IoU	F値
設定 A	92.30	95.94
設定 B	88.57	93.89

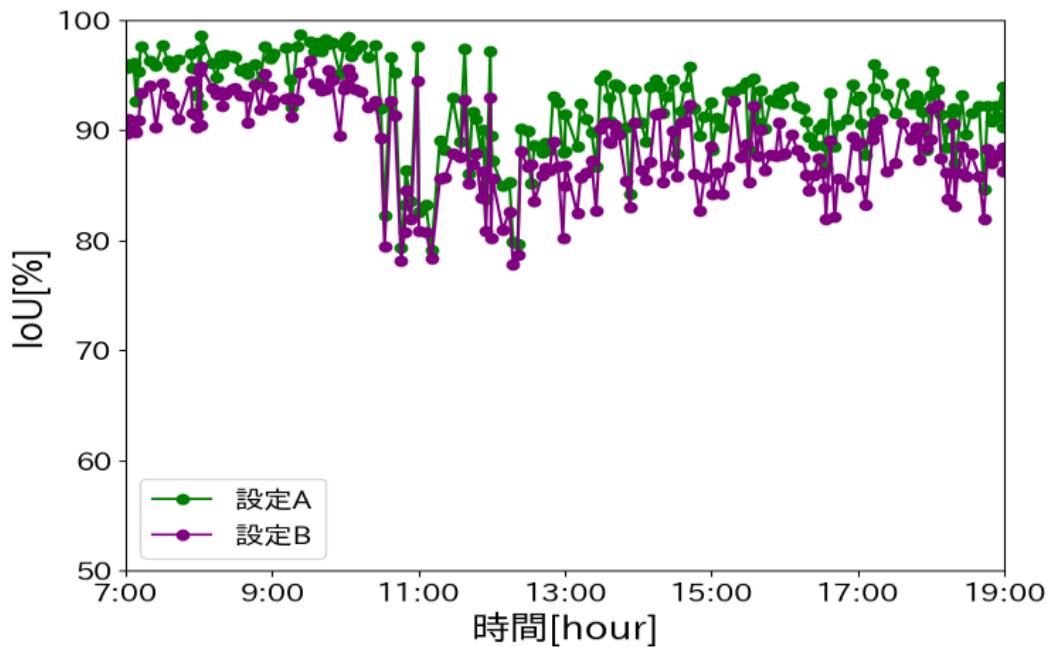


図 11: 時間別 IoU

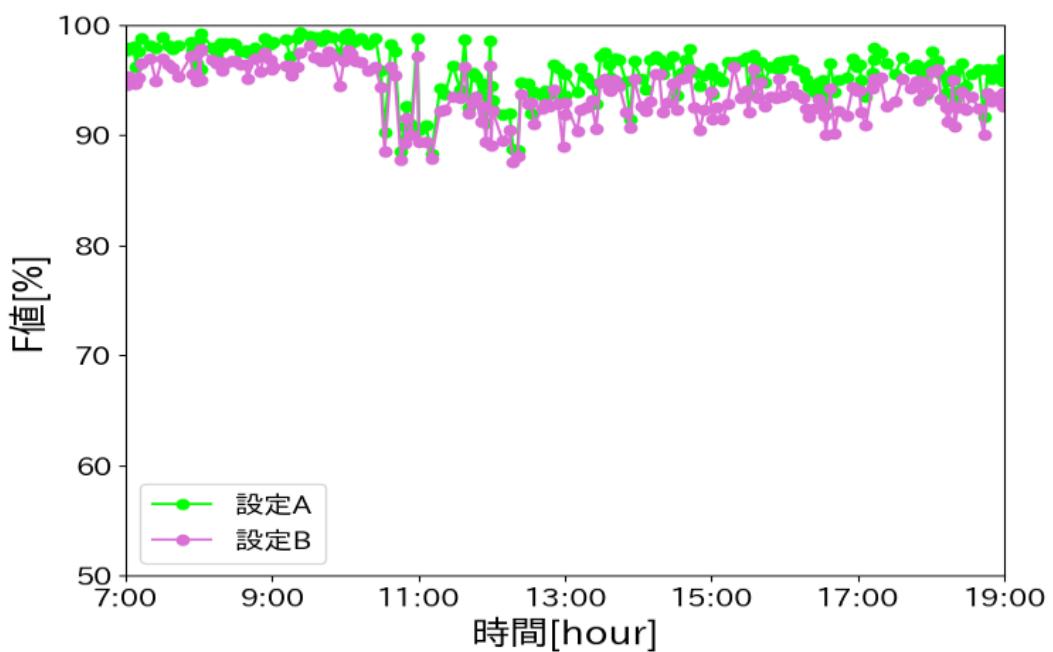
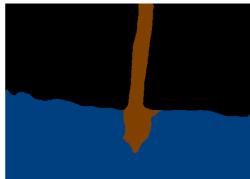
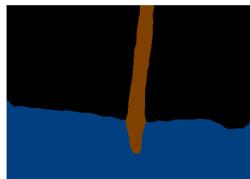
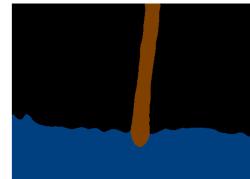


