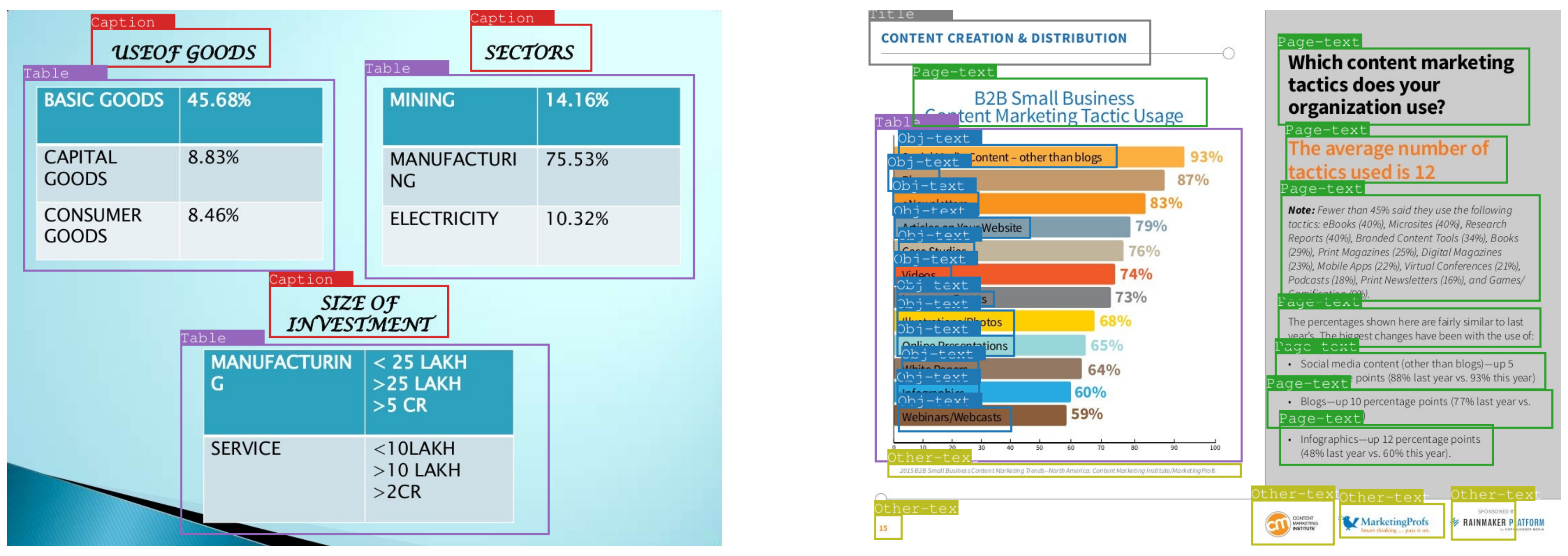
SlideCraft: Synthetic Slides Generation for Robust Slide Analysis

Travis Seng (IRIT/IPAL), Axel Carlier (IRIT/IPAL), Thomas Forgione (Polymny Studio)
Vincent Charvillat (IRIT), Wei Tsang Ooi (NUS)

Context

1. The increasing use of slide presentations in various sectors highlights the need for effective slide layout and semantic analysis.
2. Existing slide datasets (SlideVQA [4], FitVid [2]) often suffer from inconsistencies, mislabels, and incomplete annotations.
3. These issues result in suboptimal training, leading to poorer performance of deep learning models.
4. We introduce SlideCraft: a tool for generating synthetic, accurately annotated slide datasets.


