



AST: Joint detection of fences and their shadows on orthophotos of low-rise residential areas

Igor Victorovich **Vinokurov** 

Financial University under the Government of the Russian Federation, Moscow, Russia 

 igvinokurov@fa.ru

Abstract. This paper presents a method for joint detection of fences and their shadows in orthophotos of low-rise residential areas (nadir geometry), built upon the Adaptive Shadow Transformer (AST) architecture. The method relies on three key components:

- (1) learnable separate displacement fields for fences and shadows with tailored regularization;
- (2) a cross-attention mechanism that incorporates solar azimuth and elevation, as well as fence typology;
- (3) a shadow fragment connectivity loss that enables the model to perceive terrain-broken shadow segments as a unified object.

The evaluation is performed on the collected FSAD dataset, comprising 650 manually annotated images and 2,000 synthetic scenes. The proposed approach demonstrates superiority over state-of-the-art detectors (YOLO26, RT-DETR, Deformable DETR, and DINOv2+DETR), providing an absolute mAP@0.5 gain of up to 5.9 percentage points on scenes with complex terrain. (*Linked article texts in English and in Russian*).

Key words and phrases: aerial imagery, object detection, fences, shadows, vision transformers, deformable attention, cadastral mapping

2020 *Mathematics Subject Classification:* 68U10; 68T07, 68T45

Acknowledgments:

The author expresses gratitude to the students of the Financial University under the Government of the Russian Federation who participated in the annotation of the FSAD dataset. The work was carried out with the support of the Ministry of Digital Development of the Kaluga Oblast within the framework of the project on the development of artificial intelligence technologies.

For citation: Igor V. Vinokurov. *AST: Joint detection of fences and their shadows on orthophotos of low-rise residential areas*. Program Systems: Theory and Applications, 2026, **17**:4(73), pp. 57–136. (*In English, in Russian*). https://psta.psiras.ru/read/psta2026_4_57-136.pdf

Introduction

Accurate mapping of residential areas is an important task in cadastral accounting and urban planning. Fences (and similar structures) define the actual boundaries of land plots [1], but their automatic detection in aerial photographs is associated with a number of difficulties. Unlike buildings and roads, fences have a non-rigid spatial structure: sections sag between supports, and posts may be tilted under the influence of wind or seasonal ground changes. Additional complexity is created by shadows cast by fences onto the ground.

The shadow configuration depends on the height and type of the enclosure (solid, lattice, picket), the terrain, and the position of the sun, characterized by two angles: the *azimuth* φ (compass direction to the sun) and the *elevation* α (angle above the horizon). On rugged terrain, the shadow may break at ridges and slopes, creating false boundaries that standard algorithms mistakenly interpret as independent objects [2].

When surveyed at nadir, the fence itself is projected as a thin line 1–2 pixels wide, whereas its shadow forms a clearly distinguishable band whose length and structure encode the height and type of the enclosure. All input images – both real and synthetic – are reduced to an orthophotomosaic and considered in nadir geometry; the deviation angle of the optical axis from nadir does not exceed the tolerance for orthophotomosaics (no more than 3°). Thus, in orthophoto images, the shadow is not an artifact but a valuable diagnostic feature, which motivates the joint detection of fences and their shadows.

Modern approaches to object detection (RE-YOLO [3], DETR [4], Deformable DETR [5]) are oriented toward rigid structures with stable shape. They use fixed positional encoding and are unable to adapt to local deformations, which leads to errors in the detection of fences and their shadows. In this work, an adaptive shadow transformer AST is proposed, which solves these problems using three key mechanisms:

- (1) Separate offset fields for the fence and shadow with different regularization, allowing the model to independently adapt to smooth deformations of the object and sharp breaks in its projection.
- (2) A cross-attention mechanism that takes into account the azimuth and elevation of the sun, as well as the typology of the fence, improving the association of «fence–shadow» pairs.
- (3) A connectivity loss for shadow fragments, teaching the model to perceive fragments broken by terrain as a single object.

In contrast to the works [5, 6], where a common deformation field is introduced for all classes, it is proposed to use separate fields for the object and its shadow with different regularization coefficients. This allows accounting for the different physical nature of their changes – the smooth sagging of the fence and the sharp breaks of the shadow on terrain irregularities.

The article is organized as follows. Section 1 contains a review of related work. Section 2 describes the AST architecture and attention mechanisms. Section 3 is devoted to the loss function. Section 4 presents the experimental results. Section 5 provides information on the availability of the software code. Section 6 contains a discussion of limitations. Section 7 concludes.

1. Review of Related Work

1.1. Classical Methods for Line Detection in Aerial Imagery

The task of extracting linear objects from aerial photographs has a long history and is traditionally solved by methods based on the analysis of contours and gradient characteristics of the image.

The Hough transform is one of the most common approaches to straight line detection [7]. The method operates in parameter space, where each image point votes for all possible lines passing through it. Despite its widespread use, the Hough algorithm has significant drawbacks: high computational complexity and a tendency to detect false lines in areas with high pixel density [8]. In addition, the Hough transform copes poorly with lines that have gaps, which is often encountered in images of fences with frequent supports.

The LSD (*Line Segment Detection*) algorithm [9] uses gradient computation and a subsequent region growing procedure to extract line segments. The main advantage of LSD is its high speed. However, a drawback of the method is the splitting of smooth lines into a large number of isolated fragments, which complicates subsequent parameterization and object search [8].

Methods based on the form factor offer an alternative approach to line detection. In [8], an algorithm for extracting straight contour lines based on the Euclidean form factor was developed. The essence of the method consists in comparing with unity the value of the form factor,

computed as the ratio of the distance between the end points of a contour line to its actual length. The authors show that the developed algorithm improves the stability of line extraction and performance compared with the Hough transform [8]. In [10], it was proposed to use extracted lines as key elements for constructing image descriptors, which provides greater robustness to changes in brightness and contrast compared with the classical SIFT method.

For extracting straight edges in noisy images, a method was proposed that includes oriented filtering and the search for straight-line segments for each direction and scale, taking into account the sign of the gradient [11]. This approach makes it possible to distinguish positive and negative brightness jumps, which is important for detecting objects with different optical properties.

In [12], a comparison of line detection methods for aviation navigation was conducted. The authors evaluated classical methods (Hough transform, LSD) and modern deep learning-based approaches as applied to the task of recognizing runways and taxiways. The main conclusion: deep learning methods demonstrate higher robustness to noise and illumination changes, but are inferior to classical methods in processing speed.

1.2. Modern Methods for Object Detection in Aerial Photographs

Convolutional neural networks are actively used for segmentation and object detection in aerial images. The U-Net architecture [13] and its modifications show high results in semantic segmentation tasks, including the extraction of roads, buildings, and other anthropogenic objects. However, U-Net does not take into account the periodic structure of fences and does not differentiate shadows from objects.

Single-stage detectors of the YOLO family provide high accuracy and processing speed for compact objects. YOLO26 [14] is the latest model of the family, providing native end-to-end inference without post-processing. However, YOLO is ineffective for extended linear structures such as fences, since the prediction grid is not adapted to objects with large length and small width.

Transformer-based detectors offer a fundamentally different approach to detection. The DETR architecture [4] uses an attention mechanism for direct prediction of a set of objects without the need for non-maximum suppression (NMS). Deformable DETR [5] introduces deformable attention,

which allows the model to adapt to the spatial structure of objects. RT-DETR [15] provides high inference speed, approaching YOLO indicators. DINOv2 [16] is a self-supervised visual feature extraction model that can serve as a powerful feature extractor (*backbone*) for detectors, but is not itself a detector and requires attaching a detection head.

In [17], a comprehensive comparison of various detectors was conducted on the task of detecting aircraft in satellite images. The authors show that the choice of architecture significantly affects the detection accuracy of small objects, and YOLO demonstrates the best accuracy–speed trade-off.

1.3. Specialized Methods for Fence Detection

Despite the wide range of object detection methods, the task of detecting fences in aerial photographs long remained on the periphery of research. However, in recent years specialized works have appeared.

In [18], the MultiFranceFences dataset is presented, containing orthophotographs and digital surface models for semantic segmentation of fences in France. The authors use multimodal data (RGB + digital surface model) and show that adding height information significantly improves fence detection accuracy, especially in complex scenes with vegetation.

The synthetic dataset DDOS [19] includes the class «fences» (fine mesh) and is intended for training drones for navigation under obstacle conditions. The dataset contains 340 flights and 34 000 frames with annotations of thin objects.

In [20], a method for detecting fences on construction sites using UAVs was proposed. The authors developed a three-stage approach: first, fence posts are detected, then a horizontal line detector is applied to filter false positives, and finally, the distance between posts is estimated to confirm the presence of a fence.

In the field of environmental monitoring, methods for mapping fences are actively being developed to assess their impact on animal migration [21]. Studies show that fences are an important factor in landscape fragmentation, but their large-scale mapping remains a complex task.

1.4. Methods for Shadow Detection in Aerial Images

Shadow detection is an important task in aerial image processing, since shadows create false boundaries and distort the perception of objects. In [2], an evolutionary shadow correction method based on deep learning was proposed, and a benchmark dataset for UAV images was presented. The authors show that shadows from trees and architectural elements create significant difficulties for standard segmentation algorithms.

A review of methods for detecting clouds and their shadows [22] emphasizes the universality of the task of analyzing «object–projection» pairs in remote sensing. However, existing methods are oriented toward low-resolution satellite images and do not take into account the detailed geometry of roofs and fences. In [23], the MetaShadow method for object-centric detection, removal, and synthesis of shadows was presented. The method uses attention mechanisms to associate objects and their shadows, which is conceptually close to our approach. However, MetaShadow does not take into account the typology of the object and does not use geometric priors (sun position, terrain).

1.5. Datasets for Linear Object Detection

For training and evaluating linear object detection methods, the following datasets are available:

- MultiFranceFences [18] – orthophotographs with fence annotations, territory of France;
- DDOS [19] – synthetic dataset for training drones, includes the class «fences»;
- AI4Boundaries [24] – dataset for field boundary detection, contains Sentinel-2 images;
- Ciyutuo Village Dataset [1] – 65 images for cross-validation; the original polygonal annotations were re-annotated into OBB format for «fence–shadow» pairs.
- ISPRS Vaihingen [25] – benchmark for classification of urban objects and 3D reconstruction of buildings.

None of the existing datasets contains simultaneous annotation of fences and their shadows in OBB format in cottage development areas, taking into account fence typology and terrain. This justifies the need to create our own annotated dataset.

Separately, we highlight the ASD (Airborne-Shadow) dataset [26], presented by Azimi and Bahmanyar (DLR, DAGM GCPR 2023). ASD contains 1408 RGB images of size 512×512 pixels obtained by nadir aerial survey from a helicopter (DLR, Munich) with a GSD of about 13 cm/pixel. The annotation includes two shadow classes – *heavy shadow* and *light shadow* – and covers objects from buildings and bridges to small details, including poles and fences.

However, ASD is oriented toward semantic segmentation of shadows and does not contain paired «fence–shadow» annotations in oriented bounding box (OBB) format, which does not allow it to be used directly for training and quantitative evaluation of the proposed AST method. Nevertheless, ASD can be considered an important reference point for assessing cross-domain generalization ability.

1.6. Methods for Associating «Object–Shadow» Pairs

The task of associating objects and their shadows has a limited number of studies. In [27], an approach based on graph neural networks with the inclusion of physical models for predicting soft tissue deformations was proposed. The authors show that the use of physical priors improves accuracy compared with purely learnable architectures.

In the field of cloud and shadow detection [22], models of atmospheric transfer and projective geometry are used. However, these methods do not take into account the specifics of ground objects (fences, buildings) and their interaction with illumination.

In [28], a method for accounting for geometric transformations based on the theory of group equivariance was proposed, but the authors consider only rigid transformations (rotations, shifts) and do not take into account object deformations.

1.7. Comparative Analysis of Approaches

For a visual comparison of existing approaches by key criteria important for the task of detecting fences and their shadows, Table 1 has been compiled.

Analysis of the table shows that existing approaches either do not take into account object deformations, or ignore shadows, or do not associate objects with their shadows. The proposed AST is the first architecture combining all four key aspects: separate deformation fields, shadow accounting, fence typology accounting, and pair association with geometric priors.

TABLE 1. Comparison of approaches by key criteria

Approach	Deformation object	accounting shadow	Typology	Pair association	Pair detection
<i>Classical methods:</i>					
Hough [7]	—	—	—	—	±
LSD [9]	—	—	—	—	±
Form factor [8]	—	—	—	—	±
<i>Convolutional neural networks:</i>					
U-Net [13]	—	+	—	—	±
YOLO26 [14]	—	—	—	—	—
<i>Transformer-based detectors:</i>					
Deformable DETR [5]	±	—	—	—	±
DINOv2 [16]	—	—	—	—	±
<i>Specialized methods:</i>					
MetaShadow [23]	—	+	—	±	—
MultiFranceFences [18]	—	—	+	—	+
<i>Proposed method:</i>					
AST	+	+	+	+	+

Legend: + fully supports, ± partially supports, — does not support.

