

Segmentation interactive d'image avec SAM

par Ana Bernal

mentor: Samir Tanfous

Avril 2023



Dépôt github pour ce projet:

https://github.com/ana-bernal/P7_premiere-concept

Programme

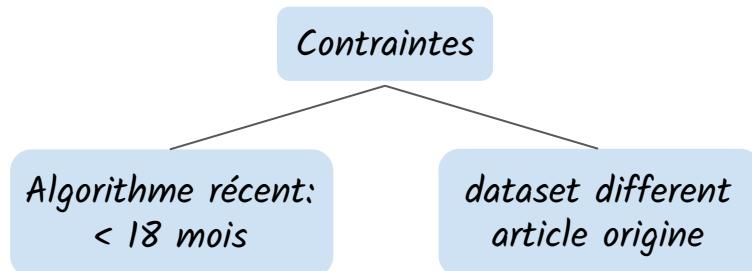
- 1** Rappel consignes
- 2** La tâche de segmentation d'image
- 3** SAM: Segment Anything Model
- 4** Notre expérience
- 5** Conclusions

1

Rappel consignes

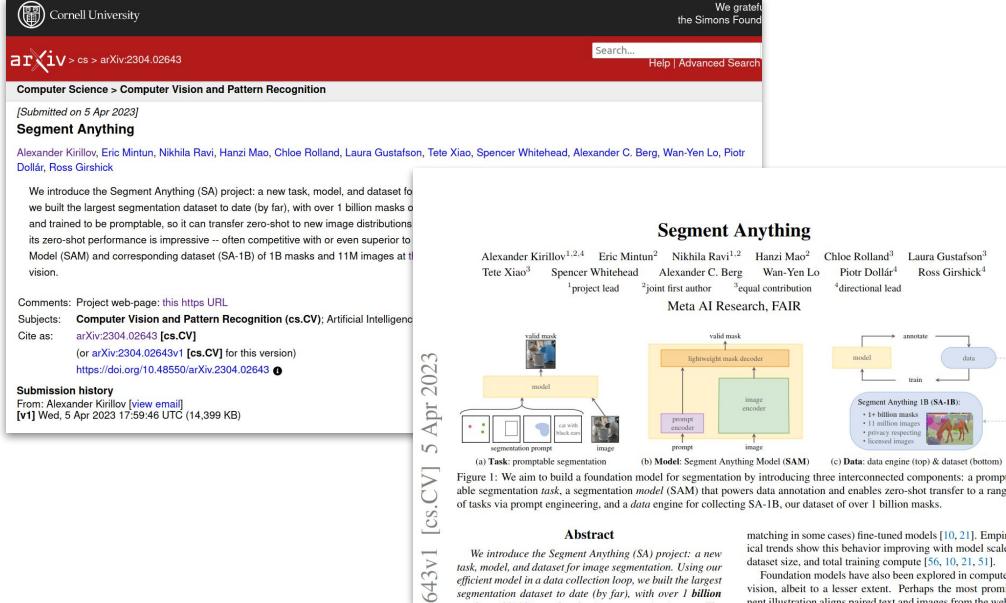
Rappel consignes

- Tour d'horizon + état de l'art → article récent SAM par  Meta
- Approfondir + choisir références → plan de travail
- Implémentation algorithme + baseline



Rappel consignes

Début avril  Meta



The screenshot shows the arXiv preprint page for 'Segment Anything'. The title is 'Segment Anything' by Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollar, and Ross Girshick. It was submitted on April 5, 2023. The abstract discusses a new task, model, and dataset for image segmentation, noting the largest dataset to date (over 1 billion masks) and its zero-shot transfer capabilities. The paper is associated with Meta AI Research, FAIR, and includes author roles like project lead, joint first author, equal contribution, and directional lead. The data engine section highlights over 1 billion masks, 1 million categories, privacy respecting licensed images, and a dataset size of over 1 billion masks.

Segment Anything

Alexander Kirillov^{1,2,4} Eric Mintun² Nikhila Ravi^{1,2} Hanzi Mao² Chloe Rolland³ Laura Gustafson³
Tete Xiao³ Spencer Whitehead¹ Alexander C. Berg¹ Wan-Yen Lo¹ Piotr Dollar⁴ Ross Girshick⁴

¹project lead ²joint first author ³equal contribution ⁴directional lead

Meta AI Research, FAIR

Comments: Project web-page: [this https URL](https://thishttpsURL)

Subjects: Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence

Cite as: [arXiv:2304.02643 \[cs.CV\]](https://arxiv.org/abs/2304.02643)
(or [arXiv:2304.02643v1 \[cs.CV\]](https://arxiv.org/abs/2304.02643v1) for this version)
<https://doi.org/10.48550/arXiv.2304.02643> 

Submission history

From: Alexander Kirillov [view email]
[v1] Wed, 5 Apr 2023 17:59:46 UTC (14,399 KB)

643v1 [cs.CV] 5 Apr 2023

Abstract

We introduce the Segment Anything (SA) project: a new task, model, and dataset for image segmentation. Using our efficient model in a data collection loop, we built the largest segmentation dataset to date (by far), with over 1 billion matching in some cases) fine-tuned models [10, 21]. Empirical trends show this behavior improving with model scale, dataset size, and total training compute [56, 10, 21, 51]. Foundation models have also been explored in computer vision, albeit to a lesser extent. Perhaps the most prominent illustration stems from text-and-image retrieval from the web.

(a) Task: promptable segmentation

(b) Model: Segment Anything Model (SAM)

(c) Data: data engine (top) & dataset (bottom)

Figure 1: We aim to build a foundation model for segmentation by introducing three interconnected components: a promptable segmentation task, a segmentation model (SAM) that powers data annotation and enables zero-shot transfer to a range of tasks via prompt engineering, and a data engine for collecting SA-1B, our dataset of over 1 billion masks.

Rappel consignes

Début avril



arXiv > cs > arXiv:2304.02643

Segment Anything

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollar, Ross Girshick

We introduce the Segment Anything (SA) project: a new task, model, and dataset for us to build the largest segmentation dataset to date (by far), with over 1 billion masks and trained to be promptable, so it can transfer zero-shot to new image distributions. Its zero-shot performance is impressive -- often competitive with or even superior to Model (SAM) and corresponding dataset (SA-1B) of 1B masks and 11M images at the vision.

