Conferência Nacional dos Agentes Produtores e Usuários de Dados - CONFEST/CONFEGE

SALVADOR, BAHIA - 3 A 5 DE DEZEMBRO DE 2025

Estratégias de segmentação de imagens para áreas urbanas

Thales Sehn Körting
thales.korting@inpe.br

INPE

Geografia e Cartografia Censitárias do IBGE: avanços conceituais e geotecnológicos e seus impactos nas políticas públicas

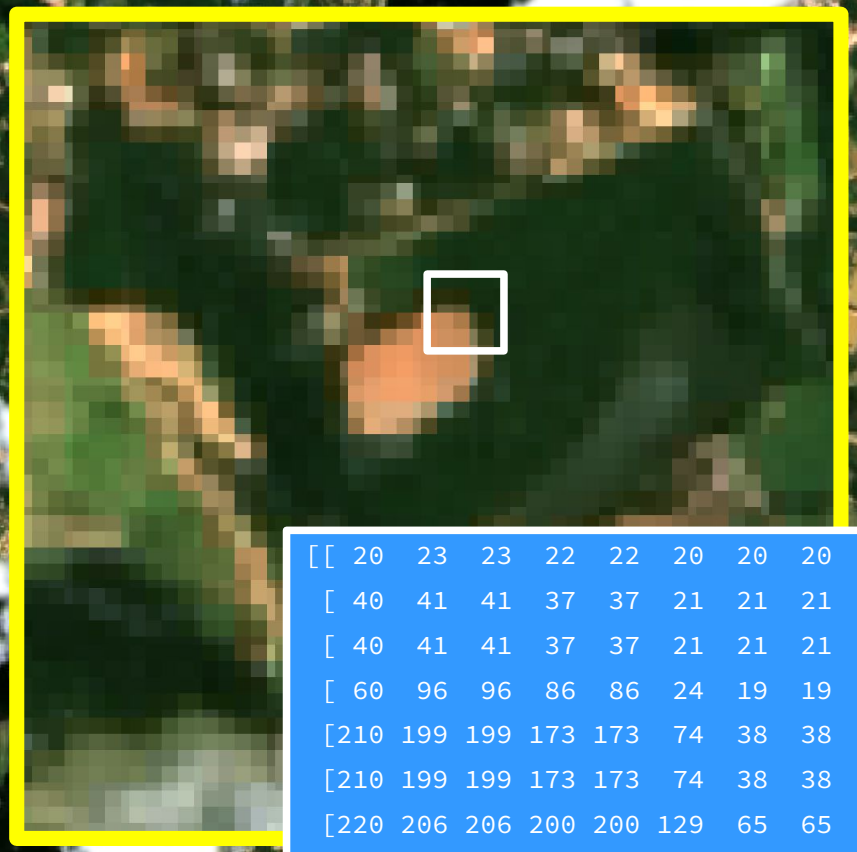
Apoio

SENAI CIMATEC UNIVERSIDADE · Sistema FIEB SESI/SENAI/IEL/CIEB · GOVERNO DO ESTADO BAHIA GOVERNO PRESENTE TRABALHA PRA GENTE

Realização

IBGE Instituto Brasileiro de Geografia e Estatística · MINISTÉRIO DO PLANEJAMENTO E ORÇAMENTO · GOVERNO DO BRASIL DO LADO DO POVO BRASILEIRO

Como construir algoritmos para processar imagens?

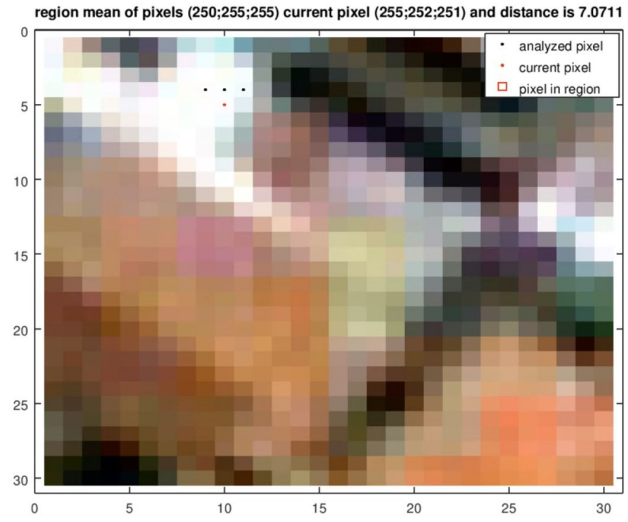


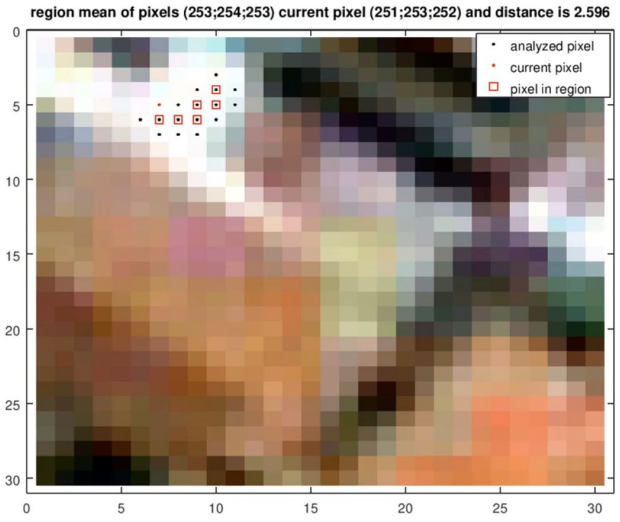
```
[[ 54  49  49  48  48  48  48  48  51  51]
 [ 45  46  46  43  43  45  45  45  44  44]
 [ 45  46  46  43  43  45  45  45  44  44]
 [ 61  90  90  82  82  48  42  42  47  47]
 [138 139 139 122 122  78  47  47  46  46]
 [138 139 139 122 122  78  47  47  46  46]
 [144 144 144 140 140 124  70  70  48  48]
 [144 144 144 140 140 124  70  70  48  48]
 [146 144 144 142 142 145  80  80  51  51]
 [146 144 144 142 142 145  80  80  51  51]]
```

