



HAL
open science

Detecting Looted Archaeological Sites from Satellite Image Time Series

Elliot Vincent, Mehraïl Saroufim, Jonathan Chemla, Yves Ubelmann, Philippe Marquis, Jean Ponce, Mathieu Aubry

► **To cite this version:**

Elliot Vincent, Mehraïl Saroufim, Jonathan Chemla, Yves Ubelmann, Philippe Marquis, et al.. Detecting Looted Archaeological Sites from Satellite Image Time Series. 2024. hal-04701017

HAL Id: hal-04701017

<https://hal.science/hal-04701017v1>

Preprint submitted on 18 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Detecting Looted Archaeological Sites from Satellite Image Time Series

Elliot Vincent^{1,2}, Mehraïl Saroufim³, Jonathan Chemla³, Yves Ubelmann³,
Philippe Marquis⁴, Jean Ponce^{5,6}, Mathieu Aubry¹

¹LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, France ²Inria Paris

³Iconem ⁴DAFA, French archaeological delegation in Afghanistan

⁴Department of Computer Science, Ecole normale supérieure (ENS-PSL, CNRS, Inria)

⁵Courant Institute of Mathematical Sciences and Center for Data Science, New York University

{elliot.vincent, mathieu.aubry}@enpc.fr, jean.ponce@ens.fr

{yu, jchemla, mehrail.saroufim}@iconem.com, marquis.dafa@hotmail.fr

Abstract

Archaeological sites are the physical remains of past human activity and one of the main sources of information about past societies and cultures. However, they are also the target of malevolent human actions, especially in countries having experienced inner turmoil and conflicts. Because monitoring these sites from space is a key step towards their preservation, we introduce the DAFA Looted Sites dataset, DAFA-LS, a labeled multi-temporal remote sensing dataset containing 55,480 images acquired monthly over 8 years across 675 Afghan archaeological sites, including 135 sites looted during the acquisition period. DAFA-LS is particularly challenging because of the limited number of training samples, the class imbalance, the weak binary annotations only available at the level of the time series, and the subtlety of relevant changes coupled with important irrelevant ones over a long time period. It is also an interesting playground to assess the performance of satellite image time series (SITS) classification methods on a real and important use case. We evaluate a large set of baselines, outline the substantial benefits of using foundation models and show the additional boost that can be provided by using complete time series instead of using single images. The code and dataset are available at <https://github.com/ElliotVincent/DAFA-LS>.

1 Introduction

Between 2003 and 2013, numerous archaeological studies were conducted in Afghanistan resulting in the identification of over a thousand sites through terrestrial surveys and satellite imagery, in particular in the northern part of the country (see Figures 1a & 1b). However, the progressive deterioration of security in Afghanistan posed challenges for continued field research and it has remained relatively unexplored by archaeologists since 2015. In 2022, the French archaeological delegation in Afghanistan (DAFA) identified the first instance of looting at the Dilberjin site [25] (Figure 1f). Leveraging their extensive knowledge of the terrain, and together with the Iconem startup, archaeologists started documenting widespread pillaging affecting more than a hundred major sites. Given the scale of this phenomenon, an automated approach becomes necessary, and would help identifying more quickly signs of looting. While some research on detecting looted archaeological sites exists (see Section 2.1), it remains limited and most existing publications do not release any data. This lack of public datasets makes it harder for machine learning researchers unfamiliar with archaeology or remote sensing data to contribute their knowledge and skill despite the theoretical and practical interest of the problem.

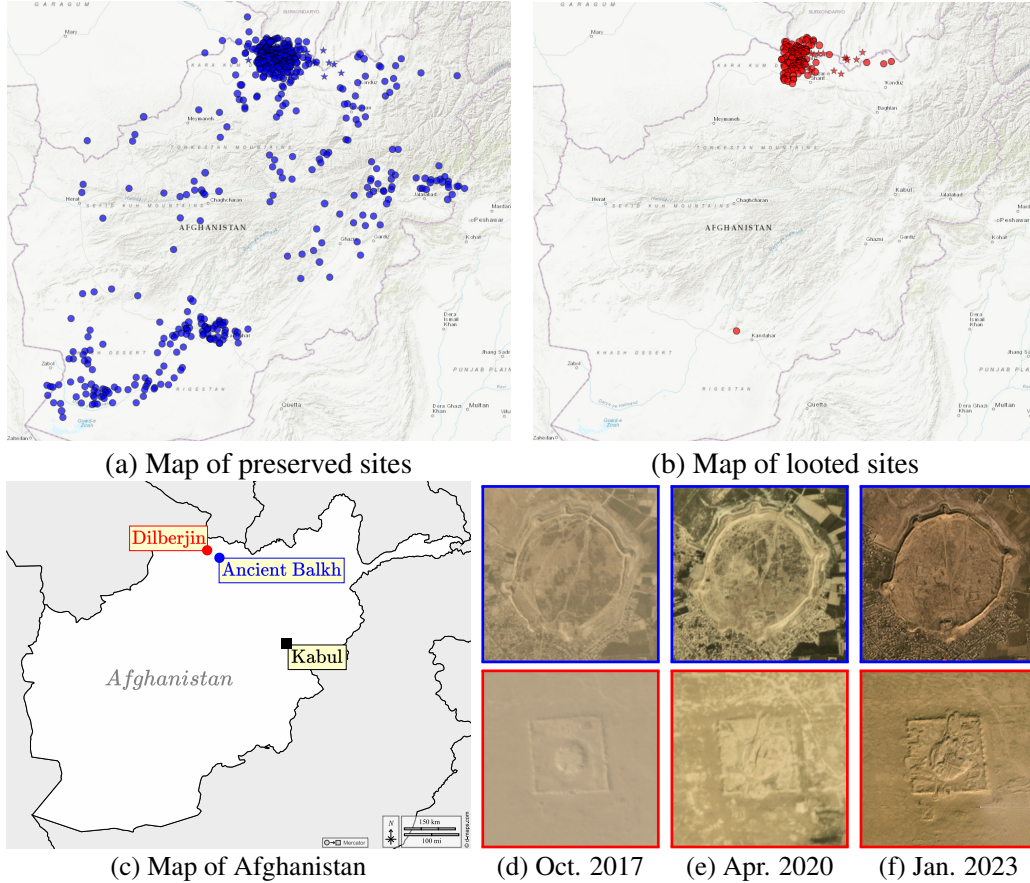


Figure 1: **DAFA Looted Sites (DAFA-LS)** contains monthly satellite image time series (SITS) of Afghan archaeological sites acquired between 2016 and 2023. We show the location of preserved (a) and looted (b) sites, adding strong random noise to their coordinates to prevent misuse of the data. Test sites are marked with a star (*). We also show images (d-f) from two sites (c): **Ancient Balkh** (in blue, top row) has been preserved from looting [80], while **Dilberjin** (in red, bottom row) suffered irreparable damage.

We introduce the DAFA Looted Sites dataset (DAFA-LS), a dataset containing 55,480 images acquired monthly over 8 years, from 2016 to 2023, across 675 Afghan archaeological sites, including 135 sites looted during this period. The dataset is organized into satellite image time series (SITS) and we distinguish between ‘looted’ and ‘preserved’ sites (see Figure 1). DAFA-LS is the first open-access dataset to make it possible for SITS classification methods to evaluate on the looting detection task. Note that looting detection is significantly different from most change detection tasks, since looting can appear as a relatively subtle appearance change, while other significant changes, e.g., in the vegetation, are not relevant.

The paper is structured as follows. In Section 2, we review related works and datasets for archaeological looting detection and SITS classification. In Section 3, we describe our dataset and its curation. In Section 4, we present diverse baseline methods and report their performance. Finally, we discuss the limitations of DAFA-LS in Section 5.

2 Related Work

We first provide an overview of the archaeological looting detection literature (Section 2.1) and then give a broad outline of satellite image time series (SITS) classification (Section 2.2).

