

Supplementary Material: CHAI, Craters in Historical Aerial Images

Marvin Burges
Computer Vision Lab
TU Wien, Austria

mburges@cvl.tuwien.ac.at

Sebastian Zambanini
Computer Vision Lab
TU Wien, Austria

zamba@cvl.tuwien.ac.at

Philipp Pirker
Luftbilddatenbank Dr. Carls GmbH
Austria

pirker@luftbilddatenbank-gmbh.at

1. Acquisition of Historical Images

The images within this dataset are sourced from our industry partner, who gathers them from various outlets, predominantly national archives. In essence, the process of procuring these images can be broken down into three key steps. First, a contractor defines the area of interest to investigate for a particular construction process. Second, preview images from the approximate region are ordered. These preliminary images are typically low-resolution scans of microfilms. Figure 1 presents an example of a microfilm (Figure 1a) and an enlarged view of the center image (Figure 1b). These microfilms serve as a reference for domain experts to evaluate aspects like image quality – for instance, identifying extensive cloud cover in the top section of the rightmost image – or to determine if an image falls within the region of interest. However, due to the extremely limited resolution of these images, intricate identification of war-related elements such as bomb craters is unfeasible. Consequently, following the assessment, high-resolution scans are requested.

Figure 4 provides an overview of the geographical distribution of the images within our dataset in Austria and Germany. In Table 6, we present a comprehensive list of the images included in the dataset, along with their rough locations, original Ground Sampling Distance (GSD), dates, image splits, and the number of annotations present in each split. To illustrate the dataset further, we showcase one example image in Figure 5. This example contains patch candidates highlighted in red, the region of interest marked in green, and crater annotations drawn in blue.

Examples of the final **light** dataset are presented in Figures 6, 7, 8, 9, 10 and 11, and display in total 60 patches per dataset split, randomly sampled. The high variety of historical aerial images is clearly visible, and one can see that image quality is consistently high, with a very low percentage of cloud coverage.

2. Inconsistent Annotations & Mitigation

Due to the variations observed in crater annotations, including slight differences and instances without craters, a decision was made to enhance the accuracy of the bounding boxes through manual adjustments. A total of 2,621 bounding boxes were either resized or repositioned, while 497 were eliminated. This refining process was carried out by the authors and validated by a domain expert. Illustrations in Figure 2 offer examples of retained and eliminated bounding boxes based on the presence of visible crater remnants. For instance, in Figures 2a, 2b and 2c, bounding boxes were retained even when craters had eroded or were filled in, yet remained discernible. Conversely, Figure 2d demonstrates the removal of a bounding box, where the central bounding box evidently lacked any representation of crater remnants.

Furthermore, we addressed inconsistent bounding boxes by resizing and adjusting them. Figure 3 illustrates an instance of this, with initial annotations shown in Figure 3a. An initial attempt was made to automatically rectify the bounding boxes utilizing the trained DINO model; however, this necessitated the establishment of a threshold for each image, as a completely automated solution would only partially rectify the bounding boxes, as exemplified in Figure 3b. Consequently, the decision was taken to manually align all bounding boxes with the prevailing standard observed in the dataset. The ultimate manually refined bounding boxes are presented in Figure 3c.

3. Training Details

Table 1 provides comprehensive details regarding the training process for each individual model. The batch size for all models, except DINO, was set to 8. Due to GPU memory constraints, DINO was trained with a reduced batch size of 2. The learning rate and optimizer were determined through experimental analysis, while the number of epochs was based on the original training configuration. Notably, we observed that most networks achieved their maximum Average Precision (AP) within 8 to 12 epochs,

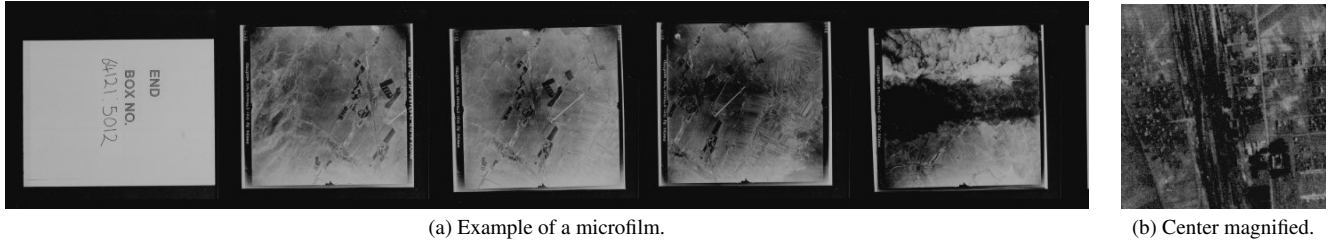


Figure 1. An illustration showcases a microfilm, available for preview from sources like a national archive, with a magnified version of the central image. Because of the significantly lower resolution in the microfilms, the train tracks in b) can only be faintly discerned.

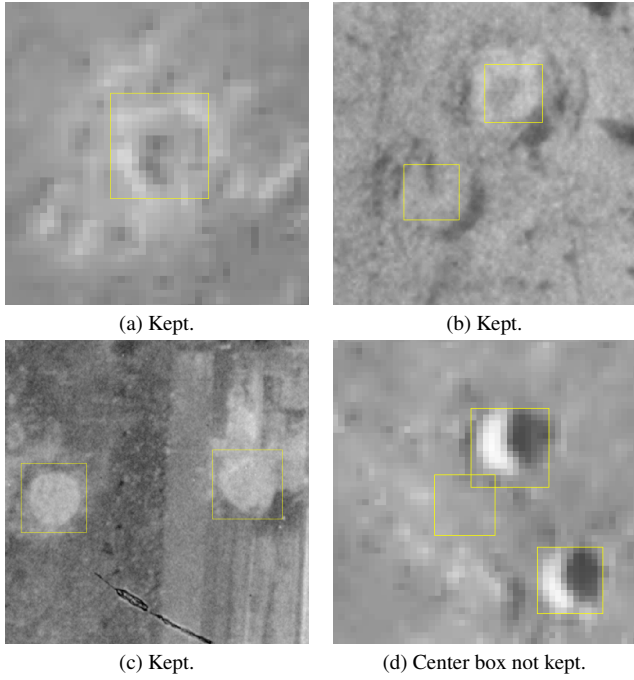


Figure 2. Examples of kept and removed bounding boxes.

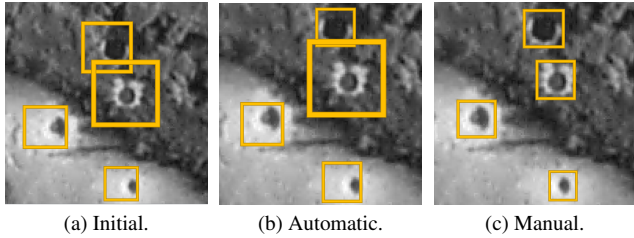


Figure 3. Examples of resized or shifted bounding boxes.

except for DETR, which continued to improve even after 100 epochs, and conditional DETR and YoloX, which required approximately 30 to 40 epochs to fully converge. It is worth mentioning that during the training of YoloF, we encountered significant instability issues. Although reducing the learning rate provided some improvement, around

20% of our training runs did not reach convergence.