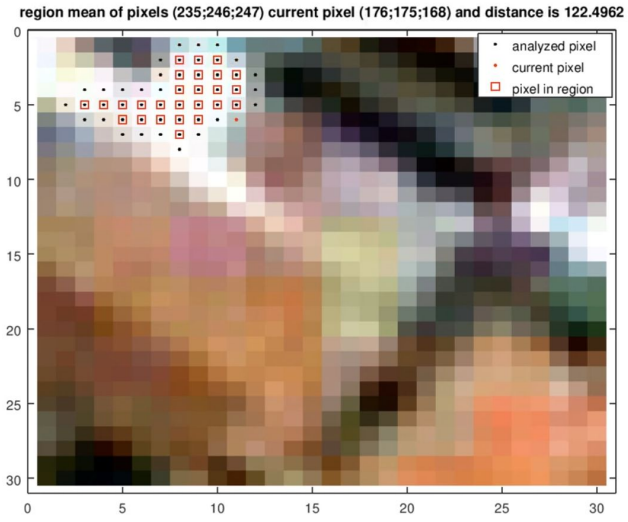
```
[[ 20  23  23  22  22  20  20  20  21  21]
 [ 40  41  41  37  37  21  21  21  18  18]
 [ 40  41  41  37  37  21  21  21  18  18]
 [ 60  96  96  86  86  24  19  19  21  21]
 [210 199 199 173 173  74  38  38  22  22]
 [210 199 199 173 173  74  38  38  22  22]
 [220 206 206 200 200 129  65  65  24  24]
 [220 206 206 200 200 129  65  65  24  24]
 [220 211 211 207 207 167  84  84  25  25]
 [220 211 211 207 207 167  84  84  25  25]]
```

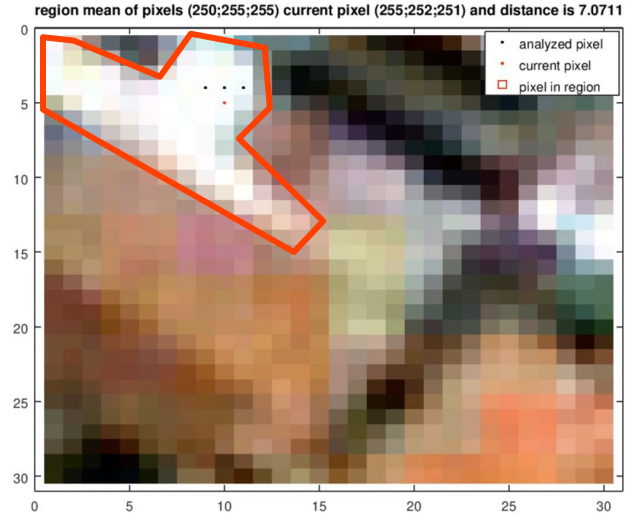
```
[[ 22  20  20  19  19  19  19  19  20  20]
 [  9  11  11   7   7  16  16  16  17  17]
 [  9  11  11   7   7  16  16  16  17  17]
 [ 25  54  54  46  46  20  14  14  20  20]
 [ 86  90  90  74  74  41  10  10  22  22]
 [ 86  90  90  74  74  41  10  10  22  22]
 [ 92  94  94  90  90  86  33  33  25  25]
 [ 92  94  94  90  90  86  33  33  25  25]
 [ 93  93  93  91  91 108  43  43  22  22]
 [ 93  93  93  91  91 108  43  43  22  22]]
```