2.1 Archaeological looting detection

SITS-based visual detection of archaeological looting. Aerial and/or satellite imagery offers a means to manually monitor archaeological sites, complementing time-consuming, expensive and sometimes hazardous ground truth surveys [57, 19]. Furthermore, human expeditions are impractical in several countries due to political and military restrictions [17]. Multi-temporal satellite imagery allows archaeologists to identify looting patterns by comparing successive images [77, 57]. The literature on using remotely sensed data to manually detect potential looting of archaeological sites is rich [72, 73, 42, 77, 78, 58, 59, 1, 14, 12, 13, 3, 87, 74, 4, 2]. Early methods relied on raw images for direct visual assessment of damage, without any image enhancement or data processing. For example, Stone [72, 73] used raw SITS to detect looting holes in Iraq, while Kennedy et al. [42] assessed site destruction by bulldozers in Jordan using aerial image time series. Several publications [1, 14, 12, 13] focus on Syrian sites located in conflict zones, visually comparing pre- and post-conflict high-resolution satellite images to identify looting, quantify damage, and assess the timing of these events. More recent work confirms the potential of visual inspection of SITS to identify looting on Syrian sites [3] or other damaging practices (ploughing, building/road/canalization constructions) in Iran [87]. Another line of work processes the images before visually analyzing them. Tapete et al. [74] apply a Gaussian filter to the images before ratioing consecutive image pairs in order to enhance morphological changes. This method is used on synthetic aperture radar (SAR) images to detect looting holes or marks on Syrian sites. To ease the process of visual interpretation of the images, Agapiou et al. [4] perform photo-enhancing operations, playing on the contrast, the brightness and the histogram of Google Earth multi-temporal images of Cyprus, in order to visually identify looting marks. Finally, Abate et al. [2] compare multi-temporal orthophotographies acquired by unmanned aerial vehicle (UAV) to the output of the maximum autocorrelation factor/multivariate alteration detection (MAF/MAD) transformations [55], commonly used in change detection studies.

Automatic looting detection. In order to speed up the detection process and allow for a fast response to potential threats, automatic identification methods have also been developed. A simple approach is to use the thresholded difference image between bi-temporal satellite acquisitions in order to reveal looting marks. Rayne et al. [64] apply this method on Sentinel-2 images a year apart to detect looting pits in Lybia and Egypt. Castilla et al. [15] compute change maps from bi-temporal satellite image pairs as the sum of their robust difference [15] (*i.e.*, difference in brightness) and the difference between their Gabor feature maps [52, 35] (*i.e.*, difference of texture). This method is applied to pre/post-disaster image pairs to identify sites damaged by terrorists in Syria and Iraq. Bowen et al. [11] localize looting pits in Egypt with a tree-based classifier using SIFT [50], SURF [9] and HOG [20] descriptors on mono-temporal mid-resolution satellite images. A series of work [45, 54, 53] investigates a clustering approach on spatial auto-correlation features for the detection of looting holes or pits with bi-temporal Syrian and Peruvian Google Earth images. El Hajj [24] classifies image patches to reveal whether they contain looting or destruction instances using an ensemble model composed of a random forest [39], an AdaBoost [27] classifier and a SMOTEBoost [16] classifier. Payntar [60] characterizes potential destruction of archaeological sites by the change in the number of different land cover classes in a given neighborhood. Quinquennially Landsat images of Peru from 1985 to 2020 are segmented independently with a random forest to obtain the land cover maps. Finally, mask R-CNNs [37] have recently been applied for the detection of looting holes on mono-temporal UAV data of several areas across the globe [5]. Recent methods are trained and evaluated on datasets that, in contrast to DAFA-LS, are not available in open-access and/or not multi-temporal, as reported in Table 1.

2.2 Satellite image time series classification

Satellite image time series datasets. Many SITS datasets have been created in recent years for applications to crop-type mapping [67, 69, 84, 31, 43, 76], wildfire spread prediction [33], tree species identification [6], wilderness mapping [23], change detection [79, 81, 82], land-cover mapping [29, 85], low-to-high resolution knowledge transfer [48], cloud removal [22], and electricity access detection [51]. DAFA-LS is the first open-access dataset for the detection of looted archaeological sites. Compared to other STIS classification datasets and applications, our benchmark presents several important specificities. First, the changes related to looting that we aim at detecting can be extremely subtle and hardly visible, while other very important but irrelevant changes are frequent. Second, because we target regular monitoring at limited cost, our dataset is built from images of

Table 1: **Aerial and satellite image datasets for archaeological site looting detection.** In Masini et al. [53], images were acquired with Google Earth with varying spatial resolution. El Hajj [24] does not provide open access to their dataset, but it can theoretically be recreated openly since (i) locations, images and labels come from open-source data, and (ii) the pre-processing steps are detailed in the paper.

	Open-access	Multi-temporal	Spatial resolution	Temporal resolution	Sensor	Location	Number of sites
Masini et al. (2020) [53]	✗	✓	Varying	Yearly	Satellite	Syria	2
El Hajj (2021) [24]	✗	✗	15m/px	—	Satellite	Syria and Iraq	9
Payntar (2023) [60]	✗	✓	30m/px	Quinquennially	Satellite	Peru	477
Altaweel et al. (2024) [5]	✓	✗	3cm/px	—	UAV	Worldwide	95
DAFA-LS (ours)	✓	✓	3m/px	Monthly	Satellite	Afghanistan	675

limited spatial resolution but available freely at high and uniform temporal resolution. Third, because of the limited number of archaeological sites, and particularly looted archaeological sites, training data is necessarily small and very imbalanced.

Satellite image time series classification approaches. Existing techniques for satellite image time series classification include pixel-wise and whole-image methods. Pixel-wise methods [61, 40, 68, 30, 83] process each pixel of the SITS independently, discarding spatial information. In practice, these methods are outperformed by whole-image approaches, like PSE+LTAE [32] or TSViT [75]. The former encodes a set of pixels before processing the temporal dimension with an attention mechanism while the latter is a fully spatio-temporal attention-based approach. Closely related to our topic are segmentation methods for SITS like 3D-Unet [70] or UTAE [31] which are U-Net [66] based architecture designed to process the temporal dimension. Very recently, several works [18, 28, 65, 86, 6, 46, 56, 34, 47] have aimed at developing a remote sensing "foundation model", introducing large models pretrained on millions of remotely sensed images, often from different sensors, modalities or spatial and temporal resolutions. These pretrained models have shown good performance on classification and segmentation downstream tasks. We evaluate both pixel-wise and whole-image methods, and demonstrate that foundation models can significantly boost performance.

3 Dataset

Sites identification and characterization. A significant number of Afghan archaeological sites are cataloged in the "Archaeological gazetteer of Afghanistan" [7] or documented by the DAFA. Because ground surveys are impractical since 2015 in Afghanistan, our team of expert archaeologists navigated through online platforms of high resolution satellite imagery (Google Earth, ESRI and Bing) in order to enrich the list with new sites and label them into the 'looted' and 'preserved' categories, depending on whether it has been damaged by malevolent human activities prior to 2023 or not. In total, 986 Afghan archaeological sites have been identified. We have gathered monthly SITS from January 2016 to December 2023 for each location. The source of our images is the Planet Basemaps product¹ from Planet Labs [63]. Images can be freely downloaded, via the "Planet education and research program", but since we do not want to release geographical locations to avoid misuse, we will instead redistribute modified versions of the images under a CC BY-NC 4.0 license². In addition to the raw acquisitions, Planet applies a proprietary post-processing algorithm in order to minimize the effects of clouds, haze and other image variability [62]. Planet images have an alpha channel indicating areas where there is no image data available. We remove all images with any pixel for which it is the case. This results in some missing time stamps, but the filtered dataset still has a median of 94 dates per time series out of 96 possible dates. Each image of the dataset covers a 1 km² area centered on an archaeological site. Images have 3 channels (RGB) with 266×266 size and a ground sampling distance (GSD) of 3 meters per pixel.

