PITA: A VISIBLE WATERMARK DATASET IS ALL YOU NEED

EPITA DEEP NEURAL NETWORK

Quentin Fisch
EPITA
quentin.fisch@epita.fr

Bastien Pouëssel
EPITA
bastien.pouessel@epita.fr

Arnaud Baradat
EPITA
arnaud.baradat@epita.fr

Théo Ripoll
EPITA
theo.ripoll@epita.fr

Tom Genlis
EPITA
tom.genlis@epita.fr

Nicolas Fidel
EPITA
nicolas.fidel@epita.fr

February 5, 2024

ABSTRACT

With the proliferation of digital content, the safeguarding of intellectual property through watermarking has become paramount. This study addresses the critical challenge of detecting visible watermarks within images, focusing on the often-overlooked need for specialized datasets. We introduce a novel dataset comprising visible watermarks, including logos and text overlays, to enhance the realism of model training and evaluation. The investigation extends to the fine-tuning of popular object detection models, such as YOLO, DeTR, and Faster R-CNN, optimizing their performance specifically for visible watermark detection. This research contributes to the advancement of practical solutions for content protection, catering to the nuances of visible watermark scenarios.

Keywords Watermark · Object detection

1 Introduction

In the growing world of digital content, using visible watermarks is an important way for content owners to claim ownership, prevent unauthorized use, and reduce the risk of others copying their work without permission.

Datasets for this purpose are not available and are quite difficult to access due to download restrictions in various countries. In addition to that, most of the watermarks in these datasets do not represent real-life scenarios. They are often completely randomly generated and don't reflect actual use cases.

We're not only creating datasets but also improving popular object detection models like YOLO, DeTR, and Faster R-CNN specifically for identifying visible watermarks. Our goal is to achieve better accuracies for this task. In the upcoming sections, we'll explain our methods for creating the dataset and finetuning the models, share our results.

2 Visible watermarks datasets

We have observed that while datasets such as COCO are available for object detection, the availability of datasets specifically designed for the detection of watermarks added to images is significantly limited. Through our research, we identified only one such dataset, which originates from the paper Wdnet: Watermark-Decomposition Network for Visible Watermark Removal [1]. This dataset provides a collection of images along with their corresponding watermark masks for the purpose of watermark removal. Additionally, we noted that accessing this dataset presented challenges in terms of data accessibility and regeneration of dataset samples.

The CLWD Dataset, introduced in Wdnet: Watermark-Decomposition Network for Visible Watermark Removal [1], comprises images sourced from the COCO Dataset (Lin et al., 2014) [2] and masks of colored watermarks featuring random positions and opacities.

2.1 PITA dataset

We decided to introduce the Pita dataset, which is based on images from the COCO dataset (Lin et al., 2014) [2] and combines these with logos from the Open Logo Detection Challenge (Su et al., 2018) [3].

The dataset introduces several changes compared to other datasets, with a focus on the task of watermark detection rather than watermark removal.

The dataset is structured into three splits: a training split, a validation split, and a test split, collectively comprising approximately 20 000 watermarked images featuring both logos and text.

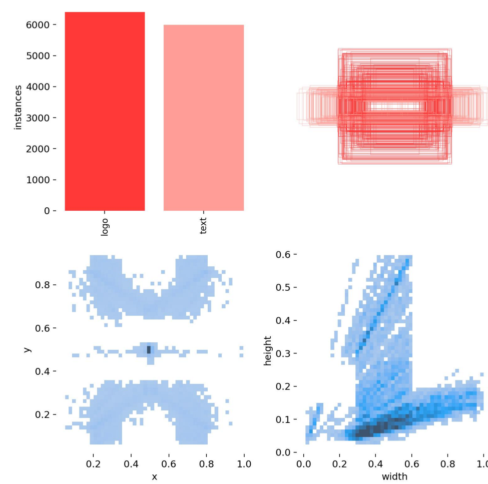
We decided to incorporate two types of labels:

- *Text*: The images are watermarked with a random font available on the computer used for generation, and the text size is also randomized.
- *Logos*: The logos are sourced from the Open Logo Detection Challenge dataset (Su et al., 2018) and are characterized by random sizes and opacities.

The position of the logo or text is randomly selected from a set of available positions, specifically corners or the center. This restriction was introduced based on the observation that watermarks on social media or stock image websites are predominantly located in these positions.

The dataset is accompanied by command-line interface tools that facilitate reproducibility. These tools support both YOLO and Hugging Face formats, allowing the download of the dataset and generation with ease.