A segmentação de imagens por crescimento de regiões

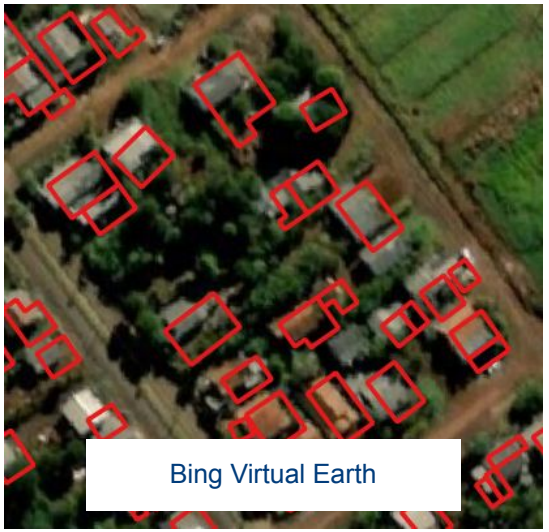


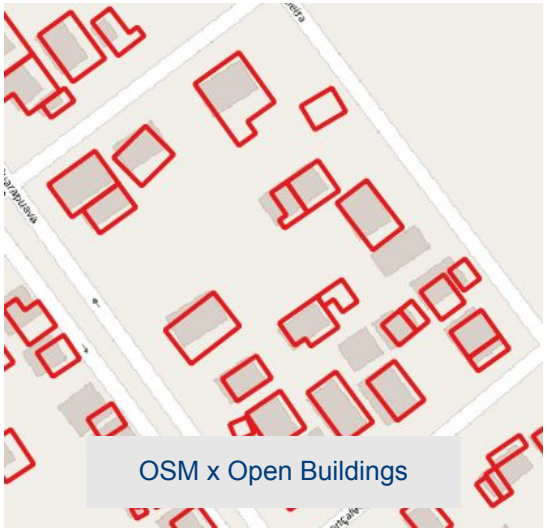




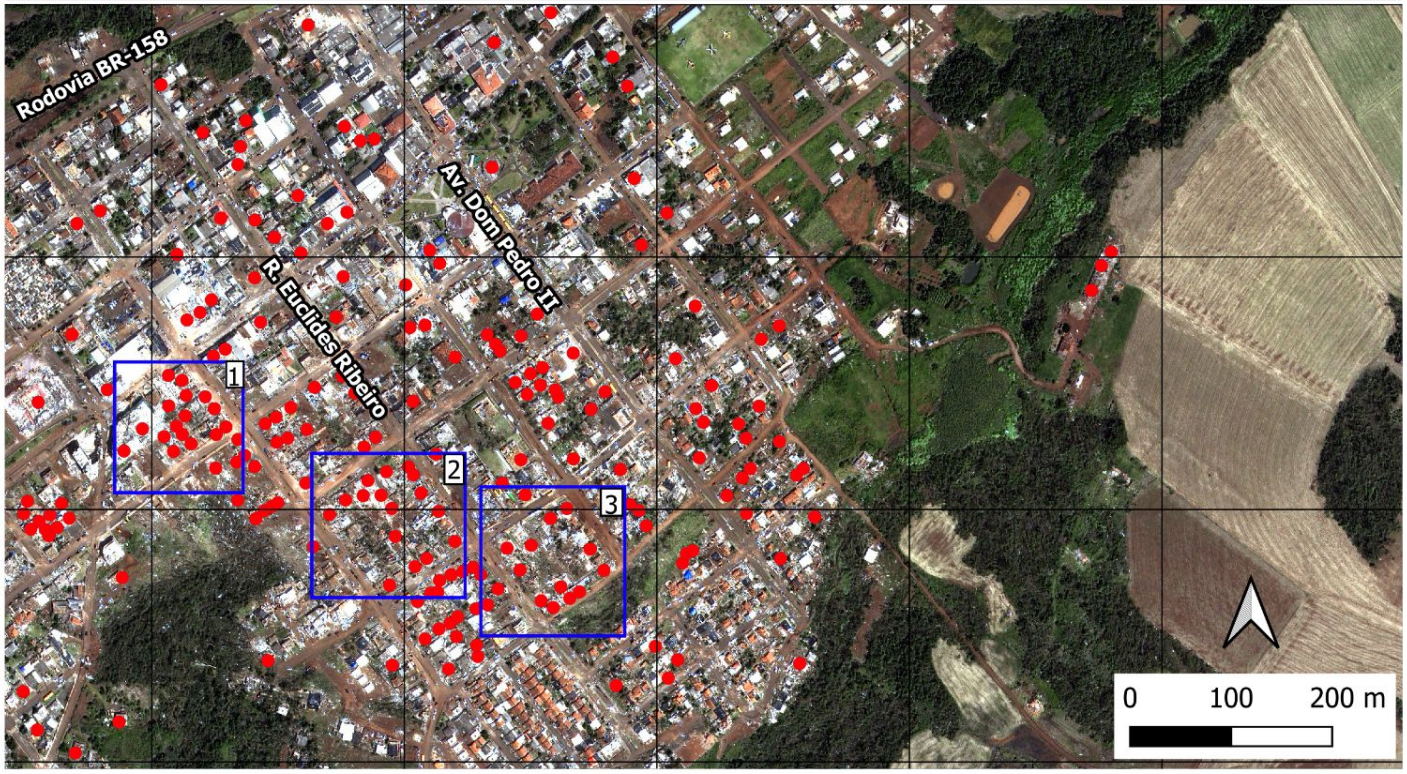


Desafios: áreas urbanas no Brasil Vs bases de dados existentes

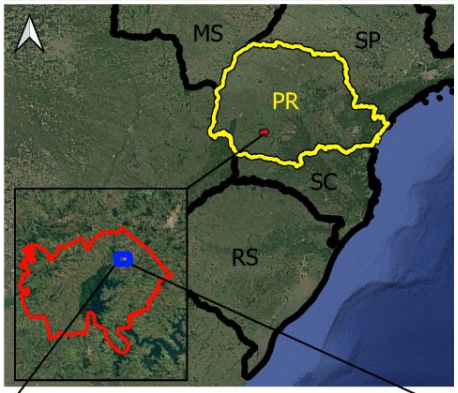




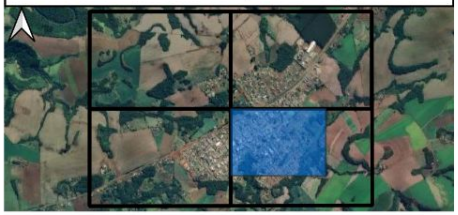
Edificações atingidas pelo tornado em Rio Bonito do Iguaçu, PR - Brasil



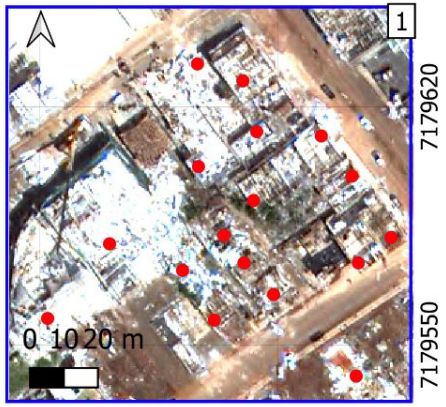
Localização

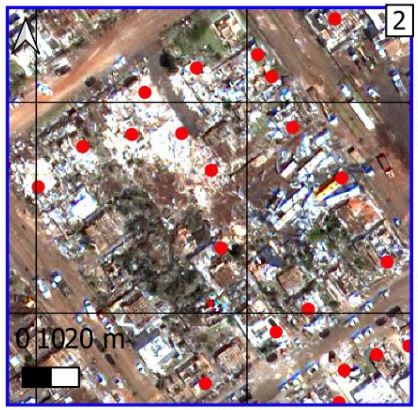


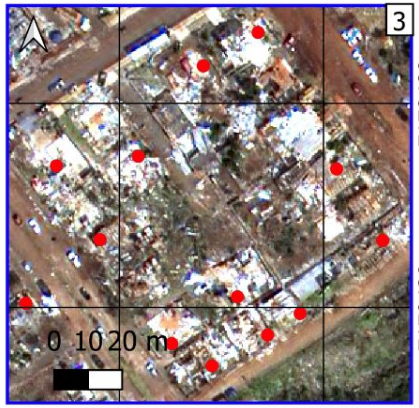
Área do município mapeada



Legenda
- Edificações atingidas
- Área mapeada
- Área total de estudo
- Município Rio Bonito do Iguaçú
- Paraná - PR







International Charter Space and Major Disasters
Activation - 1002 (Call-1146) - Tornado no Brasil