図 12: 時間別 F 値

		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			
		IoU:95.86% F値:97.89%	IoU:82.56% F値:90.45%	IoU:93.68% F値:96.74%
	設定B			
		IoU:90.28% F値:94.89%	IoU:80.88% F値:89.43%	IoU:87.90% F値:93.56%

■：背景, ■■：水面, ■■■：水位計測ポール

図 13: 領域抽出結果例

5.3 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，5.2 節より，領域抽出精度が高かった設定 A でアノテーションを行った．また，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 3 の通りとなった．推定した水位を図 14 から図 17 に示す．

表 3: 推定値と測定値の RMSE(cm)

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

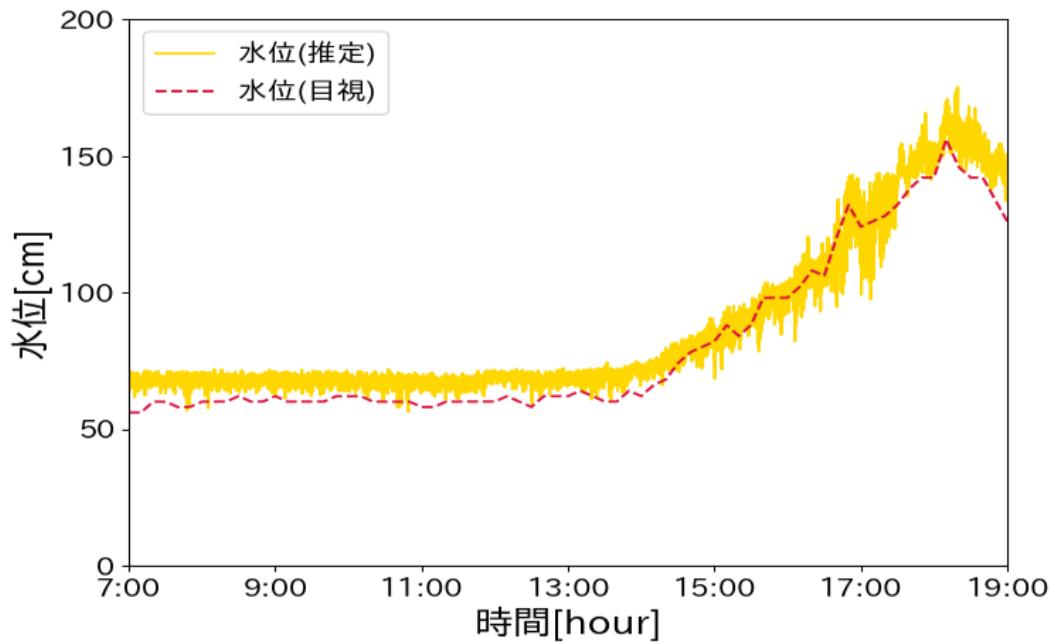


図 14: SS 手法の推定結果（テスト：7月 6 日）

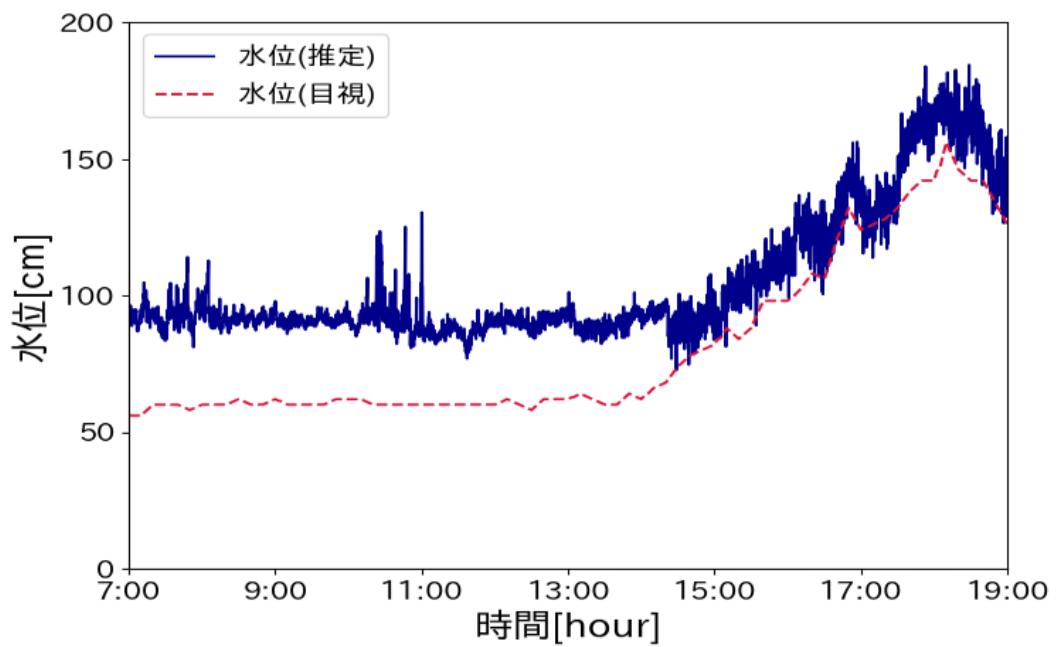


図 15: 先行研究の手法の推定結果（テスト：7月 6日）

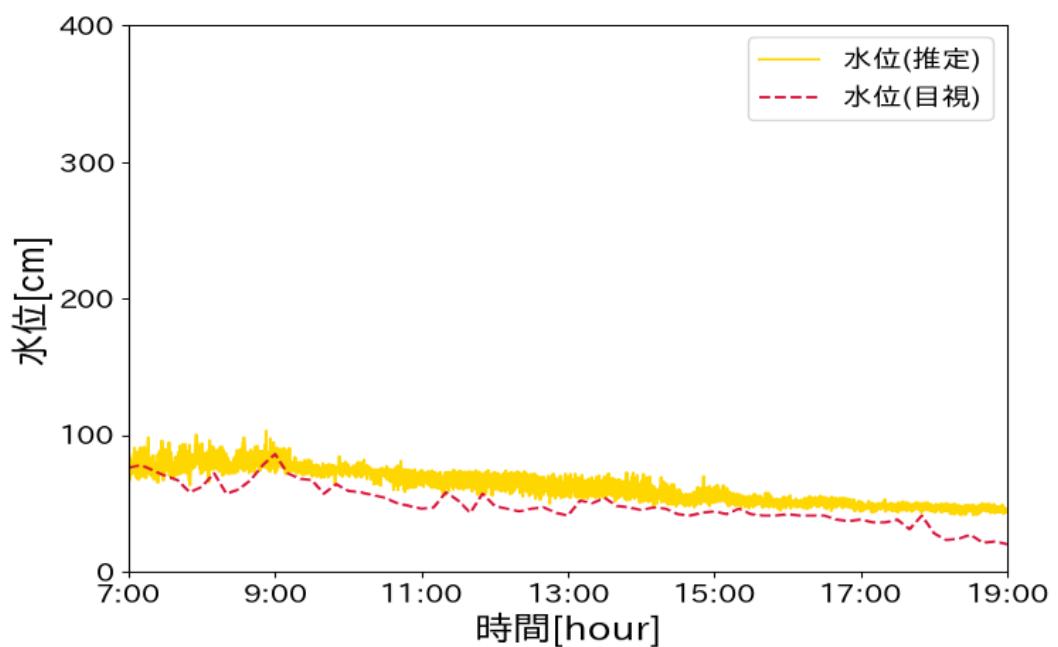


図 16: SS 手法の推定結果（テスト：7月 7日）

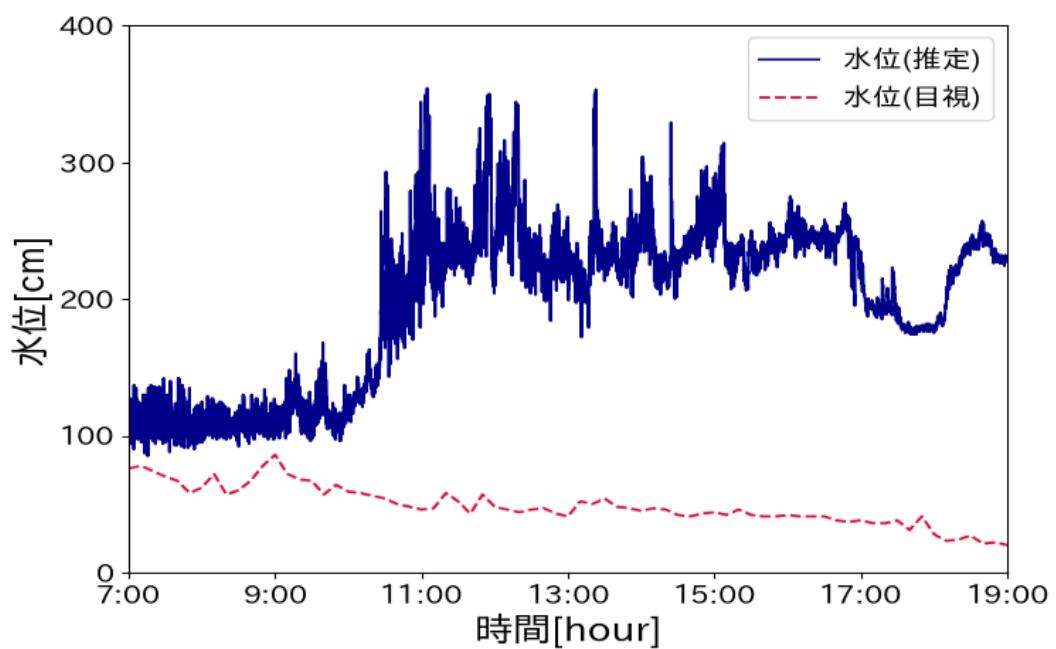


図 17: 先行研究の手法の推定結果 (テスト : 7 月 7 日)

6 考察

6.1 交差検証

2018年,7月5,6,7日の3日分の画像データからSS手法を用いて水位の推定を行った(3日分のデータの内,2日分のデータを訓練データ,残った1日分のデータをテストデータとして使用).また,アノテーション設定は設定Aを使用した.

推定した水位と測定値とのRMSEとIoU,F値は表4の通りとなった.また,それぞれのテストデータ日の抽出結果例を図24に示す.

表4から,提案手法を用いて作成したアノテーション画像を訓練に使用したモデルのIoUとF値はすべてのテスト日で,85%以上を達成したことが確認できる.

7月5日の水位推定結果を図18に,時間別のIoUとF値を図19に示す.表4より,7月5日は他の日と比較してRMSEが大きくなつた.また,7月5日の7:00から12:00の範囲で精度が低いことが図18と図19から確認できる.特に精度が低い抽出結果例を図24の(b)に示す.同じ,7月5日の画像である(a)の入力画像と(b)の入力画像から河川の状態を比べると,(b)の入力画像は水面が低く,澄んでいることが確認できる.この現象によって,水面と壁面の境界面が発見しにくくなり,領域抽出精度が下がつたと考えられる.

7月6日の水位推定結果を図20に,時間別のIoUとF値を図21に示す.図21より,15:00から17:00の間で領域抽出精度が低い瞬間があることが確認できる.図24の(d)にその瞬間の抽出結果を示す.同日かつ,高精度で抽出が行えている(c)の入力画像と(d)の入力画像から河川の状態を比べると,(d)の河川は水面が濁っていることが確認できる.この現象によって,水面の輝度値が壁面と近くなり,領域抽出精度が下がつたと考えられる.