4. Extended Results

We present the extended results in this section. In Table 2 the baseline results are presented with standard deviation. We additionally present the results of the models trained on the light dataset, but evaluated on the full test dataset in Table 3. In Table 4 all results including the standard deviation for the full and light dataset experiments are presented and in Table 5 the full results for the color experiments are presented.

Model Name	Batch Size	Learning Rate	Optimizer	Epochs	Pre-trained Model	Training Time (hrs)
Faster-RCNN	8	0.001	AdamW	24	COCO	2
SSD	8	0.001	SGD	24	COCO	3
RetinaNet	8	0.001	SGD	12	COCO	1
DETR	8	0.0001	AdamW	150	COCO	13
Cond. DETR	8	0.0001	AdamW	50	COCO	5
Def. DETR	8	0.0002	AdamW	50	COCO	8
Sparse-RCNN	8	0.00025	AdamW	12	COCO	3
YoloX-l	8	0.01	SGD	50	COCO	5
YoloF	8	0.001	SGD	12	COCO	2
TOOD	8	0.01	SGD	12	COCO	1
DDOD	8	0.001	Adam	12	COCO	1
DAB-DETR	8	0.0001	AdamW	12	COCO	6
DINO	2	0.0001	AdamW	12	COCO	9
QueryDet	8	0.001	Adam	20	COCO	9
Ori. RepPoints	8	0.008	SGD	12	DOTA	2

Table 1. Model Training Details. The time taken for training was assessed using an NVIDIA RTX 3090, employing the **light** dataset and approximated to the nearest full hour. This estimation provides a general idea, considering that training duration may vary based on the hardware used.

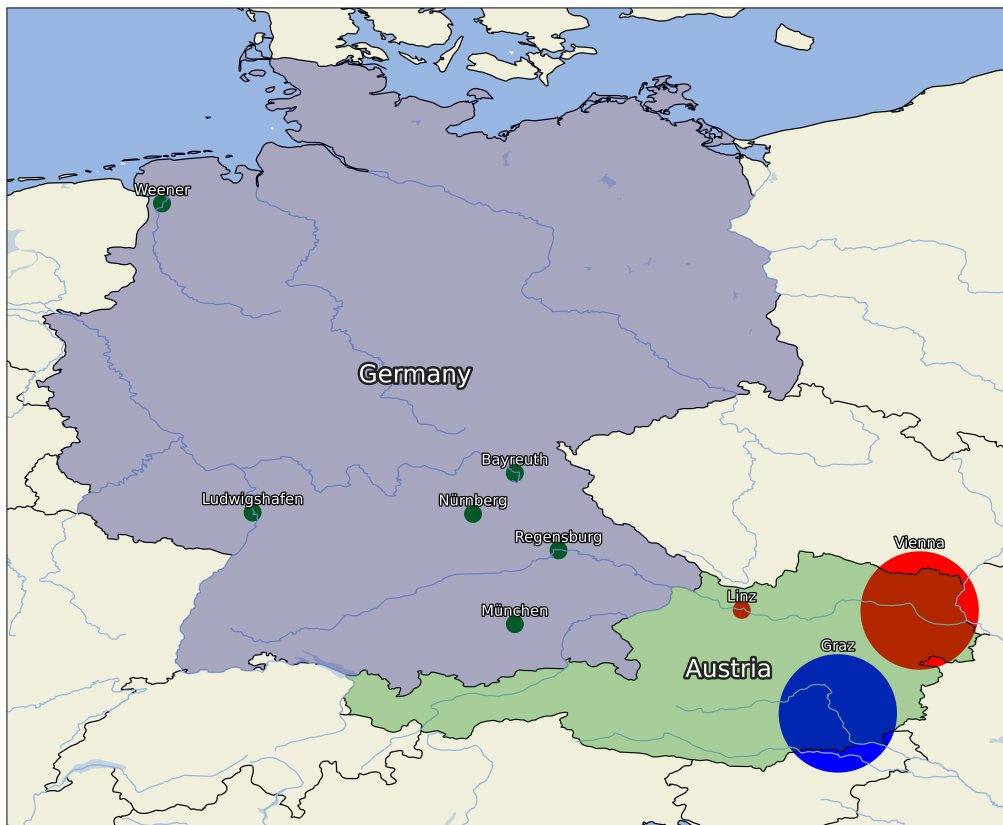


Figure 4. Image and project locations contained in our dataset. The circular depictions mark the spatial regions wherein the images were captured. The greater region of Graz is used for training, the greater region of Vienna and the city of Linz for validation, and German cities are used for testing.

	AP _s	AP _m	AP _l	AP ₂₅	AP ₅₀	AP ₇₅
Faster-RCNN	0.417 ± 0.021	0.312 ± 0.028	0.265 ± 0.018	0.741 ± 0.032	0.660 ± 0.046	0.170 ± 0.049
SSD	0.420 ± 0.019	0.327 ± 0.042	0.269 ± 0.029	0.699 ± 0.032	0.631 ± 0.032	0.226 ± 0.041
RetinaNet	0.440 ± 0.013	0.338 ± 0.004	0.293 ± 0.003	0.800 ± 0.005	0.731 ± 0.005	0.188 ± 0.025
DETR	0.430 ± 0.019	0.357 ± 0.018	0.307 ± 0.016	0.778 ± 0.019	0.702 ± 0.021	0.191 ± 0.035
Cond. DETR	0.425 ± 0.012	0.359 ± 0.017	0.324 ± 0.007	<u>0.804 ± 0.025</u>	0.722 ± 0.036	0.167 ± 0.024
Def. DETR	0.410 ± 0.031	0.334 ± 0.028	0.301 ± 0.016	<u>0.775 ± 0.032</u>	0.711 ± 0.031	0.153 ± 0.074
Sparse-RCNN	0.382 ± 0.005	0.250 ± 0.012	0.241 ± 0.011	0.726 ± 0.013	0.648 ± 0.007	0.116 ± 0.022
YoloX-l	0.456 ± 0.011	0.318 ± 0.023	0.272 ± 0.034	0.769 ± 0.020	0.619 ± 0.018	0.229 ± 0.011
YoloF	<u>0.444 ± 0.004</u>	0.336 ± 0.006	0.309 ± 0.011	0.794 ± 0.004	0.719 ± 0.012	0.194 ± 0.022
TOOD	0.440 ± 0.007	<u>0.364 ± 0.011</u>	<u>0.328 ± 0.009</u>	0.774 ± 0.019	0.714 ± 0.022	0.231 ± 0.046
DDOD	0.434 ± 0.017	<u>0.343 ± 0.026</u>	<u>0.304 ± 0.016</u>	0.775 ± 0.017	0.710 ± 0.018	<u>0.233 ± 0.046</u>
DAB-DETR	0.442 ± 0.019	<u>0.364 ± 0.022</u>	0.326 ± 0.011	0.800 ± 0.013	<u>0.739 ± 0.012</u>	0.202 ± 0.051
DINO	0.474 ± 0.007	0.393 ± 0.008	0.333 ± 0.008	0.828 ± 0.005	0.759 ± 0.015	0.273 ± 0.044
QueryDet	0.008 ± 0.004	0.166 ± 0.011	0.257 ± 0.009	0.442 ± 0.019	0.352 ± 0.015	0.075 ± 0.008
Ori. RepPoints	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.225 ± 0.025	0.177 ± 0.025	0.059 ± 0.033

Table 2. Extended training results: Baseline experiments. All values have been averaged over five runs. The best result is in **bold** and the second best is underlined. All models have been trained and tested on the **light** dataset.