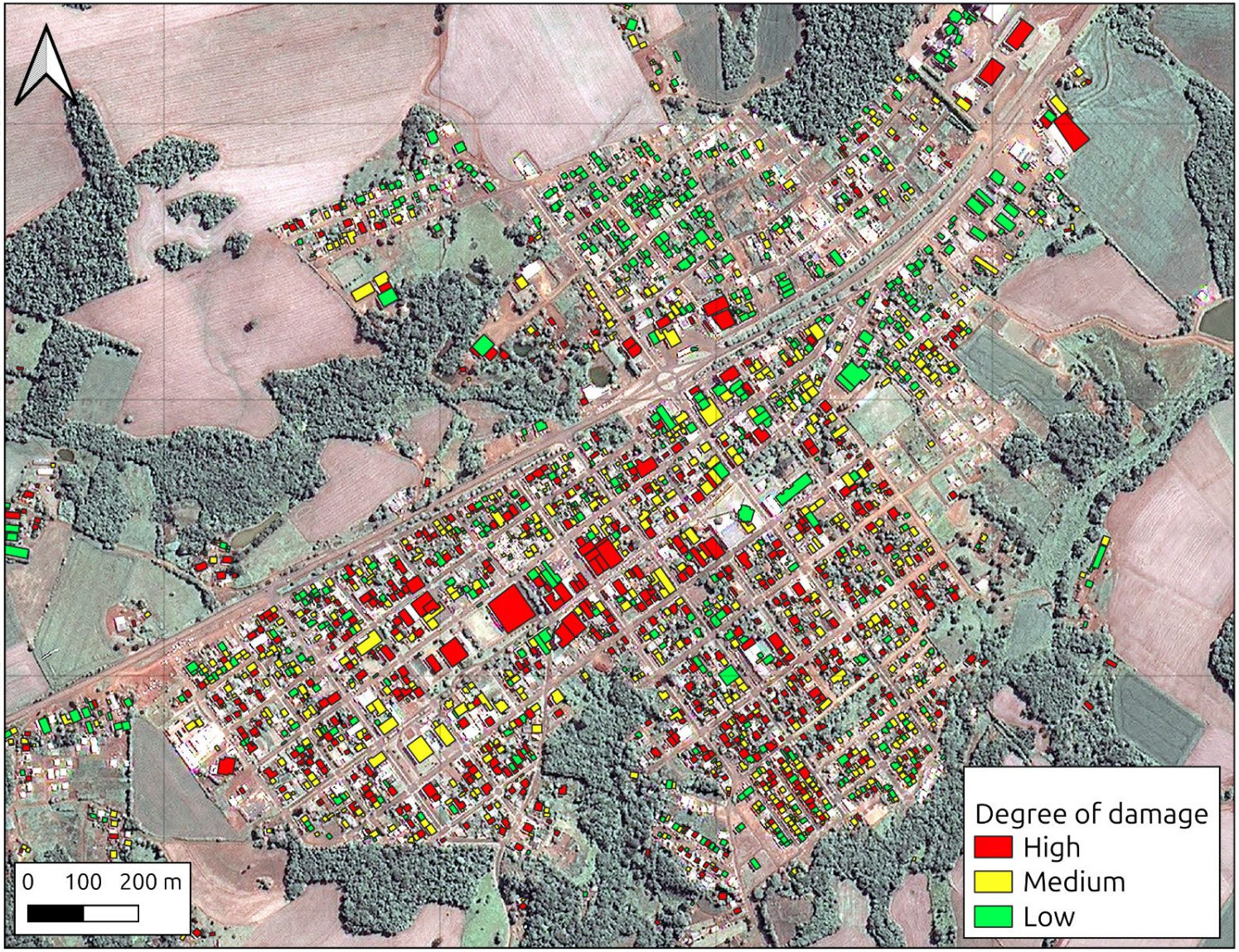
Data do evento: 07/11/2025, Data de criação do mapa: 12/11/2025, v1.

Imagem pós-evento: 9 de novembro de 2025
WorldView Legion, ©2025 Vantor, USGS/EROS SRC: 32722 WGS 84/ ZONE 22 S

As informações contidas neste produto não possuem validação de campo.

INPE · Vantor · CENAD · DEFESA CIVIL BRASIL

Estimated degree of damage to buildings hit by tornado in Rio Bonito do Iguaçu - Brazil



Degree of damage to buildings calculated from USGS/WorldView images (July 26, 2025) and CNES/Pleiades-1B images (November 11, 2025)

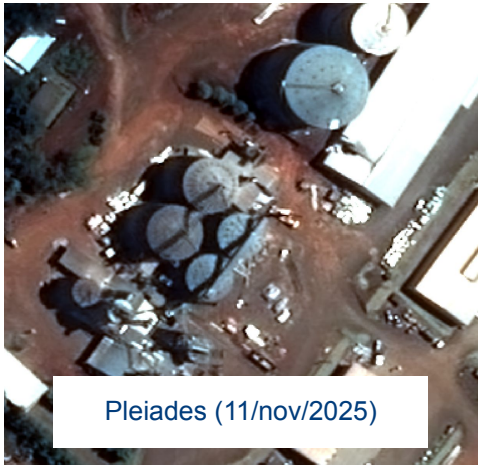
Building polygons obtained from the OpenStreetMap database

The information contained in this product has not been field validated

International Charter Space and Major Disasters Activation-1002 (Call 1146) - Tornado in Brazil	True color CBERS 4A/WPM (26.03.2025)	EPSG: 32722 WGS/UTM ZONA 22S

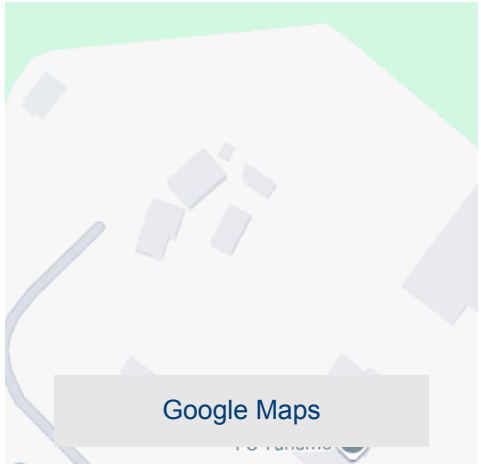
Diferentes perspectivas para o mesmo local
Todas imagens estão "registradas"