Caption
USEOF GOODS
Table
BASIC GOODS 45.68%
CAPITAL GOODS 8.83%
CONSUMER GOODS 8.46%
Caption
SECTORS
Table
MINING 14.16%
MANUFACTURING 75.53%
ELECTRICITY 10.32%
Caption
SIZE OF INVESTMENT
Table
MANUFACTURING < 25 LAKH >25 LAKH >5 CR
SERVICE <10LAKH >10 LAKH >2CR
Title
CONTENT CREATION & DISTRIBUTION
Page-text
B2B Small Business Content Marketing Tactic Usage
Table
Obj-text
Content – other than blogs 93%
87%
83%
Website 79%
76%
74%
73%
Illustrations/Photos 68%
Online Presentations 65%
64%
60%
Webinars/Webcasts 59%
Other-text
2015 B2B Small Business Content Marketing Trends—North America: Content Marketing Institute/MarketingProfs
Page-text
Which content marketing tactics does your organization use?
The average number of tactics used is 12
Note: Fewer than 45% said they use the following tactics: eBooks (40%), Microsites (40%), Research Reports (40%), Branded Content Tools (34%), Books (29%), Print Magazines (25%), Digital Magazines (23%), Mobile Apps (22%), Virtual Conferences (21%), Podcasts (18%), Print Newsletters (16%), and Games/Gamification (9%).
The percentages shown here are fairly similar to last year's. The biggest changes have been with the use of:
Social media content (other than blogs)—up 5 percentage points (88% last year vs. 93% this year)
Blogs—up 10 percentage points (77% last year vs.
Infographics—up 12 percentage points (48% last year vs. 60% this year).
Other-text
15
MarketingProfs
RAINMAKER PLATFORM

(a) Type A (b) Type B

Figure 1: SlideVQA contains some inconsistencies and errors. The Obj-Text class which represents all the text seen on any graphical elements (Image, Diagram, Table, Figure) is not always annotated.

System

SlideCraft can be broken into four components, each responsible for generating a different element: content, layout, style, and annotation.