7月7日の水位推定結果を図22に,時間別のIoUとF値を図23に示す.図23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ことが確認できる.図24の(f)にその瞬間の抽出結果を示す.同日かつ,高精度で抽出が行えている(e)の入力画像と(f)の入力画像から河川の状態を比べると,(f)の河川は水面が波立つており,水面と壁面の境界が複雑になっていることが確認できる.よつて,水面の状態が大きく変化したことが,領域抽出精度が下がつた要因と考えられる.

表4: RMSEと領域抽出精度

	7月5日	7月6日	7月8日
IoU(%)	88.33	91.05	90.36
F値(%)	93.76	95.26	94.87
RMSE(cm)	17.32	7.73	8.32

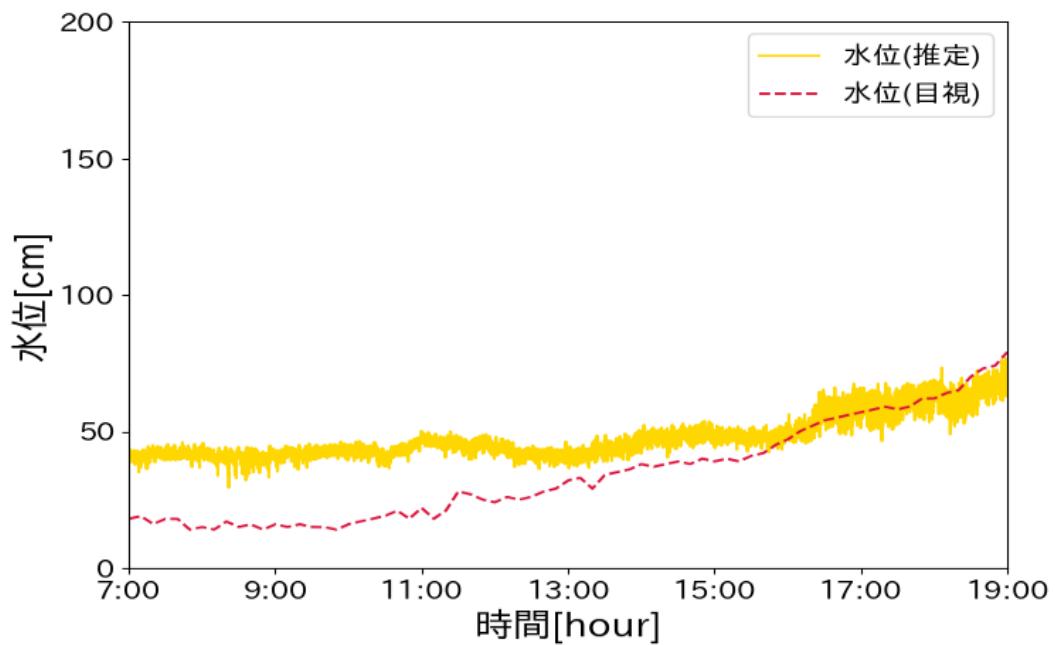


図 18: 水位の推定結果（テスト：7月5日）

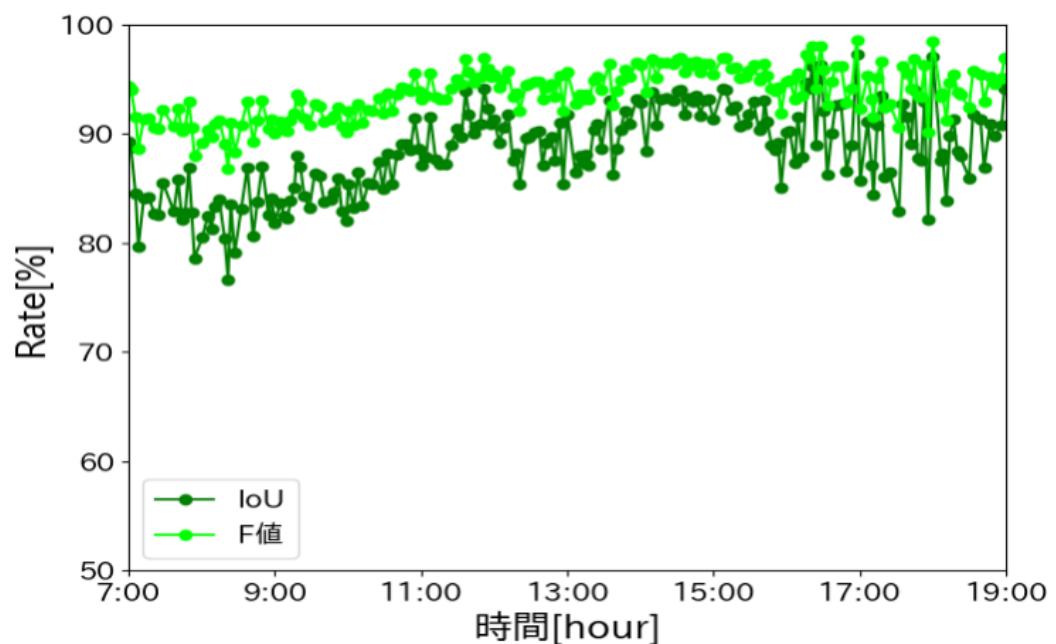


図 19: 時間別 IoU , F 値 (テスト : 7 月 5 日)