WorldView-2 (09/nov/2025)

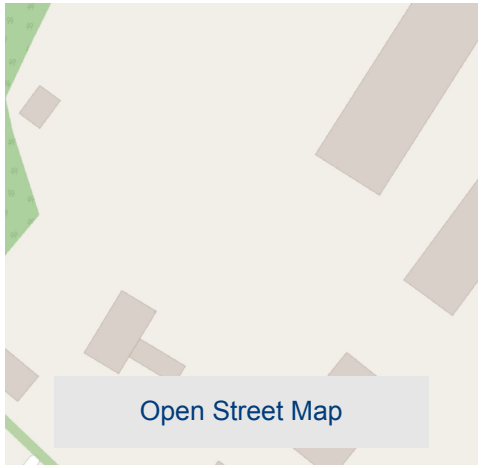

Pleiades (11/nov/2025)


WorldView-2 (26/jul/2025)


CBERS-4A (26/mar/2025)


Google Maps


Google Satellite


Open Street Map


Bing Virtual Earth

Tendência: Uso de IA para segmentação Vs o princípio FAIR nas Bases de dados para treinamento

- **F**indable: The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers.
- **A**ccessible: Once the required data is found, users need to know how can they be accessed, possibly including authentication and authorisation.
- **I**nteroperable: The data usually need to be integrated with other data, and interoperate with applications for analysis, storage, and processing.
- **R**eusable: The ultimate goal of FAIR is to optimise the reuse of data, therefore metadata and data should be well-described to be replicated/combined in different settings. Also, according to [43], data must have clear usage licenses and provide accurate information on provenance.

#bands	Top-3 OAs (%)
1 (grayscale)	99.77 [7], 99.73 [5], 99.65 [6]
3 (RGB)	99.76 [70], 99.48 [56] 99.47 [42]
3 (RGB)	99.75 [57], 99.30 [65], 98.88 [2]
3 (RGB)	99.75 [70], 97.83 [69], 97.50 [36]
3 (RGB)	99.70 [49], 95.00* [61], 69.17* [68]
4 (RGB and NIR)	99.56 [20], 99.50 [66], 99.46 [15]
3 (RGB)	99.00 [22], 98.90 [52], 98.50 [8]
3 (RG and NIR)	99.0 [56], 96.80 [30], 94.45 [42]
13	98.57 [20], 95.74 [44], 93.20 [47]
3 (RGB)	96.37 [70], 95.95 [36], 94.95 [62]
3 (RGB)	95.40 [65], 94.00 [2], 93.90 [34]
3 (RGB)	88.40 [59], 87.80 [27], 86.40 [53]
3 (RGB)	74.07* [1], 62.38* [54], 62.20* [14]
3 (RGB)	59.63 [21], 58.00 [39], 56.51 [38]
3 (RGB)	53.30 [37], 51.00 [52], 50.70 [63]

Table 2. Example of datasets used by the Deep Learning community. Column *FAIR?* indicates if the dataset complies to the FAIR principles. RGB: Red, Green, Blue. NIR: Near Infra-Red. OAs: Overall Accuracies. The datasets are ordered by the first OA. OAs marked with * were reported in the original reference as *mean Average Precision* (mAP).

Bases de dados para treinamento de algoritmos de IA

UC Merced (University of California, Merced)

Land use classification dataset with 21 classes and 100 RGB TIFF images for each class. Each image measures 256x256 pixels

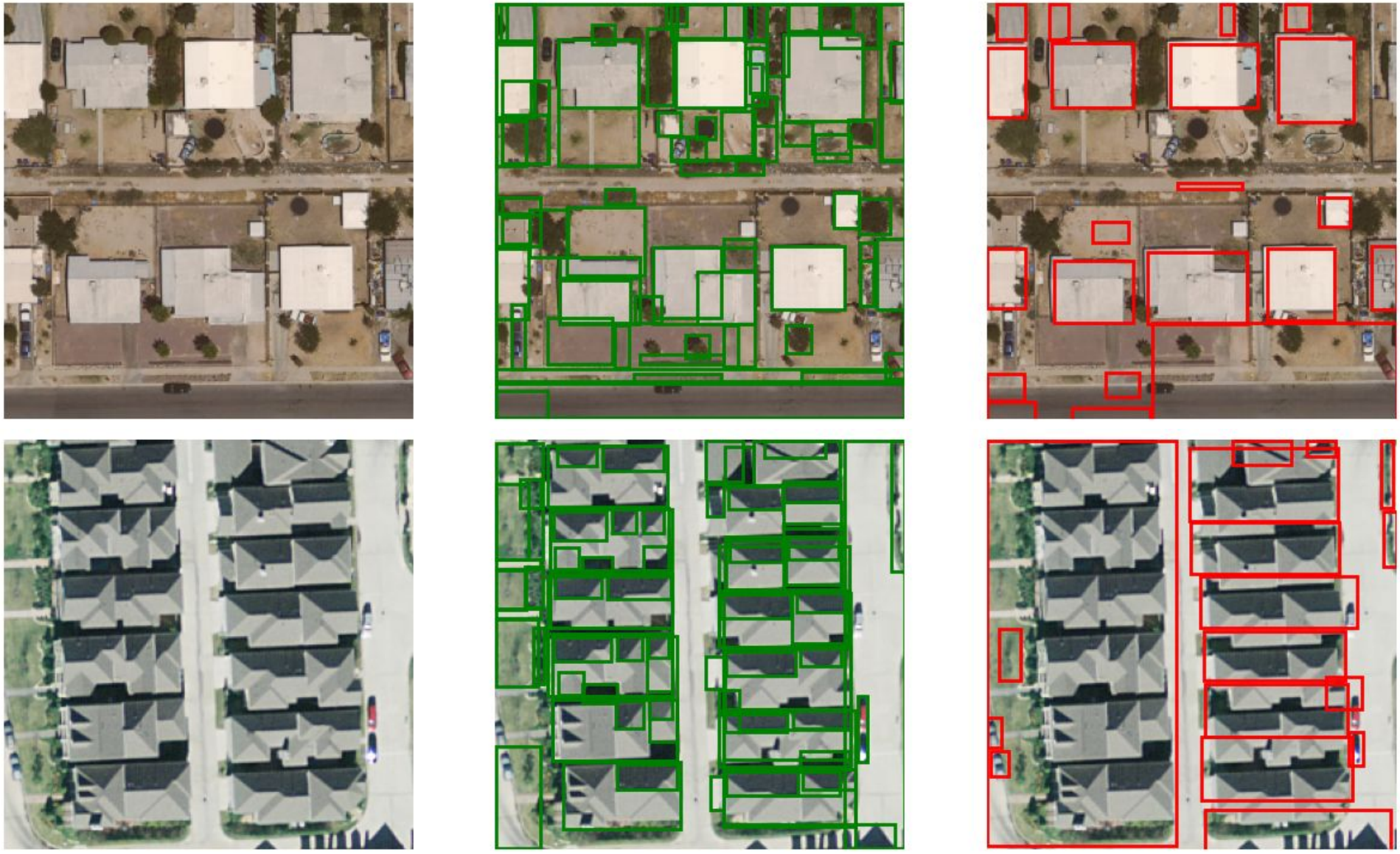
AIRS (Aerial Imagery for Roof Segmentation)