¹<https://developers.planet.com/docs/basemaps/>

²https://assets.planet.com/docs/ToS_EducationAndResearch.pdf

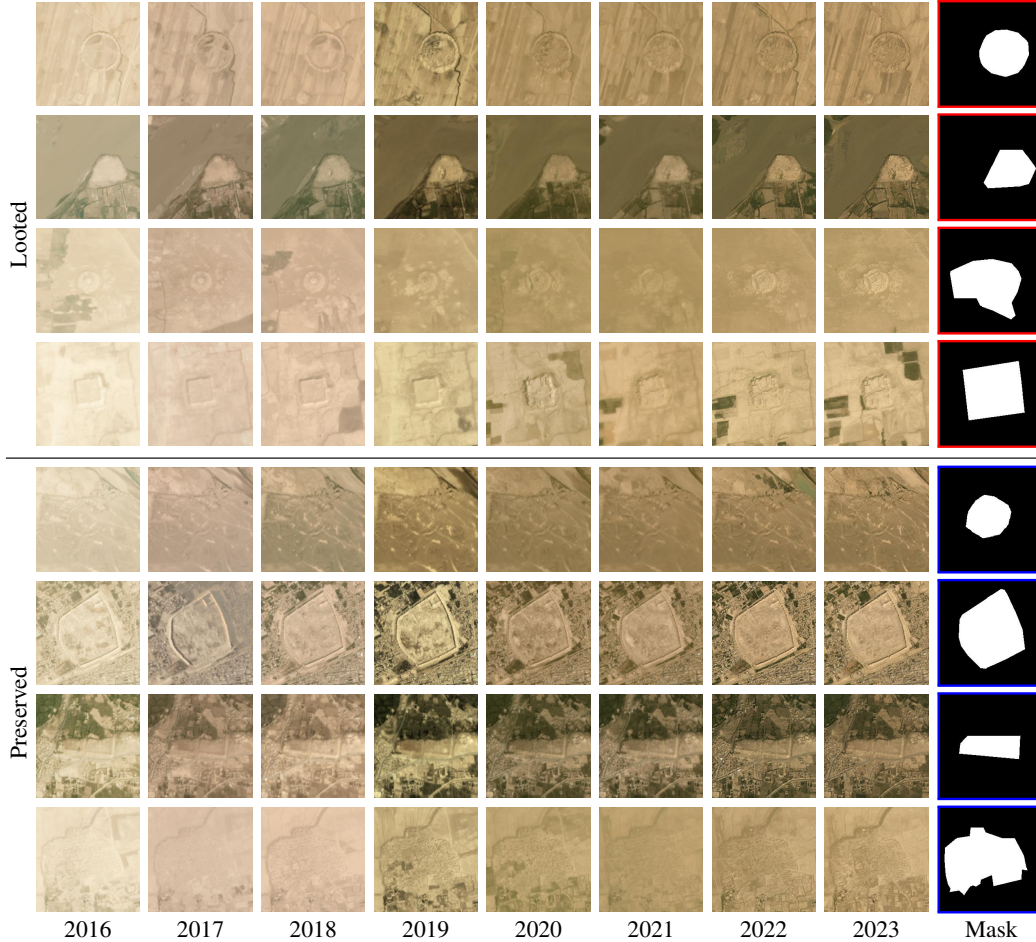


Figure 2: **Examples of time series and coarse location masks.** For each archaeological site, we show the September image for each year from 2016 to 2023 and the corresponding coarse location mask. The 4 top rows show **looted** sites (red squares) and the 4 bottom rows show **preserved** sites (blue squares).

A 1 km²-square patch centered on a site may cover other archaeological sites in varying states of preservation. Since most sites have a relatively small extent, the majority of the pixels may not provide informative data about potential looting, may contain different archaeological sites, or may contain information that is correlated with looting, such as the presence of roads. For these reasons, leveraging the most recent high-resolution Google Earth image available to date, we manually annotated a binary mask coarsely detouring the sites. A high fraction of the preserved sites could not be delineated with sufficient confidence and was eventually discarded. Several reasons may explain why a site is not visible on a SITS, for example the site may be too small to show on mid-resolution imagery or completely buried under sand or earth. The final number of sites for which a coarse location mask is available is 675 (135 looted, 540 preserved), for a total of 55,480 satellite images. We show some examples of SITS from DAFA-LS along with their coarse location mask in Figure 2.

Test set definition. As can be seen in Figures 1a & 1b, looted archaeological sites are mainly located in the northern region of Afghanistan. However, we do not want methods evaluated on our dataset to bypass the looting detection task and learn to locate the sites thanks to geographical cues. For this reason, we ensure that (i) the maximum distance between two test sites is less than 200 km (the maximum distance between two sites in the dataset is more than 1000 km), and that (ii) the maximum distance between two sites of opposite class in the test set is smaller than 40 km (the average is 7.5 km and median 5.5 km). To limit other forms of geographic bias, we also ensure that no site in the test set is closer than 1 km from a site in the rest of the dataset.

Table 2: **Distribution of areas for each class.** Classifying all sites with a 3-hectare threshold leads to a mean accuracy of 79.4%, showing the need for a balanced test set to prevent the area bias.

Surface (ha)	Looted (%)	Preserved (%)
0 <...< 1	0	16.1
1 <...< 2	2.2	34.8
2 <...< 3	12.6	22.6
3 <...< 5	22.2	14.4
5 <...< 10	34.1	8.5
10 <...< 100	28.9	3.5

We also found there was a potential bias in the sites areas, which we can estimate using the coarse location masks. The area distribution for both classes is reported in Table 2. We observe significantly different distributions between looted and preserved sites, where a simple 3-hectare threshold results in a mean classification accuracy of 79.4%. To avoid this bias, we ensure that the test set exhibits similar area distributions between the two classes by selecting the same number of sites within each area bin.

Selection using these criteria leads to a total of 61 sites (26 looted, 35 preserved) which we use as our test set.

Training and validation splits. From the remaining 614 sites we define 6 splits. First, we define 5 diverse validation splits of 18 sites each ensuring that all sites in a validation split are further than 1 km from any site out of this split. The remaining 524 sites are always used as training. We evaluate all methods using a 5-fold scheme, using one of the validation split as validation and adding the others to the 524 samples we always use for training.

4 Benchmark

4.1 Task and evaluation

Problem formalization. Let x be an RGB input time series of satellite images consisting of T images of height H and width W . Each time series corresponds to a unique archaeological site, whose coarse location is given by a binary mask m of size $H \times W$. Given a pair (x, m) , the task is to predict a label y in $\{0, 1\}$ indicating whether the corresponding site is preserved ($y = 0$) or has been looted ($y = 1$). Except when stated otherwise, we use as input to all methods the element-wise multiplication of x and m , which prevents them from leveraging information contained outside of the coarse location mask which may be correlated with looting, for example the proximity of a road.

Metrics. We evaluate the classification performance of the baselines using a set of metrics commonly used in the binary classification literature: the overall accuracy (OA), the F1-score (F1), and the area under the ROC curve (AUROC). Additionally, we use the false alarm rate (FAR), also known as the false positive rate and common in the looting detection literature [44, 45, 53, 54]. A low number of false positives is indeed desired because it limits the number of costly and time-consuming verifications and potential ground surveys on sites flagged as looted.

4.2 Baselines

We evaluated three types of methods: single frame methods, pixel-wise multi-frame methods and whole-image multi-frame methods. We summarize their key characteristics and differences in Table D1 of the appendix.

Single-frame methods. We know looted sites have been damaged prior to 2023 so that looting marks should appear on all 2023 images. Based on this observation, we train single-frame methods only looking at 2023 images and classifying them between ‘looted’ and ‘preserved’ depending on the time series they are taken from. At inference, all the 2023 images of a given time series are

input to the model separately, and the most predicted label out of the 12 predictions is selected for the corresponding site. We use as baselines ResNets [38] of various sizes trained from scratch and several frozen pre-trained foundation models: SatMAE [18], Scale-MAE [65] and DOFA [86] on top of which we train a linear classification head. They are visual transformers (ViT) [21] pretrained following the mask-autoencoding (MAE) [36] strategy. They also have their own specificities, including the dataset they have been trained on. In particular, DOFA is pretrained on more than 8 million satellite images from different sensors, modalities and spatial resolutions.

Pixel-wise multi-frame methods. We can also view DAFA-LS as a set of pixel time series and evaluate several methods designed for pixel-wise satellite image time series classification including DuPlo [40], TempCNN [61], a self-attention approach referred to as Transformer [68], and LTAE [30]. Here, only pixels $x_{i,j}$ located inside the coarse location mask (*i.e.*, $m_{i,j} = 1$), are used at training and inference for a given SITS x . At inference, we select the most predicted label between ‘looted’ and ‘preserved’ among in-mask pixels as the prediction for a given SITS.

Whole-image multi-frame methods. We evaluate several methods designed for the SITS classification task as defined in Section 4.1: PSE [32]+LTAE [30], TSViT [75], SatMAE [18]+LTAE [30], Scale-MAE [65]+LTAE [30] and DOFA [86]+LTAE [30]. LTAE is a temporal self-attention network that we apply to series of features extracted either with a pixel-set encoder (PSE) or with a frozen pretrained foundation model (SatMAE, Scale-MAE or DOFA). The Temporo-Spatial Vision Transformer (TSViT) has a fully-attentional architecture, processing the tokens first temporally then spatially. We train a small version of TSViT from scratch on DAFA-LS with a classification head.

