

# 河川スカム発生 の 早期検知 に向けた 高精度スカム検出モデルの開発

藤井 純一郎<sup>1</sup>・岡野 将大<sup>2</sup>・後藤 早苗<sup>3</sup>

<sup>1</sup> 正会員 八千代エンジニアリング株式会社 技術開発研究所 (〒111-8648 東京都台東区浅草橋 5-20-8)  
E-mail: jn-fujii@yachiyo-eng.co.jp (Corresponding Author)

<sup>2</sup> 非会員 八千代エンジニアリング株式会社 技術開発研究所 (〒111-8648 東京都台東区浅草橋 5-20-8)  
E-mail: ms-okano@yachiyo-eng.co.jp

<sup>3</sup> 正会員 八千代エンジニアリング株式会社 環境計画部 (〒111-8648 東京都台東区浅草橋 5-20-8)  
E-mail: sn-goto@yachiyo-eng.co.jp

合流式下水道越流水が流入する都市河川では、降雨後にスカムが発生し悪臭や景観の悪化、生態系への悪影響などが問題になっている。その対策として、スカムのモニタリングによる早期検知が望まれる。本研究では、スカム発生 の 早期検知 を目的として、モニタリング要件および画像認識要件を検討した上で、大阪府平野川の河川監視カメラのデータを用いて、スカム画像の特性に合わせた画像認識モデルや損失関数を実験により明らかにした。実験結果に基づき、スカム検出の精度が低い画像を教師データに追加した上で最終的な学習を行い、Scum IoU95%を達成し、要件を満足する高精度なスカム検出モデルを開発した。

**Key Words :** River Monitoring, Scum, Semantic Segmentation, Attention U-net, Loss Function

## 1. はじめに

合流式下水道越流水(CSO)が流入する河川では、降雨後にスカム(浮遊汚泥)が発生し悪臭や景観の悪化、生態系への悪影響などを及ぼすことが問題になっている<sup>1,2)</sup>。スカムの発生を抑制する予防保全的対策や、根本的な対策手法は確立していないため、当面の対応としてスカムのモニタリングを行い、スカムが発生した場合に早期に検知し、回収等の対策を行うことが有効である。

スカムのモニタリングでは、従来は現地採取による調査<sup>3,4)</sup>が行われてきたが、多大な労力を要するため調査できる地点や頻度に限りがあり、スカム発生状況を連続的に把握することは難しい。これに対し、定点カメラを用いた画像認識技術によるスカム発生 の 早期検知 や挙動把握の研究やモニタリングシステムの開発(2章で後述)も行われてきた。画像認識によるモニタリングは、調査地点や頻度を増やすことが期待でき、また統一基準で定量評価できるため調査員による個人差がない点でメリットがある。しかし、スカム発生 の 早期検知 には、発生段階の微小なスカムを検出する必要があるため、既存研究のスカム検出モデルには一層の高精度化が求められる。

そこで、本研究では画像認識によるスカム発生 の 早期検知 を実現することを目的として、モニタリング要件および画像認識要件を検討した。その上で、スカム画像の特性に合わせた画像認識モデルや損失関数の設定を実験により明らかにし、要件を満足する高精度なスカム検出モデルを開発した。

以降、2章で先行研究を整理し、3章でスカム発生 の 早期検知 のためのモニタリング要件を示す。要件を満足するスカム検出モデルの開発に向けて、4章でモデル選択、5章で損失関数の変更を行い、6章で最終モデルの学習と評価を実施する。

## 2. 先行研究

画像認識技術を用いたスカムのモニタリングに関する研究としては、エッジ抽出などの画像処理技術を用いた事例<sup>5,6)</sup>がある。しかしこれらの画像処理技術は、天候や明るさの違いなどに対してパラメータで調整する必要がある。屋外の河川監視カメラに適用するには課題が残されている。水田ら<sup>7)</sup>は機械学習手法の一つであるニューラルネットワークを用いてスカムの定量判定を

試み、一定の精度での判定に成功している。しかし、この手法は地点ごとの画素値の特性を事前に分析し、特徴量抽出方法を手動で定義する必要があることが課題として挙げられる。