Public dataset for roof segmentation from very-high-resolution aerial imagery (7.5cm). Covers almost the full area of Christchurch, the largest city in the South Island of New Zealand.

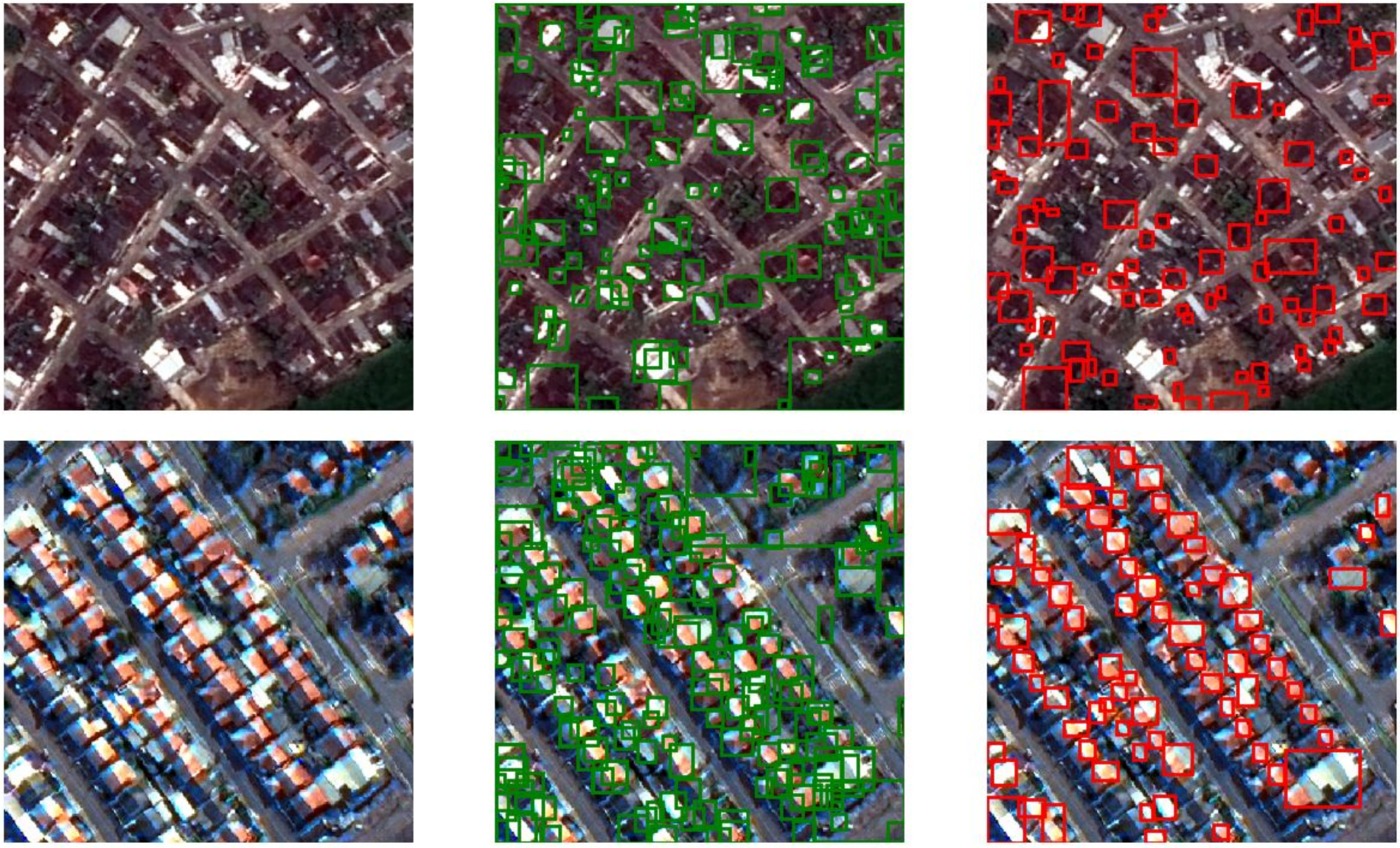
RID – Roof Information Dataset

Computer Vision-Based Photovoltaic Potential Assessment, based on Google screen shots from the German village of Wartenberg

Testes de segmentação de imagens para áreas urbanas (bounding box)



Testes de segmentação de imagens para áreas urbanas (bounding box)



Método de segmentação 1

Segment anything
A Kirillov, E Mintun, N Ravi, H Mao, C
... **Segment Anything** 1B (SA-1B): Fi
... components: a promptable segmen
☆ Salvar 99 Citar Citado por 363

Segment anything
A Kirillov, E Mintun, N Ravi, H Mao... - Proceedings of the ..., 2023 - openaccess.thecvf.com
We i **nt Anything** (SA) project: a new task, model, and dataset for image segmentation. Using efficient model in a data collection loop, we built the largest ...
☆ Salvar 99 Citar Citado por 15716 Artigos relacionados Todas as 10 versões »

2023 para 2025

Foundation model

- perform zero-shot and few-shot learning
- pre-trained on broad data
- adaptable to a wide range of tasks

Segment Anything

Alexander Kirillov[1,2,4] Eric Mintun[2] Nikhila Ravi[1,2] Hanzi Mao[2] Chloe Rolland[3] Laura Gustafson[3]
Tete Xiao[3] Spencer Whitehead Alexander C. Berg Wan-Yen Lo Piotr Dollár[4] Ross Girshick[4]
[1]project lead [2]joint first author [3]equal contribution [4]directional lead
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.