Additionally, we considered seeing the problem as a segmentation one to evaluate a wider range of methods. In this case, we do not mask the images and first train a method to segment the time series at the pixel level into three classes: ‘looted site’ (in-mask pixels for SITS of looted sites), ‘preserved site’ (in-mask pixels for SITS of preserved sites) and ‘not a site’ (out-mask pixels). At inference, the classification prediction corresponds to the majority class at the pixel-level inside the coarse location mask among ‘looted’ and ‘preserved’. This enables us to evaluate two small versions of TSViT [75] (with a segmentation head) and UTAE [31] trained from scratch. The U-Net with Temporal Attention Encoder (UTAE) consists of a U-Net architecture where a temporal attention mechanism squeezes the temporal dimension before the decoding branch.

Implementation details. We trained all methods on a single NVIDIA GeForce RTX 2080 Ti or NVIDIA V100 GPU. All methods are trained using a binary-cross entropy loss and the AdamW optimizer [49], except for the segmentation methods that have been trained using a regular cross-entropy loss. We use the implementations of DuPlo and TempCNN available in Transformer’s official public repository³ and the official implementation for all other methods. We use random resize crop, random rotate and random flip as data augmentation for whole-image methods. For whole-image time series based approaches, 24 dates (3 per year) are randomly sampled out of the 96 available at training time and the whole time stamps are used at inference. Additional details on training and architecture configurations can be found in the appendix.

4.3 Results

We report the performance of all evaluated baselines in Table 3 and make three key observations. (i) First, leveraging the DOFA foundation model outperforms other methods. Among single-frame methods, DOFA surpasses ResNet20 and ResNet18 trained from scratch, as well as the frozen SatMAE and Scale-MAE models, and is on par with the larger ResNet34 trained from scratch but with a significantly lower FAR. In the multi-frame category, DOFA+LTAE clearly outperforms PSE+LTAE, SatMAE+LTAE, and Scale-MAE+LTAE, delivering our best overall performance. Note that, although DOFA has not been trained on Planet imagery and at its particular spatial resolution, the pre-training set of DOFA includes SatlasPretrain [8] that contains images of Afghanistan in the SAR modality. DOFA (111M parameters) also has multiple orders of magnitude more parameters than ResNet20 (0.27M parameters) for example. (ii) Second, the temporal information benefits from a specific treatment, beyond a simple voting aggregation strategy: DOFA+LTAE outperforms the single-frame DOFA baseline by +7.9pt in F1 score, +3.1pt in AUROC, and +2.0pt in OA. Note that this comes at the cost of an increased false alarm rate. This observation is consistent among

³<https://github.com/MarcCoru/crop-type-mapping>

Table 3: **Classification performance.** We evaluate several methods trained on DAFA-LS and distinguish between methods that process a single image at a time (mono-temporal) and methods receiving time series as input (multi-temporal). Multi-temporal methods can be pixel-wise or whole-image based. We indicate with a star (*) methods that have been pretrained on another dataset beforehand. Best scores are highlighted in **bold** and second best are underlined. We show the standard deviations over the 5 folds in (parenthesis) and report the number of trainable parameters for each baseline.

Method	#param (x1000)	OA \uparrow	F1 \uparrow	AUROC \uparrow	FAR \downarrow
<i>Single-frame methods</i>					
ResNet20 [38]	269.2	54.7 (8.9)	54.5 (17.1)	75.3 (3.1)	54.9 (34.5)
ResNet18 [38]	11,177.5	71.8 (2.6)	64.1 (5.4)	84.5 (1.5)	19.4 (3.3)
ResNet34 [38]	21,285.7	74.1 (3.2)	<u>68.9</u> (6.3)	<u>85.2</u> (1.7)	22.3 (8.0)
SatMAE* [18]	2.1	63.6 (0.7)	41.9 (0.4)	75.3 (0.2)	<u>12.0</u> (1.1)
Scale-MAE* [65]	2.1	62.6 (0.7)	39.3 (1.9)	76.0 (0.3)	<u>12.0</u> (1.1)
DOFA* [86]	1.5	<u>76.7</u> (2.8)	67.0 (4.2)	84.0 (1.4)	7.4 (2.3)
<i>Multi-frame methods</i>					
<i>Pixel-wise methods</i>					
DuPLo [40]	86.8	52.1 (2.8)	50.4 (4.9)	50.9 (3.7)	52.0 (7.8)
TempCNN [61]	28.5	55.7 (3.4)	44.2 (9.7)	58.8 (1.8)	34.9 (9.8)
Transformer [68]	38.5	56.4 (3.7)	63.5 (3.2)	62.7 (4.1)	68.0 (10.0)
LTAE [30]	32.2	52.5 (7.8)	58.0 (4.6)	62.0 (8.5)	65.7 (18.8)
<i>Whole-image methods</i>					
PSE [32]+LTAE [30]	34.0	55.1 (9.8)	47.7 (6.2)	59.5 (6.3)	39.4 (19.4)
UTAE [31]	68.9	62.0 (3.5)	58.9 (2.3)	64.5 (4.5)	39.4 (8.6)
TSViT [75] (cls. head)	236.9	64.3 (1.2)	53.0 (3.7)	70.8 (2.3)	23.4 (4.9)
TSViT [75] (seg. head)	237.4	64.6 (3.5)	60.2 (7.1)	69.6 (4.2)	35.4 (6.9)
SatMAE* [18]+LTAE [32]	1,627.9	67.9 (4.7)	64.7 (4.0)	75.2 (3.7)	33.1 (11.1)
Scale-MAE* [65]+LTAE [32]	1,627.9	68.5 (2.4)	56.4 (7.7)	77.6 (0.8)	17.1 (4.4)
DOFA* [86]+LTAE [32]	926.1	78.7 (2.3)	74.9 (3.5)	87.1 (3.0)	18.9 (6.9)

all evaluated foundation models, with SatMAE+LTAE and Scale-MAE+LTAE outperforming their single-frame counterparts at the cost of an increased FAR. (iii) Finally, as often with SITS data, pixel-wise methods are significantly outperformed by whole-image approaches, showing the importance of spatial information for the looting detection task.

Ablation. We further evaluate whether methods can learn temporal cues from DAFA-LS despite the lack of frame-wise annotations. To achieve this, we use our best-performing baseline (DOFA+LTAE) and perform inference on subsets of the time series. We first constructed yearly time series by taking images from a specific month across all available years. As shown in Figure 3a, using yearly month-specific sub-series results in performance that is similar or worse compared to using all available time stamps. Notably, inferences made on spring or autumnal time series are comparable to those made on all-month time series in terms of F1 score. This experiment demonstrates (i) that season — and likely vegetation — affects detection performance, and (ii) that the temporal attention mechanism of LTAE can prioritize informative time stamps when using monthly time series. Second, we conducted inference on monthly year-specific time series. A similar conclusion can be drawn from the results shown in Figure 3b. We observe a performance peak using the 2020 and 2021 sub-time series, suggesting that looting-related changes (appearance of marks or scars) likely occurred during this period. A further analysis of the contribution of temporal data to performance is available in Section B of the appendix.

5 Limitations and Discussion

Our work comes with several limitations. First, all archaeological sites are located in Afghanistan, a relevant case study for which we have reliable looting annotations. Consequently, most images are of dry and desertic areas. The conclusions drawn on our dataset will likely not hold in regions with dense vegetation, where archaeological sites are hidden beneath the canopy, and other modalities like SAR or LiDAR may be required.

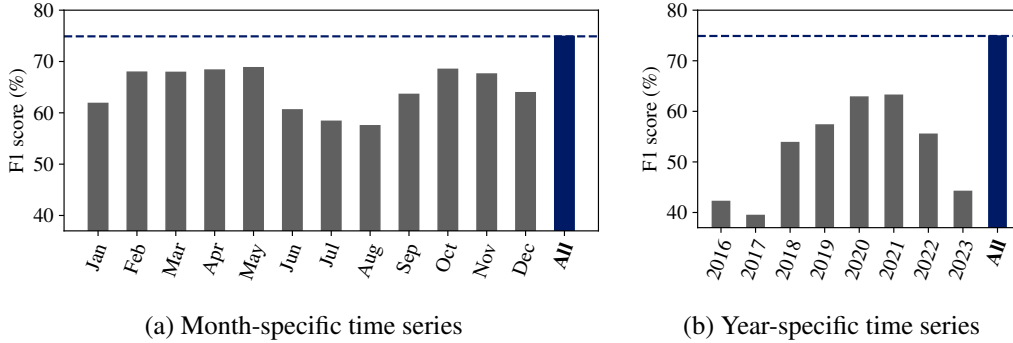


Figure 3: **Ablation of DOFA+LTAE.** We evaluate DOFA+LTAE on either month-specific (a) or year-specific (b) sub-time series of DAFA-LS. In other words, instead of taking all monthly time stamps as input, we only take images of a given month (across all years) or of a given year.