2. AST Architecture

2.1. Geometric Problem Statement and Notation

The AST architecture is parameterized by a binary «fence–shadow» system and a set of metric characteristics of their relative arrangement on the orthophotomosaic. These quantities, fixed before describing the model modules, define the structural basis of all key AST components – offset fields, cross-attention, and the consistency loss function – regardless of implementation details.

Sun position. The shadow configuration relative to the fence is uniquely determined by the sun position, characterized by two angles:

- *azimuth* $\varphi_{\text{sun}} \in [0, 2\pi)$ – the compass direction to the sun, determining the *direction* of the shadow;
- *elevation* $\alpha_{\text{sun}} \in (0, \pi/2]$ – the angle above the horizon, determining the *length* of the shadow.

For real aerial photographs, these parameters are extracted from EXIF/XMP metadata (GPS coordinates, date and time of survey) and computed using a standard astronomical solar position model [29]; for synthetic data, they are specified explicitly during generation. In the model, the quantities φ_{sun} and α_{sun} play two roles. First, as an *external geometric prior*: they directly enter the expected shadow displacement

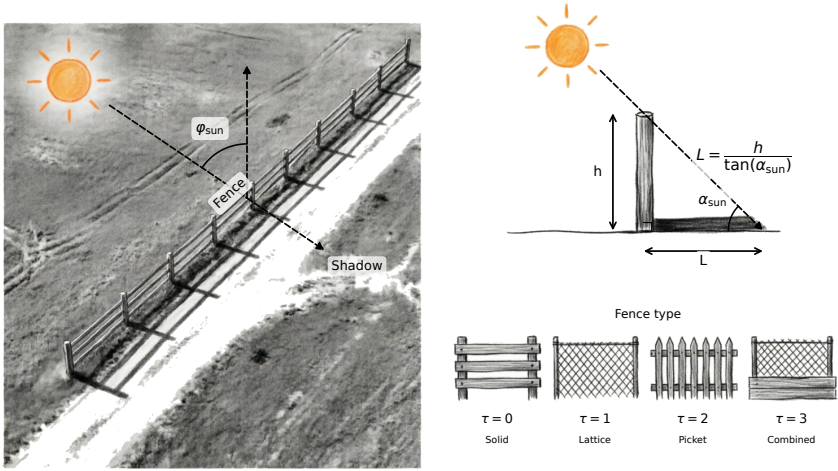


FIGURE 1. Geometry of the «fence-shadow» pair and fence types in AST

vector (subsection 3.3) and the cross-attention score (subsection 2.4), linking the direction and length of the shadow to the position of the fence. Second, for training, an *auxiliary head* is introduced for estimating φ_{sun} and α_{sun} from the image itself, solving a self-supervised task using the same metadata; it does not replace the prior from metadata, but forces the encoder features to encode illumination conditions and increases the model’s robustness to inaccuracy or partial absence of EXIF (see the limitation in subsection 6.2).

Fence type. For each object, the fence type $\tau \in \{0, 1, 2, 3\}$ is determined (solid, lattice, picket, combined). The type affects the expected shadow structure: solid fences produce clear continuous shadows, lattice fences produce intermittent ones, and picket fences produce periodic ones. This information is used further in the cross-attention mechanism (subsection 2.4) as a categorical prior that refines the correspondence between a fence and its shadow, Figure 1.

«Fence-shadow» pair and output format. The model detects both classes – **fence** and **shadow** – and associates them into matched pairs, relying on the quantities introduced above φ_{sun} , α_{sun} , and τ . For each detected object, an oriented bounding box OBB (*oriented bounding box*) is predicted – a rotated rectangle that precisely follows the orientation of the fence or its shadow; the formal definition of the OBB and the output parameters are given in subsection 2.5.

TABLE 2. Parameters of the AST feature extractor

Stage	Output size	Blocks	Attention	Field
1	$H/4 \times W/4 \times 96$	2	Window	No
2	$H/8 \times W/8 \times 192$	2	Deformable	Δ_{fence}
3	$H/16 \times W/16 \times 384$	6	Deformable	$\Delta_{\text{fence}}, \Delta_{\text{shadow}}$
4	$H/32 \times W/32 \times 768$	2	Deformable	$\Delta_{\text{fence}}, \Delta_{\text{shadow}}$

Note. By «deformable attention» we mean a modified version of the attention mechanism in which the sampling positions are shifted based on the predicted field Δ , similar to Deformable DETR [5], but adapted to work with hierarchical Swin Transformer features.

2.2. General Structure of AST

Taking into account the introduced notation, the AST architecture includes the following modules:

- A feature extractor based on Swin Transformer [30]. Unlike the standard Swin Transformer, which uses only window attention, in the proposed architecture window attention is replaced by deformable attention at stages 2–4 (Table 2). This allows the model to adapt to object deformations while preserving the hierarchical structure of Swin.
- Four heads: classification, oriented bounding box (OBB) regression, fence type classification τ , and auxiliary estimation of the sun position (azimuth φ_{sun} and elevation α_{sun}), trained in a self-supervised mode using survey metadata (see subsection 2.1).
- A module of separate offset fields for the fence and shadow.
- A cross-attention mechanism with a geometric prior on φ_{sun} , α_{sun} , and τ .

Unlike Deformable DETR [5], which uses a single offset field, AST predicts two independent fields: Δ_{fence} for the fence and Δ_{shadow} for the shadow.

Below are the specific parameters and settings of each component required to reproduce the model.

Patch embedding layer and encoder block. The input image of size $3 \times H \times W$ is transformed into a sequence of tokens by a patch embedding layer: convolution with kernel and stride 4×4 and 96 output channels, followed by LayerNorm normalization, which yields a feature map $H/4 \times W/4 \times 96$ (stage 1 in Table 2). Each encoder block is built according to the pre-norm scheme and contains two sublayers with residual connections:

- an attention module (window at stage 1 or deformable at stages 2–4) with LayerNorm before attention;

- a two-layer MLP with a hidden expansion of 4 times and GELU activation, also preceded by LayerNorm.

At stages 2–4, the spatial resolution is halved and the number of channels is doubled by a patch merging layer (linear merging of neighboring 2×2 tokens followed by LayerNorm).

Deformable attention. At stages 2–4, each query samples features at $K = 4$ learnable positions at each of $L = 4$ pyramid scales (multiscale mode, as in [5]); the position offsets and scalar level weights are predicted by a linear projector from the query. The number of attention heads in the deformable blocks is fixed and equal to 8 with a head size of 32 dimensions. The offsets are initialized near zero, so that at the beginning of training deformable attention degenerates into local neighborhood attention, and then gradually diverges to the object geometry. The window block of stage 1 uses the standard Swin configuration with a window size of 7×7 and alternation of regular and shifted window attention.

Neck and object queries. The outputs of the four stages are aligned in the number of channels to a common dimension $C = 256$ by 1×1 convolutions with GroupNorm and fed to the transformer decoder. The decoder contains 6 layers; each layer consists of self-attention of object queries, deformable cross-attention to multilevel encoder features, and an FFN. $N = 300$ learnable queries are used, each represented by a pair of *content query* and *positional query*; the latter is initialized from the query reference point and refines the geometric position of the object during decoding. On top of the decoder output for «fence–shadow» pairs, a cross-attention module with a geometric prior is applied (subsection 2.4), associating fence queries with the keys of the corresponding shadows.

Prediction heads. All heads are implemented as three-layer MLPs with a hidden dimension of 256 and ReLU activation. The final linear layers are configured separately for each task.

The classification head uses focal loss [31]. Its bias is initialized to $-\ln((1 - \pi)/\pi)$ with $\pi = 0,01$, which compensates for the imbalance between background and objects.

The OBB regression head predicts the center (x, y) , dimensions (w, h) , and angle θ . The angle is parameterized by the pair $(\sin \theta, \cos \theta)$ and recovered via atan2 . This eliminates the discontinuity at the $\pm\pi$ boundary and corresponds to the format $\mathbf{b} = (x, y, w, h, \theta)$.

The typology head is a linear layer for four classes. Training is performed with cross-entropy.

The sun position head is auxiliary in nature. It regresses the azimuth via a sine–cosine pair, and the elevation α via $\sigma(\cdot) \cdot \pi/2$, which guarantees $\alpha \in (0, \pi/2]$. The head is trained on metadata as a self-supervised task. It does not replace the geometric prior $\varphi_{\text{sun}}, \alpha_{\text{sun}}$ in the model formulas.

The module of separate offset fields (subsection 2.3) is implemented as two parallel linear projectors. Each outputs $2LK$ offsets and LK attention weights for its class $c \in \{\text{fence}, \text{shadow}\}$.

Regularization and initialization. To prevent overfitting, dropout with a drop probability of 0.1 is used in all attention and MLP blocks. In the encoder blocks, stochastic depth (*drop path*) is additionally applied: the probability of skipping a block increases linearly from 0 at the first stage to 0.2 at the fourth. Weight initialization of linear layers is performed according to the Xavier scheme [32], and absolute positional encodings are initialized with a truncated normal distribution. This combination of regularization and initialization strategies ensures stable training of two independent offset fields, preventing the collapse of one of them in the initial epochs.

2.3. Mechanism of Separate Offset Fields

In the standard Deformable DETR, for each query q_i , an offset Δ_{ij} is predicted:

$$\Delta_{ij} = s \cdot \tanh(W_{\text{offset}} q_i + b_{\text{offset}}), \quad \Delta_{ij} \in \mathbb{R}^2.$$

In AST, the offset depends on the class $c \in \{\text{fence}, \text{shadow}\}$:

$$\Delta_{ij}^{(c)} = s_c \cdot \tanh(W_{\text{offset}}^{(c)} q_i + b_{\text{offset}}^{(c)}),$$

where $s_{\text{fence}} = 0,05$ (fences deform smoothly), $s_{\text{shadow}} = 0,35$ (shadows may have sharp breaks on terrain irregularities). The scales were chosen empirically based on the analysis of object shape variability in the dataset. The deformed position for the key/value has the following form:

$$p_{j|i}^{(\text{def}, c)} = p_j + \Delta_{ij}^{(c)}.$$

2.4. Cross-Attention Mechanism with Geometric Prior

The cross-attention mechanism solves the matching problem: for each detected fence, it determines which of the shadows in the image belongs to it. The degree of correspondence of a «fence–shadow» pair is expressed by the attention weight. This weight, in addition to feature similarity, is influenced by three numerical characteristics of the scene: the fence type τ , the sun azimuth φ_{sun} , and the sun elevation α_{sun} . Their meaning and

physical justification are given in subsection 2.1. The attention weight is computed as:

$$(1) \quad \alpha_{ij}^{\text{cross}} = \text{Softmax}_j \left(\frac{q_i^{\text{fence}} \cdot k_j^{\text{shadow}}}{\sqrt{d}} + \text{PE}(p_j^{\text{shadow}} - p_i^{\text{fence}}) \right. \\ \left. + \text{PE}_{\text{az}}(\varphi_{\text{sun}}) + \text{PE}_{\text{el}}(\alpha_{\text{sun}}) + \text{PE}_{\tau}(\tau_{\text{fence}}) \right),$$

where Softmax is applied over the index j for each fixed i .

The positional encoding of the sun azimuth and elevation uses a sinusoidal representation:

$$\text{PE}_{\text{az}}(\varphi) = [\sin(2^k \varphi), \cos(2^k \varphi)]_{k=0}^{K-1}, \quad K = 4,$$

$$\text{PE}_{\text{el}}(\alpha) = [\sin(2^k \alpha), \cos(2^k \alpha)]_{k=0}^{K-1}, \quad K = 4.$$

The azimuth φ_{sun} determines the expected *direction* of the shadow relative to the fence, and the elevation α_{sun} determines the expected *length* of the shadow. The joint use of both angles allows the model to correctly associate fences with their shadows under various illumination conditions.

2.5. Output Data Format

The AST model predicts for each detected object an oriented bounding box OBB (*Oriented bounding box*) – a rotated rectangle that precisely follows the orientation of the fence or its shadow:

$$\mathbf{b} = (x, y, w, h, \theta),$$

where (x, y) are the coordinates of the box center, w is the width, h is the height, and θ is the rotation angle. For each object, the class, fence type τ , and confidence score are additionally predicted.

2.6. Inference and Grouping of Shadow Fragments

The fragment connectivity loss $\mathcal{L}_{\text{conn}}$ is used not only at the training stage but also at the inference stage. During training, an auxiliary classification task is formed for pairs of shadow fragments: whether two fragments belong to the same shadow broken by terrain, vegetation, or local occlusions.