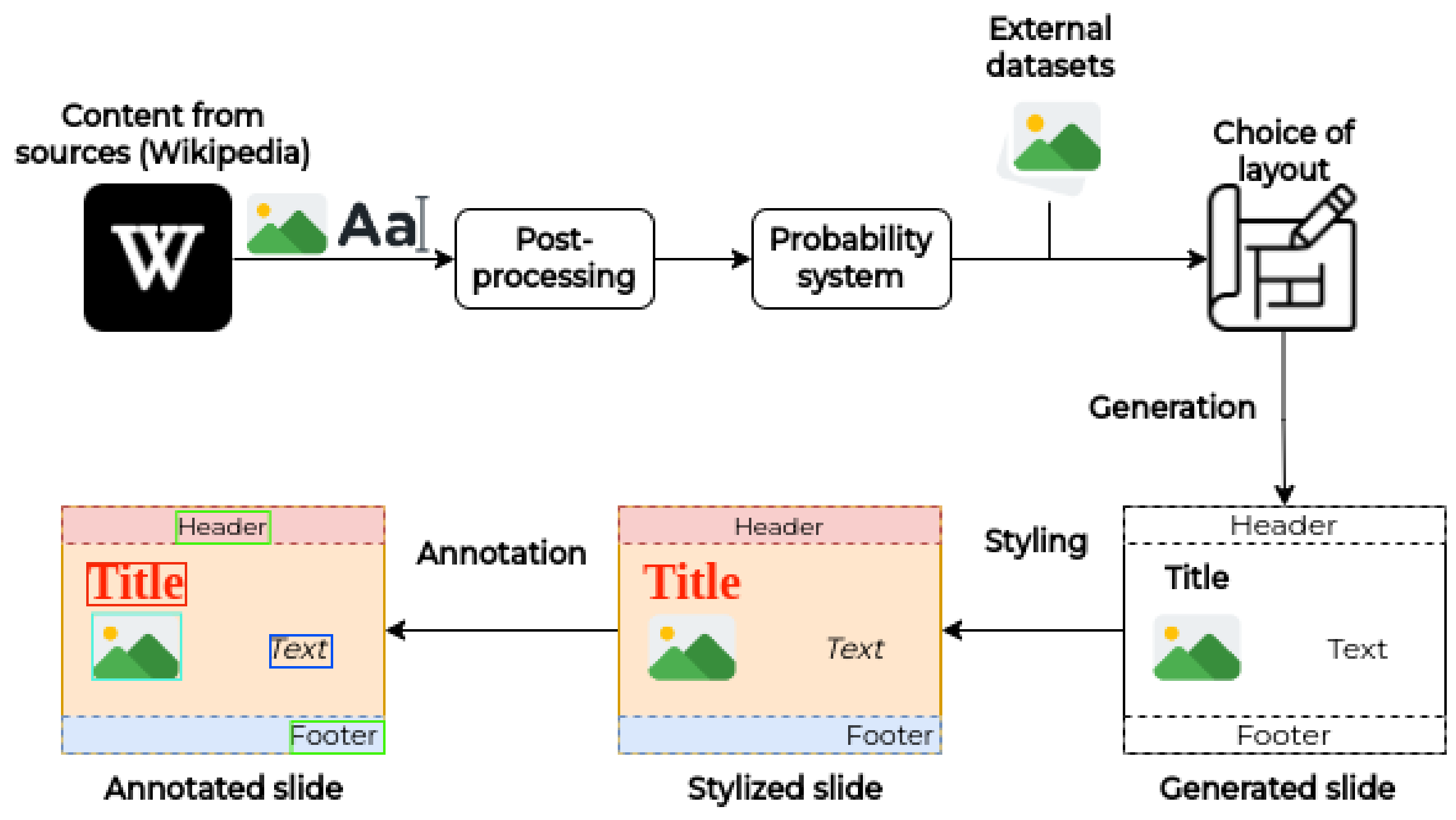


Figure 2: Overview of the multimodal educational video analysis process

Content

SlideCraft can use external sources to generate coherent slides. We currently leverage Wikipedia and external datasets to augment slides with additional elements (tables, diagrams, charts, plots, etc).

Layout

SlideCraft can generate different layouts to imitate real slides.

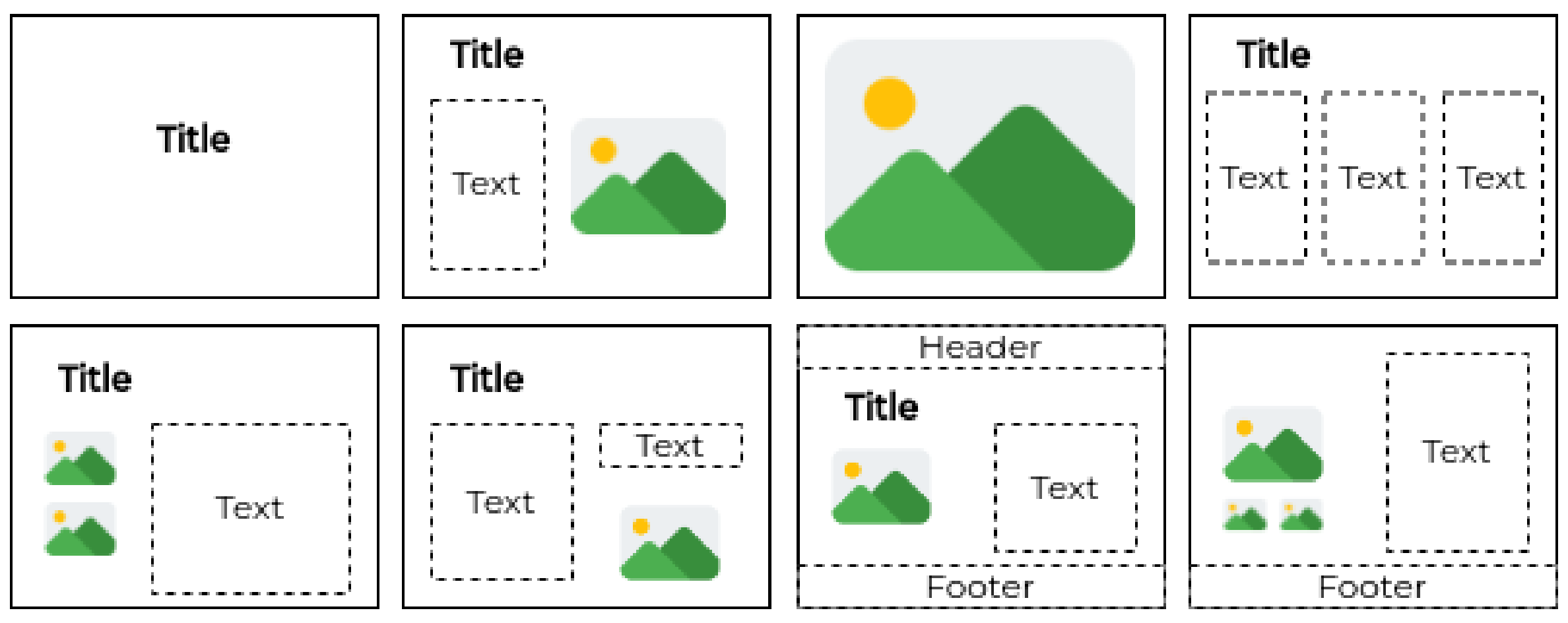


Figure 3: Examples of different layouts: different number of columns, display of header and footer, display of the title. The number of elements taken from the content can be adjusted to choose between more visual or more textual output.