(a) Batch with bounding boxes.



(b) Distribution of labels and positions of bounding boxes.

Figure 1: PITA dataset for visible watermark detection.

3 Models

In the context of our study, several modelfor object detection were fine-tuned on our dataset to assess their ability to detect watermarks, which are not included in their default classification categories.

3.1 DeTR

The Detection Transformer (DETR), introduced in End-to-End Object Detection with Transformers (Carion et al., 2020) [4], is a state-of-the-art model on the COCO dataset for object detection. This model employs a two-stage process involving region proposal networks followed by object classification, leveraging the transformer architecture.

3.2 FasterR-CNN

Faster R-CNN, an earlier model introduced in Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks (Ren et al., 2016) [5], innovated with the introduction of Region Proposal Networks (RPNs). These RPNs share convolutional layers with the feature extraction network, enhancing the efficiency and efficacy of the model.

3.3 YOLOV8

The YOLOv8 model, introduced by Ultralytics in Real-time Flying Object Detection with YOLOv8, 2023 (Reis et al., 2023) [6], is a single-stage object detection model that removes the region of interest extraction process, opting instead to directly classify and regress bounding boxes. Offered in various sizes, the most compact version begins with 3 billion parameters. The foundation of YOLOv8 is a convolutional architecture.

4 Results and Benchmark

Table 1: Benchmark on OUR dataset VS CLWD dataset

Metric	map	map_50	map_75	map_per_class
		PITA Dataset		
DeTR	–	–	–	–
FasterR-CNN	**0.9005**	**0.9839**	**0.9644**	[**0.9108**, 0.8798]
YoloV8n	0.8750	0.9690	0.9415	[0.8643, 0.8858]
YoloV8l	0.8900	0.9741	0.9473	[0.8783, **0.9018**]
		CLWD Dataset		
DeTR	–	–	–	—
FasterR-CNN	0.0005	0.0019	0.0001	—
YoloV8n	**0.0190**	**0.0394**	**0.0160**	—
YoloV8l	0.0082	0.0130	0.0063	—

5 Conclusion

In this project, we introduced a specialized dataset based on COCO, featuring two distinct labels for watermark detection, comprising approximately 20,000 samples. The dataset we created was specifically tailored for the task of object detection, not watermark removal. We facilitate the generation of datasets compatible with multiple libraries, such as Hugging Face and Ultralytics.

The benchmark provided us with good results in terms of Mean Average Precision (MaP) for both classes. As MaP accounts for both classification accuracy and Intersection over Union (IoU), it serves as a robust metric that emphasizes both aspects. Particularly, YOLOv8 demonstrated the advantages of rapid training and impressive performance on our test dataset. Training Faster R-CNN model demands more time , it surpasses other models with superior performance metrics. However, its did not translate effectively when applied to the CLWD dataset for both of your models.

We are examining multiple reasons for the underperformance of our model on this dataset:

- The dataset has only one class, consisting exclusively of logos.
- The positioning of the logos and text is different from ours. It is randomly generated across the images.
- The dataset features entirely distinct logos, including variations in rotation and opacity, which significantly differ from ours.

5.0.1 Limitations and additinal properties

Improvements could be achieved by enhancing the logo's contouring to obtain higher quality images and by adding more variations in the logo's positioning. Furthermore, an additional step could involve incorporating both logos and text in the same images or introducing multiple instances of the logo in the dataset.

References

[1] Yang Liu, Zhen Zhu, and Xiang Bai. Wdnet: Watermark-decomposition network for visible watermark removal, 2020.

[2] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Doll'a r, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014.

[3] Hang Su, Xiatian Zhu, and Shaogang Gong. Open logo detection challenge. In *British Machine Vision Conference*, 2018.

[4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers, 2020.

[5] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks, 2016.

[6] Dillon Reis, Jordan Kupec, Jacqueline Hong, and Ahmad Daoudi. Real-time flying object detection with yolov8, 2023.

[7] Danni Cheng, Xiang Li, Wei-Hong Li, Chan Lu, Fake Li, Hua Zhao, and Wei-Shi Zheng. *Large-Scale Visible Watermark Detection and Removal with Deep Convolutional Networks: First Chinese Conference, PRCV 2018, Guangzhou, China, November 23-26, 2018, Proceedings, Part III*, pages 27–40. 11 2018.

[8] Hang Su, Shaogang Gong, and Xiatian Zhu. Weblogo-2m: Scalable logo detection by deep learning from the web. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 270–279, 2017.