Second, to avoid biases, we chose to discard pixels outside the coarse location mask in most of our baselines. While these pixels may not provide direct evidence of looting, analyzing the surroundings of archaeological sites could offer insights into factors that increase the likelihood of looting. DAFA-LS includes a 1 km² area around each site, allowing for such analysis in future research.

Third, although the weak annotations in DAFA-LS make it a challenging dataset, the classification task could benefit from temporal annotations of the damage (*i.e.*, the dates when marks appear) and spatial labeling of looting scars at the pixel level. However, higher resolution imagery would likely be necessary for this purpose, but it is not freely accessible and is typically not available at high temporal resolution.

Finally, we only considered supervised baseline approaches, but given the lack of fine-grained annotations, unsupervised change detection methods might be useful for discovering looting events. We believe the variability of our image would make unsupervised approaches challenging, but we hope DAFA-LS will encourage investigations of such methods and their potential combination with the available weak annotation labels.

Potential negative societal impact. The purpose of this dataset is to encourage the development of efficient methods for detecting looted archaeological sites. The primary goal is the automatic and remote monitoring of known sites to preserve the cultural heritage they represent. To prevent misuse by malevolent individuals or organizations, we do not release the GPS coordinates of the sites in DAFA-LS, added random noise to the point coordinates before plotting the maps in Figures 1a & 1b and added random geometric transformations to the SITS we distribute.

Ethical statement. Following the template of Sandbrook et al. (2021) [71], we have ensured that DAFA-LS complies with the set of ethical principles outlined below:

(a) *Recognize and acknowledge:* We recognize and acknowledge that the release of DAFA-LS can have social impacts. In particular, a looting dataset has additional security considerations, and may cause undue harm to individuals.

(b) *Necessity and proportionality:* Archaeological sites all over the world face natural and anthropological threats. The rich literature about site monitoring confirms it is necessary to ensure their protection as they constitute a cultural heritage. The use of freely available satellite data to flag their potential looting is a proportional action to address this conservation problem. The interest of a satellite monitoring tool is to identify looting activities as quickly as possible in order to inform the country’s authorities, international experts (UNESCO, ICOMOS), and the general public.

(c) *Potential impacts on people:* In a politically unstable country such as Afghanistan, looting detection exposes the locals at risk of unforeseen legal consequences. There exists a risk to privacy, to safety and security of locals, including Human rights and to irresponsible use, and violation of legal compliance.

(d) *Consent from people*: DAFA-LS is built on authorized and official digs and observations made by Philippe Marquis (co-author), as a member of the DAFA (Délégation Archeologique Francaise en Afghanistan). French DAFA archaeologists have been helping local archaeologists since 1922, in answer to a wish from King Amanullah, which was renewed by the President of the Republic, Ashraf Ghani, in the 2010s [10]. The “*Traité d’amitié et de coopération entre la République française et la République islamique d’Afghanistan*” [26], ratified in 2012 for a duration of 20 years, reinforces the legitimacy of DAFA’s presence and action in Afghanistan. At the moment of writing, the *de facto* Taliban regime has not yet reintegrated into the international community of the UN. An institution like the DAFA therefore cannot officially initiate scientific collaborations with its Afghan counterparts. Thus, no Afghan authority can be mentioned at this stage of the work.

(e) *Transparency and accountability*: DAFA-LS is an open-access dataset created from freely available satellite images. All related code is made open-source on the official GitHub repository linked to the dataset. The spatial resolution of the images ($> 3\text{m/px}$) makes it impossible to identify individuals or their personal vehicle. The images represent an area of 1 km^2 centered on archaeological sites. Therefore, DAFA-LS cannot be used for another purpose than archaeological preservation.

(f) *Peoples’ rights and vulnerabilities*: The destruction of cultural heritage is prohibited in Afghanistan and has repeatedly been considered a war crime by the International Criminal Court [41]. These acts primarily harm the Afghan people by erasing traces of their past, which are crucial for national cohesion and the country’s future. Afghan communities living near archaeological sites could also suffer from the malicious destruction and looting of these sites.

6 Conclusion

We release DAFA-LS, a new dataset for detecting looted archaeological sites. The dataset contains monthly satellite image time series centered on sites in Afghanistan. We use this dataset to evaluate SITS classification approaches, including single-frame methods and pixel-wise or whole-image multi-frame methods. We present a strong baseline that combines the large foundation model DOFA with the LTAE architecture, specifically designed for SITS. We hope this work will encourage the community to improve classification performance on archaeological satellite data, as it is crucial for the preservation of our cultural heritage.

Acknowledgments and Disclosure of Funding

The work of Mathieu Aubry was supported by the European Research Council (ERC project DISCOVER, number 101076028). Jean Ponce was supported by the Louis Vuitton/ENS chair on artificial intelligence and the French government under management of Agence Nationale de la Recherche as part of the *Investissements d’avenir* program, reference ANR19-P3IA0001 (PRAIRIE 3IA Institute). This work was granted access to the HPC resources of IDRIS under the allocation 2024-AD011015272 made by GENCI. We thank Titien Bartette, Charlotte Fafet and Loïc Landrieu for their valuable feedbacks, and Guillaume Astruc and Lucas Ventura for their careful proofreading.

References

- [1] AAAS American Association for the Advancement of Science. Ancient history, modern destruction: Assessing the current status of syria’s world heritage sites using high-resolution satellite imagery. *aaas.org*, 2014.
- [2] D. Abate, M. Faka, C. Keleshis, C. Constantinides, A. Leonidou, and A. Papageorgiou. Aerial image-based documentation and monitoring of illegal archaeological excavations. *Heritage*, 6(5):4302–4319, 2023.
- [3] A. Agapiou. Detecting looting activity through earth observation multi-temporal analysis over the archaeological site of apamea (syria) during 2011–2012. *Journal of Computer Applications in Archaeology*, 3(1):219–237, 2020.
- [4] A. Agapiou, V. Lysandrou, and D. G. Hadjimitsis. Optical remote sensing potentials for looting detection. *Geosciences*, 7(4):98, 2017.
- [5] M. Altaweel, A. Khelifi, and M. M. Shana’ah. Monitoring looting at cultural heritage sites: Applying deep learning on optical unmanned aerial vehicles data as a solution. *Social Science Computer Review*, 42(2):480–495, 2024.

- [6] G. Astruc, N. Gonthier, C. Mallet, and L. Landrieu. Omnisat: Self-supervised modality fusion for earth observation. *arXiv preprint arXiv:2404.08351*, 2024.
- [7] W. Ball and J.-C. Gardin. *Archaeological gazetteer of Afghanistan= Catalogue des sites archéologiques d’Afghanistan*. Editions Recherche sur les civilisations, 1982.
- [8] F. Bastani, P. Wolters, R. Gupta, J. Ferdinando, and A. Kembhavi. Satlaspretrain: A large-scale dataset for remote sensing image understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16772–16782, 2023.
- [9] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7–13, 2006. Proceedings, Part I* 9, pages 404–417. Springer, 2006.
- [10] J. Bendezu-Sarmiento. Préserver le passé pour construire l’avenir : un siècle de recherches à la Délégation archéologique française en Afghanistan (DAFA). *ArchéOrient – Le Blog*, 2018. Publié le 18 mai 2018, mis à jour le 22 mai 2018.
- [11] E. F. Bowen, B. B. Tofel, S. Parcak, and R. Granger. Algorithmic identification of looted archaeological sites from space. *Frontiers in ICT*, 4:247381, 2017.
- [12] J. Casana. Satellite imagery-based analysis of archaeological looting in syria. *Near Eastern Archaeology*, 78(3):142–152, 2015.
- [13] J. Casana and E. J. Laugier. Satellite imagery-based monitoring of archaeological site damage in the syrian civil war. *PLoS one*, 12(11):e0188589, 2017.
- [14] J. Casana and M. Panahipour. Satellite-based monitoring of looting and damage to archaeological sites in syria. *Journal of Eastern Mediterranean Archaeology & Heritage Studies*, 2(2):128–151, 2014.
- [15] G. Castilla, R. H. Guthrie, and G. J. Hay. The land-cover change mapper (lcm) and its application to timber harvest monitoring in western canada. *Photogrammetric Engineering & Remote Sensing*, 75(8):941–950, 2009.
- [16] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer. Smoteboost: Improving prediction of the minority class in boosting. In *Knowledge Discovery in Databases: PKDD 2003: 7th European Conference on Principles and Practice of Knowledge Discovery in Databases, Cavtat-Dubrovnik, Croatia, September 22–26, 2003. Proceedings 7*, pages 107–119. Springer, 2003.
- [17] R. Coluzzi, R. Lasaponara, and N. Masini. Satellite imagery time series for the detection of looting activities at archaeological sites. In *EGU General Assembly Conference Abstracts*, page 10569, 2010.
- [18] Y. Cong, S. Khanna, C. Meng, P. Liu, E. Rozi, Y. He, M. Burke, D. Lobell, and S. Ermon. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35:197–211, 2022.
- [19] D. A. Contreras and N. Brodie. The utility of publicly-available satellite imagery for investigating looting of archaeological sites in jordan. *Journal of Field Archaeology*, 35(1):101–114, 2010.
- [20] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)*, volume 1, pages 886–893. Ieee, 2005.
- [21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [22] P. Ebel, Y. Xu, M. Schmitt, and X. X. Zhu. Sen12ms-cr-ts: A remote-sensing data set for multimodal multitemporal cloud removal. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.
- [23] B. Ekim, T. T. Stomberg, R. Roscher, and M. Schmitt. Mapinwild: A remote sensing dataset to address the question of what makes nature wild [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, 11(1):103–114, 2023.
- [24] H. El Hajj. Interferometric sar and machine learning: Using open source data to detect archaeological looting and destruction. *Journal of Computer Applications in Archaeology*, 4(1), 2021.
- [25] J. Follorou. Looting of Afghanistan archaeological site attributed to IS. *Le Monde*, 2023.
- [26] R. Française. LOI n° 2012-947 du 2 août 2012 autorisant la ratification du traité d’amitié et de coopération entre la république française et la république islamique d’afghanistan. *Journal Officiel de la République Française*, page 12955, 2012. Accessed from Legifrance on [date of access].
- [27] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.
- [28] A. Fuller, K. Millard, and J. R. Green. Satvit: Pretraining transformers for earth observation. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

