A. Details on generating the European Residential Building Dataset

The datasets are both acquired using free publicly available data from the respective national mapping agencies responsible for providing cartographic dataset online. While both of the dataset providers provide orthorectified imagery, the images are not true ortho (DSM-corrected imagery), and as such a challenging component of our study is to deal with varying parallax effects, a common artefact in remote sensing imagery [8].

A.1. The Danish dataset

The Danish dataset was acquired by combining publicly available datasets from the Danish national mapping institute (GeoDanmark). We combined the following datasets: GeoDanmark OrtoFoto Foraar 2021, BBR (July 2023), Danmarks Adresse Register (DAR, July 2023), GeoDanmark Vektor, Matrikel MAT.

Our process to identify detached residential houses was performing filtering that would:

1. Apply filters labelling the building as residential single family house
2. Remove buildings with more than one address on property
3. Remove buildings with more than 2 floors
4. Remove buildings which are not privately owned
5. Remove buildings which span more than 400m2

Following this, we randomly sample uniformly 120,000 buildings and perform qualitative and statistical checks, upon which we identified and excluded 44 buildings which had corrupt polygon annotations.

For the remaining set of 119,956 buildings we sample images using the GeoDanmark Ortofoto API by adding a 5m buffer around each side of the boundingbox of the polygon and acquiring the image with the buffered boundingbox.

A.2. The Dutch Dataset

The Dutch dataset was acquired by combining publicly available datasets from the Dutch mapping agency PDOK (Public Services on the Map). In order to identify residential homes we combined the following datasets: Basic Registration of Addresses and Buildings (BAG), PDOK Aerial Photo RGB (Open), TOPNL, NLExtract.

We utilize the 25cm Aerial Ortofoto dataset and generate data similarly to the Danish dataset. The dataset is in a lower spatial resolution than the Danish dataset, and the Netherlands has a higher population density than Denmark. For this reason the Dutch dataset is intended to be the harder of the two. Upon sampling the images we add a 10m buffer to each side of the boundingbox of the building polygon, and acquire an image using the PDOK Aerial Photo API. Upon visual inspection, we find most parallax artifacts occurring in the Dutch dataset.

A.3. Examples of annotations from Denmark and Netherlands

Please refer to Figures 4 and 5 below.

B. Results on the SpaceNet Vegas dataset

While not a core component of our study, we present results for GAST computed on the SpaceNet-V2 Vegas [20] dataset using ground truth bounding boxes. We have not found identified works computing identical metrics of interest, but present the ability with which GAST can learn objects in this dataset. The results can be seen in Table 4

C. Remarks on the HiSup training procedure

We trained the HiSup model using the same parameters the authors recommend in their study, in particular the same loss weights as used for the Aicrowd dataset [16]. We experimented with multiple backbones provided by the authors, and found the best backbone to be the HrNet48 as was the case for the HiSup results as well.

D. Additional qualitative results

In this section we present subsequent qualitative results to demonstrate our findings. Please refer to Figures 6,7 and 8 below.

Figure 4. Examples of ground truth from the Danish dataset

AP↑	**AP**$_{50}$↑	**AP**$_{75}$↑	**AR**↑	**AR**$_{50}$↑	**AP**$_{75}$↑	**IoU**↑	**C-IoU**↑	**MTA**↓	**PoLiS**↓	$\Delta\theta$	**N-ratio**	Θ_{Model}	**Dataset**
64.53	91.17	73.27	78.14	96.92	86.86	85.77	63.68	56.22	15.51	9.09	.95	GAST	SpaceNet Vegas