Comments: Project web-page: this https URL
Subjects: Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence
Cite as: arXiv:2304.02643 [cs.CV]
(or arXiv:2304.02643v1 [cs.CV] for this version)
https://doi.org/10.48550/arXiv.2304.02643

Submission history
From: Alexander Kirillov [view email]
[v1] Wed, 5 Apr 2023 17:59:46 UTC (14,399 KB)

643v1 [cs.CV] 5 Apr 2023

Abstract
We introduce the Segment Anything (SA) project: a new task, model, and dataset for image segmentation. Using our efficient model in a data collection loop, we built the largest segmentation dataset to date (by far), with over 1 billion

Segment Anything

Alexander Kirillov^{1,2,4} Eric Mintun² Nikhila Ravi^{1,2} Hanzi Mao²
Tete Xiao³ Spencer Whitehead¹ Alexander C. Berg¹ Wan-Yen Lo¹
¹project lead ²joint first author ³equal contribution
Meta AI Research, FAIR

(a) Task: promptable segmentation

(b) Model: Segment Anything Model (SAM)

(c) Data: data engine (top) & dataset (bottom)

Figure 1: We aim to build a foundation model for segmentation by introducing three interconnected components: a promptable segmentation **task**, a segmentation **model** (SAM) that powers data annotation and enables zero-shot transfer to a range of tasks via prompt engineering, and a **data engine** for collecting SA-1B, our dataset of over 1 billion masks.

matching, in some cases) fine-tuned models [10, 21]. Empirical trends show this behavior improving with model scale, dataset size, and total training compute [56, 10, 21, 51]. Foundation models have also been explored in computer vision, albeit to a lesser extent. Perhaps the most prominent illustration stems from text-and-images from the web,

Lifewire
TECH FOR HUMANS

NEWS HOW TO FEATURES ABOUT US

Spotlight: Streaming Sports

NEWS > SMART & CONNECTED LIFE

Meta's New Segment Anything Model for Identification Is a Big Deal, Experts Say

SAM can recognize objects it hasn't seen before

By Sascha Brodsky · Published on April 17, 2023 09:40AM EDT · Fact checked by Jerri Ledford

Key Takeaways

- Meta's new AI image segmentation tool could lead to advances like better tagging of photos on social media.
- The SAM model was trained on a vast database of images.
- There's a race to find better ways for computers to detect and recognize objects.

SMART & CONNECTED LIFE

> AI & Everyday Life News

Rappel consignes

Début avril



arXiv > cs > arXiv:2304.02643

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 5 Apr 2023]

Segment Anything

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tele Xiao, Spencer Whitehead, Alexander C. Berg, Wang Dollar, Ross Girshick

We introduce the Segment Anything (SA) project: a new task, model, and dataset for which we built the largest segmentation dataset to date (by far), with over 1 billion masks and trained to be promptable, so it can transfer zero-shot to new image distributions. Its zero-shot performance is impressive -- often competitive with or even superior to Model (SAM) and corresponding dataset (SA-1B) of 1B masks and 11M images at full vision.

Comments: Project web-page: this https URL

Subjects: Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence

Cite as: arXiv:2304.02643 [cs.CV]

(or arXiv:2304.02643v1 [cs.CV] for this version)

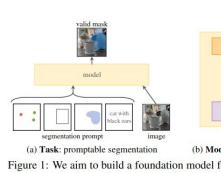
<https://doi.org/10.48550/arXiv.2304.02643>

Submission history

From: Alexander Kirillov [view email]

[v1] Wed, 5 Apr 2023 17:59:46 UTC (14,399 KB)

643v1 [cs.CV] 5 Apr 2023



Abstract

We introduce the Segment Anything (SA) project: a new task, model, and dataset for image segmentation. Unlike previous work that required collecting a large dataset and training a separate efficient model in a data collection loop, we built the largest segmentation dataset to date (by far), with over 1 billion masks and trained a promptable model that can transfer zero-shot to new image distributions.

Lifewire

TECH FOR HUMANS

NEWS HOW TO FEATURES ABOUT US

Spotlight: Streaming Sports E

NEWS > SMART & CONNECTED LIFE

Meta's New Segment Anything Model for Identification Is a Big Deal. Experts Say

HACKADAY

HOME BLOG HACKADAY.IO TINDIE HACKADAY PRIZE SUBMIT ABOUT April 26, 2023

NEED TO PICK OBJECTS OUT OF IMAGES? SEGMENT ANYTHING DOES EXACTLY THAT

by: Donald Papp

7 Comments

April 14, 2023



Rappel consignes

Début avril 

 Cornell University

arXiv > cs > arXiv:2304.02643

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 5 Apr 2023]

Segment Anything

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Dollar, Ross Girshick

We introduce the Segment Anything (SA) project: a new task, model, and dataset. We built the largest segmentation dataset to date (by far), with over 1 billion images and trained to be promptable, so it can transfer zero-shot to new image domains. Its zero-shot performance is impressive -- often competitive with or even surpassing the Mask R-CNN Model (SAM) and corresponding dataset (SA-1B) of 1B masks and 11M images in computer vision.

Comments: Project web-page: this [https URL](https://https://)
Subjects: Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence
Cite as: arXiv:2304.02643 [cs.CV]
(or arXiv:2304.02643v1 [cs.CV] for this version)
<https://doi.org/10.48550/arXiv.2304.02643>

Submission history
From: Alexander Kirillov [view email]
[v1] Wed, 5 Apr 2023 17:59:46 UTC (14,399 KB)

Lifewire
TECH FOR HUMANS

NEWS HOW TO FEATURES ABOUT US

Spotlight: Streaming Sports

NEWS > SMART & CONNECTED LIFE

Meta's New Segment Anything Model for



NEWS DEEP DIVES LEARN GG COINS VIDEOS NEWSLETTERS PODCAST

ES

• News • Artificial Intelligence

April 26, 2023

Meta Launches AI Tool That Can Identify, Separate Items in Pictures

Meta unveils "Segment Anything" AI image identification tool, but the scale of the dataset used by the social giant raises privacy concerns.