- [29] A. Garioud, N. Gonthier, L. Landrieu, A. De Wit, M. Valette, M. Poupée, S. Giordano, et al. Flair: a country-scale land cover semantic segmentation dataset from multi-source optical imagery. *Advances in Neural Information Processing Systems*, 36, 2024.
- [30] V. S. F. Garnot and L. Landrieu. Lightweight temporal self-attention for classifying satellite images time series. In *Advanced Analytics and Learning on Temporal Data: 5th ECML PKDD Workshop, AALTD 2020, Ghent, Belgium, September 18, 2020, Revised Selected Papers 6*, pages 171–181. Springer, 2020.
- [31] V. S. F. Garnot and L. Landrieu. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4872–4881, 2021.
- [32] V. S. F. Garnot, L. Landrieu, S. Giordano, and N. Chehata. Satellite image time series classification with pixel-set encoders and temporal self-attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12325–12334, 2020.
- [33] S. Gerard, Y. Zhao, and J. Sullivan. Wildfirespreads: A dataset of multi-modal time series for wildfire spread prediction. *Advances in Neural Information Processing Systems*, 36:74515–74529, 2023.
- [34] X. Guo, J. Lao, B. Dang, Y. Zhang, L. Yu, L. Ru, L. Zhong, Z. Huang, K. Wu, D. Hu, et al. Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27672–27683, 2024.
- [35] J. Han and K.-K. Ma. Rotation-invariant and scale-invariant gabor features for texture image retrieval. *Image and vision computing*, 25(9):1474–1481, 2007.
- [36] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.
- [37] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [38] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [39] T. K. Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.
- [40] R. Interdonato, D. Ienco, R. Gaetano, and K. Ose. Duplo: A dual view point deep learning architecture for time series classification. *ISPRS journal of photogrammetry and remote sensing*, 149:91–104, 2019.
- [41] International Criminal Court. Policy on cultural heritage. In *The Office of the Prosecutor*, pages 1–48, 2021.
- [42] D. Kennedy and R. Bewley. Aerial archaeology in Jordan. *Antiquity*, 83(319):69–81, 2009.
- [43] L. Kondmann, A. Toker, M. Rußwurm, A. Camero Unzueta, D. Peressuti, G. Milcinski, N. Longépé, P.-P. Mathieu, T. Davis, G. Marchisio, et al. Denethor: The dynamicearthnet dataset for harmonized, inter-operable, analysis-ready, daily crop monitoring from space. In *35th Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, pages 1–13, 2021.
- [44] R. Lasaponara, G. Leucci, N. Masini, and R. Persico. Investigating archaeological looting using satellite images and georadar: the experience in Lambayeque in north Peru. *Journal of Archaeological Science*, 42:216–230, 2014.
- [45] R. Lasaponara and N. Masini. Space-based identification of archaeological illegal excavations and a new automatic method for looting feature extraction in desert areas. *Surveys in Geophysics*, 39:1323–1346, 2018.
- [46] X. Li, D. Hong, and J. Chanussot. S2mae: A spatial-spectral pretraining foundation model for spectral remote sensing data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24088–24097, 2024.
- [47] Z. Li, B. Hou, S. Ma, Z. Wu, X. Guo, B. Ren, and L. Jiao. Masked angle-aware autoencoder for remote sensing images. *arXiv preprint arXiv:2408.01946*, 2024.
- [48] Z. Li, F. Lu, H. Zhang, L. Tu, J. Li, X. Huang, C. Robinson, N. Malkin, N. Jovic, P. Ghamisi, et al. The outcome of the 2021 IEEE GRSS Data Fusion Contest—Track MSD: Multitemporal semantic change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:1643–1655, 2022.
- [49] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [50] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. IEEE, 1999.

- [51] Y. Ma, Y. Li, K. Feng, Y. Xia, Q. Huang, H. Zhang, C. Prieur, G. Licciardi, H. Malha, J. Chanussot, et al. The outcome of the 2021 ieeegrss data fusion contest-track dse: Detection of settlements without electricity. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:12375–12385, 2021.
- [52] B. S. Manjunath and W.-Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on pattern analysis and machine intelligence*, 18(8):837–842, 1996.
- [53] N. Masini and R. Lasaponara. Recent and past archaeological looting by satellite remote sensing: Approach and application in syria. *Remote Sensing for Archaeology and Cultural Landscapes: Best Practices and Perspectives Across Europe and the Middle East*, pages 123–137, 2020.
- [54] N. Masini and R. Lasaponara. Remote and close range sensing for the automatic identification and characterization of archaeological looting. the case of peru. *Journal of Computer Applications in Archaeology*, 4(1), 2021.
- [55] A. A. Nielsen, K. Conradsen, and J. J. Simpson. Multivariate alteration detection (mad) and maf postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. *Remote Sensing of Environment*, 64(1):1–19, 1998.
- [56] M. Noman, M. Naseer, H. Cholakkal, R. M. Anwer, S. Khan, and F. S. Khan. Rethinking transformers pre-training for multi-spectral satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27811–27819, 2024.
- [57] S. Parcak. Satellite remote sensing methods for monitoring archaeological tells in the middle east. *Journal of field archaeology*, 32(1):65–81, 2007.
- [58] S. Parcak. *Satellite remote sensing for archaeology*. Routledge, 2009.
- [59] S. Parcak, D. Gathings, C. Childs, G. Mumford, and E. Cline. Satellite evidence of archaeological site looting in egypt: 2002–2013. *Antiquity*, 90(349):188–205, 2016.
- [60] N. D. Payntar. A multi-temporal analysis of archaeological site destruction using landsat satellite data and machine learning, moche valley, peru. *ACM Journal on Computing and Cultural Heritage*, 16(3):1–20, 2023.
- [61] C. Pelletier, G. I. Webb, and F. Petitjean. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, 11(5):523, 2019.
- [62] Planet Labs Inc. Planet Basemaps Product Specification. *Planet.com*, 2019.
- [63] Planet Team. Planet Application Program Interface: In Space for Life on Earth (San Francisco, CA). <https://api.planet.com>, 2024.
- [64] L. Rayne, M. C. Gatto, L. Abdulaati, M. Al-Haddad, M. Sterry, N. Sheldrick, and D. Mattingly. Detecting change at archaeological sites in north africa using open-source satellite imagery. *Remote Sensing*, 12(22):3694, 2020.
- [65] C. J. Reed, R. Gupta, S. Li, S. Brockman, C. Funk, B. Clipp, K. Keutzer, S. Candido, M. Uyttendaele, and T. Darrell. Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4088–4099, 2023.
- [66] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.
- [67] M. Rußwurm and M. Körner. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, 7(4):129, 2018.
- [68] M. Rußwurm and M. Körner. Self-attention for raw optical satellite time series classification. *ISPRS journal of photogrammetry and remote sensing*, 169:421–435, 2020.
- [69] M. Rußwurm, S. Lefèvre, and M. Körner. Breizhcrops: A satellite time series dataset for crop type identification. In *Proceedings of the International Conference on Machine Learning Time Series Workshop*, volume 3, 2019.
- [70] R. Rustowicz, R. Cheong, L. Wang, S. Ermon, M. Burke, and D. Lobell. Semantic segmentation of crop type in africa: A novel dataset and analysis of deep learning methods. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 75–82, 2019.
- [71] C. Sandbrook, D. Clark, T. Toivonen, T. Simlai, S. O’Donnell, J. Cobbe, and W. Adams. Principles for the socially responsible use of conservation monitoring technology and data. *Conservation Science and Practice*, 3(5):e374, 2021.
- [72] E. C. Stone. Patterns of looting in southern iraq. *Antiquity*, 82(315):125–138, 2008.
- [73] E. C. Stone. An update on the looting of archaeological sites in iraq. *Near Eastern Archaeology*, 78(3):178–186, 2015.

