No Annotations for Object Detection in Art through Stable Diffusion
Supplementary Material

A. Datasets

An overview of the two art object dection datasets, ArtDL 2.0 [2] and IconArt [1], is provided in Tab. 1. Both of the datasets consist of images of paintings containing Christian icons.

B. WSCP training hyperparameters

We present the hyperparameters for training the lightweight MLP in the WSCP in Tab. 2.

C. Prompts

We detail the various prompts used in NADA.

C.1. ZSCP

We present the prompts (choice and score) used to prompt the VLM to classify images in the ZSCP in Tab. 3.

C.2. Prompt construction

We present the classes, prompts, templates used in the prompt construction for image reconstruction in the class-conditioned detector.

Class names For each class in ArtDL, we use the title of its equivalent Wikipedia article, resulting in the following classes:

> *Anthony of Padua*; *John the Baptist*; *Paul the Apostle*; *Francis of Assisi*; *Mary Magdalene*; *Saint Jerome*; *Saint Dominic*; *Mary, mother of Jesus*; *Saint Peter*; *Saint Sebastian*

Meanwhile for IconArt, we use the following texts for the classes:

> *person* (equivalent to *Saint Sebastian*), *crucifixion of jesus*, *angel*, *mary*, *baby* (equivalent to *child jesus*), *naked person* (equivalent to *nudity*), *ruins*

Template By default we insert the class in the simple prompt `A painting of [CLASS]`, where `[CLASS]` is the class being detected. For classes *person*, *baby*, and *naked person*, we use `A painting of a [CLASS].`

Caption We prompt the same VLM used to classify the images in NADA (with ZSCP) to instead caption the images using the prompt `Describe the visual elements in the image in one sentence. Include the term "[CLASS]".` If the class is not found in the caption or is located at a part of the caption that is beyond the maximum input length of the diffusion model, we prepend the caption with the prompt `A painting of [CLASS].` formatted with the class name.

D. Per-class detection results

We present the AP_{50} per class for ArtDL 2.0 in Tab. 4. No class is detected the easiest or hardest across all experimental settings. When comparing methods, NADA (with WSCP) provides near consistent gains in AP_{50} over NADA (with ZSCP), improving AP_{50} in all classes except for Mary and boosting detection performance within the same class by 24.1 AP_{50} on average. Intuitively, Oracle has the best performance across all classes.

Per-class IconArt results are provided in Tab. 5. NADA consistently detects Crucifixion of Jesus the best, but struggles to detect nudity and angel relative to other clases in all experimental settings. Furthermore, NADA (with ZSCP) outperforms NADA (with WSCP) on only four of the seven classes, with both methods having the same AP_{50} on angel. Differences between class proposer are smaller, as NADA (with ZSCP) provides only a 1.2 AP_{50} improvement over NADA (with WSCP). While Oracle proves the best overall AP_{50}, it actually underperforms NADA on Crucufixion of Jesus, angel, and Mary.

E. Qualitative analysis

In Figure 4 of the main paper, from left to right, top to bottom: samples 1, 2, 5, and 6 are from ArtDL 2.0 and samples 3, 4, 7, and 8 are from IconArt.

Table 1. Details of the evaluation datasets. ArtDL 2.0 and IconArt provide different splits for classification and detection evaluation.

	ArtDL 2.0 [2]	IconArt [1]
Type of art	Paintings	Paintings
Type of objects	Christian icons	Christian icons
Num. object classes	10	7
Num. train images - classification	21,673	1,421
Num. test images - classification	2,632	2,031
Num. test images - detection	808	1,480
Num. validation images - classification	2,628	610
Num. validation images - detection	1,625	-

Table 2. Hyperparameters for training the MLP classifier in NADA (with WSCP). LR is learning rate and WD is weight decay.

Dataset	Layers	Classification	Loss	LR	WD	Classes
ArtDL 2.0 [2]	2	single-label	cross-entropy	1e−4	0	10
IconArt [1]	3	multi-label	binary cross-entropy	1e−3	1e−3	7

Table 3. Prompts used in the ZSCP of NADA (with ZSCP). `[CLASSES]` refers to the list of classes.

Prompt	Dataset	Contents
Choice	ArtDL 2.0 [2]	`Who is in the painting? Choose from the following: [CLASSES]`
Choice	IconArt [1]	`Which of the options are in the painting? Choose from the following: [CLASSES]`
Score	all datasets	`Which of the Christian iconographic symbols are in the painting? Choose from the following: [CLASSES] For each symbol, give a score from 0 to 1 of how confident you are. Put your answer in a dictionary first and then reason your answer. Be as accurate as possible. If none of the symbols are present, output ’None’`

Table 4. AP$_{50}$ for each class in ArtDL 2.0. *Mean* refers to the overall AP$_{50}$ reported in the main paper.

Class Proposal	Antony of Padua	John the Baptist	Paul	Francis	Mary Magdalene	Jerome	Dominic	Mary	Peter	Sebastian	Mean
NADA (with WSCP)	29.5	35.1	26.7	50.7	60.1	58.3	51.3	55.5	40.2	51.5	45.8
NADA (with ZSCP)	7.6	21.1	2.5	15.6	24.3	30.2	7.7	60.0	3.9	45.5	21.8
Oracle	42.0	40.8	79.3	56.2	80.8	68.3	55.8	68.5	54.5	66.5	61.3

Table 5. AP$_{50}$ for each class in IconArt 2.0.

Class Proposal	Saint Sebastian	Crucifixion of Jesus	Angel	Mary	Child Jesus	Nudity	Ruins	Mean
NADA (with WSCP)	6.8	47.9	0.4	15.2	14.0	3.4	9.1	13.8
NADA (with ZSCP)	11.7	43.1	0.4	20.7	15.0	2.2	12.3	15.1
Oracle	21.0	45.8	0.3	20.3	17.5	5.4	20.3	18.7

References

[1] Nicolas Gonthier, Yann Gousseau, Said Ladjal, and Olivier Bonfait. Weakly supervised object detection in artworks. In *VISART Workshop at ECCV*, 2018. 1, 2

[2] Federico Milani, Nicolò Oreste Pinciroli Vago, and Piero Fraternali. Proposals generation for weakly supervised object detection in artwork images. *J. Imaging*, 8(8), 2022. 1, 2