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*** *masks on 11M licensed and privacy respecting images. The model is designed and trained to be promptable, so it can transfer zero-shot to new image distributions and tasks. We evaluate its capabilities on numerous tasks and find that its zero-shot performance is impressive – often competitive with or even superior to prior fully supervised results. We are releasing the Segment Anything Model (SAM) and corresponding dataset (SA-1B) of 1B masks and 11M images at https://segment-anything.com to foster research into foundation models for computer vision.*

1. Introduction

Large language models pre-trained on web-scale datasets are revolutionizing NLP with strong zero-shot and few-shot generalization [10]. These "foundation models" [8] can generalize to tasks and data distributions beyond those seen during training. This capability is often implemented with *prompt engineering* in which hand-crafted text is used to prompt the language model to generate a valid textual response for the task at hand. When scaled and trained with abundant text corpora from the web, these models' zero and few-shot performance compares surprisingly well to (even 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 aligns paired text and images from the web. For example, CLIP [82] and ALIGN [55] use contrastive learning to train text and image encoders that align the two modalities. Once trained, engineered text prompts enable zero-shot generalization to novel visual concepts and data distributions. Such encoders also compose effectively with other modules to enable downstream tasks, such as image generation (*e.g.*, DALL·E [83]). While much progress has been made on vision and language encoders, computer vision includes a wide range of problems beyond this scope, and for many of these, abundant training data does not exist.

In this work, our goal is to build *a foundation model for image segmentation*. That is, we seek to develop a promptable model and pre-train it on a broad dataset using a task that enables powerful generalization. With this model, we aim to solve a range of downstream segmentation problems on new data distributions using prompt engineering.

The success of this plan hinges on three components: **task**, **model**, and **data**. To develop them, we address the following questions about image segmentation:

1. What **task** will enable zero-shot generalization?
2. What is the corresponding **model** architecture?
3. What **data** can power this task and model?

Segment Anything (SA)

11M licensed and privacy respecting images

- Professional annotators labeled masks by clicking foreground/background object points using a browser based interactive segmentation tool
- Masks were refined using pixel-precise *brush* and *eraser* tools

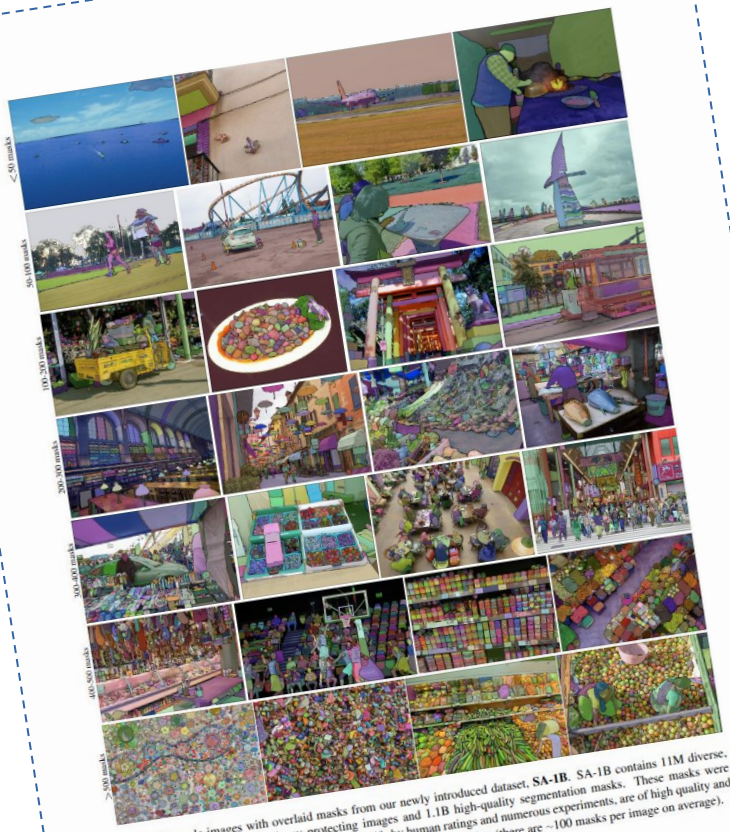


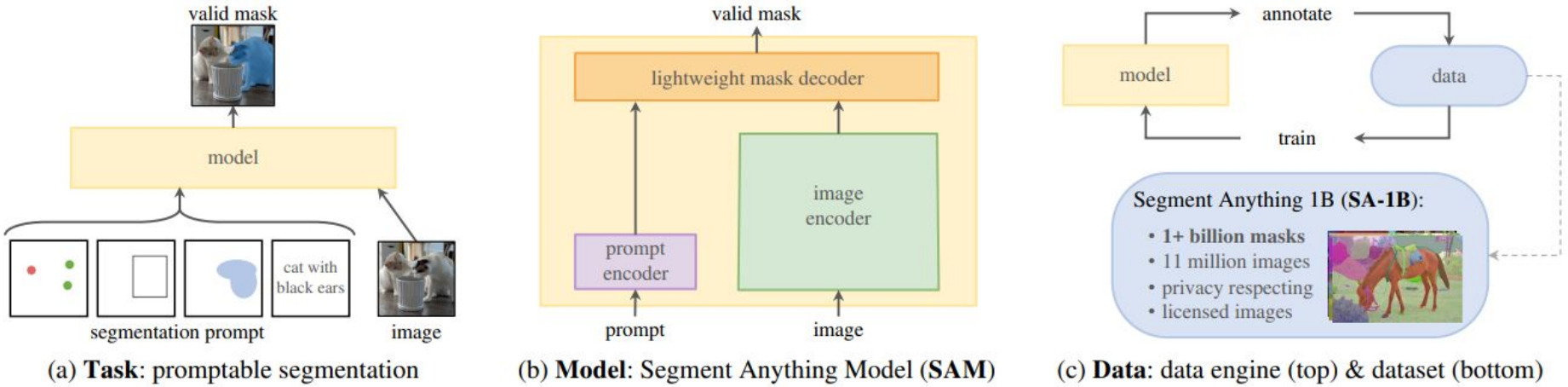
Figure 2: Example images with overlaid masks from our newly introduced dataset, **SA-1B**. SA-1B contains 11M diverse, high-resolution, licensed, and privacy protecting images and 1.1B high-quality segmentation masks. These masks were annotated *fully automatically* by SAM, and as we verify by human ratings and numerous experiments, are of high quality and diversity. We group images by number of masks per image for visualization (there are ~100 masks per image on average).