Table 4. Results of GAST on the SpaceNet V2 Buildings dataset

Figure 5. Examples of ground truth from the Dutch dataset

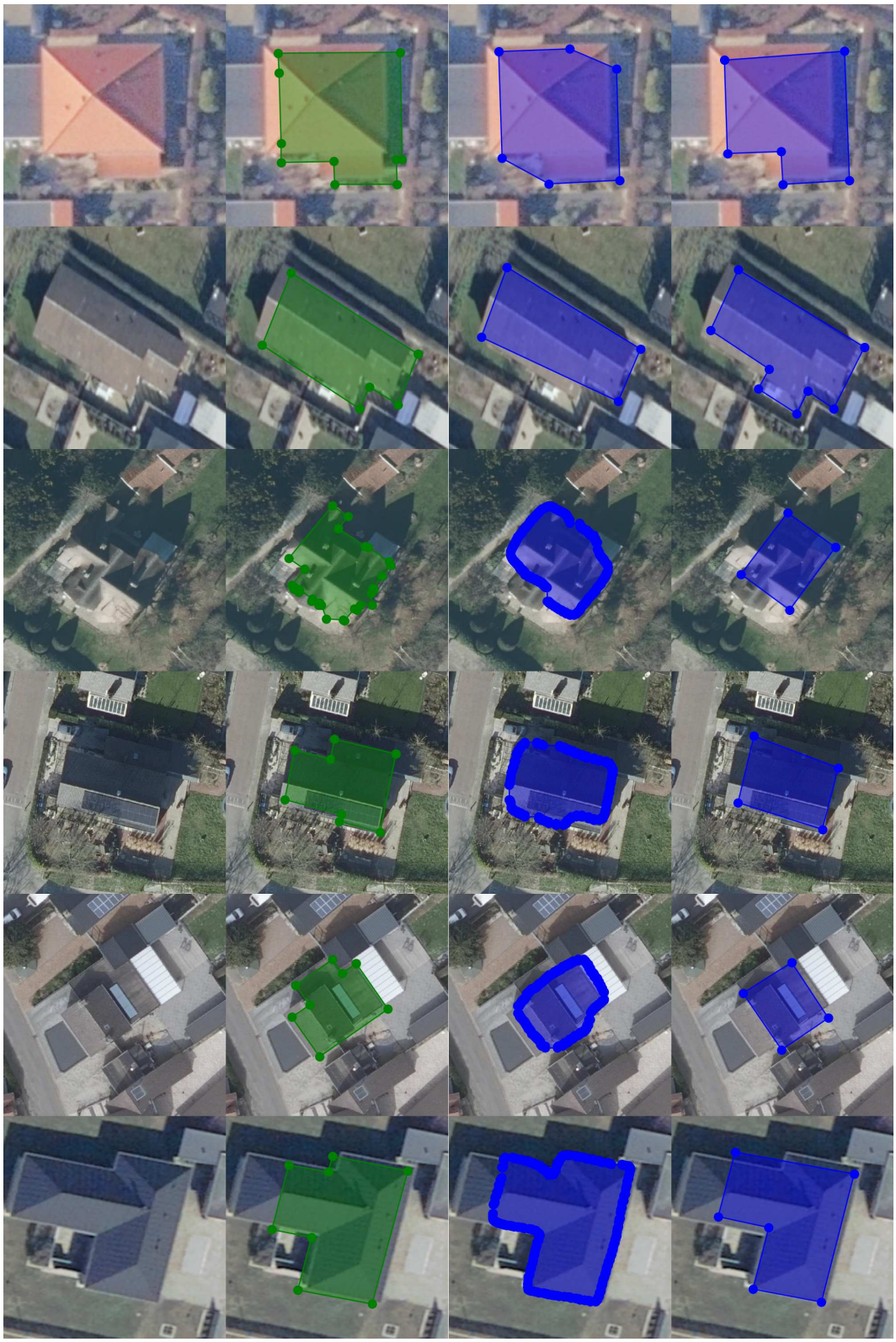

Figure 6. Qualitative results from left to right: Ground truth, ground truth with annotation, HiSup prediction, GAST (ours) prediction. From the top row down, for each two rows the figures correspond to models trained and performing inference on the Danish, Dutch and Danish and Dutch datasets combined.