Style

SlideCraft allows for the customization of font, font style, font size, background and colors.

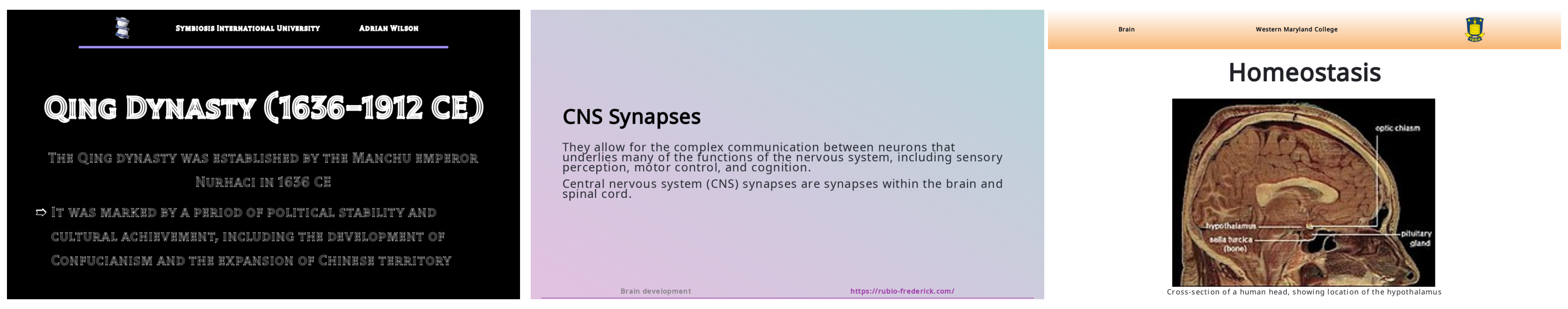

Qing Dynasty (1636–1912 CE)
The Qing dynasty was established by the Manchu emperor Nurhaci in 1636 CE
It was marked by a period of political stability and cultural achievement, including the development of Confucianism and the expansion of Chinese territory
CNS Synapses
They allow for the complex communication between neurons that underlies many of the functions of the nervous system, including sensory perception, motor control, and cognition.
Central nervous system (CNS) synapses are synapses within the brain and spinal cord.
Homeostasis

Figure 4: Examples of SlideCraft's generated slides. Different styles and layouts are shown. Colors, backgrounds, font size, and font style are chosen randomly to increase the diversity of the generation.

Annotation

SlideCraft can generate bounding boxes and segmentation masks with custom classes to fit existing datasets.