At the inference stage, the model first predicts a set of detections of the classes **fence** and **shadow**. For each shadow, a feature vector is additionally formed, used by the auxiliary connectivity head. For a pair of shadow fragments i and j , the probability of their belonging to the same

shadow is computed as:

$\hat{p}_{ij} = \sigma \left(\text{MLP}([q_i^{\text{shadow}}, q_j^{\text{shadow}}, p_i^{\text{shadow}} - p_j^{\text{shadow}}, \Delta_i^{\text{shadow}}, \Delta_j^{\text{shadow}}]) \right)$,
 where q_i^{shadow} and q_j^{shadow} are the features of the corresponding object queries, $p_i^{\text{shadow}}, p_j^{\text{shadow}}$ are the centers of the predicted OBBs, and $\Delta_i^{\text{shadow}}, \Delta_j^{\text{shadow}}$ are the corresponding offset fields.

Next, a connectivity graph $G = (V, E)$ is constructed, where the vertices V are the detected shadow fragments, and edges are drawn only between pairs of fragments whose center distance does not exceed the admissible radius $R_{\max}$. Edges with $\hat{p}_{ij} < \tau_{\text{conn}}$ are removed. The connected components of the graph are interpreted as separate shadows. In the implementation, by default, $\tau_{\text{conn}} = 0,5$ and $R_{\max} = 3,5$ m are used, converted to pixels taking into account the GSD. The value of $R_{\max}$ is justified by the physical constraints of the scene: at a maximum fence height $h \approx 2,5$ m and a minimum sun elevation $\alpha_{\text{sun}} = 35^\circ$ (a typical constraint for aerial surveying to avoid excessively long shadows overlapping neighboring plots), the maximum theoretical shadow length is $2,5 / \tan(35^\circ) \approx 3,57$ m. Thus, $R_{\max}$ covers all physically possible shadow breaks for the given survey range. For images with a different GSD or at a different time of day, $R_{\max}$ is automatically scaled proportionally to the length of the reference shadow.

This procedure allows preserving the advantages of end-to-end detection while simultaneously ensuring the grouping of broken fragments. If fragments of the same shadow have a significant physical break caused by terrain or a tree crown, the model does not merge them into a single continuous box, but assigns them a single group identifier `shadow_group_id`. If subsequent vectorization is required, the group can be approximated by a single polyline or a minimal oriented rectangle.

3. Loss Function

The final loss function is a weighted sum of four components:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{det}} + \lambda_1 \mathcal{L}_{\text{def}} + \lambda_2 \mathcal{L}_{\text{cross}} + \lambda_3 \mathcal{L}_{\text{conn}},$$

where $\lambda_1 = 0,1$, $\lambda_2 = 0,5$, $\lambda_3 = 0,3$. The coefficients were selected using grid search with a step of 0.1 for λ_1 and 0.05 for λ_2, λ_3 on the validation set (10% of the data). The optimal values correspond to the best mAP on validation.

3.1. Detection Loss

The detection loss combines three components:

$$\mathcal{L}_{\text{det}} = \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{box}} + \mathcal{L}_{\text{giou}},$$

where $\mathcal{L}_{\text{cls}}$ is the focal loss [31] for classification, $\mathcal{L}_{\text{box}}$ is the L1 loss for OBB parameter regression, and $\mathcal{L}_{\text{giou}}$ is the generalized IoU loss computed for oriented boxes R-GIoU (*Rotated GIoU*) [33]. Unlike the standard horizontal GIoU, R-GIoU computes the intersection area of rotated rectangles, which is correct for linear objects of arbitrary orientation.

3.2. Regularization of Offset Fields

Since fences and shadows have different natures of deformations, the work uses *different* regularization coefficients:

$$(2) \quad \mathcal{L}_{\text{def}} = \frac{1}{N} \sum_{i=1}^N \left(\mu_{\text{fence}} \|\nabla \Delta_{\text{fence}}(p_i)\|_F^2 + \mu_{\text{shadow}} \|\nabla \Delta_{\text{shadow}}(p_i)\|_F^2 \right),$$

where $\mu_{\text{fence}} = 1,0$ and $\mu_{\text{shadow}} = 0,2$. A weaker coefficient for Δ_{shadow} allows the shadow offset field to preserve sharp breaks at the boundaries of terrain irregularities, while the fence field remains smooth. This separation eliminates the contradiction between the requirement of smoothness for the fence and the admissibility of breaks for the shadow.

3.3. «Fence–Shadow» Consistency Loss

The consistency loss ensures a geometric link between matched pairs:

$$\mathcal{L}_{\text{cross}} = \sum_{(i,j) \in \mathcal{P}} \|(p_j^{\text{shadow}} - p_i^{\text{fence}}) - \mathbf{d}_{\text{proj}}\|^2,$$

where $\mathcal{P}$ is the set of matched «fence–shadow» pairs, and $\mathbf{d}_{\text{proj}}$ is the expected shadow displacement vector computed from the geometric model.

For flat terrain, the shadow displacement vector is determined by the fence height h_{fence} , the sun elevation α_{sun} , and the azimuth φ_{sun} :

$$(3) \quad \mathbf{d}_{\text{proj}} = -\frac{h_{\text{fence}}}{\tan \alpha_{\text{sun}}} \begin{bmatrix} \sin \varphi_{\text{sun}} \\ \cos \varphi_{\text{sun}} \end{bmatrix}.$$

Here, the factor $h_{\text{fence}} / \tan \alpha_{\text{sun}}$ determines the *length* of the shadow (the lower the sun, the longer the shadow); the unit vector $[-\sin \varphi, -\cos \varphi]^\top$ sets the *direction* opposite to the sun azimuth.

Adaptation to complex terrain. On rugged terrain, the real shadow projection deviates from the theoretical one, which could lead to incorrect penalization of the model (*conflicting gradients*). To solve this problem, an

adaptive weight coefficient ω_{relief} is introduced, depending on the local surface slope, and the consistency loss is modified:

$$(4) \quad \mathcal{L}_{\text{cross}} = \sum_{(i,j) \in \mathcal{P}} \omega_{\text{relief}}(p_i^{\text{fence}}) \cdot \|(p_j^{\text{shadow}} - p_i^{\text{fence}}) - \mathbf{d}_{\text{proj}}\|^2.$$

The weight $\omega_{\text{relief}} \in [0, 1]$ is computed based on the local elevation gradient (available at the training stage from a synthetic DEM or approximated from textures on real data). On flat areas, $\omega_{\text{relief}} \rightarrow 1$, providing a rigid geometric constraint. In zones of strong elevation change (slopes, ravines), $\omega_{\text{relief}} \rightarrow 0$, allowing the model to rely on visual features and the connectivity loss $\mathcal{L}_{\text{conn}}$, which is specifically designed to work with broken shadows on complex terrain.

Formula (3) is also valid for orthorectified images (nadir survey, constant scale). For oblique images, it must be generalized taking into account the homography linking the ground plane to the image plane.

The fence height h_{fence} is estimated from a digital surface model (DSM) obtained from a stereo pair or lidar data. In the absence of a DSM, the average value for the given fence type is used: $h_{\text{solid}} = 1,8$ m (solid fence), $h_{\text{lattice}} = 1,5$ m (lattice fence), $h_{\text{picket}} = 1,2$ m (picket fence), $h_{\text{comb}} = 1,6$ m (combined fence).

Separate attention is required for the coordination of coordinate systems. The sun azimuth φ_{sun} is specified relative to geographic north, whereas formula (3) uses the image orientation (X to the right, Y downward). The correspondence is achieved by taking into account the yaw angle of the tile extracted from EXIF metadata, which ensures correct recalculation of the shadow direction in pixel coordinates.

3.4. Shadow Fragment Connectivity Loss

The connectivity loss teaches the model to perceive fragments broken by terrain as a single object:

$$\mathcal{L}_{\text{conn}} = - \sum_{(i,j) \in \mathcal{F}} [y_{ij} \log p_{ij} + (1 - y_{ij}) \log(1 - p_{ij})],$$

where $\mathcal{F}$ is the set of shadow fragment pairs, $y_{ij} = 1$ if fragments i and j are parts of the same shadow broken by terrain, and $y_{ij} = 0$ otherwise. The pairs $\mathcal{F}$ are formed at the annotation stage: annotators mark fragments belonging to the same shadow with a single identifier `shadow_group_id`. At the training stage, for each image, up to 32 fragment pairs are randomly selected in a balanced manner (50% positive, 50% negative). The quantity p_{ij} is the probability predicted by the model that fragments i and j belong to the same shadow.

TABLE 3. Composition of the FSAD dataset

Fence type	Images	Objects	Pairs
Solid	220	440	380
Lattice	180	320	280
Picket	150	300	250
Combined	100	180	150
Total	650	1240	1060

4. Experimental Results

4.1. Datasets and Training Procedure

Training and validation were conducted on a combination of data:

- (1) The proprietary FSAD dataset (*Fence & Shadow Aerial Dataset*), containing 650 aerial photographs of cottage development with manual annotation of fences and their shadows. The images were obtained from UAVs (resolution 5280×3956 , flight altitude 80–120 m, camera oriented strictly at nadir) in 2023–2025 in four regions of Russia (Moscow, Kaluga, Tula, and Tver oblasts) and orthorectified in Agisoft Metashape using RTK-GPS data and ground control points, which ensures constant scale and the absence of perspective distortions; the average ground sampling distance (GSD) is 2,1–2,8 cm/pixel.

Annotation was performed in *Supervisely*^{URL} by three annotators with subsequent verification. For each object, the following were indicated: oriented bounding box (OBB), class (fence/shadow), shadow group identifier (`shadow_group_id`), fence type. Survey metadata (date, time, GPS coordinates) were saved for subsequent computation of the sun position. The composition of the dataset is given in Table 3.

- (2) 2000 images of synthetic data generated in *Unreal Engine 5*^{URL} using the open-source plugin *Colosseum*^{URL} (a fork of the archived AirSim). Four types of fences were modeled, terrain with slope angles of 0–30°, sun azimuth 30–150°, and sun elevation 20–60°. The virtual camera was oriented strictly at nadir, and the rendering resolution was selected so that the GSD corresponded to real images (2,1–2,8 cm/pixel). This ensures consistency of shadow geometry between synthetic data and the projection formula (3) used in $\mathcal{L}_{\text{cross}}$.

For each frame, the exact values of φ_{sun} , α_{sun} , and the elevation map were saved. To minimize the domain gap between synthetic and real data, a set of tonal and stylistic adaptation measures was used. In particular, for synthetic images, histogram matching was performed to align the global brightness and contrast distribution with real FSAD orthophotos. Additionally, random color jittering and haze simulation were used in the Unreal Engine 5 render pipeline, which imitates the optical distortions of a real camera and smooths out the unnaturally sharp boundaries of synthetic shadows.

Figure 2 presents a visual comparison of a synthetic image before (Figure 2a) and after (Figure 2b) applying the adaptation

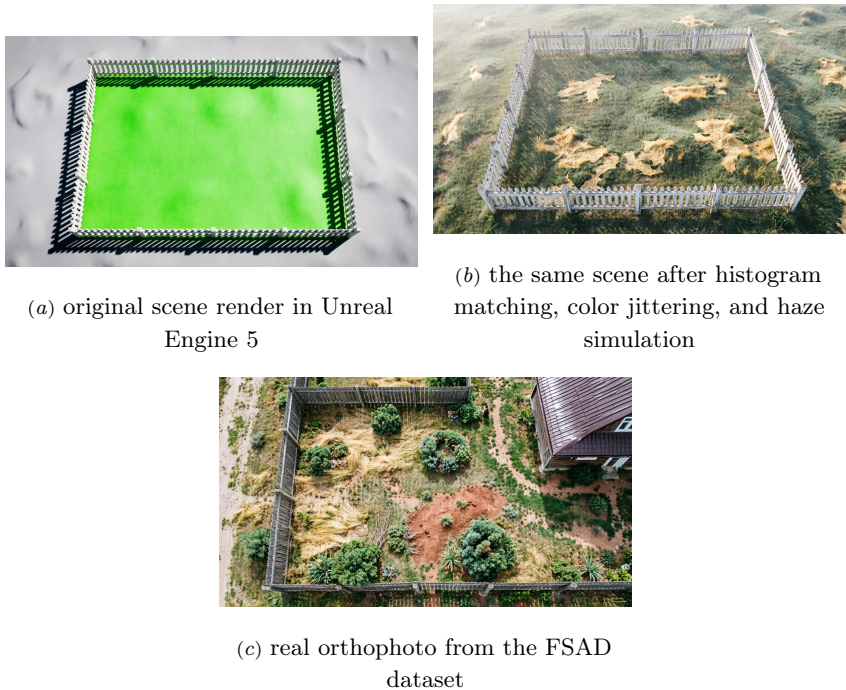


FIGURE 2. Visual assessment of domain gap minimization

complex with a real FSAD orthophoto (Figure 2c). All images are shown after orthorectification (nadir geometry). Brightness distributions are shown in Figure 3.