深層学習を用いた画像認識技術では、上記の特徴量抽出まで含めて学習することが可能となった。中谷ら<sup>8)</sup>は深層学習を用いた Semantic Segmentation により、ピクセルレベルでのスカム領域の検出を行うモデルを提案している。モデルには畳み込みニューラルネットワークを用いた Semantic Segmentation モデルである U-Net<sup>9)</sup>を採用し、画像特性の影響を抑えることに成功した。また後続研究<sup>10)</sup>として、教師画像の作成コスト削減と精度向上を実現するダミー画像を用いた学習方法を提案している。しかし、上記の研究ではモデルを U-Net に固定しており、河川スカムのような水面の一部に点在するような検出対象に対して有効と思われる、その他の Semantic Segmentation モデルとの比較検討はなされていない。また精度は mIoU で 0.464 と報告されており、スカムの早期検出のためには更なる精度向上が求められる。

### 3. スカムモニタリングの要件

#### (1) 対象河川

本研究では、大阪府平野川を対象とした。大阪市内を流下する一級河川平野川は、低平地にある感潮河川であること、また、合流式下水道が整備されているため降雨に伴い下水が放流されること等から、30年以上にわたってスカムの発生が問題になっている。

スカム監視のため、沿川に 18 台のカメラが設置され、「スカムアラート」と呼ばれるモニタリングシステムが運用され、画像の蓄積もされていることから対象河川として選定した。

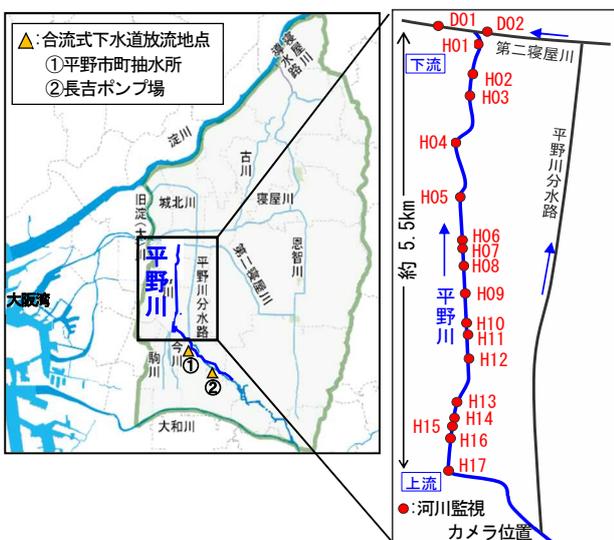


図-1 対象箇所

#### (2) 対象カメラ

河川監視カメラの設置状況を図-1に示す。平野川沿川に設置されている 18 台に対し、各地点において 1 分間に 1 枚撮影して AVI 形式で保存された画像データを対象とした。河川監視カメラの画像例を図-2に示す。画像サイズは 1280\*720 ピクセル、3チャンネルの RGB 画像である。

本研究では 2020 年 1 月～2022 年 12 月の画像を河川管理者から借用して使用した。

#### (3) スカム画像の特徴

スカムが浮遊する水域や周辺背景など解析対象ではない背景領域が占める割合が大きく、対象物であるスカムの領域は小さいことから不均衡データである。

また、スカムはまとまった領域ではなく、水面上に点在しつつ面的な広がりを示すことが多い点の特徴である。またそれらが水面の揺らぎなどと映りが類似する場合があります。画像認識上スカムの抽出を困難とさせている。

#### (4) モニタリング要件

河川管理者が監視カメラを用いてスカムの確認を行っている現状を踏まえ、画像認識によるモニタリング要件を以下の通り設定した。

- 固定カメラを対象に毎分河川スカムの領域を検出すること（動画でのリアルタイム処理は不要）
- スカム発生のみし（図-3）を検出するため微小（被覆率 5%以内目標）なスカムを検出できること
- 検出結果に基づき水面に対する河川スカムの被覆率を算出すること



図-2 カメラ画像例

(左：スカムなし/右：スカムあり)



