Segment Anything Is Not Always Perfect: An Investigation of SAM on Different Real-world Applications [1]

Wei Ji[1], Jingjing Li[1], Qi Bi[2], Tingwei Liu[3], Wenbo Li[4], Li Cheng[1]

[1]University of Alberta [2]Wuhan University [3]Dalian University of Technology [4]Samsung Research America

Background

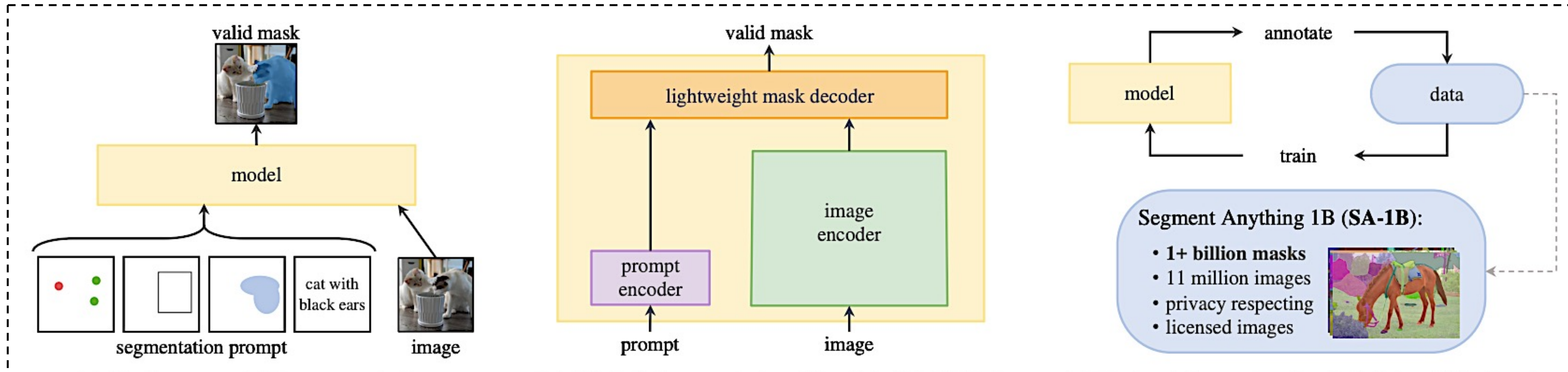

Figure 1. Illustration of segment anything model [2]. Meta AI approaches the vision foundation model for segmentation by introducing three interconnected components: a promptable segmentation task, a segmentation model (SAM) that powers data annotation and enables zero-shot transfer to a range of tasks via prompt engineering, and a data engine for collecting SA-1B, the dataset of over 1 billion masks.

Motivation

Motivation: *How does SAM perform in various real-world applications, e.g., industry and healthcare?*

- We conduct a series of intriguing investigations into the performance of SAM across various applications, particularly in the field of *industrial defect segmentation* and its three closely-related segmentation tasks in *natural images*, including concealed object detection, dichotomous image segmentation, and shadow detection. The concealed, fine-grained, and non-salient characteristics of these tasks offer valuable insights for the development and refinement of industrial segmentation.
- We further analyze and discuss the ***advantages and limitations*** of SAM in these applications.
- Based on these studies, we have made some ***observations and insights*** toward promoting the development of segmentation models in industrial inspection.

Outlook and Opportunity

Outlook: *We provide an outlook on future development of segmentation models in industrial inspection. Several promising potential directions are as follows:*

- Industry-focused SAM & dedicated large-scale dataset.
- Pretraining strategy (tailored for industrial segmentation).
- Multi-modal SAM (e.g., depth or thermal)
- Semi-supervised application (e.g., weak scribble)

