

# Iterative Inpainting Augmentation による 不十分なアノテーションからのインスタンスセグメンテーション

小島 和輝† 飯山 将晃†

滋賀大学大学院 データサイエンス研究科†

## 1. はじめに

インスタンスセグメンテーションの問題点として、推論時に画像中の一部の物体が検出されない現象（過少検出）がある。この原因の一つは、不十分なアノテーションを用いることである。インスタンスセグメンテーションの既存手法[1, 2, 3]の多くは画像中の全てのインスタンスにアノテーションが付与されていることを前提としている。不十分なアノテーションを用いた場合は学習が安定して行われず、過少検出が引き起こされる。もう一つの原因は、画像中の物体同士の隣接である。画像中で隣接する物体領域から得られる画像特徴量が影響し、隣接する他の物体の検出を妨げる場合がある。また、物体同士が隣接することにより、各物体の正確な物体領域の特定も困難になる。

## 2. 提案手法

本研究では、不十分なアノテーションを用いた場合でも物体の過少検出を抑制し正確なインスタンスセグメンテーションを行うため、Iterative Inpainting Augmentation を提案する。具体的には、検出済の領域をインペインティングにより除去する Cut-and-Inpainting、推論時に繰り返しインスタンスセグメンテーションを行う Iterative Inference の 2 つの処理から構成される。

Cut-and-Inpainting では物体同士が隣接することによる影響を物体領域の除去によって低減させるとともに、画像中で検出すべきインスタンス数を減らし、未検出物体を検出しやすい状況を作り出す。Iterative Inference では、少数インスタンスの画像に対しても安定した検出を行うために、インペインティングした画像を学習したモデルを用意し、Cut-and-Inpainting に

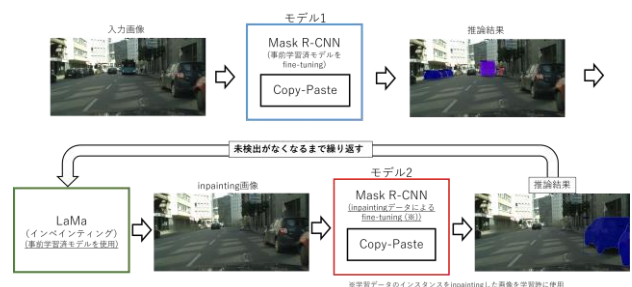


図 1. 提案手法の概要

より生成した画像に対して推論を繰り返し行う。これらの処理により未検出の物体を少しずつ検出させ、過少検出を抑制させる。

提案手法の概要を図 1 に示す。既存のインスタンスセグメンテーションでも用いられている Copy-Paste Augmentation[4]を適用することで、インスタンス数の偏りによる精度低下を防ぐ。また、インスタンスセグメンテーションモデルとして Mask R-CNN[1]をベースにし、通常画像を用いたモデル（モデル 1）と学習データのいくつかの物体領域に対しインペインティングを行った画像を作成し、それらを用いてモデル 1 を fine-tuning したモデル（モデル 2）を作成する。

なお、インペインティングモデルには LaMa[5]を用いた。以降、提案手法を Iteration-IpAug と呼ぶ。

## 3. 実験

本研究では 2 種類のデータセットを使用した。1 つは、DeepFish Dataset[6]と呼ばれる、水産資源管理を目的とした魚識別、尾数計測用のデータセットである。このデータセットはアノテーションが全ての魚に対し完全には付与されていないため、評価指標には尾数の MAE を用いて正解データと検出物体数の誤差を評価する。なお、正解尾数については筆者が手動でアノテーションを行った。もう 1 つは、Cityscapes Dataset[7]である。市街地景観の画像データセットであり、アノテーションが完全に付与され

Instance segmentation from insufficient annotation using iterative inpainting augmentation

Kazuki Ojima† Masaaki Iiyama†  
Graduate School of Data Science, Shiga University†