図-3 スカム発生のみし例

## (5) 画像認識要件

モニタリング要件に基づき、画像認識の技術要件を検討した。スカムは水面上に点在しつつ面的な広がりを示すことから、ピクセル単位の Semantic Segmentation タスクを適用することとした。河川スカムの Semantic Segmentation を行うための公開モデルやデータは存在しないため、本研究では独自に教師データを作成し、深層学習を行った。画像認識要件は以下の通りである。

- Semantic Segmentation モデルによりスカムの領域検出（スカムかそれ以外の2クラス）
- Semantic Segmentation の後処理として水面領域（固定）に対するスカム割合を算出
- 1枚の推論時間 10秒以内
- スカムクラス IoU 精度 95%目標

## 4. モデルの選定

スカム検出モデルの開発に当たり、少量の教師データを用いて同一の条件で複数モデルの学習を行い、スカム検出のベースとなるモデルの選定を行った。

### (1) 使用データ

モデル選定のための教師データとしてスカムが映った画像（以下 Positive 画像）から、カメラ・時間帯・被覆率の異なる 738 枚を目視で選定し、アノテーションした（図-4）。アノテーションはピクセルレベルでスカム領域のみを塗りつぶして行った。また、先行研究で水面の波や映りこみの誤認識が課題となっていたため、スカムが映っていない画像（以下 Negative 画像）を 3 : 2 の割合で混在させた。結果的に Positive 画像 738 枚と Negative 画像 492 枚を合わせた 1230 枚を教師データとした。

学習時はこれらの画像を Positive/Negative ごとに train : validation : test = 81:9:10 に分割し、それぞれ 1,009 : 98 : 123 枚を用いた。モデルの入力は 1280\*720 ピクセルの画像を 256\*448 ピクセルにリサイズした（DeepLabv3+Xceptionv2 のみモデルの制約により 300\*512 ピクセル）。

### (2) モデルの比較実験

先行研究で採用している U-Net をベースラインとして、不均衡データや微小な領域の Segmentation を改善する可能性のある表-1 に示すモデルを比較した。

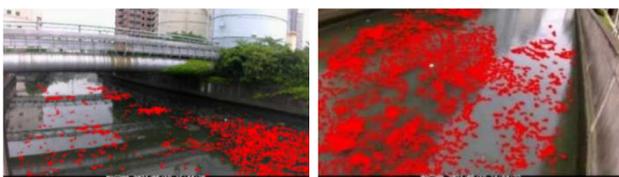


図-4 アノテーションの例

比較したモデルの概要を以下に示す。

Attention U-Net<sup>[11]</sup>は U-Net を基として、不均衡データに対応するため対象領域の特徴量を強調し、背景領域の特徴量の活性化を抑制する Attention 機能を実装したものである。SegNet<sup>[12]</sup>は U-Net と同時期に発表されたエンコーダ-デコーダ形式のモデルで、U-Net と異なりスキップ接続を使わないモデルである。DeepLabv3+<sup>[13]</sup>は空間ピラミッドプーリングと呼ばれる機構により、複数のスケールのコンテキストを用いて高精細な Segmentation が行えるとされる。DeepLabv3+はバックボーンネットワークが複数用意されているため、実験ではバックボーンの異なる 4 種類の DeepLabv3+を比較した。

学習条件は表-2 に示す設定値で、すべてのモデルで共通の条件で学習を行った。

### (3) モデルの評価

Semantic Segmentation では一般的な IoU (Intersect of Union) を評価指標とし、test 画像に対するスカムクラスの IoU(Scum IoU)、スカム以外の IoU(Background IoU)、それらの平均(mIoU)を算出した。評価に当たっては、モニタリング要件としてスカムを高精度に検出できることが求められることから Scum IoU を重視して評価を行った。

評価の結果、Attention U-Net が他のモデルに比べて大幅に高い Scum IoU を示したため、ベースモデルとして採用することとした。Attention U-Net が高精度にスカムを検出できた理由としては、領域割合が少ないスカムに対して Attention 機構が有効に働いたためと考えた。