Histogram matching and haze simulation smooth out the unnatural

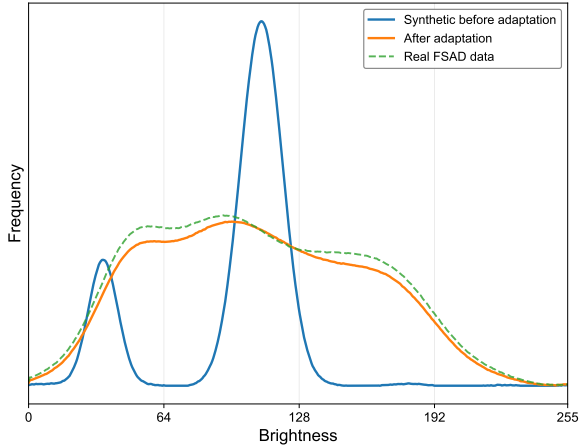


FIGURE 3. Brightness distributions of synthetic data before and after adaptation in comparison with real FSAD data (illustration for Figure 2)

sharpness of synthetic shadows and align the global brightness distribution (Figure 3), which confirms the effectiveness of the chosen strategy for minimizing the domain gap.

③ Ciyutuo Village Dataset [1] – 65 images for cross-validation.

The dataset was divided into training (80%), validation (10%), and test (10%) sets, stratified by fence type and region. Training was conducted in two stages: pretraining on synthetic data (30 epochs), fine-tuning on real data (90 epochs). Optimizer AdamW, initial learning rate $\eta = 10^{-4}$ with cosine decay, batch size 8, input image resolution 1024×1024 . Training was performed on $2 \times \text{NVIDIA A100}$ (80 GB), total time – about 8 hours. All results are reported as the mean over three independent runs with different initializations.

To control overfitting on the small real sample, the following measures were applied. The number of epochs at both stages was fixed (30 and 90). Early stopping was not used, since cosine learning rate decay combined with regularization (dropout and stochastic depth, Section 2), as well as weight decay = 0.01 in AdamW, ensured monotonic growth of mAP@0.5 on validation up to the last epoch without divergence of the training and validation curves (Figure 5). In the first 500 iterations of each stage, linear

warmup of the learning rate from 0 to η was applied. For testing, in each of the three runs, the checkpoint with the best mAP@0.5 on the validation set was saved. These weights were used in all tables of [Section 4](#).

Augmentation was limited to photometric transformations (brightness, contrast, color noise, blur) and geometric transformations preserving the nadir projection: rotations by multiples of 90° and reflections. With each such rotation/reflection, the sun azimuth value φ_{sun} is corrected analytically so that the geometric prior in (1) and (3) remains valid. Perspective and arbitrary affine distortions are not applied, since they violate the flat shadow geometry underlying the loss $\mathcal{L}_{\text{cross}}$.

When using random crops, which can break a «fence–shadow» pair, the «orphaned objects» strategy is applied: if one element of a pair falls into the crop and the other does not, the remaining object continues to participate in training of the standard detection loss $\mathcal{L}_{\text{det}}$, but the association component $\mathcal{L}_{\text{cross}}$ for this pair is forcibly zeroed. This allows the model to learn object detection independently of the presence of their shadows in the frame.

4.2. Baseline Models for Comparison

For an objective evaluation of the effectiveness of the proposed AST method, four competitor models representing different architectural approaches to object detection were used in the experiments:

- YOLO26 [14] – the latest version of the YOLO family, providing native end-to-end inference without post-processing. It was chosen as a representative of single-stage detectors focused on high speed.
- RT-DETR [15] – a real-time transformer detector combining attention mechanisms with optimizations for fast inference.
- Deformable DETR [5] – a detector with deformable attention, which serves as the direct base architecture for the proposed AST. Its choice allows evaluating the contribution of the proposed modifications.
- DINOv2+DETR – a combination of the powerful self-supervised feature extractor DINOv2 (ViT-B/14) [16] and the standard DETR detection head [4] with 300 queries. DINOv2 itself is not a detector, so for a fair comparison, a detection head similar to that used in the original DETR was attached to it. The DINOv2 weights were not

TABLE 4. Comparison of detection accuracy (mAP@0.5 $\uparrow$ based on Rotated IoU, mean $\pm$ standard deviation over three runs)

Model	Fences	Shadows on terrain flat complex		Average*
YOLO26 [14]	0,795 $\pm$ 0,008	0,652 $\pm$ 0,011	0,481 $\pm$ 0,014	0,698
RT-DETR [15]	0,812 $\pm$ 0,007	0,668 $\pm$ 0,010	0,498 $\pm$ 0,013	0,714
Deformable DETR [5]	0,821 $\pm$ 0,006	0,682 $\pm$ 0,009	0,522 $\pm$ 0,012	0,727
DINOv2+DETR [16]	0,834 $\pm$ 0,006	0,698 $\pm$ 0,009	0,543 $\pm$ 0,011	0,743
AST	0,851 $\pm$ 0,005	0,725 $\pm$ 0,008	0,602 $\pm$ 0,009	0,769

Note. Average* is weighted by the number of objects in the test set: fences – $N_1 = 106$, shadows (flat terrain) – $N_2 = 74$, shadows (complex terrain) – $N_3 = 32$.

frozen; full joint training with the detection head was performed, but a reduced learning rate ($\eta_{\text{backbone}} = 10^{-5}$) was used for the backbone layers, which is standard practice to prevent catastrophic forgetting of pretrained visual features.

All listed models were trained on the same FSAD dataset with an identical split into training, validation, and test sets. To ensure a fair comparison, all training conditions were unified: input image resolution (1024 $\times$ 1024), augmentation set, optimizer, and number of epochs. This allows direct comparison of results, since all differences are due solely to the architectural features of the models.

4.3. Comparison of Detection Accuracy

The results of the detection accuracy comparison are presented in Table 4. The values of the mAP@0.5 metric (Rotated IoU was used) were computed separately for fences, shadows on flat terrain, and shadows on complex terrain. For each model, the mean values and standard deviations over three independent runs are given. The «Average» column contains the mAP@0.5 value weighted by the number of objects in the test set, taking into account the different number of objects in each scenario; the corresponding weights N_1 , N_2 , N_3 are given in the note to the table.

TABLE 5. Error analysis (shadows on complex terrain)

Model	FNR ↓	FPR ↓	Breaks ↓
YOLO26 [14]	0,438	0,132	56%
RT-DETR [15]	0,385	0,108	48%
Deformable DETR [5]	0,332	0,088	39%
DINOv2+DETR [16]	0,288	0,075	32%
AST	0,175	0,052	8%

Note. Here and below, arrows indicate the direction of metric improvement:
 ↑ – larger values are better, ↓ – smaller values are better.

The proposed AST method outperforms all baseline models in all categories. The most significant advantage is achieved in the detection of shadows on complex terrain: the absolute gain in mAP@0.5 over the best baseline model (DINOv2+DETR) is 5.9 p.p. (percentage points) – 0.602 versus 0.543, which corresponds to a relative improvement of approximately 10.9%. For fences, the absolute gain of AST over DINOv2+DETR is 1.7 p.p.

For a more detailed analysis of errors in shadow detection on complex terrain, Table 5 provides three additional indicators: false negative rate FNR, false positive rate FPR, and the proportion of shadows broken into fragments. All three metrics are minimized, i.e., the lower the value, the better.

The «Breaks» metric in Table 5 shows the *proportion* of ground-truth shadows on complex terrain that the model erroneously split into several disconnected OBB boxes. For example, the value of 0.56 for YOLO26 means that 56% of shadows were detected with breaks (as several separate objects). The closer the value is to 0, the better the model preserves the geometric integrity of shadows. AST demonstrates a significant reduction in all three types of errors. Particularly indicative is the proportion of broken shadows: thanks to the fragment connectivity loss ($\mathcal{L}_{\text{conn}}$), this indicator is reduced from 32% for DINOv2+DETR to 8% for AST. The significant decrease in FNR (from 28.8 to 17.5%) confirms that the model misses objects less often, and the decrease in FPR (from 7.5 to 5.2%) indicates a reduction in false detections.

TABLE 6. Contribution of components (mAP@0.5 for shadows on complex terrain; values in parentheses are gains in p.p.)

Configuration	mAP $\uparrow$
Deformable DETR (base)	0,522
+ separate fields $\Delta_{\text{fence}}, \Delta_{\text{shadow}}$	0,548 (+2,6)
+ separate regularization ($\mu_{\text{fence}} \neq \mu_{\text{shadow}}$)	0,556 (+0,8)
+ fence typology module	0,574 (+1,8)
+ cross-attention with $\varphi_{\text{sun}}, \alpha_{\text{sun}}$	0,588 (+1,4)
+ connectivity loss $\mathcal{L}_{\text{conn}}$	0,602 (+1,4)

4.4. Ablation Analysis

To evaluate the contribution of each proposed component, an ablation analysis was conducted. The baseline is Deformable DETR, to which modifications implemented in AST were sequentially added. The results are presented in Table 6.

The largest gain comes from the separation of deformation fields (+2.6 p.p.), which confirms the hypothesis about the different nature of fence and shadow deformations. Separate regularization (2) additionally improves the result by 0.8 p.p., allowing the shadow field to preserve breaks without losing the smoothness of the fence field. The addition of the fence typology module and cross-attention taking into account the sun position gives gains of 1.8 and 1.4 p.p., respectively. The shadow fragment connectivity loss $\mathcal{L}_{\text{conn}}$ provides another 1.4 p.p., which is especially important for scenes with complex terrain.

As already noted, the fence height h_{fence} directly determines the expected shadow length in $\mathcal{L}_{\text{cross}}$ (see formula (3)). In this regard, it is of interest to evaluate the model’s robustness to errors in specifying this parameter. During testing, the fence height was varied in the range of $\pm 30\%$ relative to the base values (for a solid fence, for example, values from 1.26 to 2.34 m were considered). According to the obtained data, the decrease in mAP@0.5 for shadows on complex terrain did not exceed 1.8 p.p. Such an insignificant degradation is explained by the fact that $\mathcal{L}_{\text{cross}}$ acts as a soft prior rather than a hard constraint, so the model compensates for the inaccuracy of the fence height through visual features and the cross-attention mechanism.

TABLE 7. mAP@0.5 for shadows on complex terrain under degraded conditions (values in parentheses are drops in p.p.)

Model	Original	+ Noise ($\sigma=25$)	+ $\Delta\varphi = 15^\circ$
YOLO26 [14]	0,481	0,358 (−12,3)	0,372 (−10,9)
RT-DETR [15]	0,498	0,375 (−12,3)	0,389 (−10,9)
Deformable DETR [5]	0,522	0,402 (−12,0)	0,418 (−10,4)
DINOv2+DETR [16]	0,543	0,428 (−11,5)	0,445 (−9,8)
AST	0,602	0,541 (−6,1)	0,558 (−4,4)

4.5. Robustness to Changes in Survey Conditions

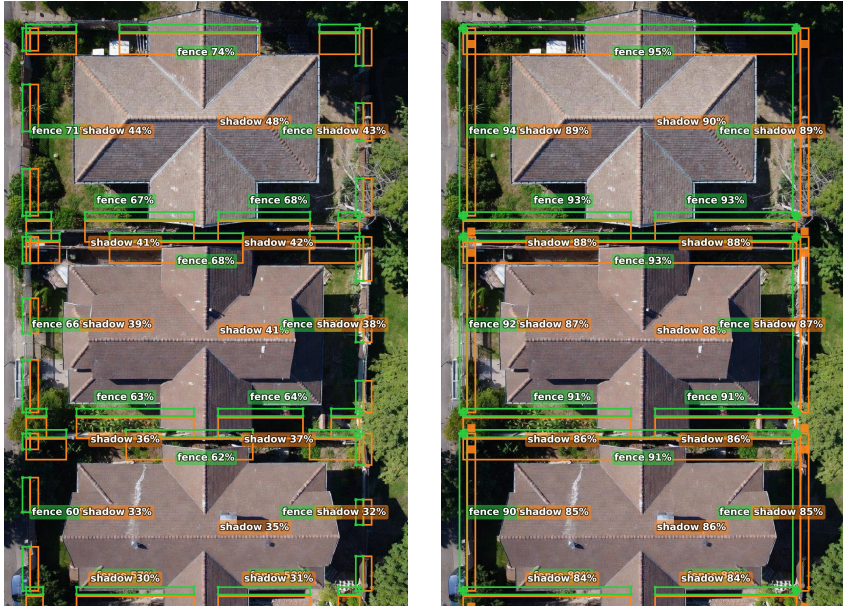
To evaluate the robustness of the models to data quality degradation, experiments were conducted with the addition of Gaussian noise ($\sigma = 25$) and a change in the sun azimuth by $\Delta\varphi = 15^\circ$. The results are presented in Table 7.

AST demonstrates a significantly smaller drop in accuracy under image quality degradation and changes in illumination conditions. When noise is added, the mAP drop for AST is 6.1 p.p. versus 11.5–12.3 p.p. for baseline models. When the azimuth changes by 15° , the AST drop is 4.4 p.p. versus 9.8–10.9 p.p. for competitors. Such robustness is explained by the explicit accounting of the sun position in the cross-attention mechanism (1), which allows the model to adapt to illumination changes without retraining for specific survey conditions.