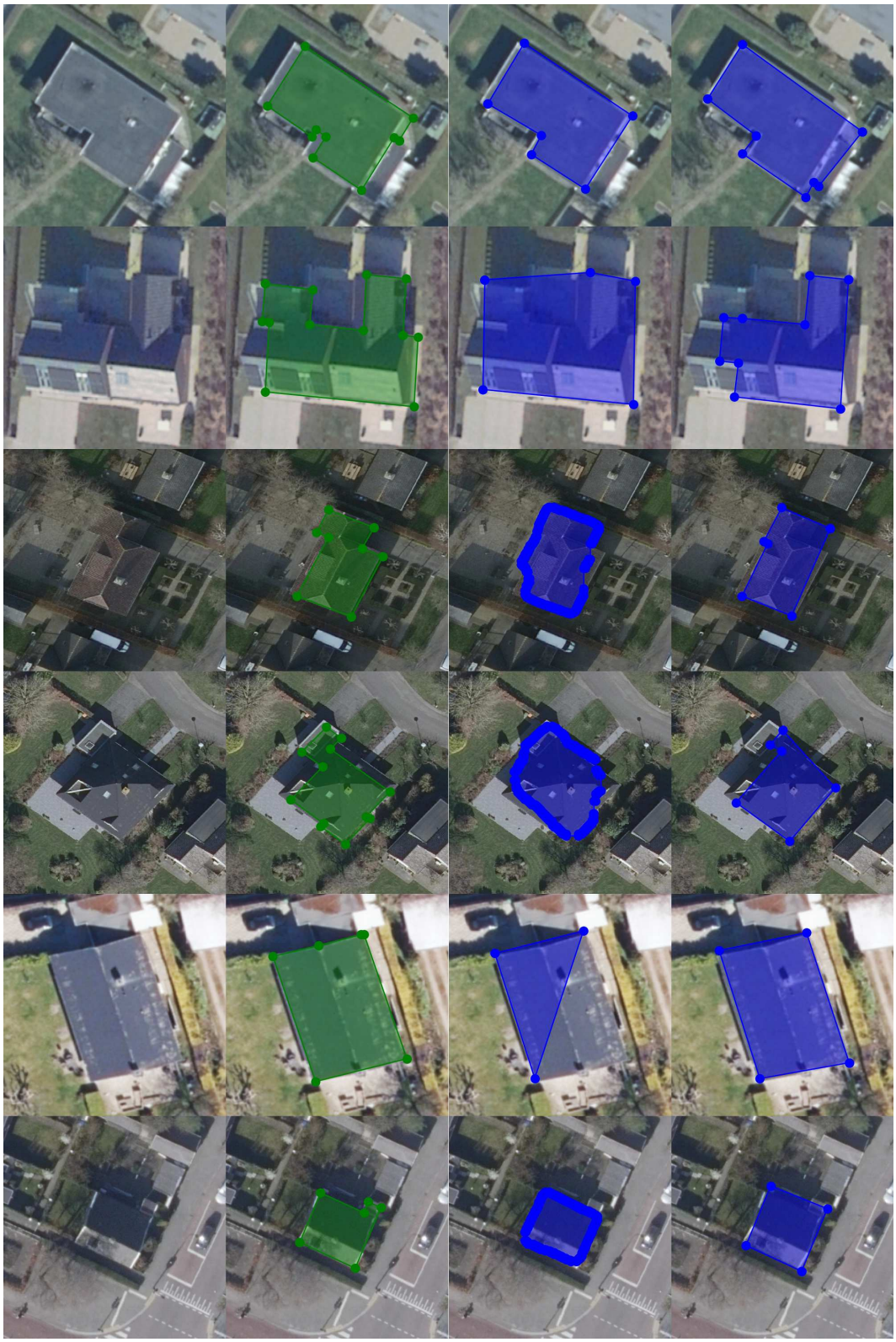

Figure 7. Qualitative results from left to right: Ground truth, ground truth with annotation, HiSup prediction, GAST (ours) prediction. From the top row down, for each two rows the figures correspond to models trained and performing inference on the Danish, Dutch and Danish and Dutch datasets combined.