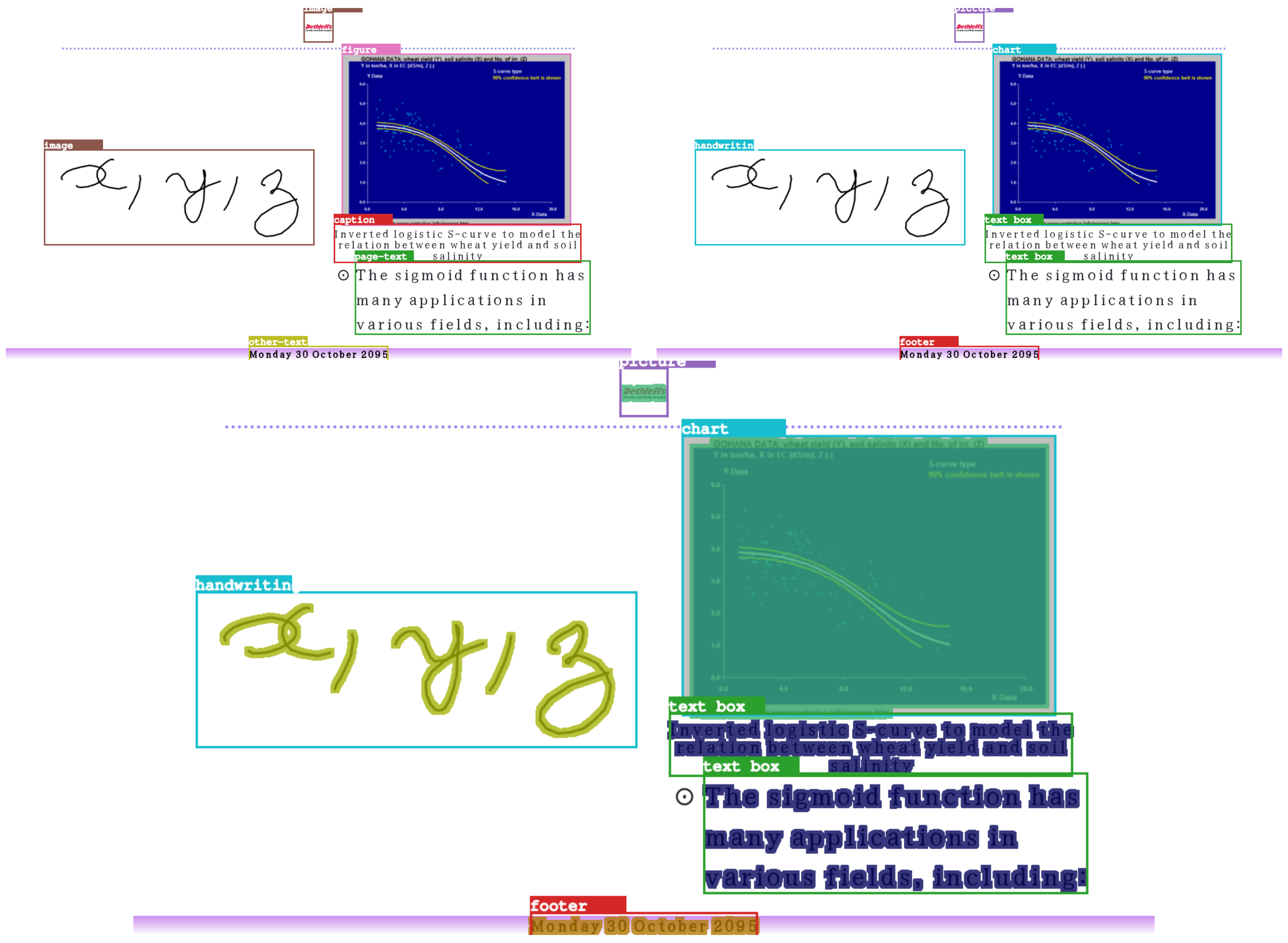


figure
image
caption
page-text
other-text
Monday 30 October 2095
chart
handwriting
text box
footer
Inverted logistic S-curve to model the relation between wheat yield and soil salinity
The sigmoid function has many applications in various fields, including:

Figure 5: Example of generated slides for SlideVQA annotations (Top Left), FitVid annotations (Top Right), segmentation masks (Bottom). Handwriting, header, footer, and equation, called Figure in FitVid, do not exist in SlideVQA.

Experiments

Extending datasets with SlideCraft data enhances performance of object detection, particularly for smaller datasets.

Table 1: mAP50 of FasterRCNN and YoloV8 on SlideVQA test set and FitVid test set, training with and without the slides generated by SlideCraft.

Dataset	FasterRCNN [3]			Yolov8 [1]		
	Original mAP50	Mix mAP50	Improvement	Original mAP50	Mix mAP50	Improvement
SlideVQA	0.641	**0.645**	0.59%	0.682	**0.685**	0.44%
FitVid	0.485	**0.530**	9.22%	0.513	**0.581**	13.26%

SlideCraft reduces the need for manual labeling in training object detection models while potentially boosting performance.