By Jason Nelson

Apr 7, 2023

4 min read



Rappel consignes

Début avril 

 Cornell University

 arXiv > cs > arXiv:2304.02643

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 5 Apr 2023]

Segment Anything

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollar, Ross Girshick

We introduce the Segment Anything (SA) project: a new task, model, and dataset for us to build the largest segmentation dataset to date (by far), with over 1 billion masks and trained to be promptable, so it can transfer zero-shot to new image distributions. Its zero-shot performance is impressive -- often competitive with or even superior to Model (SAM) and corresponding dataset (SA-1B) of 1B masks and 11M images at the vision.

Comments: Project web-page: this https URL

Subjects: Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence

Cite as: arXiv:2304.02643 [cs.CV] (or arXiv:2304.02643v1 [cs.CV] for this version)
<https://doi.org/10.48550/arXiv.2304.02643> 

Submission history

From: Alexander Kirillov [[view email](#)]
[v1] Wed, 5 Apr 2023 17:59:46 UTC (14,399 KB)

643v1 [cs.CV] 5 Apr 2023

Segment Anything

Alexander Kirillov^{1,2,4}, Eric Mintun², Nikhila Ravi^{1,2}, Hanzi Mao², Chloe Rolland³, Laura Gustafson³, Tete Xiao³, Spencer Whitehead¹, Alexander C. Berg¹, Wan-Yen Lo¹, Piotr Dollar⁴, Ross Girshick⁴

¹project lead ²joint first author ³equal contribution ⁴directional lead

Meta AI Research, FAIR

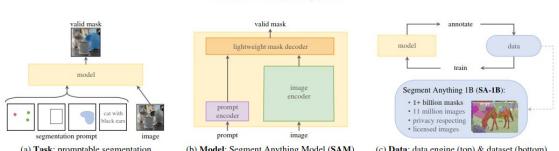


Figure 1: We aim to build a foundation model for segmentation by introducing three interconnected components: a promptable segmentation *task*, a segmentation *model* (SAM) that powers data annotation and enables zero-shot transfer to a range of tasks via prompt engineering, and a *data engine* for collecting SA-1B, our dataset of over 1 billion masks.

Abstract

We introduce the Segment Anything (SA) project: a new task, model, and dataset for image segmentation. Using our efficient model in a data collection loop, we built the largest segmentation dataset to date (by far), with over 1 billion matching, in some cases, fine-tuned models [10, 21]. Empirical trends show this behavior improving with model scale, dataset size, and total training compute [56, 10, 21, 51]. Foundation models have also been explored in computer vision, albeit to a lesser extent. Perhaps the most prominent illustration stems from text-and-images from the web.

Choix d'article principal



2

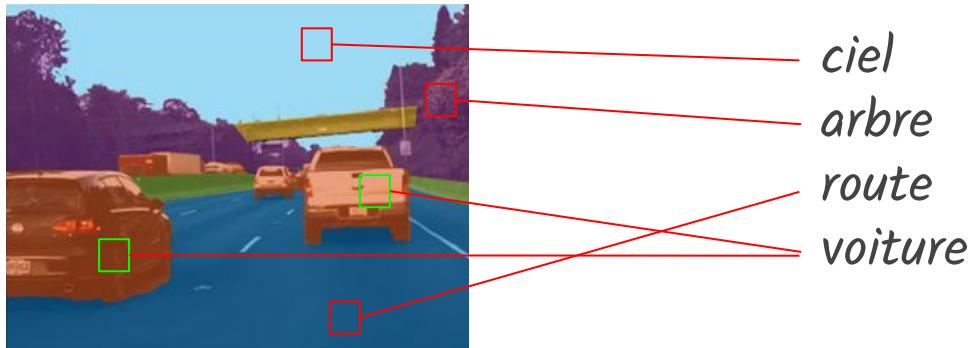
La segmentation d'image

La segmentation d'image

- Partitionner une image digitale dans plusieurs segments/régions
- **But:** simplifier/mieux comprendre
- **Applications:** Recherche d'image par le contenu, imagerie médicale (localiser tumeurs, mesurer volume de tissus, etc), détection de piétons, détections de feu sur la route (conduite autonome), etc...

La segmentation d'image

Concrètement : **classification** des pixels



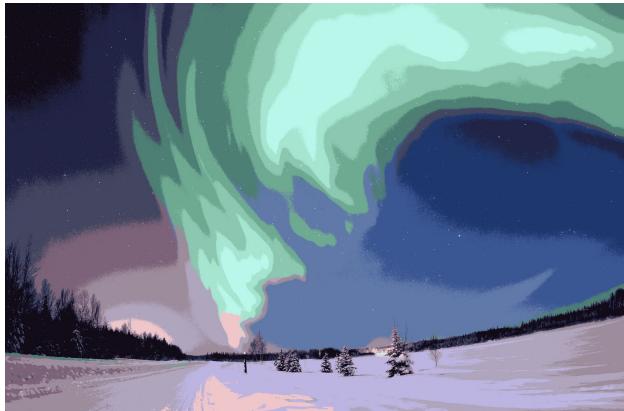
La segmentation d'image: méthodes classiques

Thresholding (seuillage d'image)



La segmentation d'image: méthodes classiques

K-means



La segmentation d'image: méthodes classiques

Détection de bords

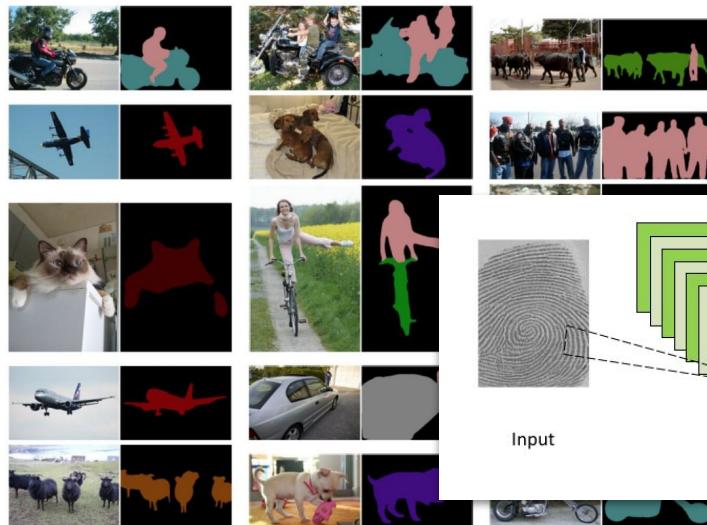


La segmentation d'image:

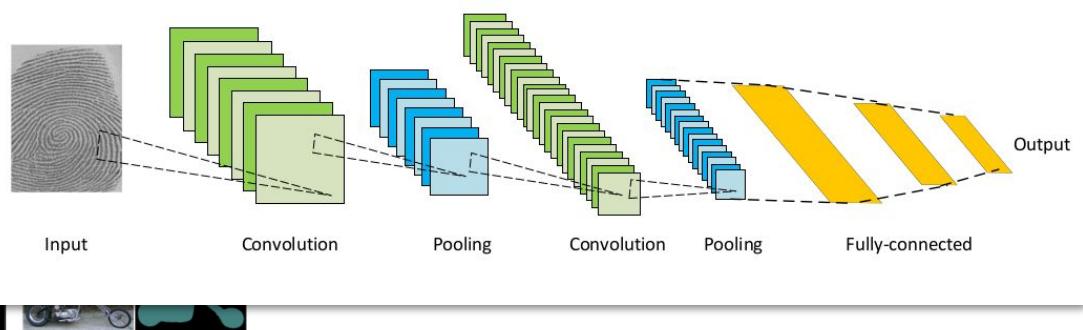
Réseaux des neurones



Extraction automatique des features (bords, volumes ...) →
facilitent la tâche



- 1) Fully convolutional networks
- 2) Convolutional models with graphical models
- 3) Encoder-decoder based models
- 4) Multi-scale and pyramid network based models
- 5) R-CNN based models (for instance segmentation)