	AP _s	AP _m	AP _l	AP ₂₅	AP ₅₀	AP ₇₅
Faster-RCNN	0.414 ± 0.044	0.336 ± 0.027	0.275 ± 0.012	0.592 ± 0.034	0.512 ± 0.049	0.194 ± 0.028
SSD	0.408 ± 0.038	0.346 ± 0.030	0.278 ± 0.038	0.519 ± 0.023	0.473 ± 0.028	0.239 ± 0.018
RetinaNet	0.456 ± 0.003	0.361 ± 0.006	0.302 ± 0.003	0.645 ± 0.007	0.590 ± 0.019	0.222 ± 0.022
DETR	0.427 ± 0.021	0.367 ± 0.015	0.317 ± 0.017	0.621 ± 0.014	0.579 ± 0.024	0.159 ± 0.053
Cond. DETR	0.445 ± 0.017	0.365 ± 0.009	0.330 ± 0.012	0.625 ± 0.022	0.574 ± 0.024	0.162 ± 0.017
Def. DETR	0.410 ± 0.039	0.347 ± 0.027	0.302 ± 0.026	0.592 ± 0.039	0.555 ± 0.032	0.176 ± 0.032
Sparse-RCNN	0.358 ± 0.020	0.286 ± 0.006	0.225 ± 0.014	0.533 ± 0.009	0.478 ± 0.014	0.191 ± 0.014
YoloX-l	0.436 ± 0.022	0.357 ± 0.018	0.281 ± 0.029	<u>0.693 ± 0.024</u>	0.515 ± 0.010	0.239 ± 0.025
YoloF	0.453 ± 0.006	0.367 ± 0.007	0.312 ± 0.010	<u>0.690 ± 0.013</u>	<u>0.620 ± 0.022</u>	0.231 ± 0.019
TOOD	0.446 ± 0.022	0.373 ± 0.008	0.335 ± 0.012	0.655 ± 0.007	0.605 ± 0.015	<u>0.240 ± 0.026</u>
DDOD	0.455 ± 0.012	0.359 ± 0.020	0.307 ± 0.023	0.614 ± 0.048	0.555 ± 0.049	0.234 ± 0.033
DAB-DETR	<u>0.457 ± 0.029</u>	<u>0.378 ± 0.025</u>	<u>0.339 ± 0.021</u>	0.629 ± 0.008	0.590 ± 0.013	0.192 ± 0.045
DINO	0.484 ± 0.022	0.408 ± 0.007	0.341 ± 0.013	0.702 ± 0.023	0.659 ± 0.025	0.274 ± 0.035
QueryDet	0.008 ± 0.004	0.149 ± 0.013	0.249 ± 0.009	0.418 ± 0.02	0.334 ± 0.014	0.072 ± 0.007
Ori. RepPoints	-	-	-	-	-	-

Table 3. Additional training results: Baseline experiment, trained on the **light** dataset, but tested on the **full** dataset. All values have been averaged over five runs. The best result is in **bold** and the second best is underlined.

	AP _s	AP _m	AP _l	AP ₂₅	AP ₅₀	AP ₇₅
Light	0.484 ± 0.022	0.408 ± 0.007	0.341 ± 0.013	0.702 ± 0.023	0.659 ± 0.025	0.274 ± 0.035
Full	0.458 ± 0.027	0.374 ± 0.017	0.302 ± 0.030	0.699 ± 0.029	0.653 ± 0.016	0.216 ± 0.036
FT-1	0.468 ± 0.031	0.392 ± 0.022	0.320 ± 0.018	0.680 ± 0.035	0.622 ± 0.03	0.225 ± 0.008
FT-3	0.450 ± 0.014	0.372 ± 0.009	0.294 ± 0.013	0.676 ± 0.033	0.618 ± 0.032	0.183 ± 0.050
FT-6	0.452 ± 0.008	0.366 ± 0.007	0.275 ± 0.022	0.658 ± 0.044	0.595 ± 0.038	0.182 ± 0.030

Table 4. Extended training results: Full, light and finetuning experiment. All values have been averaged over five runs. The best result is in **bold**. All models have been tested on the **full** dataset.

	AP _s	AP _m	AP _l	AP ₂₅	AP ₅₀	AP ₇₅
Grey	0.474 ± 0.007	0.393 ± 0.008	0.333 ± 0.008	0.828 ± 0.005	0.759 ± 0.015	0.273 ± 0.044
RGB	0.490 ± 0.008	0.396 ± 0.029	0.318 ± 0.020	0.821 ± 0.012	0.741 ± 0.012	0.356 ± 0.063

Table 5. Extended training results: Grey and RGB experiment. All values have been averaged over five runs. The best result is in **bold**. Colored model tested on the colored light dataset, grey model tested on the grey version.