- [74] D. Tapete, F. Cigna, and D. N. Donoghue. ‘looting marks’ in space-borne sar imagery: Measuring rates of archaeological looting in apamea (syria) with terrasars-x staring spotlight. *Remote Sensing of Environment*, 178:42–58, 2016.
- [75] M. Tarasiou, E. Chavez, and S. Zafeiriou. Vits for sits: Vision transformers for satellite image time series. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10418–10428, 2023.
- [76] M. Tarasiou, R. A. Güler, and S. Zafeiriou. Context-self contrastive pretraining for crop type semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022.
- [77] D. Thomas and A. Gascoigne. Recent archaeological investigations of looting around the minaret of jam. In *Art and Archaeology of Afghanistan*, pages 155–167. Brill, 2006.
- [78] D. Thomas, F. Kidd, S. Nikolovski, and C. Zipfel. The archaeological sites of afghanistan in google earth. *AARGnews*, 37(September):22–30, 2008.
- [79] A. Toker, L. Kondmann, M. Weber, M. Eisenberger, A. Camero, J. Hu, A. P. Hoderlein, Ç. Şenaras, T. Davis, D. Cremers, et al. Dynamicearthnet: Daily multi-spectral satellite dataset for semantic change segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21158–21167, 2022.
- [80] UNESCO World Heritage Centre. 1928: City of Balkh (antique Bactria). *World Heritage Tentative List*, 2004.
- [81] A. Van Etten, D. Hogan, J. M. Manso, J. Shermeyer, N. Weir, and R. Lewis. The multi-temporal urban development spacenet dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6398–6407, 2021.
- [82] S. Verma, A. Panigrahi, and S. Gupta. Qfabric: Multi-task change detection dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1052–1061, 2021.
- [83] E. Vincent, J. Ponce, and M. Aubry. Pixel-wise agricultural image time series classification: Comparisons and a deformable prototype-based approach. *arXiv preprint arXiv:2303.12533*, 2023.
- [84] G. Weikmann, C. Paris, and L. Bruzzone. Timesen2crop: A million labeled samples dataset of sentinel 2 image time series for crop-type classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:4699–4708, 2021.
- [85] R. Wenger, A. Puissant, J. Weber, L. Idoumghar, and G. Forestier. Multisenge: a multimodal and multitemporal benchmark dataset for land use/land cover remote sensing applications. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:635–640, 2022.
- [86] Z. Xiong, Y. Wang, F. Zhang, A. J. Stewart, J. Hanna, D. Borth, I. Papoutsis, B. L. Saux, G. Camps-Valls, and X. X. Zhu. Neural plasticity-inspired foundation model for observing the Earth crossing modalities. *arXiv preprint arXiv:2403.15356*, 2024.
- [87] F. Zaina and Y. Nabati Mazloumi. A multi-temporal satellite-based risk analysis of archaeological sites in gazvin plain (iran). *Archaeological prospection*, 28(4):467–483, 2021.

Appendix

We first provide additional visualizations (Section A) and temporal analysis (Section B), and then complete the benchmark with implementations details (Section C) and further clarifications (Section D). Our official repository and information on how to download the dataset can be found at <https://github.com/ElliottVincent/DAFA-LS>.

A Additional visualizations

A.1 Examples of looting marks

We show in Figure A1 three examples of visible looted marks. For the selected sites, we show images before and after the looting, with a zoom on the damaged area. The scars are typical of mechanical looting performed with bulldozers for example.

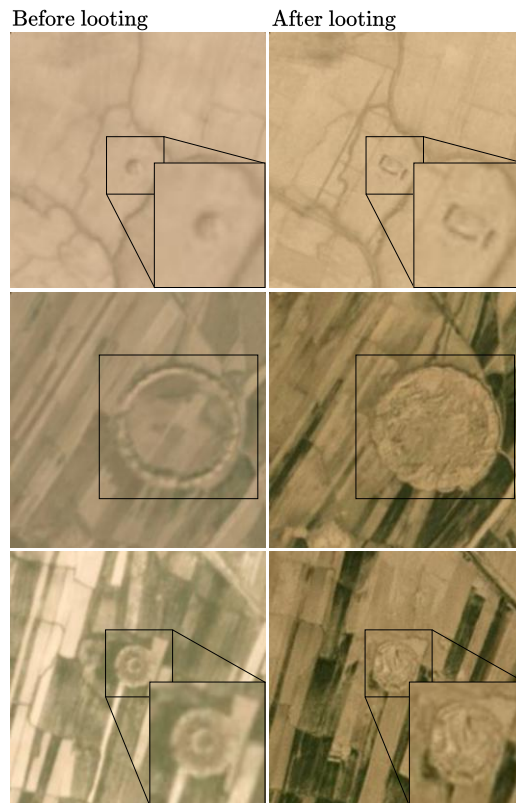


Figure A1: **Example of visible looting marks.**

A.2 Failure cases

We report in Figure A2 all the time series for which our best baseline (DOFA+LTAE) predicts the wrong label with a confidence higher than 95%. We note that, for these sites, it is very difficult for a non-expert human eye to identify looting marks if any, or even to clearly see the structure of an archaeological site.