表-1 比較対象モデルと精度評価結果（太字は最良）

	Scum IoU [%]	BG IoU [%]	mIoU [%]
U-Net	68.77	96.04	82.39
Attention U-Net	<b>87.41</b>	86.59	<b>87.00</b>
SegNet-VGG16	77.68	96.70	84.69
DeepLabv3+ ResNet18	73.94	96.89	85.42
DeepLabv3+ ResNet50	72.95	96.84	84.90
DeepLabv3+ Xceptionv2	75.25	97.05	86.15
DeepLabv3+ MobileNetv2	75.16	<b>97.09</b>	86.12

表-2 学習条件

学習条件	設定値
ミニバッチ	4
反復回数	13,200 回, 52 エポック
損失関数	Focal Loss ( $\gamma=2.5$ )
最適化手法	Adam

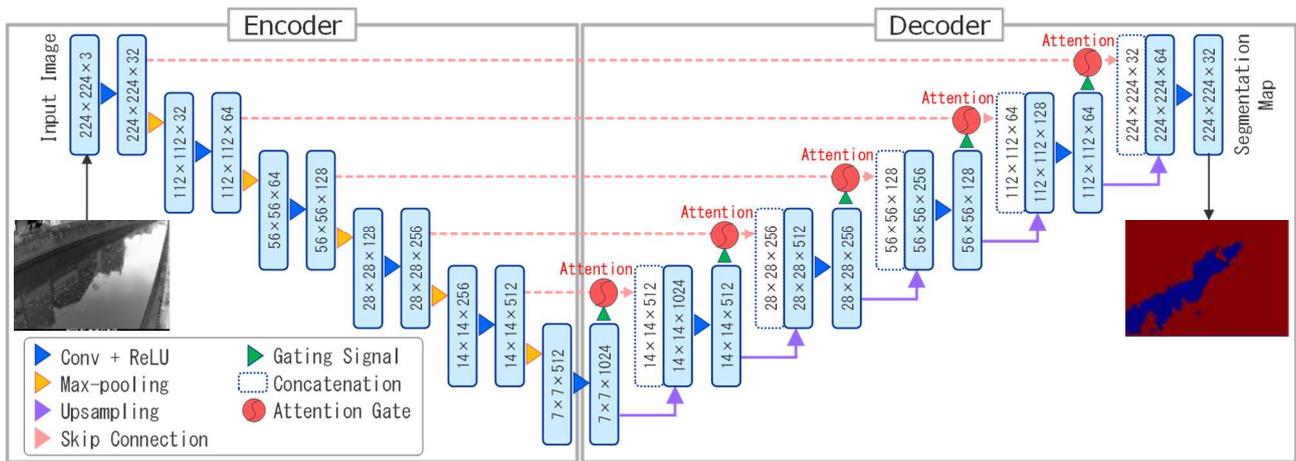


図-5 Attention U-Net ネットワーク構造

図-5に Attention U-Net のネットワーク構造を示す。Attention 機構は、文章や画像の特定の部分に注意を向けるようにモデルを学習させる仕組みである。Attention U-Net では U-Net のスキップ接続に Attention 機構を追加し、Encoder の特徴量を Decoder へ連結する前に Attention Gate を通すモデルである。これにより特定のオブジェクトをセグメント化する際に、どの領域に焦点を当てるかについて、スキップ接続に追加の情報を与える。Attention Gate は Encoder の特徴量  $x$  と Decoder の特徴量  $g$  を受け取るが、特徴量  $g$  はネットワークの次の最下層から取得されるため次元が小さく、特徴表現が優れている。特徴量  $x$  と  $g$  をそれぞれ畳み込み処理を行って計算した特徴量マップの和を ReLU 関数にかけた後、さらに畳み込みを行いシグモイド関数により Attention 係数(重み)を生成する。

スキップ接続に Attention Gate を通すのは、入力値と予測値の関連度が高い特徴量に対して空間方向の Attention を採用し、畳み込み処理によって抽出した特徴量マップのどの領域に注意を向けるかを重みづけする手法であり、その特徴量を重視した予測モデルを構築することができる。