4.6. Qualitative Results

Figure 4 presents examples of detection on an orthophoto of a cottage plot (strictly top-down view, nadir), illustrating the advantage of the proposed approach in preserving the integrity of the fence and its shadow. The scene from the test images of the FSAD dataset contains a rectangular picket fence perimeter; solar illumination forms a striped shadow along each solid side of the perimeter on the shaded side.

Figure 4a shows the result of the baseline Deformable DETR model. The model perceives the fence and its shadow as a set of disparate fragments along individual sides of the perimeter, without linking them through corners: the shadow is split into many disconnected detections with low



(a) Deformable DETR: fence and shadow split into disconnected fragments

(b) AST: holistic perception of the fence and shadow as a single object along the shaded side

FIGURE 4. Qualitative comparison on an orthophoto of a cottage plot (nadir geometry, after orthorectification)

confidence (30–48%), and the confidence monotonically decreases from the top cottage to the bottom one (44–48% on the top, 36–42% on the middle, and 30–35% on the bottom).

The fence is also detected only partially, in short segments along individual sections of the fence, with a noticeable spread across cottages: 67–74% on the top, 63–68% on the middle, and 60–62% on the bottom. The integrity of the perimeter is completely lost: the contour does not close at the corners, and the shadow breaks up into separate rectangular fragments at each side.

Figure 4b presents the AST result. The model restores a continuous fence contour along the entire perimeter with closed corners (confidence 90–95% across the three cottages) and perceives the shadow as a single object – a closed band along the shaded side (confidence 84–90%). At

TABLE 8. Comparison of computational complexity of models (resolution 1024×1024)

Model	Parameters (M)	FLOPs	FPS $\uparrow$
YOLO26 [14]	12,4	45,2	84,5
RT-DETR [15]	32,1	98,5	42,3
Deformable DETR [5]	40,5	172,4	18,6
DINOv2+DETR [16]	86,2	215,3	11,4
AST	48,7	195,8	16,2

the same time, the confidence also monotonically decreases from the top cottage to the bottom one (95/90% on the top, 93/88% on the middle, 91/84% on the bottom), which reflects the deteriorating visibility of the fence and shadow in the lower part of the image. Unlike Deformable DETR, the fence contour remains continuous on all four sides of each plot, and the shadow is reproduced as a solid band without breaks or false splits. Several more examples of the Deformable DETR and AST models are given in the [Appendix](#).

The integrity of the result is achieved through two mechanisms. First, the connectivity loss $\mathcal{L}_{\text{conn}}$ teaches the network to perceive all fragments of one shadow as a single whole, which reduces the number of erroneous breaks from 32% to 8% ([Table 5](#)); this also holds for scenes with complex terrain. The second mechanism is separate offset fields for the fence and shadow with different regularization, which allows preserving the fence shape at bends and accurately reproducing the shadow geometry. The visual improvement is confirmed numerically: the AST confidence in the fence is 90–95% and in the shadow – 84–90%, whereas for the baseline Deformable DETR these values are 67–74% and 30–48%, respectively, which is consistent with the data in [Table 4](#).

4.7. Computational Complexity and Inference Speed

For an objective comparison of different detector architectures, [Table 8](#) provides computational complexity metrics: the number of parameters, the number of floating-point operations (FLOPs), and inference speed (FPS). Along with lightweight detectors (YOLO26, RT-DETR), the comparison includes heavier models: Deformable DETR, as well as the combination of DINOv2 with DETR, which allows evaluating the position of the proposed AST model among the considered architectures.

As can be seen from Table 8, the presented models demonstrate a natural trade-off between computational efficiency and architectural complexity. Lightweight detectors (YOLO26, RT-DETR) expectedly show higher inference speed due to fewer parameters and FLOPs. At the same time, transformer architectures, including Deformable DETR and DINOv2+DETR, have significantly greater computational capacity due to self-attention mechanisms and multistage feature processing.

The proposed AST model holds an intermediate position: in terms of parameter count (48.7 M) and FLOPs (195.8), it is comparable to Deformable DETR, but is somewhat slower at inference (16.2 FPS versus 18.6). This is attributable to the additional shadow-connectivity graph construction performed at the post-processing stage. Nevertheless, the accuracy gain on complex terrain (+5.9 p.p. in mAP) over the best baseline model (DINOv2+DETR) justifies the additional computational cost for cadastral accounting tasks, where inference is performed offline and speed requirements are not critical.

The additional computational costs of AST at the inference stage are associated with the construction of the shadow connectivity graph (subsection 2.6). The theoretical complexity of pairwise comparison of fragments is $O(N^2)$, where N is the number of detected shadows. However, in practice, this operation is not a bottleneck for two reasons:

- (1) Spatial pruning – graph edges are built only for pairs whose distance does not exceed $R_{\max}$. The use of a spatial index (e.g., a KD-tree) reduces the complexity of candidate search to $O(N \log N)$.
- (2) Limited number of shadows: on a typical 1024×1024 tile of cottage development, the number of detected shadows N rarely exceeds 30–50, which makes the computation of the MLP for pairwise features negligible (less than 2 ms on GPU).

Thus, the grouping procedure does not impair the overall performance of the system.

4.8. Cross-Validation on an External Dataset

To evaluate the generalization ability of the model, testing was conducted on the external Ciyutuo Village dataset [1], containing 65 images from China. The original annotation of the dataset does not contain the classes «fence» and «shadow» and is made in polygon format, so for the

TABLE 9. Cross-validation results on the Ciyutuo Village dataset (mAP@0.5, mean $\pm$ standard deviation over three runs)

Model	Fences	Shadows
YOLO26 [14]	0,595 $\pm$ 0,016	0,351 $\pm$ 0,018
RT-DETR [15]	0,622 $\pm$ 0,015	0,378 $\pm$ 0,017
Deformable DETR [5]	0,641 $\pm$ 0,014	0,412 $\pm$ 0,016
DINOV2+DETR [16]	0,679 $\pm$ 0,013	0,453 $\pm$ 0,015
AST	0,714 $\pm$ 0,011	0,538 $\pm$ 0,013

purposes of this work, all 65 images were re-annotated into OBB format for «fence–shadow» pairs according to the same rules as FSAD (oriented box, class, fence type, shadow group identifier). Re-annotation was performed manually; none of these images were used in the training and validation of AST. The results are presented in Table 9.

Since the Ciyutuo Village scenes contain predominantly rugged terrain, the drop in shadow accuracy in the text is calculated relative to the «complex terrain» column of Table 4.

AST outperforms all four baseline models in both fence and shadow detection. The absolute gain over the best baseline model (DINOV2+DETR) is 3.5 p.p. for fences and 8.5 p.p. for shadows; over Deformable DETR – 7.3 and 12.6 p.p. respectively. The indicators on Ciyutuo Village are lower than on FSAD for all models without exception, which is expected when the domain changes. At the same time, the magnitude of the accuracy drop correlates with the robustness of the model: YOLO26 loses 20.0 p.p. in fences and 13.0 p.p. in shadows relative to its results on FSAD, whereas AST loses only 13.7 and 6.4 p.p. This is consistent with the results of Table 7 and confirms that explicit accounting of shadow geometry makes the model less sensitive to the data distribution.

The decrease in absolute accuracy is explained by differences in the architecture of development, fence types, and survey conditions between the Russian regions (FSAD) and the territory of China (Ciyutuo Village). The consistent gain of AST over the entire spectrum of baseline models shows that the proposed mechanisms – separate offset fields, fence typology accounting, and geometric priors of the sun position – are effective not only on the data used for training but also on new scenes with a different feature distribution.

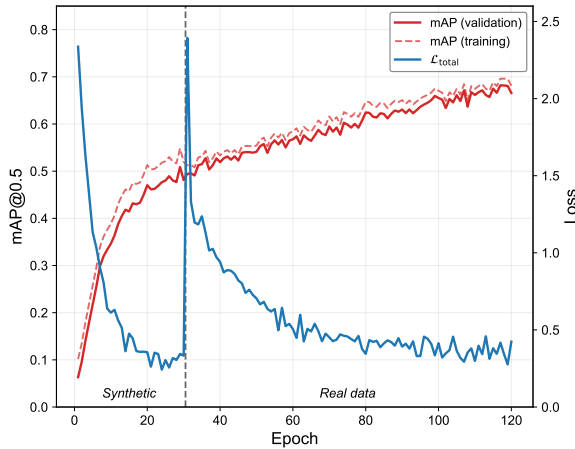


FIGURE 5. Training dynamics of AST: $mAP@0.5$ on the validation (solid red) and training (dashed red) sets and the total loss $\mathcal{L}_{total}$ (blue). The vertical line is the boundary between the synthetic pretraining stage and the real-data fine-tuning stage

Additionally, let us consider the ASD (Airborne-Shadow) dataset [26]. Its annotation is oriented toward semantic segmentation of shadows (classes *heavy* and *light shadow*) and does not contain paired «fence–shadow» annotation in OBB format. Therefore, direct quantitative comparison with AST using the $mAP@0.5$ metric is impossible without re-annotating the dataset. Nevertheless, ASD confirms the relevance of the task of fine-grained shadow detection in nadir aerial images (ASD, DLR) and is considered a promising direction for adapting the model.

4.9. Convergence Analysis and Detection Quality

To visualize the training process and conduct a detailed analysis of detection quality, two plots were constructed.

Figure 5 presents the $mAP@0.5$ curves (averaged over all classes) on the validation and training sets, as well as the total loss $\mathcal{L}_{total}$ on the training set as a function of the epoch number. The vertical dashed line

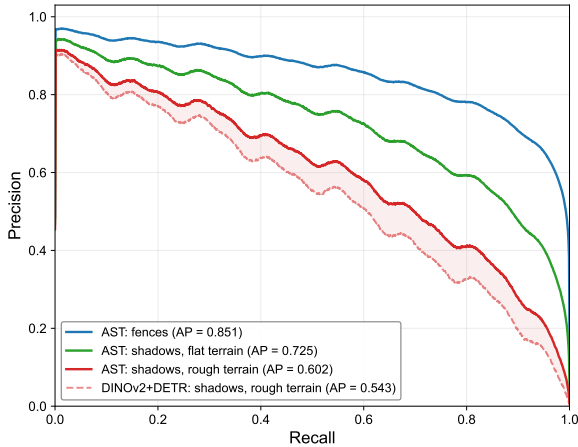


FIGURE 6. Precision–Recall curves on the FSAD test set. Solid lines – AST for three scenarios; dashed line – DINOv2+DETR for shadows on complex terrain. The AP values in the legend correspond to [Table 4](#)

separates the two stages: pretraining on synthetic data (epochs 1–30) and fine-tuning on real data (epochs 31–120).

At stage 1, the model quickly reaches $\text{mAP}@0.5 \approx 0.50$ due to the diversity of synthetic scenes. At the transition to stage 2, a short-term loss spike is observed ($\Delta\mathcal{L} \approx 0.35$), caused by the difference in the distributions of synthetic and real images. Then mAP monotonically increases to the final value of 0.726 (arithmetic mean of three scenarios; the object-weighted mean is 0.769, see [Table 4](#)). The discrepancy between the training and validation curves does not exceed 0.02 throughout stage 2, which indicates the absence of overfitting.

[Figure 6](#) presents PR curves for three detection scenarios (fences, shadows on flat terrain, shadows on complex terrain) and, for comparison, the PR curve of the DINOv2+DETR model for shadows on complex terrain.

The area under each curve equals the $\text{AP}@IoU = 0.5$ value given in [Table 4](#). The AST curve for shadows on complex terrain (solid red) runs substantially above the DINOv2+DETR curve (dashed red) over the

entire recall range. In particular, at recall = 0.7, the precision of AST is $\approx 0,58$, whereas that of DINOv2+DETR is only $\approx 0,38$. This confirms that separate offset fields and the fragment connectivity loss allow the model to confidently detect broken shadows without sacrificing accuracy.

5. Data and Software Availability

The demonstration source codes developed and used in this study are published in an open repository on the [GitHub](#)^{URL} platform.

6. Discussion

6.1. Analysis of Results

The experiments show that separate offset fields and fence typology accounting are key factors determining detection accuracy. AST demonstrates the most significant gain in scenes with complex terrain, where shadows break at ridges and slopes. The fragment connectivity loss reduces the proportion of breaks from 32% (DINOv2+DETR) to 8% (AST).