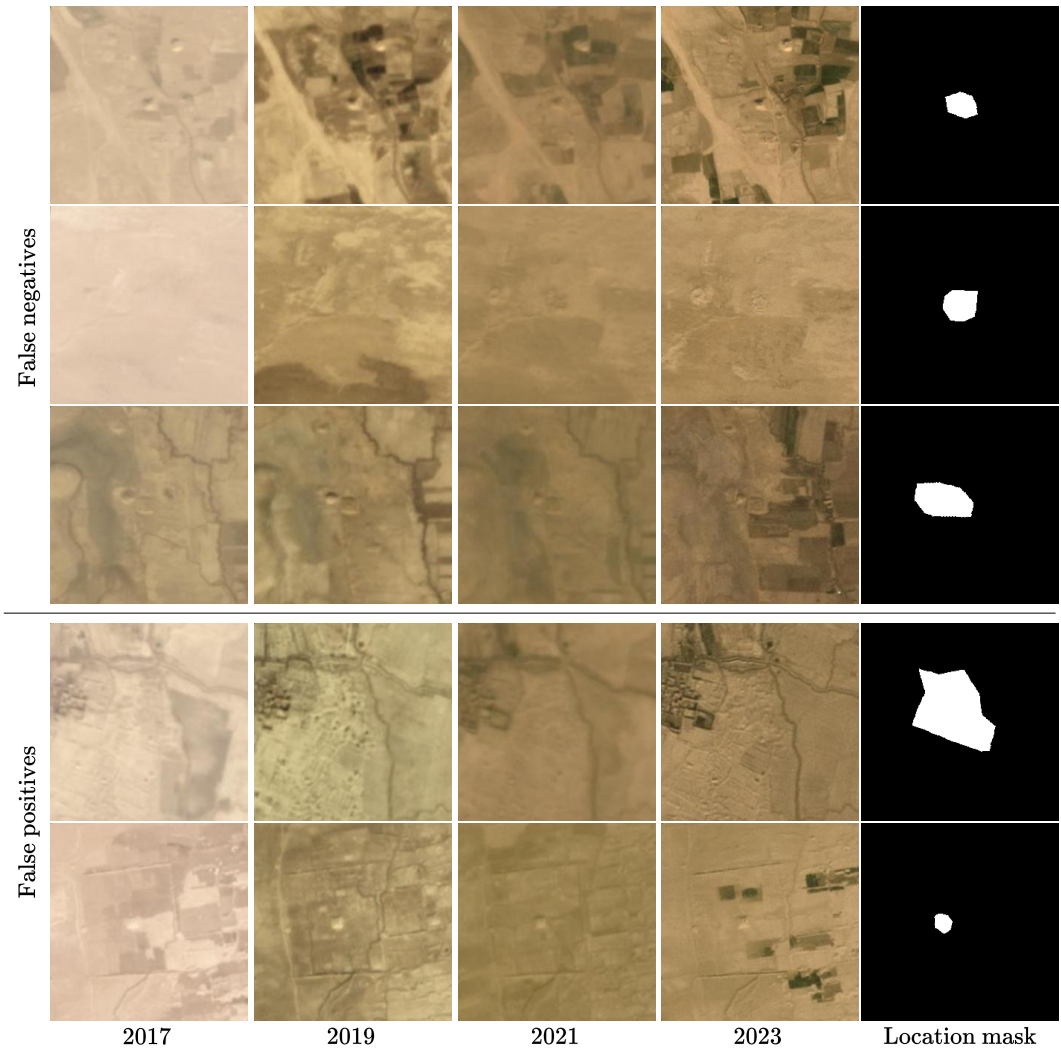


Figure A2: **Example of failure cases.** We show all the time series for which our best baseline (DOFA+LTAE) predicts the wrong label with a confidence higher than 95%.

B Temporal analysis

B.1 Additional ablation

We provide in Figure B1 a third ablation, using time series of various lengths at inference for our best performing baseline, DOFA+LTAE. Time series span periods from 20XX to 2023. We observe that the performance consistently decreases as the series becomes shorter.

B.2 Attention weights

We visualize in Figure B2 the temporal attention weights, gathered by year and averaged over all looted test sites, for two attention heads of DOFA+LTAE. We can see that the years 2020 and 2021 draw more attention from the model.

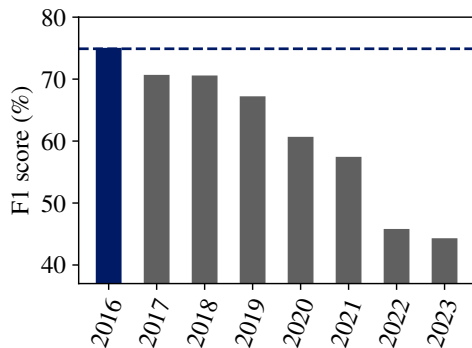


Figure B1: **Inference with shorter sequence length.** We perform inferences with various temporal range using DOFA+LTAE. We use time series spanning period from 20XX to 2023 with 20XX being the year indicated on the x-axis.

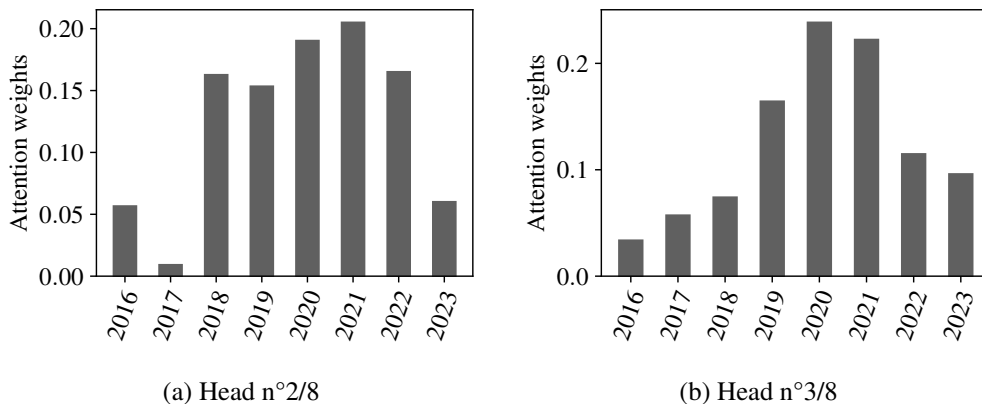


Figure B2: **Temporal attention.** We report the attention weights for two of the eight heads of DOFA+LTAE. We report the mean of the weights for all looted test sites. We sum all the monthly weights for each year.

C Implementation details

C.1 Single-frame methods

We use the PyTorch implementation of ResNet20 for CIFAR-10 by Yerlan Idelbayev⁴, the torchvision implementation of ResNet18 and ResNet34, and the official PyTorch implementation of SatMAE⁵, Scale-MAE⁶ and DOFA⁷. We use a base version of foundations models when available (DOFA) and a large version otherwise (SatMAE, Scale-MAE). The models are trained with a learning rate of 10^{-3} for 60 epochs and a batch size of 32 (13,380 iterations). ResNet20 is trained from scratch. We train a single linear layer on top of a pretrained frozen DOFA.

C.2 Pixel-wise multi-frame methods

We use the PyTorch implementations of DuPLO and TempCNN available in Transformer’s official public repository⁸ and the PyTorch implementation of LTAE available in UTAE’s official public

⁴https://github.com/akamaster/pytorch_resnet_cifar10

⁵<https://github.com/sustainlab-group/SatMAE>

⁶<https://github.com/bair-climate-initiative/scale-mae>

⁷<https://github.com/zhu-xlab/DOFA>

⁸<https://github.com/MarcCoru/crop-type-mappin>

repository⁹. These models are trained with a learning rate of 10^{-4} for 2 epochs and a batch size of 128 (25,000 iterations). Each batch contains 128 pixel-wise time series, sampled from possibly different SITS. For all these methods, we use smaller versions of the architecture compared to their default setting, taking into account the relative small size of DAFA-LS and to limit over-fitting. The used configurations can be found in our official repository¹⁰.

C.3 Whole-image multi-frame methods

We use the official PyTorch implementations of UTAE¹¹ and TSViT¹². These models are trained with a learning rate of 10^{-4} for 100 epochs and a batch size of 4 (14,900 iterations). For PSE+LTAE, SatMAE+LTAE, Scale-MAE+LTAE and DOFA+LTAE, we use SatMAE, Scale-MAE, DOFA and LTAE implementations mentioned above and the official implementation of PSE¹³. These models are trained with a learning rate of 10^{-4} for 200 epochs and a batch size of 8 (14,900 iterations). For PSE+LTAE, 1024 randomly sampled in-mask pixels are used during training and all in-mask pixels are used at inference. A majority voting rule is applied at inference to determine the final prediction for a given SITS. For SatMAE+LTAE, Scale-MAE+LTAE and DOFA+LTAE, the backbone is loaded with pretrained weights and frozen during training. We use smaller versions UTAE and TSViT compared to their default setting, taking into account the relative small size of DAFA-LS and to limit over-fitting. The used configurations can be found in our official repository¹⁴.

D Benchmark: additional details

We detail in Table D1 the main characteristics of the evaluated baselines. In particular, we report the dates available at training time (*Dates used*), the learning task (*Task*), and the inference procedure (*Inference strategy*).

Table D1: **Categorization of baseline methods.** We explicit the main characteristics of the different evaluated baseline methods.

Id	Category	Image level	Dates used	Input type	Task	Inference strategy
(i)	Single-frame	Whole-image	2023	Image	Classification	12-month image voting
(ii)	Multi-frame	Pixel-wise	2016-2023	SITS	Classification	In-mask pixel voting
(iii)	Multi-frame	Whole-image	2016-2023	SITS	Segmentation	In-mask pixel voting
(iv)	Multi-frame	Whole-image	2016-2023	SITS	Classification	Direct prediction
Id	Model name					
(i)	ResNet20/18/34 [38], SatMAE [18], Scale-MAE [65], DOFA [86]					
(ii)	DuPLo [40], Transformer [68], LTAE [30], TempCNN [61]					
(iii)	UTAE [31], TSViT (seg. head) [75]					
(iv)	{PSE [32], SatMAE [18], Scale-MAE [65], DOFA [86]}+LTAE [30], TSViT (cls. head) [75]					

⁹<https://github.com/VSainteuf/utae-paps>

¹⁰<https://github.com/ElIiotVincent/DAFA-LS>

¹¹See footnote 4.

¹²<https://github.com/michaeltrs/DeepSatModels>

¹³<https://github.com/VSainteuf/pytorch-psetae>

¹⁴See footnote 5