Investigation

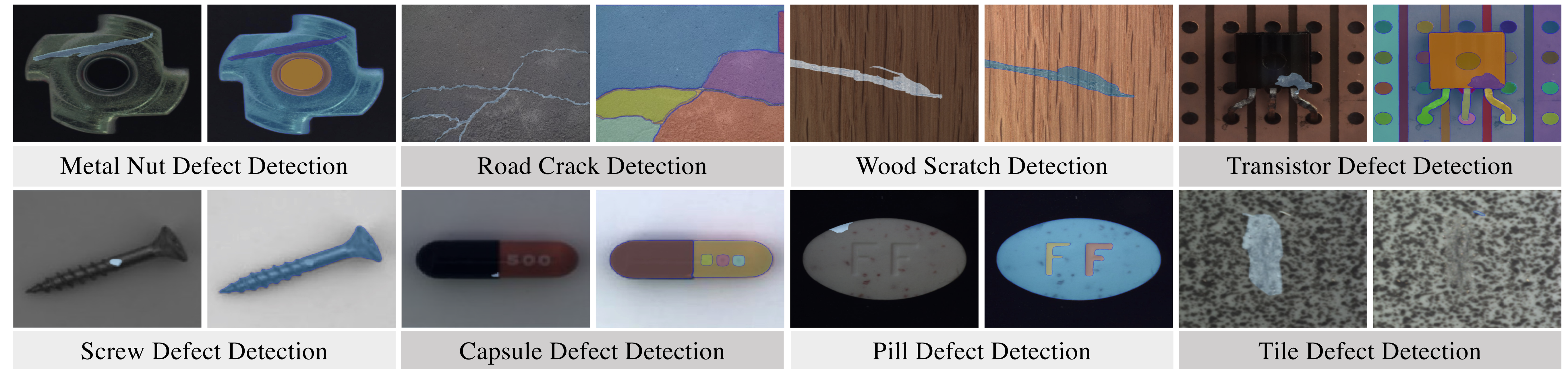

Figure 2. Results of Segment Anything Model (SAM) on various industrial applications, where we adopt Everything mode to obtain SAM segmentations (right). The ground truth is masked with image for reference purpose (left).

(a) Results on CDS2K [5].

Model	Backbone	$\mathcal{M}\downarrow$
SINetV2$_{22}$ [3]	Res2Net50	0.102
HitNet$_{23}$ [7]	PVTv2-B2	0.118
CamF-P$_{23}$ [20]	PVTv2-B4	0.100
DGNet$_{23}$ [9]	EffiNet-B4	**0.089**
SAM$_{23}$	ViT-B	0.372
	Δ*diff*	+28.3%
	ViT-L	0.281
	Δ*diff*	+19.2%
	ViT-H	0.265
	Δ*diff*	+17.6%

(b) Results on COD10K [4].

Model	Backbone	$\mathcal{M}\downarrow$
PreyNet$_{22}$ [23]	ResNet50	0.034
SegMaR$_{22}$ [10]	ResNet50	0.034
PFNet+$_{23}$ [14]	ResNet50	0.037
ZoomNet$_{22}$ [15]	ResNet50	**0.029**
SAM$_{23}$	ViT-B	0.108
	Δ*diff*	+7.9%
	ViT-L	0.065
	Δ*diff*	+3.6%
	ViT-H	0.054
	Δ*diff*	+2.5%

(c) Results on DIS-TE4 [16].

Model	Backbone	$\mathcal{M}\downarrow$
Gate$_{20}$ [24]	ResNet50	0.109
PFNet$_{21}$ [13]	ResNet50	0.107
IS-Net$_{22}$ [16]	U2Net	**0.072**
SAM$_{23}$	ViT-B	0.179
	Δ*diff*	+10.7%
	ViT-L	0.166
	Δ*diff*	+9.4%
	ViT-H	0.166
	Δ*diff*	+9.4%

(d) Results on SBU [18].

Model	Backbone	$\mathcal{M}\downarrow$
DSC$_{18}$ [8]	VGG-16	0.032
DSDNet$_{19}$ [25]	ResNext	0.036
MTMT$_{20}$ [2]	ResNext	**0.029**
SAM$_{23}$	ViT-B	0.203
	Δ*diff*	+17.4%
	ViT-L	0.187
	Δ*diff*	+15.8%
	ViT-H	0.183
	Δ*diff*	+15.4%

Table 1. Quantitative results of SAM on applications of (a) industrial defect detection, (b) concealed object detection, (c) dichotomous image segmentation, and (d) shadow detection. "Everything" mode is used here. M represents mean absolute error (the lower the better). Δ shows the performance gaps between SAMs and the best performing state-of-the-art models.