Name	Project	Location	GSD in m	Year	Month	Day	Split	Bombload
1945.03.15.Wienxxiii.98	Wienxxiii	Vienna Area	0.169	1945	3	15	Val	10
1944.06.13.Wienxxiii.97	Wienxxiii	Vienna Area	0.182	1944	6	13	Val	10
1945.04.05.Wienxxi.96	Wienxxi	Vienna Area	0.159	1945	4	5	Val	61
1945.03.23.Wienxxi.95	Wienxxi	Vienna Area	0.222	1945	3	23	Val	90
1945.02.16.Wienxxi.94	Wienxxi	Vienna Area	0.271	1945	2	16	Val	53
1945.04.05.Wienpenzing.93	Wienpenzing	Vienna Area	0.159	1945	4	5	Val	13
1945.02.28.Wienpenzing.92	Wienpenzing	Vienna Area	0.265	1945	2	28	Val	74
1945.02.27.Wienpenzing.91	Wienpenzing	Vienna Area	0.275	1945	2	27	Val	68
1945.02.20.Wienpenzing.90	Wienpenzing	Vienna Area	0.243	1945	2	20	Val	63
1945.04.05.Wieniii.89	Wieniii	Vienna Area	0.159	1945	4	5	Val	16
1945.02.28.Wieniii.88	Wieniii	Vienna Area	0.265	1945	2	28	Val	8
1944.10.07.Weener.86	Weener	Germany	0.349	1944	10	7	Test	7
1944.10.07.Weener.87	Weener	Germany	0.349	1944	10	7	Test	7
1945.04.02.Schwechatparkplatz.85	Schwechatparkplatz	Vienna Area	0.176	1945	4	2	Val	15
1945.03.21.Schwechatparkplatz.84	Schwechatparkplatz	Vienna Area	0.243	1945	3	21	Val	10
1944.08.07.Schwechatparkplatz.83	Schwechatparkplatz	Vienna Area	0.330	1944	8	7	Val	14
1944.04.24.Schwechatparkplatz.82	Schwechatparkplatz	Vienna Area	0.169	1944	4	24	Val	5
1945.03.26.Schwechatflughafen.81	Schwechatflughafen	Vienna Area	0.180	1945	3	26	Val	14
1945.03.21.Schwechatflughafen.80	Schwechatflughafen	Vienna Area	0.243	1945	3	21	Val	8
1944.12.06.Schwechatflughafen.79	Schwechatflughafen	Vienna Area	0.212	1944	12	6	Val	40
1944.12.06.Schwechatflughafen.78	Schwechatflughafen	Vienna Area	0.212	1944	12	6	Val	30
1944.04.24.Schwechatflughafen.77	Schwechatflughafen	Vienna Area	0.169	1944	4	24	Val	28
1944.02.08.Schwechatflughafen.76	Schwechatflughafen	Vienna Area	0.180	1944	2	8	Val	12
1945.04.17.Regensburg.74	Regensburg	Germany	0.169	1945	4	17	Test	20
1945.03.16.Regensburg.73	Regensburg	Germany	0.254	1945	3	16	Test	20
1945.09.02.Regensburg.75	Regensburg	Germany	0.318	1945	9	2	Test	20
1944.02.25.Nuernberg.71	Nuernberg	Germany	0.260	1944	2	25	Test	10
1945.03.21.Nuernberg.72	Nuernberg	Germany	0.254	1945	3	21	Test	21
1944.07.20.Muenchen.70	Muenchen	Germany	0.381	1944	7	20	Test	26
1944.07.04.Mannsworther.63	Mannsworther	Vienna Area	0.191	1944	7	4	Val	28
1945.04.01.Mannsworther.69	Mannsworther	Vienna Area	0.201	1945	4	1	Val	103
1945.03.16.Mannsworther.68	Mannsworther	Vienna Area	0.180	1945	3	16	Val	105
1945.02.28.Mannsworther.67	Mannsworther	Vienna Area	0.254	1945	2	28	Val	98
1945.01.04.Mannsworther.66	Mannsworther	Vienna Area	0.254	1945	1	4	Val	86
1944.10.07.Mannsworther.65	Mannsworther	Vienna Area	0.635	1944	10	7	Val	40
1944.09.13.Mannsworther.64	Mannsworther	Vienna Area	0.184	1944	9	13	Val	28
1944.11.01.Ludwigshafen2017.61	Ludwigshafen2017	Germany	0.265	1944	11	1	Test	445
1943.09.24.Ludwigshafen2017.60	Ludwigshafen2017	Germany	0.265	1943	9	24	Test	29
1945.02.15.Ludwigshafen2017.62	Ludwigshafen2017	Germany	0.265	1945	2	15	Test	1,708
1945.03.22.Ludwigshafen.59	Ludwigshafen	Germany	0.265	1945	3	22	Test	261
1945.01.14.Ludwigshafen.58	Ludwigshafen	Germany	0.265	1945	1	14	Test	244
1944.10.12.Ludwigshafen.57	Ludwigshafen	Germany	0.265	1944	10	12	Test	74
1945.03.23.Lobau.56	Lobau	Vienna Area	0.222	1945	3	23	Val	42
1945.02.14.Lobau.55	Lobau	Vienna Area	0.191	1945	2	14	Val	41
1944.10.07.Lobau.54	Lobau	Vienna Area	0.286	1944	10	7	Val	29
1945.04.20.Linz.52	Linz	Linz	0.191	1945	4	20	Val	46
1945.02.25.Linz.51	Linz	Linz	0.243	1945	2	25	Val	46
1945.05.14.Linz.53	Linz	Linz	0.212	1945	5	14	Val	46
1945.04.20.Graz.50	Graz	Graz Area	0.176	1945	4	20	Train	427
1945.04.20.Graz.49	Graz	Graz Area	0.176	1945	4	20	Train	231
1944.11.17.Graz.1	Graz	Graz Area	0.265	1944	11	17	Train	119
1945.04.05.Graz.24	Graz	Graz Area	0.180	1945	4	5	Train	1,198
1945.04.02.Graz.22	Graz	Graz Area	0.212	1945	4	2	Train	463
1945.04.02.Graz.21	Graz	Graz Area	0.212	1945	4	2	Train	241
1945.04.02.Graz.20	Graz	Graz Area	0.212	1945	4	2	Train	223
1945.04.02.Graz.19	Graz	Graz Area	0.212	1945	4	2	Train	294
1945.04.02.Graz.18	Graz	Graz Area	0.212	1945	4	2	Train	64
1945.03.22.Graz.17	Graz	Graz Area	0.159	1945	3	22	Train	69
1945.03.22.Graz.16	Graz	Graz Area	0.159	1945	3	22	Train	40
1945.03.22.Graz.15	Graz	Graz Area	0.159	1945	3	22	Train	22
1945.03.15.Graz.14	Graz	Graz Area	0.169	1945	3	15	Train	630
1945.03.15.Graz.13	Graz	Graz Area	0.169	1945	3	15	Train	285
1945.03.02.Graz.12	Graz	Graz Area	0.286	1945	3	2	Train	209
1945.02.28.Graz.11	Graz	Graz Area	0.265	1945	2	28	Train	186
1945.02.23.Graz.10	Graz	Graz Area	0.275	1945	2	23	Train	37
1945.02.23.Graz.9	Graz	Graz Area	0.275	1945	2	23	Train	118
1945.02.21.Graz.8	Graz	Graz Area	0.243	1945	2	21	Train	30
1945.02.21.Graz.7	Graz	Graz Area	0.243	1945	2	21	Train	38
1945.02.21.Graz.6	Graz	Graz Area	0.243	1945	2	21	Train	10
1945.02.21.Graz.5	Graz	Graz Area	0.243	1945	2	21	Train	41
1945.02.21.Graz.4	Graz	Graz Area	0.243	1945	2	21	Train	28
1945.01.05.Graz.3	Graz	Graz Area	0.254	1945	1	5	Train	121
1944.11.17.Graz.2	Graz	Graz Area	0.265	1944	11	17	Train	120
1945.04.05.Graz.23	Graz	Graz Area	0.159	1945	4	5	Train	578
1945.04.05.Graz.25	Graz	Graz Area	0.180	1945	4	5	Train	1,916
1945.04.20.Graz.48	Graz	Graz Area	0.176	1945	4	20	Train	1,135
1945.04.08.Graz.26	Graz	Graz Area	0.318	1945	4	8	Train	166
1945.04.20.Graz.47	Graz	Graz Area	0.176	1945	4	20	Train	103
1945.04.20.Graz.46	Graz	Graz Area	0.176	1945	4	20	Train	8
1945.04.20.Graz.45	Graz	Graz Area	0.176	1945	4	20	Train	218
1945.04.20.Graz.44	Graz	Graz Area	0.176	1945	4	20	Train	290
1945.04.20.Graz.43	Graz	Graz Area	0.176	1945	4	20	Train	130
1945.04.20.Graz.42	Graz	Graz Area	0.176	1945	4	20	Train	518
1945.04.20.Graz.41	Graz	Graz Area	0.176	1945	4	20	Train	150
1945.04.20.Graz.40	Graz	Graz Area	0.176	1945	4	20	Train	411
1945.04.20.Graz.39	Graz	Graz Area	0.176	1945	4	20	Train	4
1945.04.20.Graz.38	Graz	Graz Area	0.176	1945	4	20	Train	52
1945.04.20.Graz.37	Graz	Graz Area	0.176	1945	4	20	Train	267
1945.04.20.Graz.36	Graz	Graz Area	0.176	1945	4	20	Train	190
1945.04.20.Graz.35	Graz	Graz Area	0.176	1945	4	20	Train	797
1945.04.20.Graz.34	Graz	Graz Area	0.176	1945	4	20	Train	206
1945.04.20.Graz.33	Graz	Graz Area	0.176	1945	4	20	Train	850
1945.04.20.Graz.32	Graz	Graz Area	0.176	1945	4	20	Train	162
1945.04.20.Graz.31	Graz	Graz Area	0.176	1945	4	20	Train	127
1945.04.20.Graz.30	Graz	Graz Area	0.176	1945	4	20	Train	54
1945.04.12.Graz.29	Graz	Graz Area	0.180	1945	4	12	Train	56
1945.04.11.Graz.28	Graz	Graz Area	0.254	1945	4	11	Train	1,397
1945.04.08.Graz.27	Graz	Graz Area	0.318	1945	4	8	Train	286
1945.04.11.Bayreuth_0	Bayreuth	Germany	0.176	1945	4	11	Test	31