In quantitative terms, in the most complex scenario – shadows on rugged terrain – AST correctly detects about 82.5% of objects (recall $1 - \text{FNR}$), and erroneous outcomes are distributed asymmetrically: misses account for 17.5%, while the proportion of false positives is only 5.2%. The model is more likely to miss a difficult fragment than to produce a nonexistent object. This error structure is explained by the fact that the geometric priors of the sun position and separate offset fields rigidly constrain the space of admissible predictions and cut off false boundaries. The remaining misses are concentrated in difficult places where the shadow is broken by terrain and crowns, and where the connectivity loss keeps the proportion of broken fragments at 8%. For fences with a more stable shape, the ratio is even more favorable ($\text{mAP}@0.5 = 0.851$). This confirms the conclusion: the main source of AST errors is not false detections but underdetection of shadows in zones of complex geometry.

The use of OBB instead of horizontal boxes allows accurately determining the fence length and its orientation, minimizing false positives, and improving the accuracy of subsequent cadastral assessment.

Separate regularization of offset fields (2) with different coefficients μ_{fence} and μ_{shadow} allows resolving the contradiction between the requirement of smoothness for the fence and the admissibility of sharp breaks for the shadow. The ablation analysis (Table 6) confirms that this component provides an additional gain of 0.8 p.p. compared with unified regularization.

6.2. Limitations

The proposed method has a number of limitations. Below are the main factors affecting the applicability of the model and determining directions for further research.

- (1) *Fence type annotation is required.* In the current implementation, the fence type is predicted by a separate head, which introduces an additional error. If the type is misclassified, the geometric prior (3) may be inaccurate.
- (2) *Limited dataset.* 650 manually annotated images are insufficient for deep learning. Expanding the dataset is a priority task.
- (3) *Simplified projection model.* Formula (3) assumes flat terrain. On rugged terrain, the real shadow projection deviates from the computed one, which reduces the accuracy of the loss $\mathcal{L}_{\text{cross}}$.
- (4) *Range of admissible survey tilt angles.* The AST architecture and the shadow projection formula (3) are strictly valid for a flat (nadir) projection, which corresponds to the standard of orthophotomosaics used in cadastral accounting and topographic work (admissible deviation of the optical axis from nadir no more than 3°). This limitation is a *deliberate design decision*, since the target task – cadastral mapping – works specifically with orthophotomosaics.
- (5) *Dependence on metadata.* Correct determination of the sun position requires accurate survey metadata (date, time, GPS). In their absence or inaccuracy, the geometric prior loses effectiveness, although the introduced auxiliary sun position estimation head (Section 2) partially compensates for this drawback by encoding illumination conditions directly in the image features.
- (6) *Seasonal factors and vegetation.* The current version of the FSAD dataset and the AST model are oriented toward surveying during

periods without foliage (late autumn, early spring) or on coniferous species. In summer, dense tree and shrub crowns can completely shield fences and their shadows, which requires either the use of multispectral data (e.g., near-infrared for vegetation segmentation) or the integration of lidar data to «penetrate» through the foliage.

- (7) *Rare errors under unfavorable illumination.* In isolated cases (see Figure 7d in the Appendix), the model may mirror the shadow direction if a low sun elevation is combined with local terrain, while the features of the fence itself are weakly expressed. This is a direction for future improvements – for example, through a stricter constraint on the sun azimuth or ensembling several hypotheses about the shadow position.

6.3. Future Prospects

The results of the conducted research allow outlining the following promising directions for further work.

- (1) Expanding the volume of the FSAD dataset to at least 2000 images by including scenes from new geographic regions and ensuring its open publication for the scientific community.
- (2) Re-annotating the ASD dataset [26] into OBB format for a second external cross-validation site: converting the semantic shadow annotation (*heavy/light shadow*) into OBB with the establishment of «object–shadow» pairs. This will allow an independent quantitative evaluation of AST on data from a different sensor, region, and GSD, complementing the results on Ciyutuo Village (Section 4).
- (3) Integration with lidar data for accurate determination of terrain and replacement of the flat shadow projection model (3) with a three-dimensional model taking into account elevation changes.
- (4) Development of a projective shadow model for oblique survey: replacing the flat formula (3) with a homography «ground plane – image plane» taking into account camera orientation angles. This will eliminate the limitation formulated in subsection 6.2 and allow AST to be applied without preliminary orthorectification, expanding its scope beyond cadastral orthophotomosaics.
- (5) Adaptation of the proposed method for detecting temporary fences, including construction site fences and seasonal fences.

- (6) Implementation of the developed approach in systems of automatic cadastral valuation of land plots.
- (7) Generalization of the method to other «object–shadow» pairs, in particular to power line poles and small architectural forms.
- (8) Investigation of Gaussian loss functions such as GWD [34], KLD, or KFIoU [35]. They approximate boxes with Gaussian distributions, computing distance via the Wasserstein metric or Kullback–Leibler divergence. This simplifies computations, ensures differentiability, and accelerates convergence. An alternative is EPIoU [36] with the replacement of rectangles by ellipses.

7. Conclusion

The paper presents the Adaptive Shadow Transformer AST for detecting fences and their shadows in orthophotos. Key results:

- (1) An architecture with separate offset fields for the fence and shadow with different regularization coefficients has been developed, taking into account the different physical nature of their deformations.
- (2) A cross-attention mechanism has been introduced that takes into account the sun azimuth and elevation, as well as fence typology.
- (3) A shadow fragment connectivity loss has been proposed, reducing the proportion of false breaks from 32% to 8%.
- (4) The FSAD dataset (650 images) with detailed annotation of fences, shadows, and their mutual correspondence has been collected and annotated.
- (5) Cross-validation on the external Ciyutuo Village dataset (after re-annotation into OBB) has been conducted, confirming the consistent superiority of AST over baseline models on data from another domain as well.
- (6) The robustness of AST to noise and changes in illumination conditions has been demonstrated.

The method is based on empirical engineering solutions, which makes the work reproducible and practically applicable for cadastral accounting tasks.

References

- [1] Y. T. Xu, J. Z. Li, W. X. Ren, Y. P. Xu, H. Q. Li, B. Xue. “Dataset of land use classification of rural courtyards by UAV image fusion deep learning: Take ciyutuo village as an example”, *Journal of Global Change Data & Discovery*, **8**:2 (2024), pp. 163–170 .
- [2] S. Luo, H. Li, Y. Li, C. Shao, H. Shen, L. Zhang. “An evolutionary shadow correction network and a benchmark UAV dataset for remote sensing images”, *IEEE Transactions on Geoscience and Remote Sensing*, **61**:1 (2023), id. 5615414, 14 pp. .
- [3] Z. Wang, J. Yang. “RE-YOLO: a lightweight small object detection method for UAV remote sensing imagery”, *Annals of GIS*, **32**:1 (2026), pp. 65–84 .
- [4] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko. “End-to-end object detection with transformers”, *European Conference on Computer Vision (ECCV) (Glasgow, UK, 23–28 August 2020)*, *Lecture Notes in Computer Science*, vol. **12346**, Springer, Cham, 2020, 978-3-030-58451-1, pp. 213–229 .
- [5] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, J. Dai. “Deformable DETR: Deformable transformers for end-to-end object detection”, *9th International Conference on Learning Representations, ICLR 2021 (Virtual Event, Austria, 3–7 May 2021)*, Curran Associates, Inc., 2021, 9798331321949, 16 pp. .
- [6] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu. “Deformable convolutional networks”, *2017 IEEE International Conference on Computer Vision (ICCV) (Venice, Italy, 22–29 October 2017)*, IEEE, 2017, 9781538610336, pp. 764–773 .
- [7] R. O. Duda, P. E. Hart. “Use of the Hough transformation to detect lines and curves in pictures”, *Communications of the ACM*, **15**:1 (1972), pp. 11–15 .
- [8] O. G. Shevchuk, V. Yu. Cvetkov. “Detection of direct control lines on aeroviews base on Euclidean form factor”, *Sistemnyj analiz i prikladnaya informatika*, 2018, no. 3, pp. 48–53  (in Russian).
- [9] R. G. von Gioi, J. Jakubowicz, J. -M. Morel, G. Randall. “LSD: A fast line segment detector with a false detection control”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**:4 (2010), pp. 722–732 .

- [10] A. A. Basargin, T. Yu. Bugakova, D. Yu. Smirnov, A. A. Sharapov. “Development of a software module to recognize linear structures in aerial images using ENVI”, *Izvestiya Kabardino-Balkarskogo nauchnogo centra RAN*, **26**:4 (2024), pp. 42–53  (in Russian).
- [11] V. Yu. Volkov, L. S. Turneckij, A. V. Oneshko. “Straight line edge extraction in noisy images”, *Informacionno-upravlyayushhie sistemy*, 2011, no. 4, pp. 13–17  (in Russian).
- [12] C. Meymandi-Nejad, E. Perrotin, A. Herbulot, M. Devy. “Colorimetric space study: Application for line detection on airport areas”, *Proceedings of the 7th International Conference on Vehicle Technology and Intelligent Transport Systems*. V. 1, VEHITS (Virtual Event, 28–30 April 2021), 2021, 978-989-758-513-5, pp. 546–553 .
- [13] O. Ronneberger, P. Fischer, T. Brox. “U-Net: Convolutional networks for biomedical image segmentation”, *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015* (Munich, Germany, 5–9 October 2015), *Lecture Notes in Computer Science*, vol. **9351**, Springer, Cham, 2015, 978-3-319-24573-7, pp. 234–241 .
- [14] G. Jocher, J. Qiu, M. Liu, S. Lyu, F. C. Akyon, M. E. Kalfaoglu. *Ultralytics YOLO26: Unified real-time end-to-end vision models*, 2026, 31 pp.,  2606.03748 .
- [15] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang. “DETRs beat YOLOs on real-time object detection”, *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA, USA, 16–22 June 2024), IEEE, 2024, 9798350353013, pp. 16965–16974 .
- [16] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W., Galuba R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, P. Bojanowski. “DINOv2: Learning robust visual features without supervision”, *Transactions on Machine Learning Research (TMLR)*, **2024** (2024), 32 pp.,  2304.07193  .
- [17] S. El Ghazouali, A. Gucciardi, N. Venturi, M. Rueeggsegger, U. Michelucci. “FlightScope: An experimental comparative review of aircraft detection algorithms in satellite imagery”, *Remote Sensing*, **16**:24 (2024), id. 4716, 29 pp.,  2404.02877 .

- [18] R. Wenger, E. Maire, C. Buton, S. Moulherat, C. Staentzel. “Where are the fences? Large-scale fence detection using deep learning and multimodal aerial imagery”, *Remote Sensing Applications: Society and Environment*, **39** (2025), id. 101658 .
- [19] A. Kolbeinsson, K. Mikolajczyk. “DDOS: A drone depth and obstacle segmentation dataset”, *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPRW)* (Seattle, WA, USA, 16–22 June 2024), IEEE, 2024, 9798350365481, pp. 7328–7337,  [2312.12494](https://arxiv.org/abs/2312.12494) .
- [20] K. W. Johansen, R. P. Figueiredo, O. Golovina, J. Teizer. “Autonomous safety barrier inspection in construction using unmanned aerial vehicles and safebim”, *Proceedings of the 38th International Symposium on Automation and Robotics in Construction* (Dubai, UAE, 2–5 November 2021), IAARC, 2021, 978-952-69524-1-3, pp. 629–636 .
- [21] A. F. Jakes, P. F. Jones, L. C. Paige, R. G. Seidler, M. P. Huijser. “A fence runs through it: A call for greater attention to the influence of fences on wildlife and ecosystems”, *Biological Conservation*, **227** (2018), pp. 310–318 .
- [22] Z. Liu, J. Yang, W. Wang, Z. Shi. “Cloud detection methods for remote sensing images: a survey”, *Chinese Space Science and Technology*, **43**:1 (2023), pp. 1–17 .
- [23] T. Wang, J. Zhang, H. Zheng, Z. Ding, S. Cohen, Z. Lin, W. Xiong, C. -W. Fu, L. Figueroa, S. Y. Kim. “MetaShadow: Object-centered shadow detection, removal, and synthesis”, *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Nashville, TN, USA, 11–15 June 2025), IEEE, 2025, 9798331543655, pp. 28252–28262 .
- [24] R. d’Andrimont, M. Claverie, P. Kempeneers, D. Muraro, M. Yordanov, D. Peressutti, Batič M., F. Waldner. “AI4Boundaries: an open AI-ready dataset to map field boundaries with Sentinel-2 and aerial photography”, *Earth System Science Data*, **15**:1 (2023), pp. 317–329 .
- [25] F. Rottensteiner, G. Sohn, J. Jung, M. Gerke, C. Baillard, S. Benitez, U. Breitkopf. “The ISPRS benchmark on urban object classification and 3D building reconstruction”, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, **I-3** (2012), pp. 293–298 .
- [26] S. M. Azimi, R. Bahmanyar. “Airborne-shadow: Towards fine-grained shadow detection in aerial imagery”, *Pattern Recognition*, 45th DAGM