3

Segment Anything (SA)

Dataset with 1 billion masks (SA-1B)

- *Assisted-manual (up to 34 seconds/mask)*
  - High resolution images (3300x4950)
  - downsampled to ~1500x2250
  - 4.3 million masks from 120k images (1B ?)
- *Semi-automatic*
- *Fully automatic*
  - *SA-1B only includes automatically generated masks*



(a) **Task**: promptable segmentation (b) **Model**: Segment Anything Model (**SAM**) (c) **Data**: data engine (top) & dataset (bottom)

Promptable segmentation returns a valid segmentation mask given any *prompt*

A *prompt* simply specifies a starting point for the segmentation

- Point
- Box
- Mask
- *Free-form text*

Método de Segmentação 2

A CLUSTERING TECHNIQUE FOR SUMMARIZING MULTIVARIATE DATA

By Geoffrey H. Ball and David J. Hall

Stanford Research Institute, Menlo Park, California

Scientific measurements frequently involve large numbers of variables whose complex interactions are not easily found. A practical computing method termed ISODATA, which finds the cluster structure of such data, is described. The resulting description of the data provides a fit to the data of a set of cluster centers that tends to minimize the sum of the squared distances of each data point from its closest cluster center. An application to the grouping or clustering of the answers of 209 people to an 80-question sociological survey illustrates the utility of the method.

The work reported here grew out of efforts to develop adaptive preprocessing for pattern-recognition machines (Ball and Hall, 1964). It became apparent, however, that the techniques developed were not limited to this task. In fact, they seem applicable to a recurring basic problem in many disciplines —that of analyzing large numbers of samples of multivariate data.

This technique, termed ISODATA, groups patterns in a way determined primarily by the data itself. (ISODATA represents Iterative Self-Organizing Data Analysis Technique (A).[1]

This technique automatically clusters all of the data into distinct and independent groups. An average response pattern is used to represent a group of patterns, and the iterative process creates new average response patterns to improve the accuracy of trial or existing average response patterns. The process also combines average response patterns that are so similar that their being separate fails to provide a significant amount of information about the structure of the response patterns. Computation time for this technique grows *linearly* with the number of response patterns and the number of dimensions and the number of clusters. An ALGOL program implementing this procedure on a Burroughs B-5500 digital computer has been applied to two actual research problems as well as to artificial problems (Ball, Brain, Burch, and Hall, 1964). (Run times were typically 15 to 20 minutes.) In outline form, the essential steps of the technique are:

(1) Select a "typical" set of response patterns to be used as initial "cluster points."

(2) Group together patterns that lie closest in Euclidian distance to the same cluster point. (The patterns and these cluster points are both vectors.)

(3) Split into two points the average response pattern of each group found in Step (2) if the group's associated "within-group variability" exceeds a threshold, θ_s, which is set by the operator. The group is split by forming two new cluster points from the average response pattern. The new cluster points are identical to the average response pattern, except for that response to a single item, having

[1] The application of ISODATA to sociological data was aided by contracts from the Behavioral Sciences Division of the Air Force Office of Scientific Research, Washington, D.C. The development of the technique was aided by the Graphical Data Transducer Branch, Data Division, Communications Department, USAEL, Fort Monmouth, New Jersey, and by internal funding from Stanford Research Institute. The questionnaires described in this paper were designed and administered by the Systems Analysis Laboratory of the Management Sciences Division of Stanford Research Institute, under the direction of Dr. Howard Vollmer. We would like to thank him for providing the data and for his assistance in interpreting the results we obtained. (His interpretation appears in Vollmer, 1964.) Miss Wanda Fedurek and Mrs. Klara Miracle rendered valuable assistance in plotting and tabulating data, for which we thank them.

A clustering technique for summarizing multivariate data

GH **Ball**, DJ Hall - **Behavioral science**, 1967 - Wiley Online Library

... This can be compared to the word profile of Group 6, where 20 percent are under 30 years of age; have a moderate amount of education (33 percent MS, **67** percent BS) ; are civilians; ...

☆ Salvar 99 Citar Citado por 1356 Artigos relacionados Todas as 6 versões

Referências

Datasets vetoriais

- Open Buildings - Google Research
- OpenStreetMap

Rio Bonito do Iguaçu

- Tornado in Brazil - Charter Activations

Datasets para IA

- Datasets for deep learning with satellite & aerial imagery
- http://weegee.vision.ucmerced.edu/datasets/landuse.html
- https://github.com/MasterSkepticista/Rooftop-Instance-Segmentation
- https://github.com/TUMFTM/RID

Métodos de IA

- Ball, Geoffrey H., and David J. Hall. "A clustering technique for summarizing multivariate data." Behavioral science 12.2 (1967): 153-155. https://doi.org/10.1002/bs.3830120210
- Kirillov, Alexander, et al. "Segment anything." Proceedings of the IEEE/CVF international conference on computer vision. 2023. https://arxiv.org/pdf/2304.02643

Obrigad@!

Agradecimentos:

- FAPESP #2023/09118-6
- CNPq #302205/2023-3

Para mais informações, acesse o QRcode

Apoio

Realização