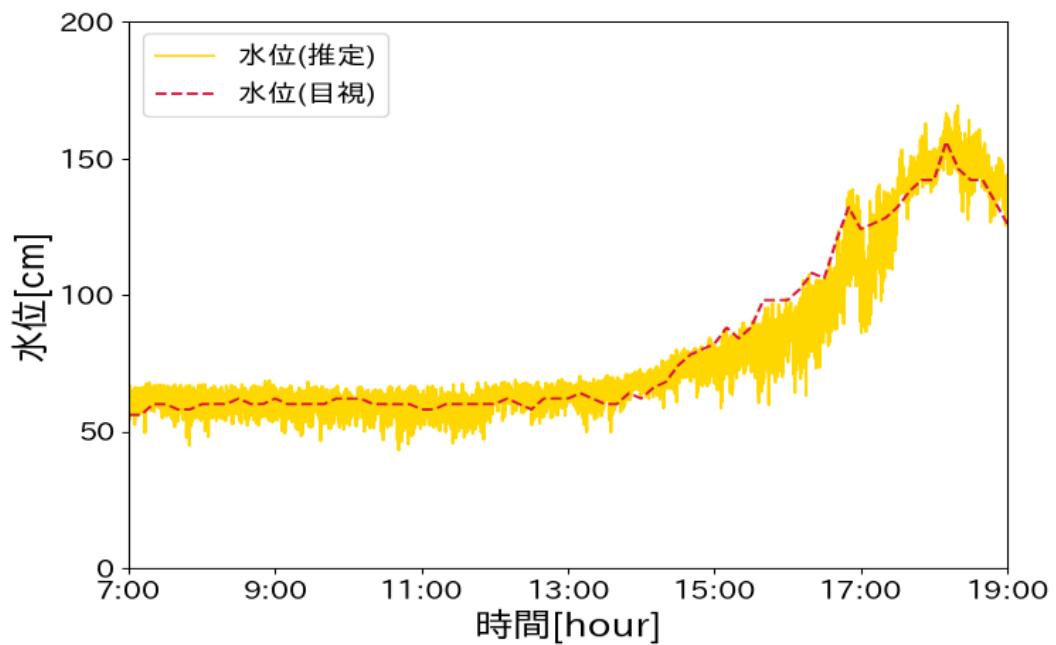


図 20: 水位の推定結果（テスト：7月6日）

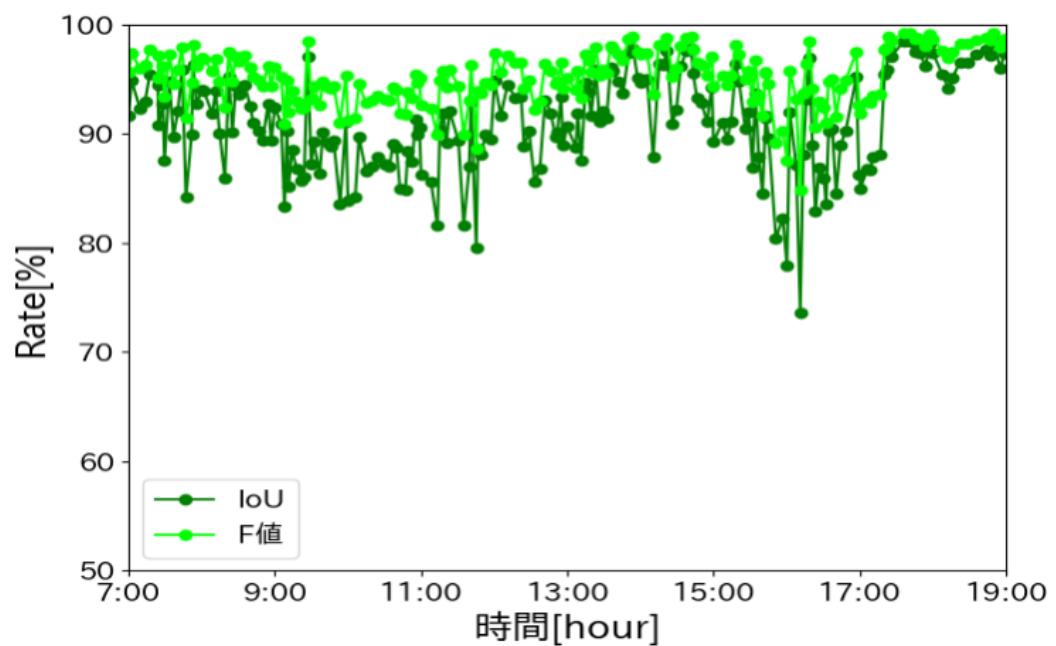


図 21: 時間別 IoU , F 値 (テスト : 7 月 6 日)