Table 6. Detailed information about the images contained in the data set

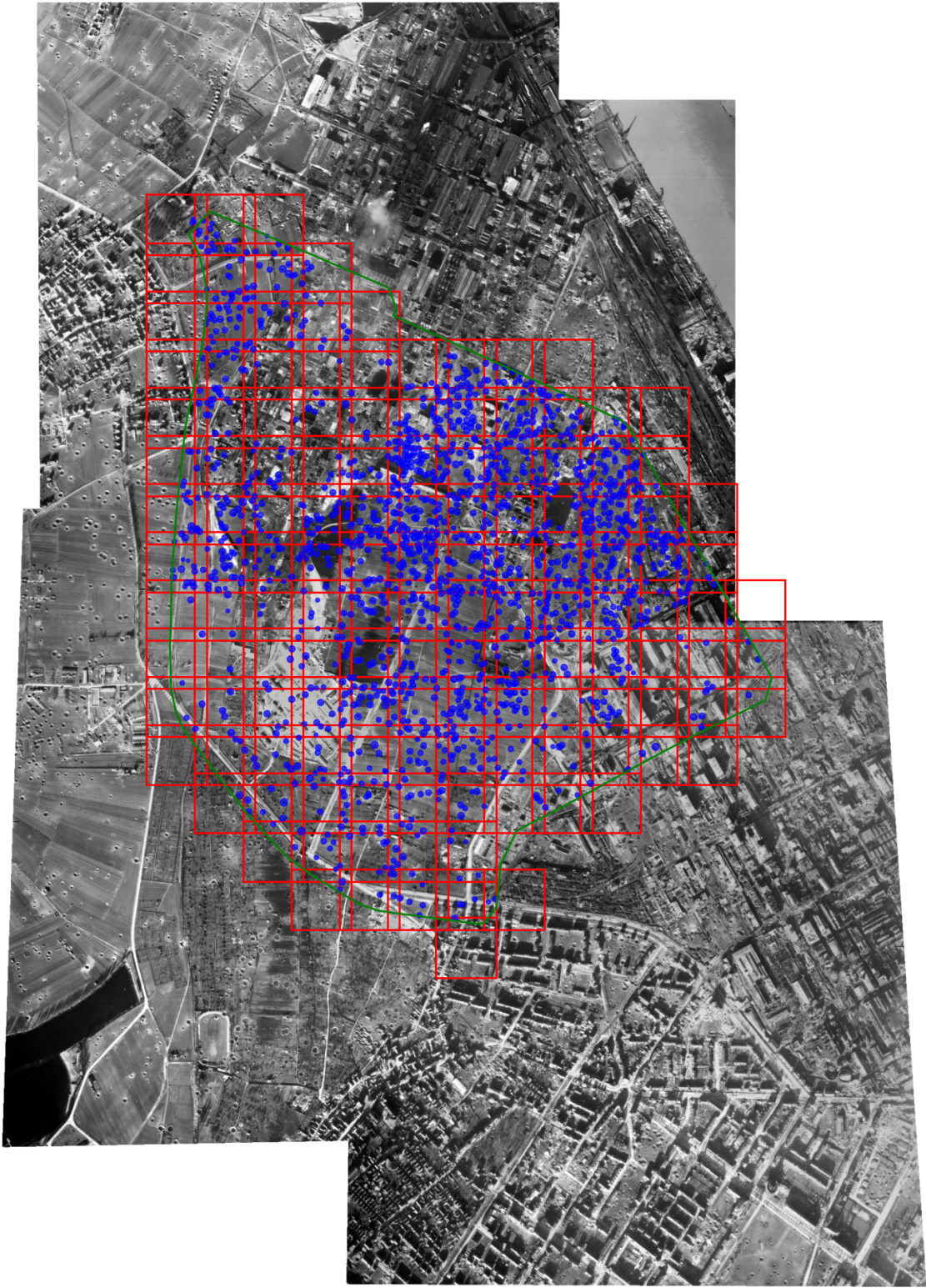


Figure 5. An example image from the dataset in full size. Red: Extracted patches, Green: Area of interest, Blue: Crater annotations.