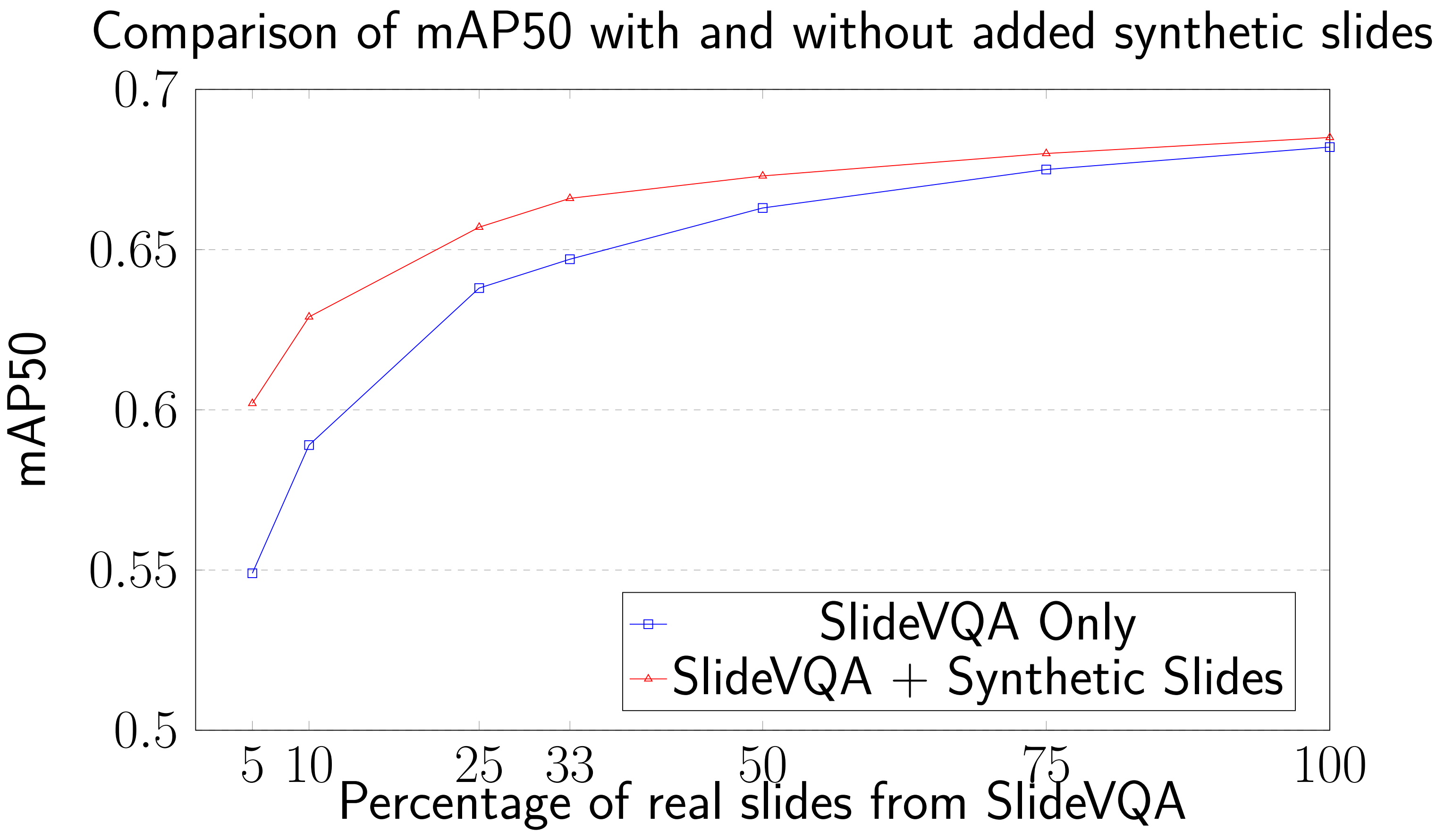


Figure 6: Comparison of mAP50 obtained after training YoloV8 with and without added 25,000 SlideCraft's generated slides. Testing on SlideVQA test dataset. The results show a direct correlation between dataset size and the impact from adding synthetic slides: smaller datasets see greater benefits from the inclusion of synthetic slides.

References

[1] Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLO (Jan 2023), https://github.com/ultralytics/ultralytics

[2] Kim, J., Choi, Y., Kahng, M., Kim, J.: FitVid: Responsive and Flexible Video Content Adaptation. In: CHI Conference on Human Factors in Computing Systems. pp. 1–16. ACM, New Orleans LA USA (Apr 2022). https://doi.org/10.1145/3491102.3501948, https://dl.acm.org/doi/10.1145/3491102.3501948

[3] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 28. Curran Associates, Inc. (2015), https://proceedings.neurips.cc/paper$_{f}iles/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf$

[4] Tanaka, R., Nishida, K., Nishida, K., Hasegawa, T., Saito, I., Saito, K.: SlideVQA: A Dataset for Document Visual Question Answering on Multiple Images (Jan 2023). https://doi.org/10.48550/arXiv.2301.04883, http://arxiv.org/abs/2301.04883, arXiv:2301.04883 [cs]