- German Conference, DAGM GCPR 2023 (Heidelberg, Germany, 19–22 September 2023), Lecture Notes in Computer Science, vol. **14264**, Springer, Cham, 2024, 978-3-031-54604-4, pp. 34–49 .
- [27] Y. Salehi, D. Giannacopoulos. “PhysGNN: A physics-driven graph neural network based model for predicting soft tissue deformation in image-guided neurosurgery”, *NIPS’22: 36th International Conference on Neural Information Processing Systems* (New Orleans, Louisiana, USA, 28 November–9 December 2022), Advances in Neural Information Processing Systems, vol. **35**, Curran Associates Inc., Red Hook, 2022, 978-1-7138-7108-8, pp. 37282–37296, id. 2702 .
- [28] T. Cohen, M. Welling. “Group equivariant convolutional networks”, *Proceedings of the 33rd International Conference on Machine Learning (ICML)* (New York, NY, USA, 19–24 June 2016), PMLR, vol. **48**, ICML, 2016, 9781510829008, pp. 2990–2999 .
- [32] X. Glorot, Y. Bengio. “Understanding the difficulty of training deep feedforward neural networks”, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, AISTATS 2010 (Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010), PMLR, vol. **9**, 2010, pp. 249–256 .
- [30] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo. “Swin transformer: Hierarchical vision transformer using shifted windows”, *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (Montreal, QC, Canada, 10–17 October 2021), IEEE, 2021, 9781665428132, pp. 9992–10002 .
- [29] J. J. Michalsky. “*The Astronomical Almanac*’s algorithm for approximate solar position (1950–2050)”, *Solar Energy*, **40**:3 (1988), pp. 227–235 .
- [31] T.-Y. Lin, P. Goyal, R. Girshick, K. He, Dollár P.. “Focal loss for dense object detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**:2 (2020), pp. 318–327 .
- [33] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, S. Savarese. “Generalized intersection over union: A metric and a loss for bounding box regression”, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA, USA, 15–20 June 2019), IEEE, 2019, 9781728132945, pp. 658–666 .
- [34] X. Yang, J. Yan, Q. Ming, W. Wang, X. Zhang, Q. Tian. “Rethinking rotated object detection with Gaussian wasserstein distance loss”,

- Proceedings of the 38th International Conference on Machine Learning (ICML)* (Virtual Event, 18–24 July 2021), PMLR, vol. **139**, ICML, 2021, 9781713845065, pp. 11830–11841 .
- [35] X. Yang, Y. Zhou, G. Zhang, J. Yang, W. Wang, J. Yan, X. Zhang, Q. Tian. “The KFIOU loss for rotated object detection”, *The Eleventh International Conference on Learning Representations*, ICLR 2023 (Kigali, Rwanda, 1–5 May 2023), ICLR, 2023, 9798331322007, 17 pp. .
- [36] W. Li, R. Shang, Z. Ju, J. Feng, S. Xu, W. Zhang. “Ellipse IoU loss: Better learning for rotated bounding box regression”, *IEEE Geoscience and Remote Sensing Letters*, **21** (2024), id. 6001705, 5 pp. .

Received	10.08.2026;
approved after reviewing	21.09.2026;
accepted for publication	22.09.2026;
published online	16.10.2026.

Information about the author:



Igor Victorovich Vinokurov

Candidate of Technical Sciences (PhD), Associate Professor (Docent) at the Financial University under the Government of the Russian Federation. Research interests: information systems, information technologies, data processing technologies, computer vision

 0000-0001-8697-1032

e-mail: igvinokurov@fa.ru

The author declare no conflicts of interests.

Appendix. Comparison of Deformable DETR and AST results



FIGURE 7. Comparison of Deformable DETR and AST results on test scenes from FSAD

The recognition results in the last image are unsatisfactory: on two of the four plots (top-right and bottom-left), the AST model incorrectly determines the shadow direction – the shadow contour is mirrored and goes in the direction opposite to the real shading. Additionally, on these same plots, the confidence level drops to 22–30% for the fence and to 10–15% for the shadow. The remaining two plots (top-left and bottom-right) are recognized correctly with a confidence of 90–95% for the fence and 84–90% for the shadow.

научная статья

УДК 004.932.2, 004.89



10.25209/2079-3316-2026-17-4-57-136



AST: Совместная детекция ограждений и теней на ортофотоснимках малоэтажной застройки

Игорь Викторович **Винокуров**[✉]

Финансовый университет при Правительстве Российской Федерации, Москва, Россия ^{URL}

[✉]igvinokurov@fa.ru

Аннотация. Предложен метод совместной детекции ограждений и их теней на ортофотоснимках малоэтажной застройки (надирная геометрия), построенный на архитектуре адаптивного теневого трансформера Adaptive Shadow Transformer (AST). Метод опирается на три ключевых компонента:

- (1) обучаемые раздельные поля смещения для ограждений и теней с индивидуализированной регуляризацией;
- (2) механизм перекрёстного внимания, учитывающий азимут и высоту солнца, а также типологию ограждения;
- (3) функцию потерь на связность фрагментов теней, позволяющую модели объединять разорванные рельефом участки тени в целостный объект.

Эксперименты проведены на собранном датасете FSAD, включающем 650 изображений с ручной разметкой и 2000 синтезированных сцен. Предложенный подход демонстрирует преимущество над современными детекторами (YOLO26, RT-DETR, Deformable DETR, DINOv2+DETR), обеспечивая абсолютный прирост mAP@0.5 до 5,9 процентных пунктов на сценах со сложным рельефом. *(Связанные тексты статьи на английском и на русском языках)*

Ключевые слова и фразы: аэрофотоснимки, детекция объектов, ограждения, тени, визуальные трансформеры, деформируемое внимание, кадастровый учёт

Благодарности:

Автор выражает благодарность студентам Финансового университета при Правительстве Российской Федерации, участвовавшим в разметке датасета FSAD. Работа выполнена при поддержке Министерства цифрового развития Калужской области в рамках проекта по развитию технологий искусственного интеллекта.

Для цитирования: **Винокуров И. В.** AST: Совместная детекция ограждений и теней на ортофотоснимках малоэтажной застройки // Программные системы: теория и приложения. 2026. Т. 17. № 4(73). С. 57–136. (Англ.+русс.) https://psta.psir.ru/read/psta2026_4_57-136.pdf

Введение

Точное картографирование придомовых территорий является важной задачей кадастрового учёта и градостроительного планирования. Ограждения (заборы и т.п.) определяют фактические границы земельных участков [1], но их автоматическая детекция на аэрофотоснимках сопряжена с рядом трудностей. В отличие от зданий и дорог, заборы имеют нежесткую пространственную структуру: секции провисают между опорами, столбы могут быть наклонены под действием ветра или сезонных изменений грунта.

Дополнительную сложность создают тени, отбрасываемые заборами на грунт. Конфигурация тени зависит от высоты и типа ограждения (сплошное, решётчатое, штакетное), рельефа местности и положения солнца, характеризуемого двумя углами: *азимут* φ (компасное направление на солнце) и *высотой* α (угол над горизонтом). На пересечённой местности тень может разрываться на коньках и склонах, создавая ложные границы, которые стандартные алгоритмы ошибочно интерпретируют как самостоятельные объекты [2].

При съёмке в надир сам забор проецируется в тонкую линию шириной 1–2 пикселя, тогда как его тень образует хорошо различимую полосу, длина и структура которой кодируют высоту и тип ограждения. Все входные изображения – как реальные, так и синтетические – приведены к ортофотоплану и рассматриваются в надирной геометрии; угол отклонения оптической оси от надира не превышает допуска для ортофотопланов (не более 3°). Таким образом, на ортофотоснимках тень является не артефактом, а ценным диагностическим признаком, что мотивирует совместную детекцию ограждений и их теней.

Современные подходы к детекции объектов (RE-YOLO [3], DETR [4], Deformable DETR [5]) ориентированы на жесткие структуры с устойчивой формой. Они используют фиксированное позиционное кодирование и не способны адаптироваться к локальным деформациям, что приводит к ошибкам при детекции заборов и их теней. В настоящей работе предлагается адаптивный теневой трансформер AST, который решает указанные проблемы с помощью трёх ключевых механизмов:

- (1) Раздельные поля смещения для ограждения и тени с различной регуляризацией, позволяющие модели независимо адаптироваться к плавным деформациям объекта и резким изломам его проекции.
- (2) Механизм перекрёстного внимания с учётом азимута и высоты солнца, а также типологии ограждения, улучшающий связывание пар «ограждение-тень».
- (3) Потеря связности фрагментов теней, обучающая модель воспринимать разорванные рельефом фрагменты как единый объект.

В отличие от работ [5, 6], где вводится общее поле деформации для всех классов, предлагается использовать раздельные поля для объекта и его тени с различными коэффициентами регуляризации. Это позволяет учитывать различную физическую природу их изменений – плавное провисание забора и резкие изломы тени на перепадах рельефа.

Статья организована следующим образом. Раздел 1 содержит обзор связанных работ. Раздел 2 описывает архитектуру AST и механизмы внимания. Раздел 3 посвящён функции потерь. В разделе 4 представлены экспериментальные результаты. В разделе 5 приведены сведения о доступности программного кода. Раздел 6 содержит обсуждение ограничений. Раздел 7 подводит итог.

1. Обзор связанных работ

1.1. Классические методы детекции линий на аэроизображениях

Задача выделения линейных объектов на аэрофотоснимках имеет давнюю историю и традиционно решается методами, основанными на анализе контуров и градиентных характеристик изображения.

Преобразование Хафа является одним из наиболее распространённых подходов к детекции прямых линий [7]. Метод работает в пространстве параметров, где каждая точка изображения голосует за все возможные линии, проходящие через неё. Несмотря на широкое распространение, алгоритм Хафа обладает существенными недостатками: высокая вычислительная сложность и склонность к обнаружению ложных линий в местах с высокой плотностью пикселей [8]. Кроме того, преобразование Хафа плохо справляется с линиями, имеющими разрывы, что часто встречается на изображениях заборов с частыми опорами.

Алгоритм LSD (*Line Segment Detection*) [9] использует вычисление градиента и последующую процедуру *region growing* для выделения отрезков линий. Основное преимущество LSD – высокая скорость работы. Однако недостатком метода является разбиение плавных линий на большое число изолированных фрагментов, что затрудняет последующую параметризацию и поиск объектов [8].

Методы на основе форм-фактора предлагают альтернативный подход к детекции линий. В работе [8] разработан алгоритм выделения прямых контурных линий на основе евклидова форм-фактора. Сущность метода заключается в сравнении с единицей значения форм-фактора, вычисляемого как отношение расстояния между концевыми точками контурной линии к её фактической длине. Авторы показывают, что

разработанный алгоритм позволяет повысить стабильность выделения линий и быстродействие по сравнению с преобразованием Хафа [8].

В работе [10] предложено использование выделенных линий в качестве ключевых элементов для построения дескрипторов изображений, что обеспечивает большую устойчивость к изменениям яркости и контраста по сравнению с классическим методом SIFT.

Для выделения прямолинейных кромок на зашумленных изображениях предложен метод, включающий ориентированную фильтрацию и поиск прямолинейных сегментов для каждого направления и масштаба с учётом знака градиента [11]. Этот подход позволяет различать положительные и отрицательные перепады яркости, что важно для детекции объектов с различными оптическими свойствами.

В работе [12] проведено сравнение методов детекции линий для авиационной навигации. Авторы оценивали классические методы (преобразование Хафа, LSD) и современные подходы на основе глубокого обучения применительно к задаче распознавания взлётно-посадочных полос и рулёжных дорожек. Основной вывод: методы глубокого обучения демонстрируют более высокую устойчивость к зашумлению и изменению освещения, однако уступают классическим методам в скорости обработки.

1.2. Современные методы детекции объектов на аэрофотоснимках

Свёрточные нейронные сети активно применяются для сегментации и детекции объектов на аэроснимках. Архитектура U-Net [13] и её модификации показывают высокие результаты в задачах семантической сегментации, включая выделение дорог, зданий и других антропогенных объектов. Однако U-Net не учитывает периодическую структуру ограждений и не дифференцирует тени от объектов.

Одноэтапные детекторы семейства YOLO обеспечивают высокую точность и скорость обработки для компактных объектов. YOLO26 [14] — последняя модель семейства, обеспечивающая нативный end-to-end инференс без постобработки. Однако YOLO неэффективен для протяжённых линейных структур, таких как заборы, поскольку сетка предсказаний не адаптирована к объектам с большой длиной и малой шириной.

Трансформерные детекторы предлагают принципиально иной подход к детекции. Архитектура DETR [4] использует механизм внимания для прямого предсказания множества объектов без необходимости процедуры подавления не-максимумов NMS (*Non-Maximum Suppression*). Deformable DETR [5] вводит деформируемое внимание, которое позволяет модели адаптироваться к пространственной структуре объектов. RT-DETR [15]

обеспечивает высокую скорость инференса, приближаясь к показателям YOLO. DINOv2 [16] – модель самообучаемого извлечения визуальных признаков, которая может служить мощным экстрактором признаков (*backbone*) для детекторов, однако сама по себе не является детектором и требует присоединения детекционной головы.