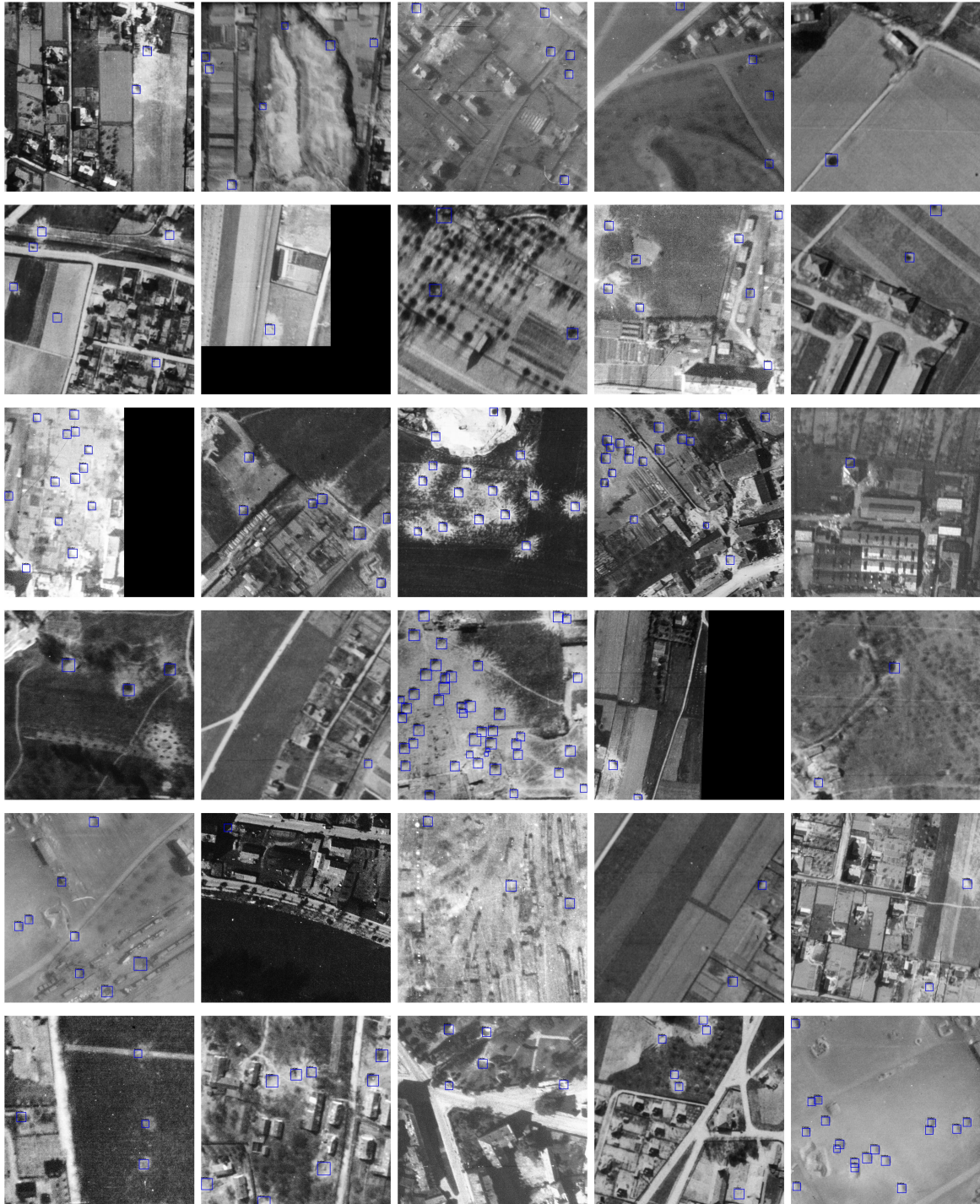


Figure 6. 30 examples from the training split of the dataset.

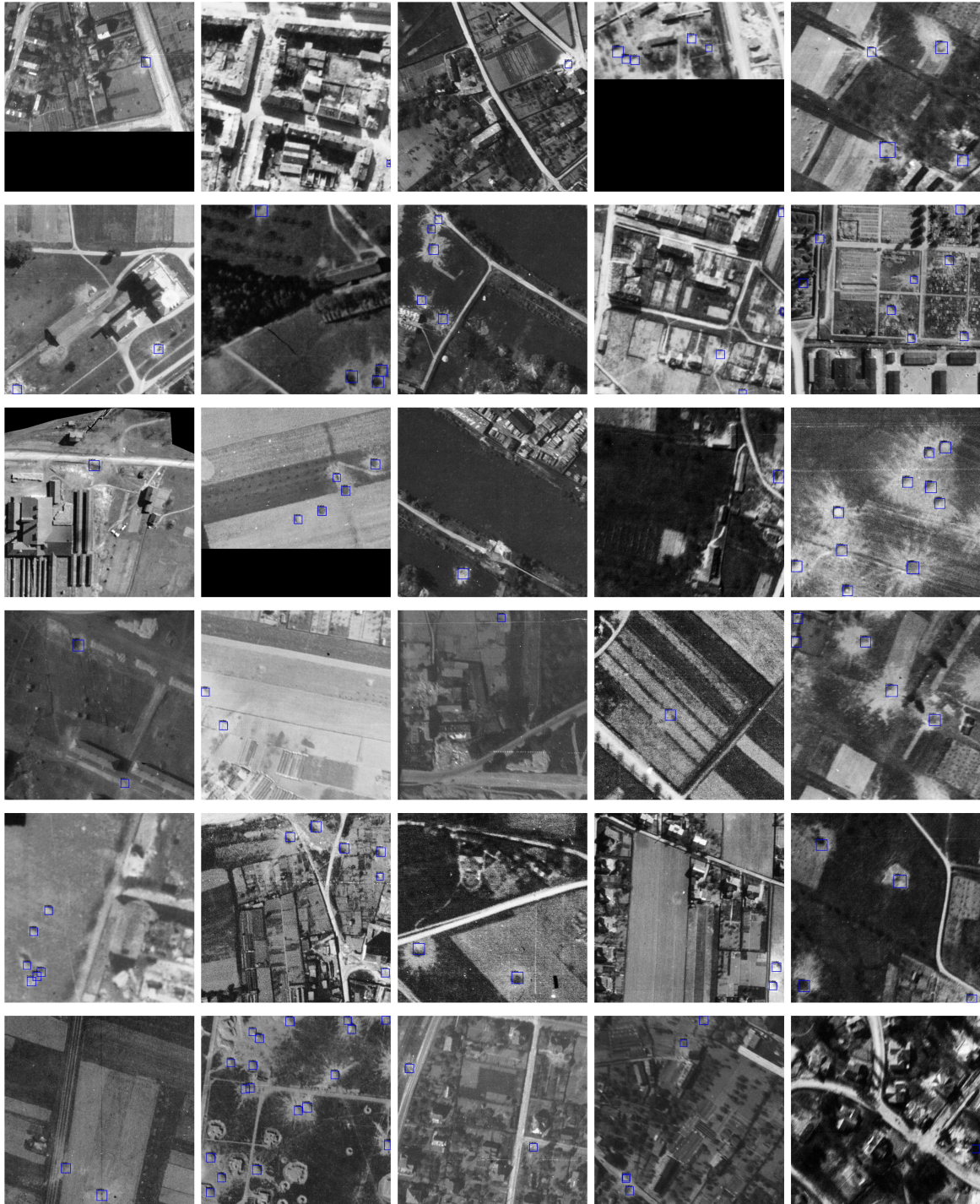


Figure 7. Further 30 examples from the training split of the dataset.