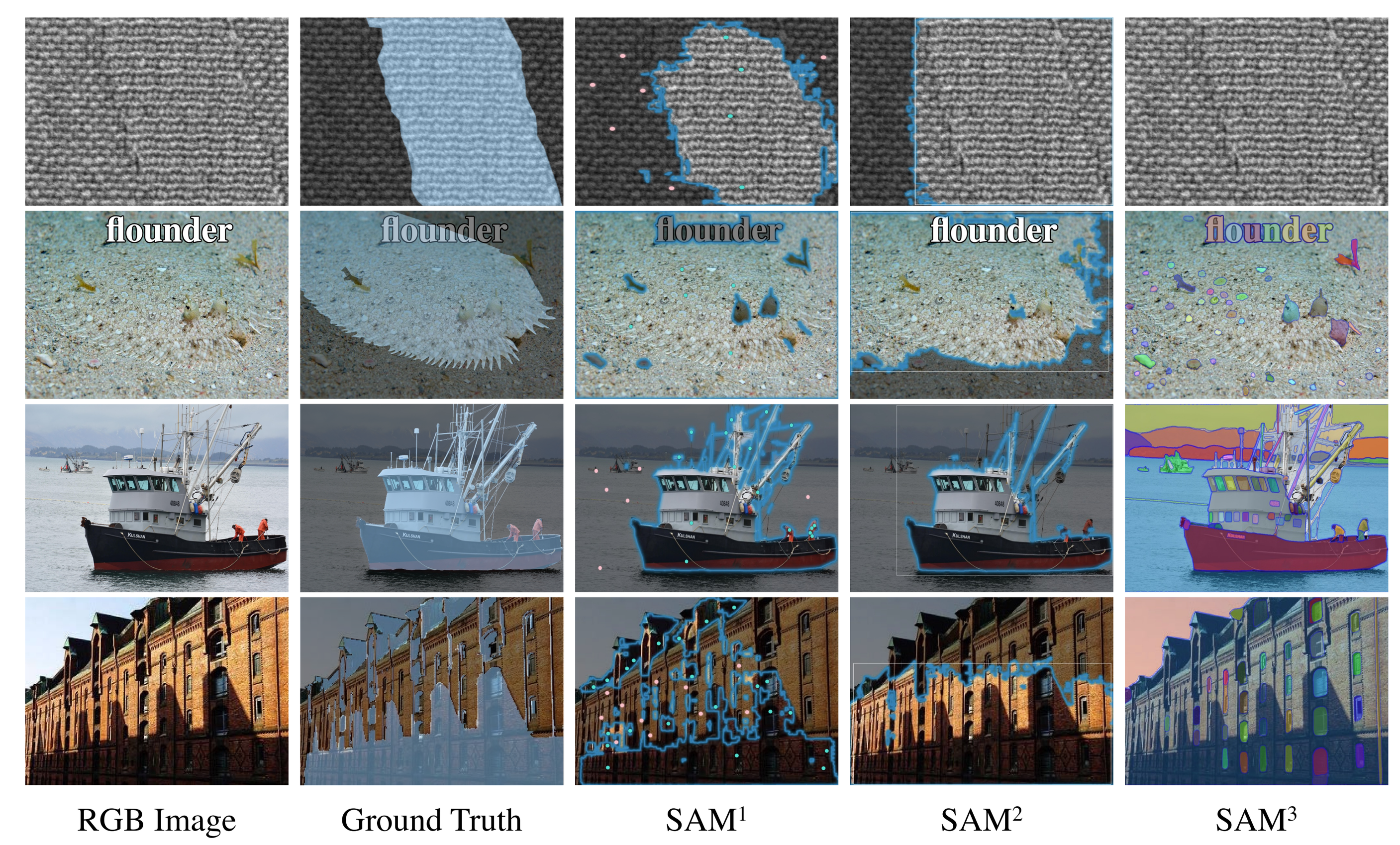

Figure 3. Qualitative results of SAM on applications of (a) industrial defect detection, (b) concealed object detection, (c) dichotomous image segmentation, and (d) shadow detection. SAM$^{1/2/3}$ mean using Click, Box, and Everything modes respectively.

Discussion

Discussion: *We discuss the advantages and limitations of SAM in practice.*

- Excellent generalization on common scenes.
- Require strong prior knowledge.
- Less effective in low-contrast applications.
- Limited understanding of professional data.
- Smaller and fine-grained objects can pose challenges for SAM.

[1] Wei Ji, Jingjing Li, Qi Bi, Tingwei Liu, Wenbo Li, and Li Cheng. "Segment anything is not always perfect: An investigation of sam on different real-world applications." *arXiv preprint arXiv:2304.05750* (2023).
[2] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao et al. "Segment anything." *arXiv preprint arXiv:2304.02643* (2023).