本研究においては、スキップ接続時に画像内に点在する小さなスカムの領域に注意が向けられたことで、効率的にスカムの特徴を学習できたと考えた。その結果、非常に小さなオブジェクトでもうまく検出できるようになり、Scum IoU が向上したと考えた。

## 5. 損失関数の変更

ベースモデルとして選定した Attention U-Net は他モデルに比べて精度が高いものの、Scum IoU が 87% とまだ改善の余地がある。表-2 に示す条件のうち、精度に影響が大きい損失関数の変更による改善を試みた。

表-3 比較対象損失関数と精度評価結果 (太字は最良)

	Scum IoU [%]	BG IoU [%]	mIoU [%]
<b>Base: Focal Loss</b>	87.41	86.59	87.00
<b>1: Focal Tversky Loss</b>	88.91	89.05	88.98
<b>2: 1 + Triplet Loss</b>	90.38	89.73	90.05
<b>3: 2 + IoU Loss</b>	89.05	89.15	89.11
<b>4: 2 + Dice Loss</b>	<b>91.35</b>	<b>90.69</b>	<b>91.02</b>

### (1) 使用データ

モデルの選定時と同一の 1230 枚の教師データを使用し、train : validation : test 分割、リサイズの条件も踏襲した。

### (2) 損失関数の比較実験

モデルの選定時は損失関数を Focal Loss に固定して実験を行ったが、スカム画像の特徴に合わせて精度改善が見込める、表-3 に示す損失関数で学習を行った。Focal Loss をベースラインとして、それぞれの損失関数による改善効果を定量的に確認した。なお、損失関数以外の学習条件は、モデルの選定時と同様とした。

比較した損失関数の概要を以下に示す。

著者らはコンクリートのひび割れを対象とした研究<sup>14)</sup>で、不均衡かつ微小な領域の Semantic Segmentation において Focal Tversky Loss が有効であることを示した。スカムも同様の特徴を有するため、本研究においても Focal Tversky Loss の適用を試みた。また、4章の実験結果の推論画像を見たところ、スカムが発生していない Negative 画像の一部の領域をスカムと誤って検出していることが確認された。各ピクセルを Positive と Negative に十分に分離できていないことが示唆されるため、クラス間の距離を離す効果がある Triplet Loss を追加した。次に、画像認識要件で目標とする精度指標の IoU を直接的に損失関数に組み込むため、IoU Loss とその派生形の Dice Loss を追加して実験を行った。

### (3) 損失関数の評価

モデルの選定時と同様に ScumIoU で評価した (表-3) . 評価の結果, 4 の Focal Tversky Loss + Triplet Loss + Dice Loss のケースが最良の結果を示した.

Focal Tversky Loss では FP(偽陽性)と FN(偽陰性)に重みをつけた Tversky Index と呼ぶ指標が有効に働き, 微小なスカム領域をセグメント化することができたと考えた. また, 損失関数が Focal Loss と同様に指数関数で定義されていることで, 分類が容易な背景領域が損失関数に与える影響を抑制し, 分類が困難なスカム領域に比重が置かれる効果が出たと推察した.

Triplet Loss は一般的には距離学習で使われる損失関数で, Query/Positive/Negative の 3 つの画像を入力として, 同一のクラスは近づけ, 異なるクラスを遠ざけるように学習が行われる. 本研究ではこれをピクセル単位で適用した結果, スカム領域と背景領域の特徴量距離が離れるように学習が進み, IoU が向上したと考えた.

IoU Loss と Dice Loss はいずれも正解領域と予測領域の重なりを評価する損失関数である. 本研究では IoU Loss を加えると精度が低下したのに対し, Dice Loss では精度が向上した. 損失関数の式からは理由が不明であり, 教師データ依存の偶発的な事象の可能性がある.

## 6. 教師データ追加による最終モデル学習

これまでの実験結果に基づき, Attention U-Net を Focal Tversky Loss + Triplet Loss + Dice Loss の損失関数で最終モデルの学習を行った. 教師データを追加してバリエーションを増強するとともに, 学習回数を増やして精度向上を図った.