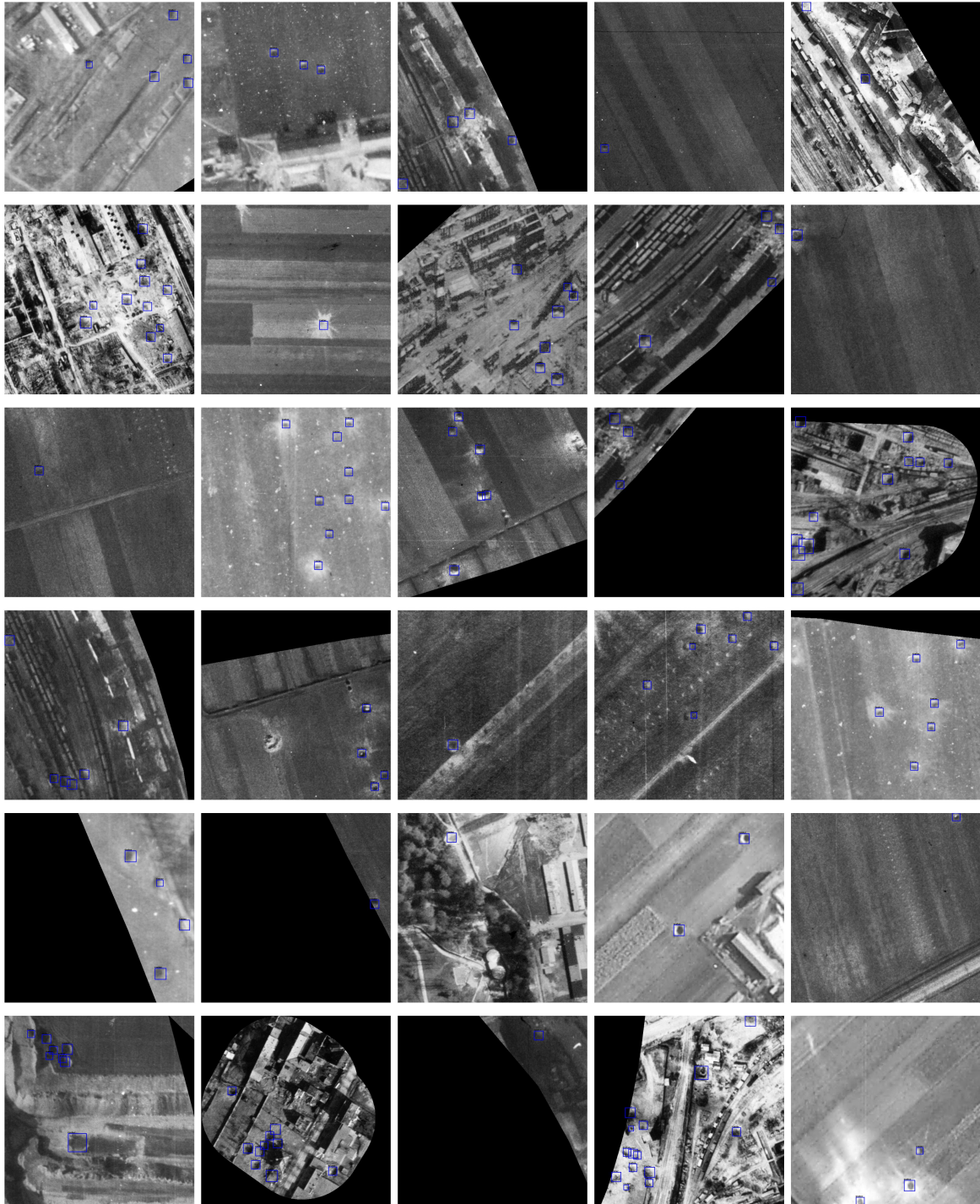


Figure 8. 30 examples from the validation split of the dataset.

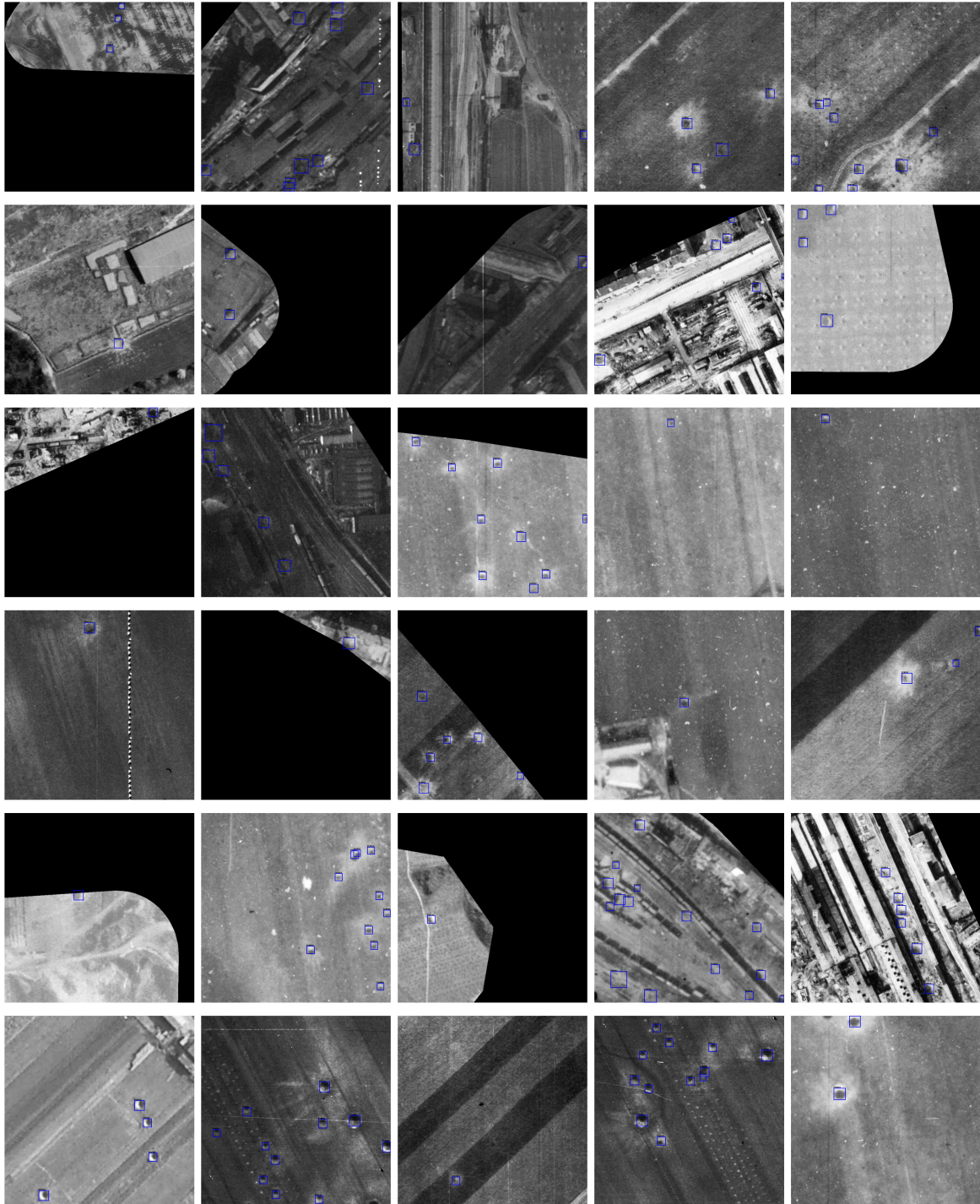


Figure 9. Further 30 examples from the validation split of the dataset.