La segmentation d'image

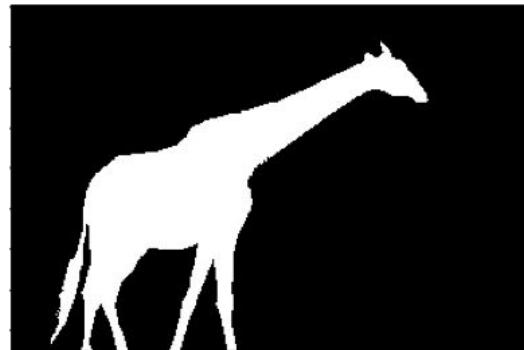
2

Nous: cartes de segmentation (masque, **mask**) binaire

Image original



Masque

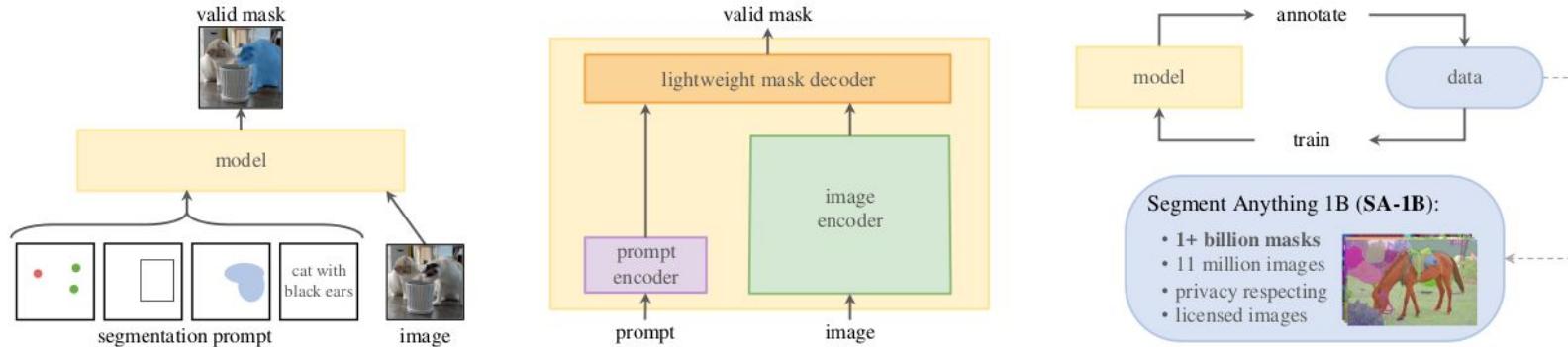


3

SAM

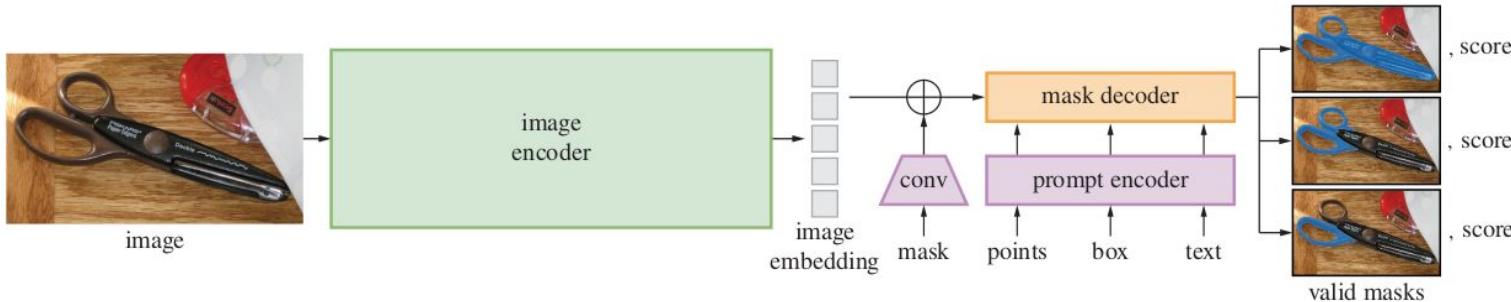
Segment Anything Model

Tâche + modèle de fondation + dataset

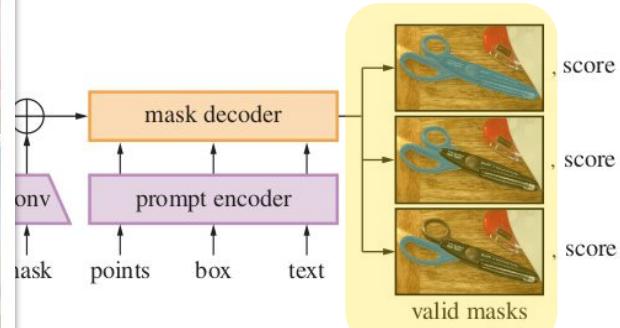


Dataset **SA-1B**

Architecture du modèle

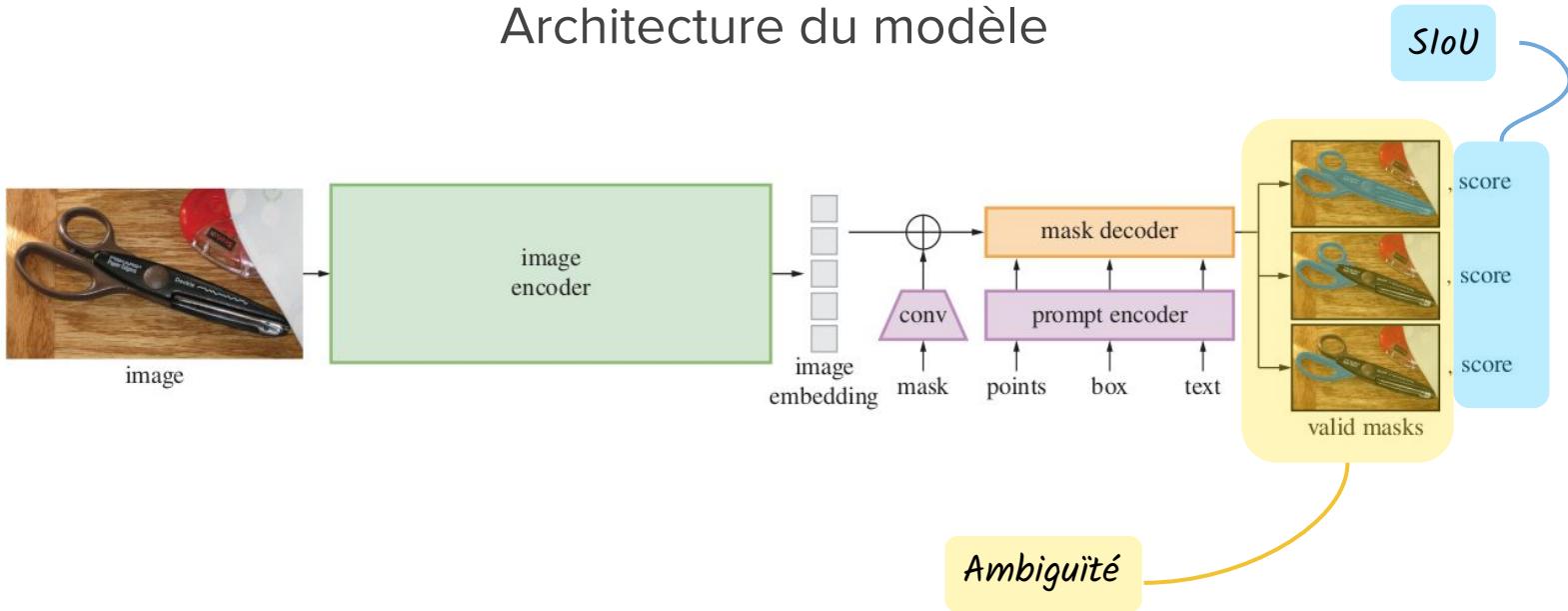


Architecture du modèle



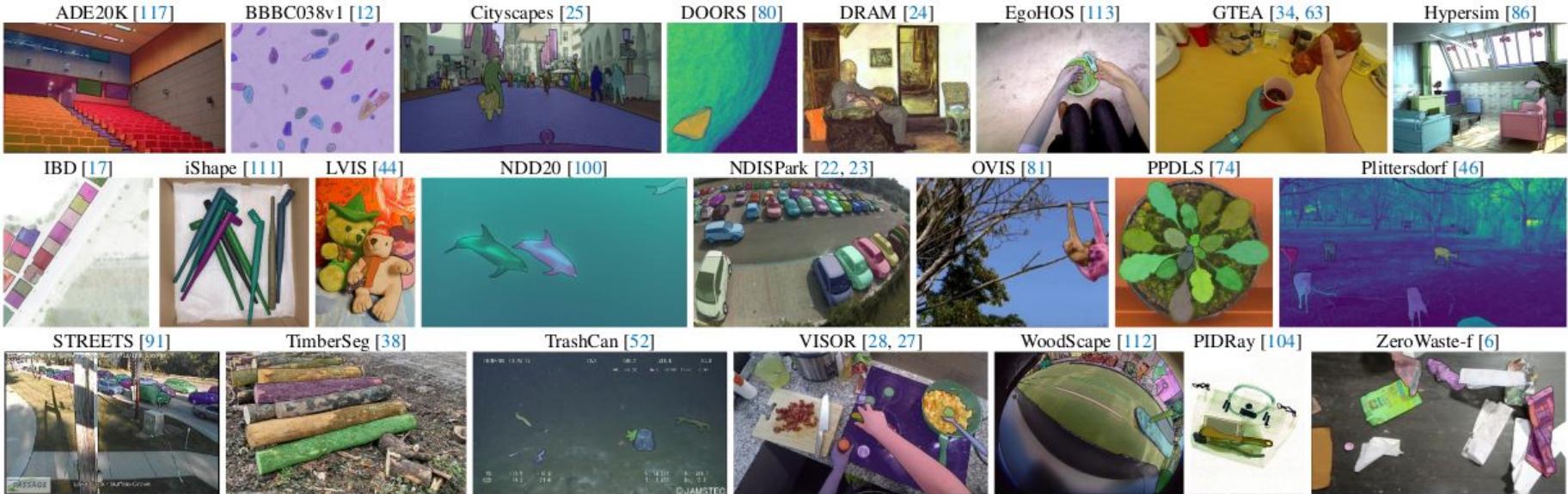
Ambiguité

Architecture du modèle



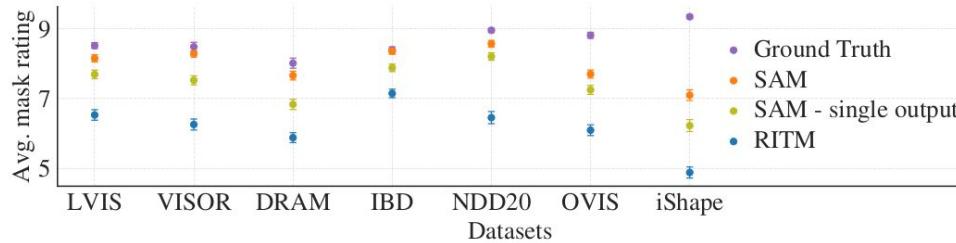
Évaluation du modèle

23 datasets différents → différents domaines

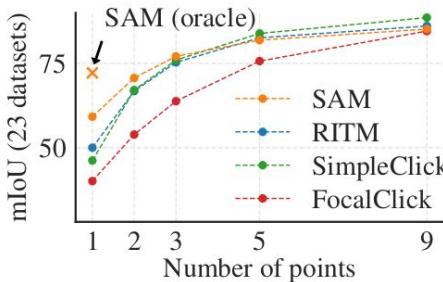


Évaluation du modèle

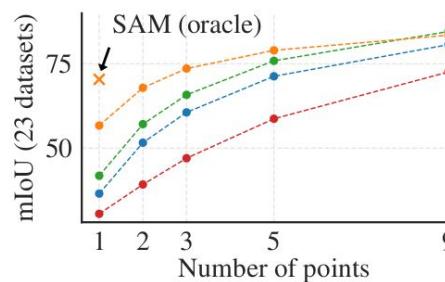
23 datasets différents → différents domaines



(b) Mask quality ratings by human annotators



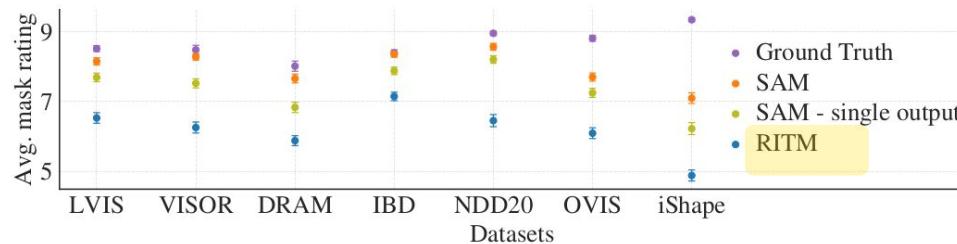
(c) Center points (default)



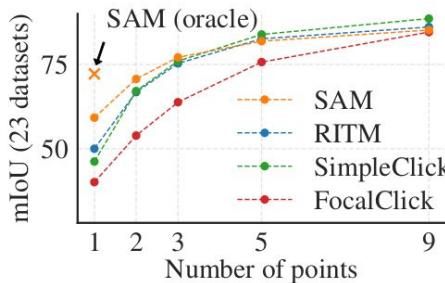
(d) Random points

Évaluation du modèle

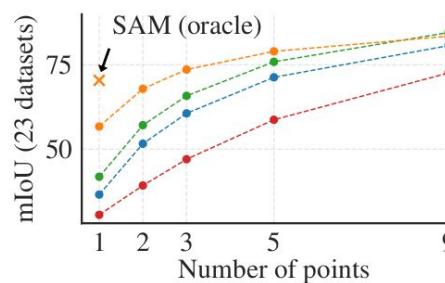
23 datasets différents → différents domaines



(b) Mask quality ratings by human annotators



(c) Center points (default)



(d) Random points

4

Notre expérience

Expérience

4

- Dataset
- Baseline
- Prédictions + évaluation

Expérience

Dataset

The image shows two side-by-side screenshots of academic and machine learning platforms.

IEEE Xplore: A screenshot of the IEEE Xplore digital library interface. It displays a search result for "A Large-Scale Dataset for Fish Segmentation and Classification". The results page includes a navigation bar with "All" selected, a search bar, and a sidebar for "kaggle". Key details shown are the publisher (IEEE), citation count (12 Paper Citations, 1257 Full Text), and authors (Oguzhan Ulucan, Diclehan Karakaya, Mehmet Turkan). Below this is a grid of 32 small images of various fish species.

Kaggle: A screenshot of the Kaggle platform. It shows the same dataset under the title "A Large Scale Fish Dataset". The page includes a search bar, user profile information for "OGUZHAN ULUCAN - UPDATED 2 YEARS AGO" (896 notebooks), a download button ("Download (3 GB)"), and a preview section showing a 4x8 grid of fish images. Below the preview are links for "Data Card", "Code (197)", and "Discussion (5)".

Bottom Right Text: The text "+ masques binaires" is overlaid at the bottom right of the image.