### (1) 使用データ

18 地点のカメラ映像で, 時間帯, 四季を通じた画像データを対象に, スカム画像の追加アノテーションを実施した. また, 4 章および 5 章の実験結果から誤検知しやすい紛らわしい画像も「Negative 画像」として追加した (図-6) . Negative 画像の例としては, 桜の花びら, 植物やごみ等の流下浮遊物, 船影・船による波, 水しぶき・雨, さざなみ, きらめき, 空や周囲の映り込み, 鳥類・魚類等である.

最終的に表-4 に示す 3,193 枚の教師データを得た.

### (2) 学習結果

選定したモデルと損失関数を用いて, 3193 枚の教師データを学習した. 教師データは train:validation = 9:1 に分割し, validation データでの損失が最小となった時点を終最終モデルとして採用した.

最終モデルの validation データに対する精度は表-5 の通りであった. 教師データ追加により, いずれの指標も精度向上が見られ, ScumIoU も目標の 95% に到達した.

また推論結果画像を目視で確認し, 教師データと比較してスカム領域を精度よく検出できていることを確認した (図-7) .



図-6 Negative 画像例

表-4 最終教師データ数

Positive画像	2,180枚
Negative画像	1,013枚
合計	3,193枚

表-5 最終モデルと精度評価結果

	Scum IoU [%]	BG IoU [%]	mIoU [%]
最終モデル	95.13	95.15	95.14

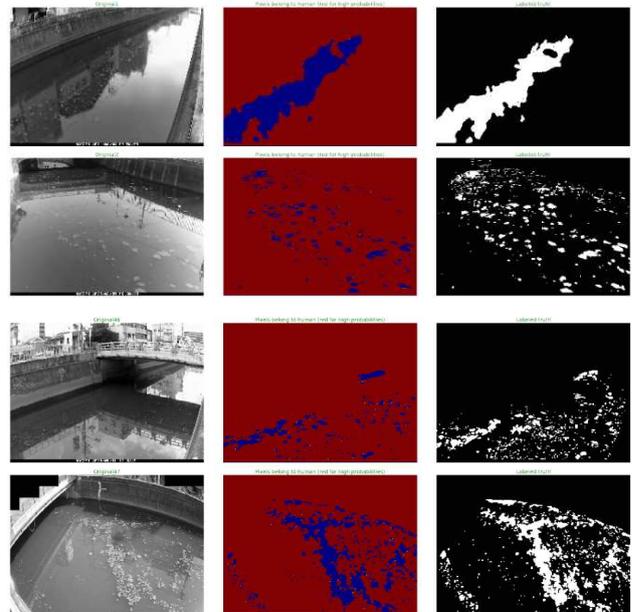


図-7 推論結果の例 (左から元画像, 推論結果, 正解画像)

## 7. おわりに

### (1) 本研究の成果

本研究では、スカム発生の早期検知を実現することを目的として、画像認識技術の要件を定義した。その上で、大阪府平野川の河川監視カメラのデータを用いて、スカム画像の特性に合わせた画像認識モデルや損失関数の設定を実験により明らかにした。

さらに、実験時点でスカム検出の精度が低い画像を教師データに追加した上で最終的な学習を行い、Scum IoU95%を達成し、要件を満足する高精度なスカム検出モデルを開発した。特に、モニタリング上重要なスカム発生の兆し(被覆率 5%以内の微小なスカム)を捉える精度を実現したことが本研究の貢献である。

### (2) 今後の課題

本研究では日中の画像を対象としたが、夜間に発生して夜明けとともに大量に発生しているスカムが確認されたケースも見られたことから、夜間のモニタリングおよび検知の仕組みを構築することが望ましい。対象河川は住宅地に存在し、夜間に光源を使用し続けることは難しく、検知手法に工夫が必要である。

また、研究成果の高精度スカム検出モデルを河川監視カメラシステムに実装することが今後の課題である。その上で、モニタリング結果に基づきスカム発生のメカニズムが解明され、スカムの発生抑制につながることを期待する。