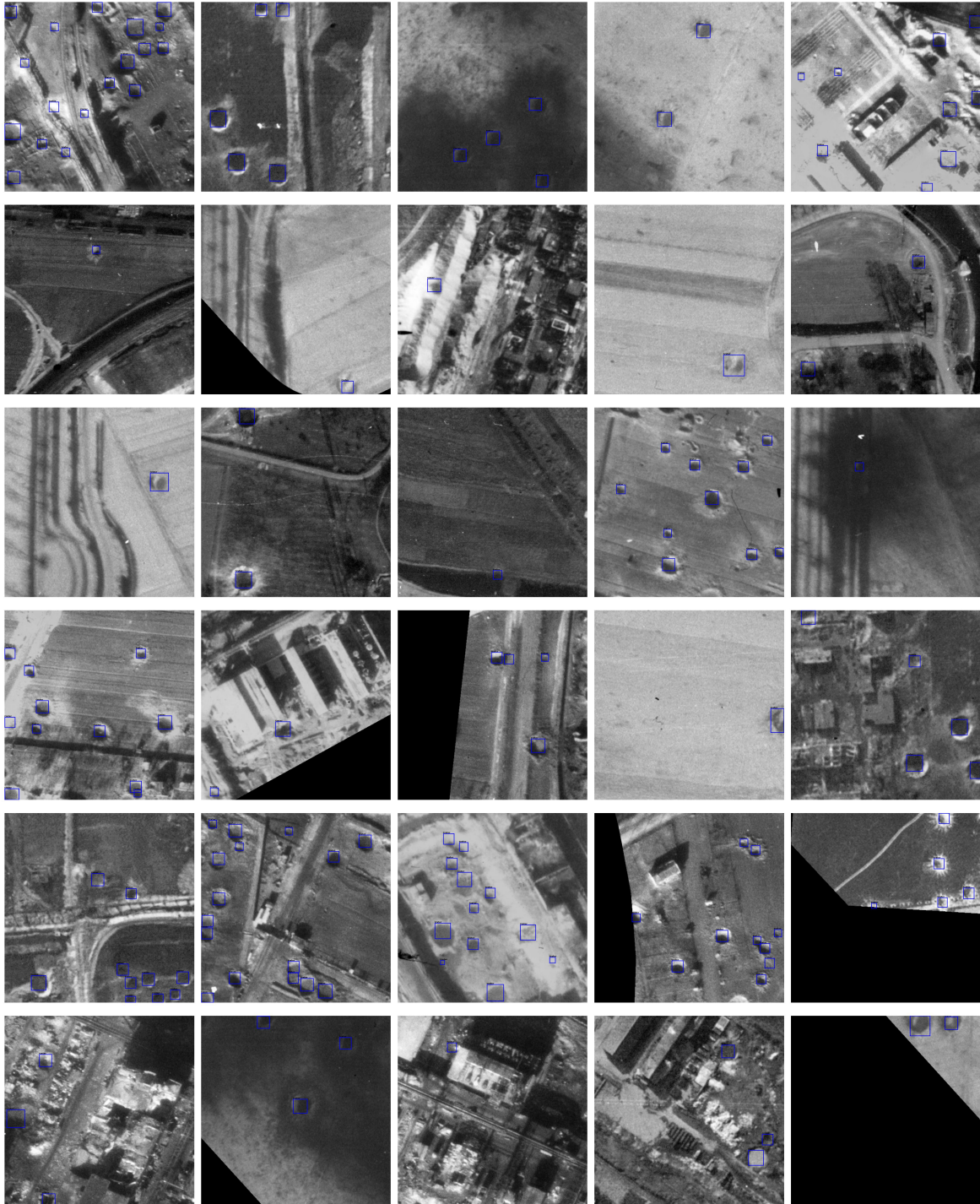


Figure 10. 30 examples from the test split of the dataset.

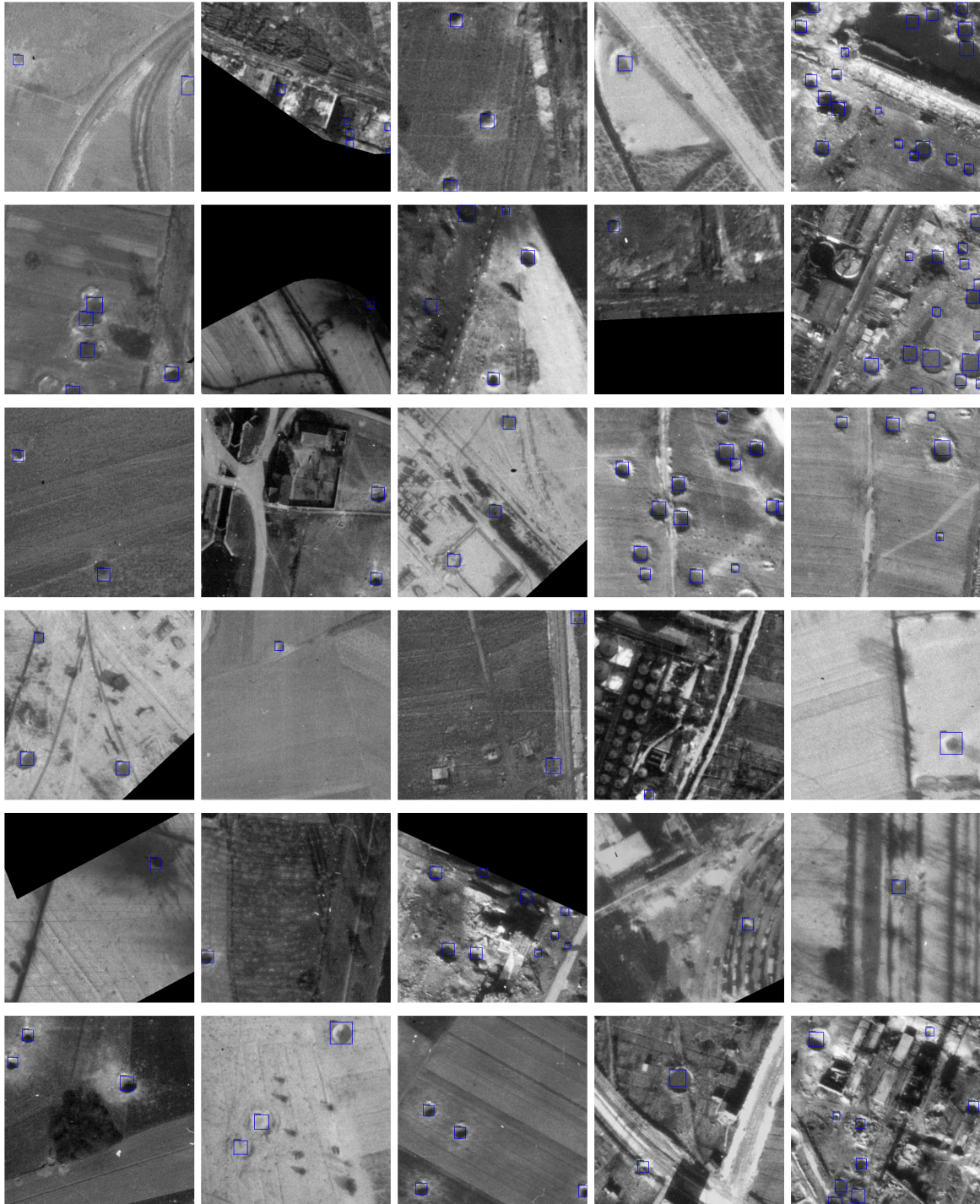


Figure 11. Further 30 examples from the test split of the dataset.