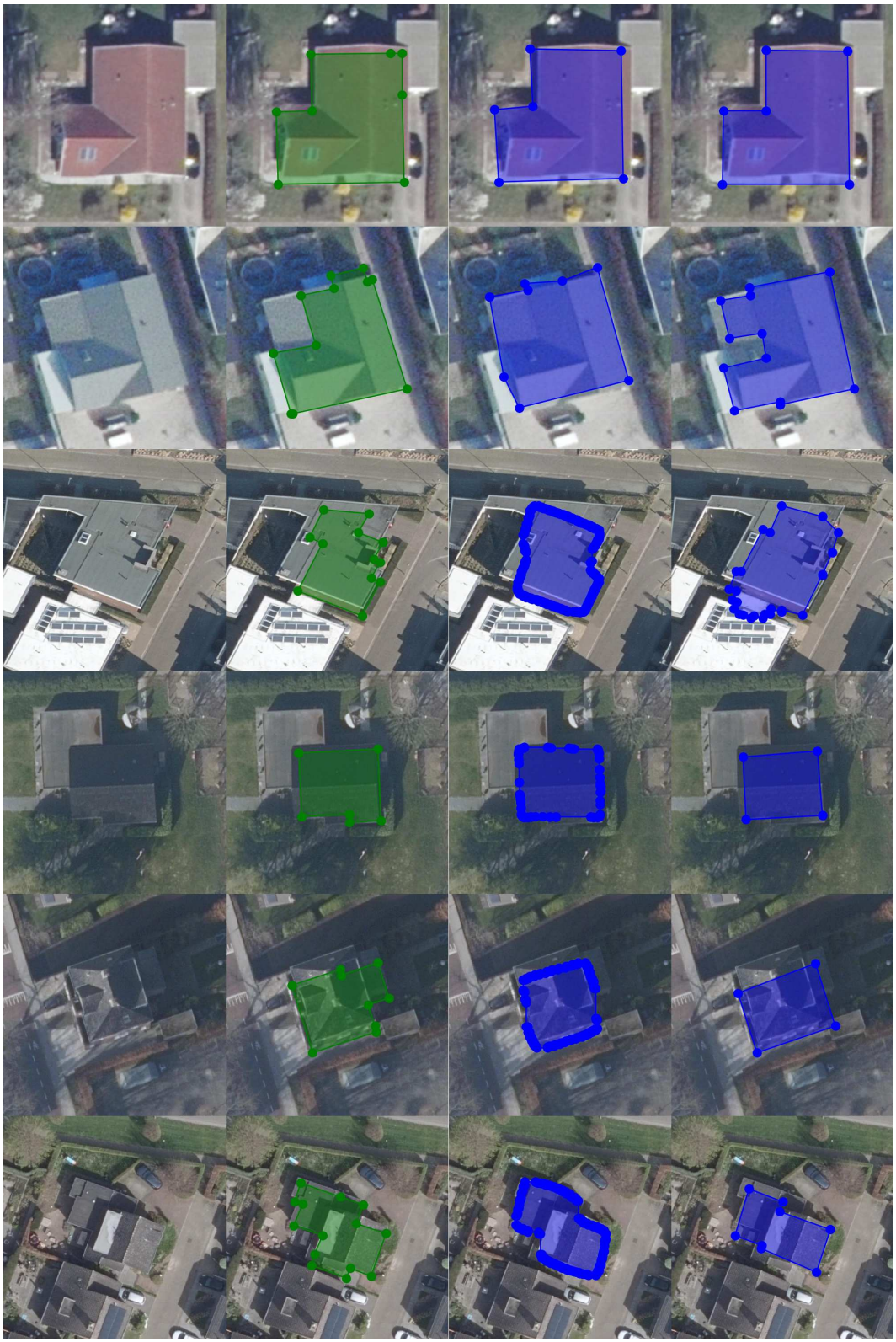

Figure 8. Qualitative results from left to right: Ground truth, ground truth with annotation, HiSup prediction, GAST (ours) prediction. From the top row down, for each two rows the figures correspond to models trained and performing inference on the Danish, Dutch and Danish and Dutch datasets combined.