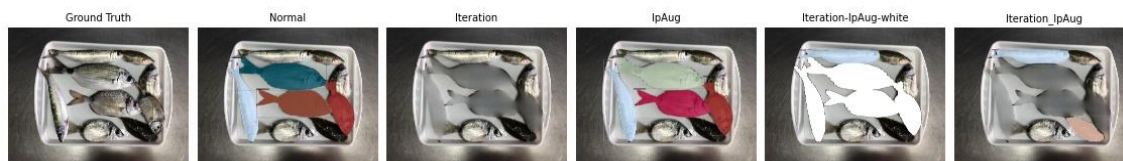


図 2. DeepFish Dataset の推論結果



図 3. Cityscapes Dataset の推論結果

表 1. DeepFish Dataset での実験結果

手法/繰り返し回数	MAE	
	1	2
Normal	0.995	0.995
Iteration	0.900	0.910
IpAug	1.440	1.440
Iteration-IpAug-white	0.904	0.904
<b>Iteration-IpAug(ours)</b>	<b>0.781</b>	<b>0.768</b>

ているため、評価には MS COCO 評価指標を用いた。なお、本研究では不十分なアノテーションを再現するため、アノテーションの削除割合を 95%としたデータを使用した。

性能比較のため、インペインティングを行わず通常の推論を行う手法 (Normal), Normal の推論結果に基づき検出部分にインペインティングを行った画像を作成し、それを対象に推論を繰り返す手法 (Iteration), モデル 2 をもちいで繰り返さず推論を行う手法 (IpAug), 提案手法において LaMa の代わりに検出領域を白く塗りつぶすことでインペインティングを行う手法 (Iteration-IpAug-white) との比較を行った。

各データセットの実験結果を表 1, 2 および図 2, 3 に示す。実験では、推論時の反復回数を DeepFish Dataset では 2 回, Cityscapes Dataset では 3 回とした。反復推論を取り入れた手法 (Iteration, Iteration-IpAug-white, Iteration-IpAug) で従来手法 (Normal) よりも精度が向上するが、その中でも提案手法が最も良い精度を示した。図 2, 3 から、提案手法を用いた場合は比較手法よりも検出物体数が増加していることが分かる。また、Iteration-IpAug-white は Normal よりも高い精度を示しているものの、提案手法には劣る結果となっている。これより、ただ物体領域を塗りつぶすのではなく、インペインティングにより物体そのものを画像中から除去することで物体同士の隣接による影響を低減でき、検出精度の向上に繋がると考えられる。

#### 4. まとめ

本研究では、インスタンスセグメンテーションの過少検出を抑制する手法を提案した。事件の結果、提案手法には過小検出を改善し、検出精度を向上させる効果があることが分かった。最新アーキテクチャの導入、さらに検出困難なデータセットを用いた汎化性能の検証を行うことが今後の課題である。

表 2. Cityscapes Dataset での実験結果

IoU= 0.50:0.95	Normal	Iter.	IpAug	Iter-IpAug white	<b>Iter. IpAug (ours)</b>
mAP (all)	0.212	0.149	0.125	0.212	<b>0.244</b>
mAP (large)	0.471	0.353	0.278	0.472	<b>0.513</b>
AR (all)	0.288	0.291	0.236	0.290	<b>0.328</b>
AR (large)	0.574	0.576	0.455	0.576	<b>0.614</b>

#### 参考文献

- [1]He, Kaiming, et al. "Mask r-cnn." *ICCV*, 2017.
- [2]Chen, Kai, et al. "Hybrid task cascade for instance segmentation." *CVPR*, 2019.
- [3]Wang, Xinlong, et al. "Solo: Segmenting objects by locations." *ECCV*, 2020.
- [4]Ghiasi, Golnaz, et al. "Simple copy-paste is a strong data augmentation method for instance segmentation." *CVPR*, 2021.
- [5]Suvorov, Roman, et al. "Resolution-robust large mask inpainting with fourier convolutions." *WACV*, 2022.
- [6] Garcia-d' Urso, Nahuel, et al. "The DeepFish computer vision dataset for fish instance segmentation, classification, and size estimation." *Sci. Data* Vol.9, No.1, pp.287, 2022.
- [7]Cordts, Marius, et al. "The cityscapes dataset for semantic urban scene understanding." *CVPR*, 2016.