Expérience

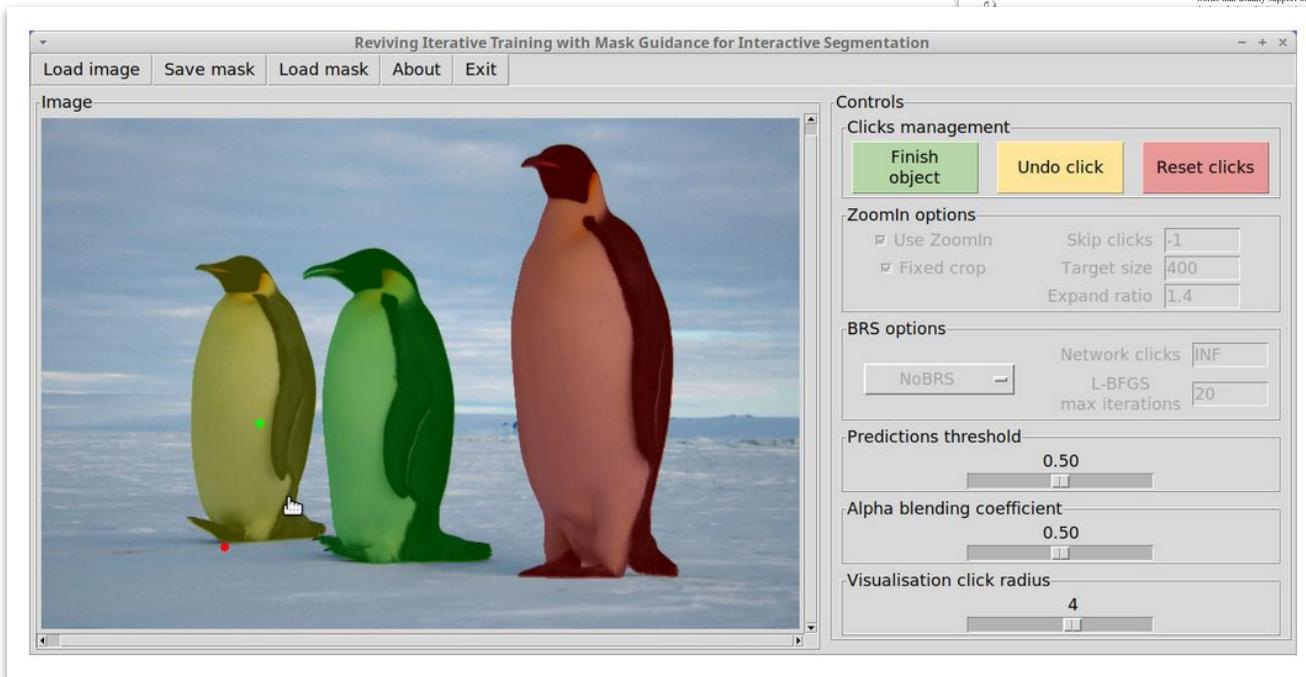


Dataset

Espèce (nom original)	Espèce (nom en français)	Nombre d'images
Hourse Mackerel	Maquereau	50
Black Sea Sprat	<i>Clupeonella cultriventris</i>	50
Striped Red Mullet	Rouget de roche	50
Gilt-Head Bream	Dorade royale	50
Red Mullet	Rouget	50
Sea Bass	Bar commun	50
Shrimp	Crevette	50
Trout	Truite	50
Red Sea Bream	Dorade japonaise	50
		Total: 450

Expérience

Baseline: RITM



Reviving Iterative Training with Mask Guidance for Interactive Segmentation

Konstantin Sofiiuk, Ilia A. Petrov* and Anton Konushin**

Visual Understanding lab., AI Center Moscow, Samsung Electronics Co., Lesnaya 5C, Moscow, Russia

ARTICLE INFO

Keywords: interactive segmentation, segmentation, mask refinement

ABSTRACT

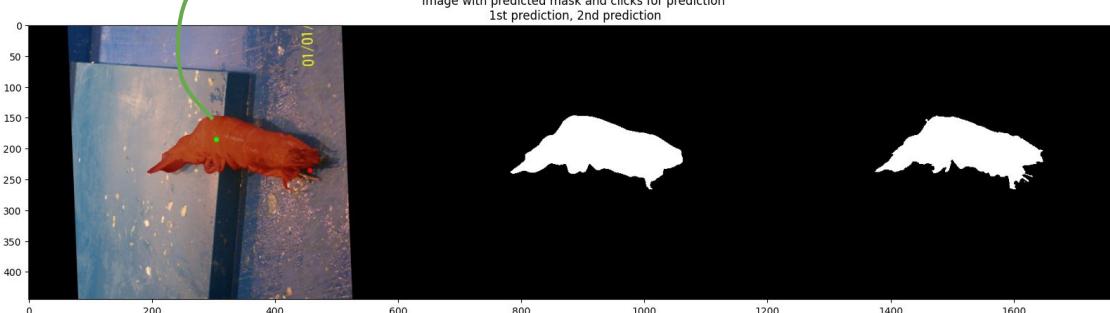
Recent works on click-based interactive segmentation have demonstrated state-of-the-art results by training deep learning models on a few user interactions. These approaches are more computationally expensive compared to feedback-based approaches, as they require performing backward passes through a network during inference and are hard to deploy on mobile frameworks that usually support only forward passes. In this paper, we extensively evaluate various segmentation and discovery that new state-of-the-art results can be achieved by employing a novel segmentation framework that employs the segmentation masks from previous an entirely new object, but also to start with an external dataset greatly impacts the quality of interactive segmentation on a combination of COCO and LVIS with diverse and high-quality images. The code and trained models are available at https://github.com/ailic-vui/ritm_interactive_segmentation.

Expérience

Prédictions: **RITM**

entrée pour SAM

clic !



Expérience

Prédictions: **SAM**

Ambiguous task results: SAM

Mask 1, SAM Score: 0.852



Mask 2, SAM Score: 1.002



Mask 3, SAM Score: 0.818



Expérience

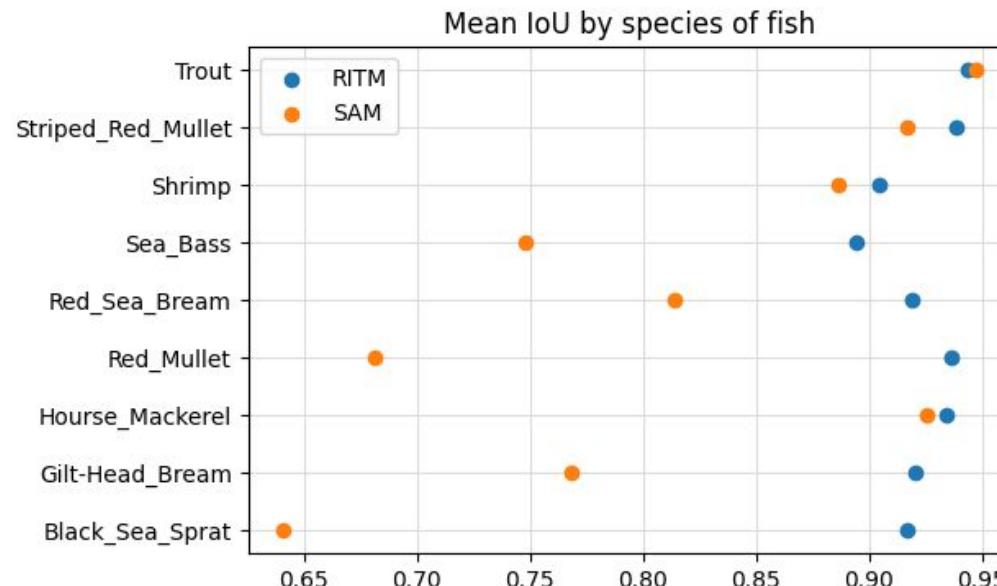
Évaluation: comparaison des IoU

	iou_RITM	iou_SAM
count	450.000000	450.000000
mean	0.922935	0.814156
std	0.043442	0.259960
min	0.526922	0.023829
25%	0.910158	0.823418
50%	0.925036	0.920395
75%	0.943318	0.958745
max	0.992594	0.987818

	iou_RITM	iou_SAM
species		
Black_Sea_Sprat	0.916618	0.640691
Gilt-Head_Bream	0.920119	0.768064
Hourse_Mackerel	0.934028	0.925069
Red_Mullet	0.936137	0.681345
Red_Sea_Bream	0.918904	0.813904
Sea_Bass	0.894340	0.748280
Shrimp	0.904728	0.886251
Striped_Red_Mullet	0.938162	0.916985
Trout	0.943377	0.946816