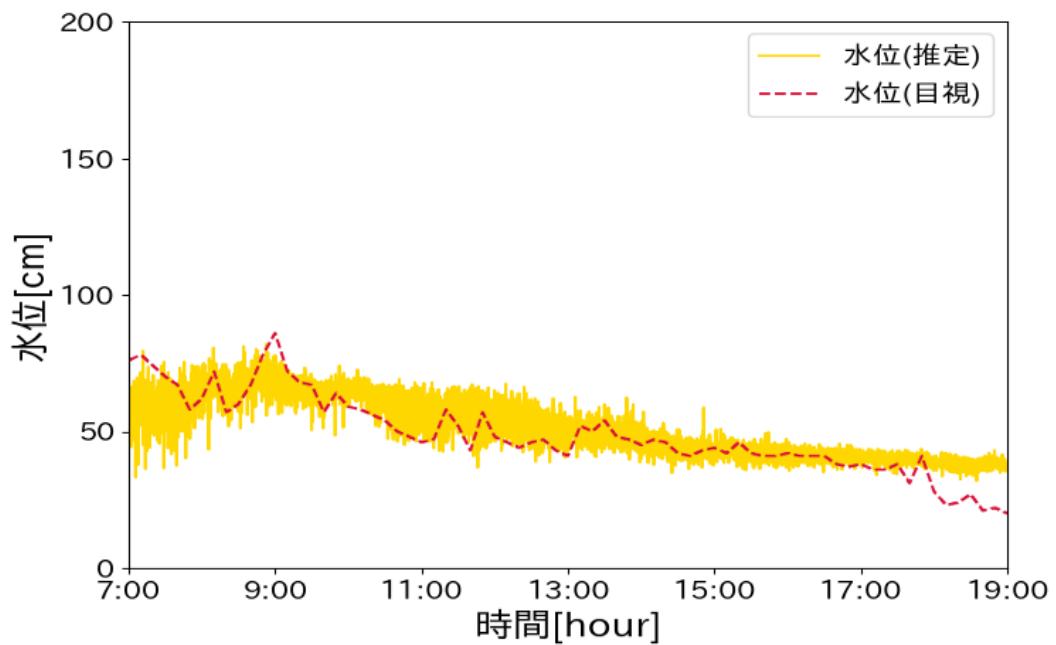


図 22: 水位の推定結果（テスト：7月7日）

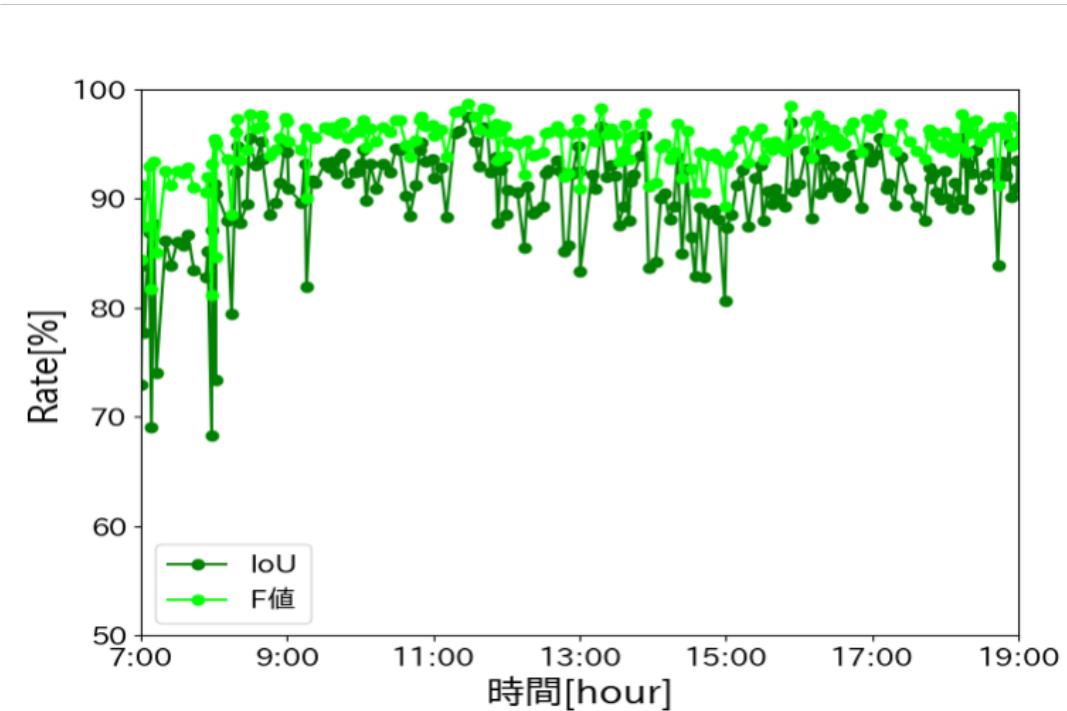


図 23: 時間別 IoU , F 値 (テスト : 7 月 7 日)

	入力画像	正解画像	領域抽出結果 画像	
7月5日				
(a)				IoU : 92.58% F値 : 96.15%
(b)				IoU : 78.61% F値 : 88.02%
7月6日				
(c)				IoU : 94.02% F値 : 96.92%
(d)				IoU : 73.63% F値 : 84.82%
7月7日				
(e)				IoU : 90.09% F値 : 94.79%
(f)				IoU : 69.09% F値 : 81.71%
				

図 24: 領域抽出結果例

6.2 評価実験について

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とF値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、2種類のアノテーション設定で構築したモデルの領域抽出精度をIoUとF値で比較した。表2、図11、図12より、設定Aの方がIoUとF値が良い結果になることが確認できた。図13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っていることわかる。このような推定により、IoUとF値が低くなったと考えられる。

5.3 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3より、SS手法を用いた場合の方がRMSEは小さくなかった。

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テスト日で IoU と F 値が 85 %以上となつた。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水位を推定し、目視で測定した水位との誤差を利用して評価を行つた。その結果、SS 手法の RMSE が 15.02cm、先行手法が 163.44cm となり、SS 手法を用いた場合の誤差の方が小さくなつた。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, Applied Sciences,vol.11,issue.20,(2021).
- [5] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. : Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818).(2018).
- [6] Badrinarayanan, V., Kendall, A.,Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.(2017).
- [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. AI・データサイエンス論文集, 1(J1), 481-490.(2020年).
- [8] 山根達郎, 全邦釤: 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 A, 65, 130-138.(2019年).

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情報科学研究所システム工学科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科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。