**謝辞**：本研究において、西大阪治水事務所児島 直美氏、茨木土木事務所奥野 喜久氏に多大なるご助言を頂いた。ここに記して、深甚なる謝意を表します。

### 参考文献

- 1) 山崎正夫, 津久井公昭: 河川におけるスカムの発生に関する研究 (その1) スカムの期限に関する研究, 東京環境科学研究所年報, vol. 1, 174-179, 1991.
- 2) 菅原正孝, 石川宗孝, 西田一雄, 北村誠, 梶智裕, 木本茂宏, 大西敏夫: 汚濁河川に発生するスカムの生成機構について, 環境技術, vol. 24 (7), 402-405, 1995.
- 3) 二瓶泰雄, 若月宣人: 洪水時河川における浮遊ゴミ輸送量計測の試み, 土木学会論文集 b, vol. 66 (1), 19-24, 2010.

- 4) 二瓶泰雄, 白川明宏, 鈴木達裕, 赤松良久: 出水時大河川における浮遊ゴミ輸送特性に関する現地観測と湾内 do 環境へ及ぼす影響, 土木学会論文集 b2(海岸工学), vol. 66 (1), 2010.
- 5) 南まさし, 尾ノ井龍仁, 二瓶泰雄, 西島拓駿, 堀田琢哉, 船本優月, 金子博, 大谷明, 片岡智哉, 日向博文: 河川漂流ゴミ輸送量の自動連続モニタリング手法の開発と最上川観測への適用, 土木学会論文集 b1 (水工学), vol. 71 (4), I\_1225-I\_1230, 2015.
- 6) Kataoka, T. and Nihei, Y.: Quantification of Floating Riverine Macro-Debris Transport Using an Image Processing Approach. *Sci Rep*, vol. 10 (1), 2198, 2020.
- 7) 水田周作, 高崎忠勝, 河村明, 天口英雄, 石原成幸: 定点カメラ画像を用いたニューラルネットワークによる都市河川のスカム自動判別, 土木学会論文集 b1 (水工学), vol. 71 (4), I\_1231-I\_1236, 2015.
- 8) 中谷祐介, 奥村素生, 岩岡慶晃, 西田修三: U-Netを用いた河川浮遊物の連続観測手法, 土木学会論文集 b1 (水工学), vol. 76 (2), I\_997-I\_1002, 2020.
- 9) Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, <https://doi.org/10.48550/arXiv.1505.04597>, 2015.
- 10) 中谷祐介, 懸樋洗大: U-Netを用いた河川スカム連続検出手法の改良, 土木学会論文集 b1 (水工学), vol. 77 (2), I\_895-I\_900, 2021.
- 11) Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B. and Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas, <https://doi.org/10.48550/arXiv.1804.03999>, 2018.
- 12) Badrinarayanan, V., Kendall, A. and Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, <https://doi.org/10.48550/arXiv.1511.00561>, 2016.
- 13) Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, <https://doi.org/10.48550/arXiv.1802.02611>, 2018.
- (14) 都築幸乃, 吉田龍人, 大久保順一, 藤井純一郎, 山下隆義: Focal Tversky Loss による護岸のひび割れセグメンテーションの精度向上, 人工知能学会全国大会論文, vol.36, 2022,

(Received June 30, 2024)  
(Accepted August 31, 2024)

## Development of high-precision scum detection model for early warning of river scum

Junichiro FUJII, Masahiro OKANO and Sanae GOTO

In urban rivers where combined sewer overflow inflow occurs, scum forms after rainfall, causing problems such as bad smells, deterioration of the landscape, and adverse effects on the ecosystem. As a countermeasure, early warning through monitoring of scum is desired. In this study, we examined the monitoring and image recognition requirements with the aim of detecting scum formation at an early stage, and then used data from a river monitoring camera on the Hirano River in Osaka Prefecture to experimentally clarify the image recognition model and loss function that match the characteristics of scum images. Based on the experimental results, we added images with low accuracy in detecting scum to the training data and conducted final training, and developed a high-precision scum detection model that satisfied the requirements by achieving Scum IoU95%.