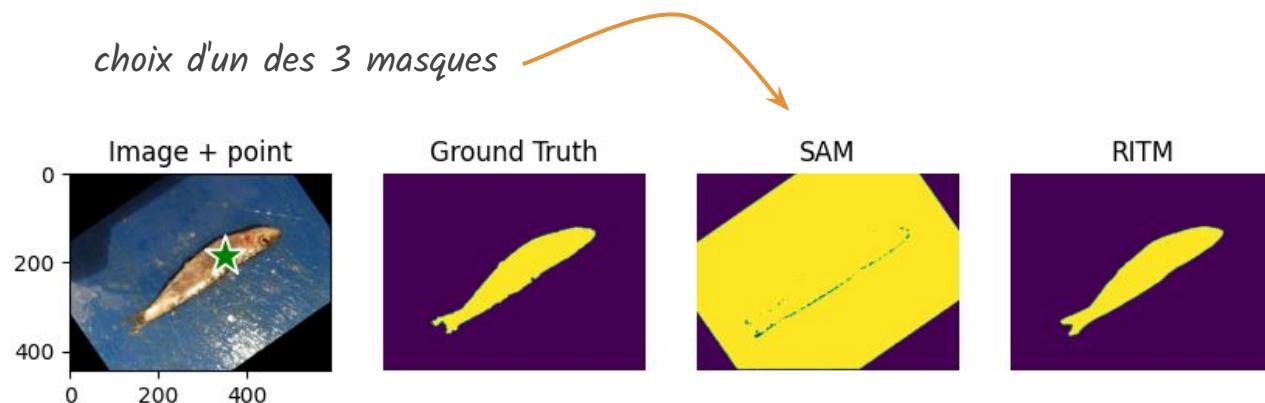
Expérience

Évaluation: comparaison des IoU



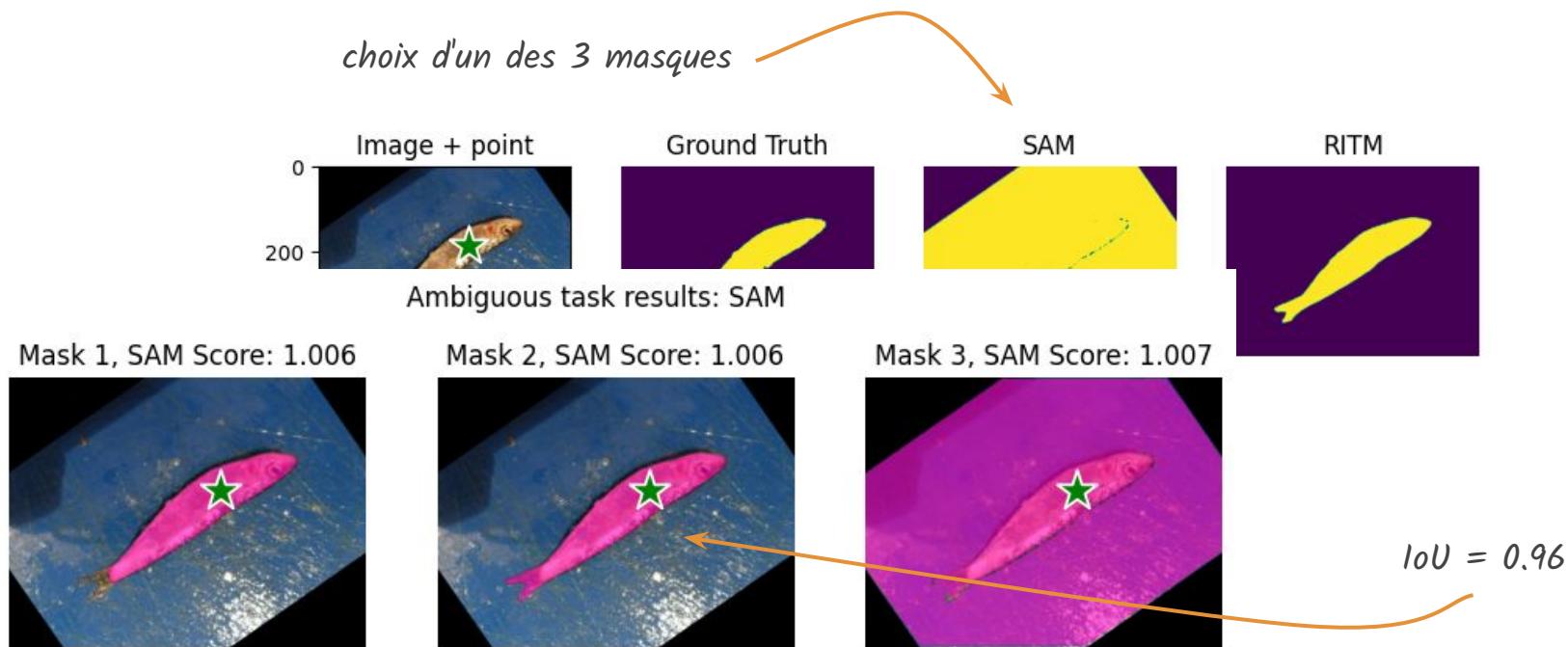
Expérience

Possibles explications du résultat inattendu



Expérience

Possibles explications du résultat inattendu



5

Conclusions

Conclusions

Contrairement aux attentes, dans cette expérience, nous obtenons de meilleurs résultats avec RITM (le modèle de base) qu'avec SAM (le modèle principal). De mon point de vue de non-expert, je pense que les raisons en sont les suivantes :

- Possible manque de qualité des étiquettes de l'ensemble de données : Certains masques de vérité au sol sont de très mauvaise qualité par rapport à l'image originale. Nous ne savons pas comment ces masques ont été générés et malheureusement l'accès à l'article source de ce jeu de données n'est pas disponible gratuitement.
- Comme SAM est entraîné à effectuer des tâches ambiguës, nous n'avons choisi qu'un seul des masques prédits pour notre évaluation. Parfois, avec notre unique clic choisi, le masque choisi a sélectionné un "objet" correspondant à la quasi-totalité de l'image (voir les exemples ci-dessus).

Merci