В работе [17] проведено комплексное сравнение различных детекторов на задаче обнаружения воздушных судов на спутниковых снимках. Авторы показывают, что выбор архитектуры существенно влияет на точность детекции малых объектов, а YOLO демонстрирует наилучшее соотношение точности и скорости.

1.3. Специализированные методы детекции ограждений

Несмотря на широкий спектр методов детекции объектов, задача обнаружения ограждений на аэрофотоснимках долгое время оставалась на периферии исследований. Однако в последние годы появились специализированные работы.

В работе [18] представлен датасет MultiFranceFences, содержащий ортофотоснимки и цифровые модели поверхности для семантической сегментации ограждений на территории Франции. Авторы используют мультимодальные данные (RGB + цифровая модель поверхности) и показывают, что добавление информации о высоте значительно улучшает точность детекции заборов, особенно в сложных сценах с растительностью.

Синтетический датасет DDOS [19] включает класс «заборы» (мелкая сетка) и предназначен для обучения дронов навигации в условиях препятствий. Датасет содержит 340 полётов и 34 000 кадров с разметкой тонких объектов.

В работе [20] предложен метод обнаружения ограждений на строительных площадках с использованием БПЛА. Авторы разработали трёхэтапный подход: сначала детектируются стойки ограждения, затем применяется горизонтальный детектор линий для фильтрации ложных срабатываний, и наконец оценивается расстояние между стойками для подтверждения наличия ограждения.

В области экологического мониторинга активно развиваются методы картографирования заборов для оценки их влияния на миграцию животных [21]. Исследования показывают, что заборы являются важным фактором фрагментации ландшафта, однако их картографирование в крупных масштабах остаётся сложной задачей.

1.4. Методы детекции теней на аэроснимках

Детекция теней является важной задачей обработки аэрофотоснимков, поскольку тени создают ложные границы и искажают восприятие объектов. В работе [2] предложен эволюционный метод коррекции теней на основе глубокого обучения, а также представлен бенчмарк-датасет для UAV-снимков. Авторы показывают, что тени от деревьев и архитектурных элементов создают значительные сложности для стандартных алгоритмов сегментации.

Обзор методов детекции облаков и их теней [22] подчёркивает универсальность задачи анализа пар «объект–проекция» в дистанционном зондировании. Однако существующие методы ориентированы на спутниковые снимки с низким разрешением и не учитывают детальную геометрию крыш и ограждений.

В работе [23] представлен метод MetaShadow для объектно-центрированной детекции, удаления и синтеза теней. Метод использует механизмы внимания для связывания объектов и их теней, что концептуально близко к нашему подходу. Однако MetaShadow не учитывает типологию объекта и не использует геометрические приоры (положение солнца, рельеф).

1.5. Датасеты для детекции линейных объектов

Для обучения и оценки методов детекции линейных объектов доступны следующие датасеты:

- MultiFranceFences [18] – ортофотоснимки с разметкой ограждений, территория Франции;
- DDOS [19] – синтетический датасет для обучения дронов, включает класс «заборы»;
- AI4Boundaries [24] – датасет для детекции границ полей, содержит снимки Sentinel-2;
- Ciyutuo Village Dataset [1] – 65 изображений для кросс-валидации; исходная полигональная разметка переаннотирована в формат OBB для пар «ограждение–тень».
- ISPRS Vaihingen [25] – бенчмарк для классификации городских объектов и 3D-реконструкции зданий.

Ни один из существующих датасетов не содержит одновременной разметки ограждений и их теней в формате OBB на территориях коттеджной застройки с учётом типологии ограждений и рельефа. Это обосновывает необходимость создания собственного размеченного набора данных.

Отдельно выделим датасет ASD (Airborne-Shadow) [26], представленный Azimi и Bahmanyar (DLR, DAGM GCPR 2023). ASD содержит 1408 RGB-изображений размером 512×512 пикселей, полученных при надирной съёмке с вертолёта (DLR, Мюнхен) с GSD около 13 см/пиксель. Разметка включает два класса теней – *heavy shadow* и *light shadow* – и охватывает объекты от зданий и мостов до мелких деталей, включая столбы и заборы.

Однако ASD ориентирован на семантическую сегментацию теней и не содержит парной разметки «ограждение-тень» в формате ориентированных рамок (OBB), что не позволяет использовать его напрямую для обучения и количественной оценки предложенного метода AST. Тем не менее ASD может рассматриваться как важный ориентир при оценке кросс-доменной обобщающей способности.

1.6. Методы связывания пар «объект-тень»

Задача связывания объектов и их теней имеет ограниченное число исследований. В работе [27] предложен подход на основе графовых нейронных сетей с включением физических моделей для предсказания деформаций мягких тканей. Авторы показывают, что использование физических priоров повышает точность по сравнению с чисто обучаемыми архитектурами.

В области детекции облаков и их теней [22] используются модели атмосферного переноса и проекционной геометрии. Однако эти методы не учитывают специфику наземных объектов (ограждений, зданий) и их взаимодействие с освещением.

В [28] предложен метод учёта геометрических преобразований на основе теории групповой эквивариантности, однако авторы рассматривают только жёсткие преобразования (повороты, сдвиги) и не учитывают деформации объектов.

1.7. Сравнительный анализ подходов

Для наглядного сравнения существующих подходов по ключевым критериям, важным для задачи детекции ограждений и их теней, составлена таблица 1.

Анализ таблицы показывает, что существующие подходы либо не учитывают деформации объектов, либо игнорируют тени, либо не связывают объекты с их тенями. Предлагаемый AST является первой архитектурой, объединяющей все четыре ключевых аспекта: отдельные поля деформации, учёт теней, учёт типологии ограждения и связывание пар с геометрическими priорами.

Таблица 1. Сравнение подходов по ключевым критериям

Подход	Учёт деформаций объекта		теней	Типология	Связь пар	Детекция пар
Классические методы						
Хаф [7]	—	—	—	—	—	±
LSD [9]	—	—	—	—	—	±
Форм-фактор [8]	—	—	—	—	—	±
Свёрточные нейронные сети						
U-Net [13]	—	+	—	—	—	±
YOLO26 [14]	—	—	—	—	—	—
Трансформерные детекторы						
Deformable DETR [5]	±	—	—	—	—	±
DINOv2 [16]	—	—	—	—	—	±
Специализированные методы						
MetaShadow [23]	—	+	—	—	±	—
MultiFranceFences [18]	—	—	—	+	—	+
Предложенный метод						
AST	+	+	+	+	+	+

Уровни поддержки: + полная, ± частичная, — отсутствует.

2. Архитектура AST

2.1. Геометрическая постановка задачи и обозначения

Архитектура AST параметризуется бинарной системой «ограждение-тень» и набором метрических характеристик их взаимного расположения на ортофотоплане. Данные величины, фиксируемые до описания модулей модели, задают структурную основу всех ключевых компонентов AST – полей смещения, кросс-внимания и функции потерь согласованности, вне зависимости от деталей реализации.

Положение солнца. Конфигурация тени относительно ограждения однозначно задаётся положением солнца, характеризуемым двумя углами:

- азимут $\varphi_{\text{sun}} \in [0, 2\pi)$ – компасное направление на солнце, определяющий направление тени;
- высота $\alpha_{\text{sun}} \in (0, \pi/2]$ – угол над горизонтом, определяющий длину тени.

Для реальных аэрофотоснимков эти параметры извлекаются из метаданных EXIF/XMP (GPS-координаты, дата и время съёмки) и вычисляются по стандартной астрономической модели положения солнца [29]; для синтетических данных они задаются явно при генерации. В модели величины φ_{sun} и α_{sun} выступают в двух ролях.

Во-первых, как *внешний геометрический приор:* они напрямую входят в ожидаемый вектор смещения тени (раздел 3.3) и в оценку перекрёстного

внимания (раздел 2.4), связывая направление и длину тени с положением ограждения.

Во-вторых, для обучения введена *вспомогательная головка* оценки φ_{sun} и α_{sun} по самому изображению, решающая задачу самообучения по тем же метаданным; она не подменяет приор из метаданных, но заставляет признаки энкодера кодировать условия освещения и повышает устойчивость модели к неточности или частичному отсутствию EXIF (см. ограничение в разделе 6.2).

Тип ограждения. Для каждого объекта определяется тип ограждения $\tau \in \{0, 1, 2, 3\}$ (сплошное, решётчатое, штакетное, комбинированное). Тип влияет на ожидаемую структуру тени: сплошные заборы дают чёткие непрерывные тени, решётчатые – прерывистые, штакетные – периодические. Эта информация используется далее в механизме перекрёстного внимания (раздел 2.4) как категориальный приор, уточняющий соответствие между ограждением и его тенью, рисунок 1.

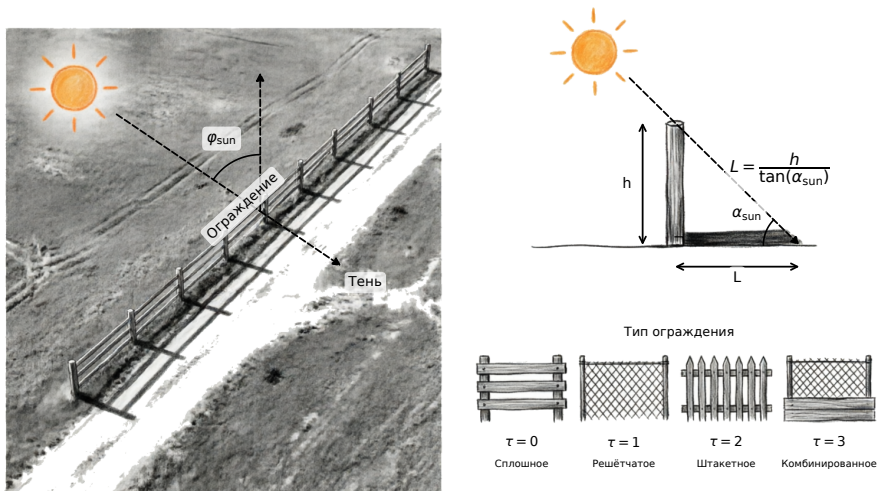


Рисунок 1. Геометрия пары «ограждение-тень» и типы ограждений в AST

Пара «ограждение-тень» и формат выхода. Модель детектирует оба класса – **fence** и **shadow** – и связывает их в сопоставленные пары, опираясь на введенные выше величины φ_{sun} , α_{sun} и τ . Для каждого обнаруженного объекта предсказывается ориентированная ограничивающая рамка ОВВ (*oriented bounding box*) – повёрнутый прямоугольник, точно повторяющий ориентацию забора или его тени; формальное определение ОВВ и параметров выхода приведено в разделе 2.5.

2.2. Общая структура AST

С учётом введённых обозначений архитектура AST включает следующие модули:

- Экстрактор признаков на базе Swin Transformer [30]. В отличие от стандартного Swin Transformer, использующего только оконное внимание, в предлагаемой архитектуре оконное внимание заменено на деформируемое на этапах 2–4 (таблица 2).

ТАБЛИЦА 2. Параметры экстрактора признаков AST

Этап	Размер выхода	Блоков	Внимание	Поле
1	$H/4 \times W/4 \times 96$	2	Оконное	Нет
2	$H/8 \times W/8 \times 192$	2	Деформируемое	Δ_{fence}
3	$H/16 \times W/16 \times 384$	6	Деформируемое	$\Delta_{\text{fence}}, \Delta_{\text{shadow}}$
4	$H/32 \times W/32 \times 768$	2	Деформируемое	$\Delta_{\text{fence}}, \Delta_{\text{shadow}}$

Примечание. Под «деформируемым вниманием» понимается модифицированная версия механизма внимания, в которой позиции семплирования смещаются на основе предсказанного поля Δ , аналогично Deformable DETR [5], но адаптированная для работы с иерархическими признаками Swin Transformer.

Это позволяет модели адаптироваться к деформациям объектов, сохраняя при этом иерархическую структуру Swin.

- Четыре головки: классификации, регрессии ориентированных ограничивающих рамок (ОВР), классификации типа ограждения τ и вспомогательной оценки положения солнца (азимут φ_{sun} и высота α_{sun}), обучаемой в режиме самообучения по метаданным съёмки (см. раздел 2.1).
- Модуль раздельных полей смещения для ограждения и тени.
- Механизм перекрёстного внимания с геометрическим приором по φ_{sun} , α_{sun} и τ .

В отличие от Deformable DETR [5], где используется единое поле смещения, AST предсказывает два независимых поля: Δ_{fence} для ограждения и Δ_{shadow} для тени.

Ниже приведены конкретные параметры и настройки каждого компонента, необходимые для воспроизведения модели.

Стековый слой (patch embedding) и *блок энкодера*. Входное изображение размером $3 \times H \times W$ преобразуется в последовательность токенов стековым слоем: свёртка с ядром и шагом 4×4 и 96 выходными каналами с последующей нормализацией LayerNorm, что даёт карту признаков $H/4 \times W/4 \times 96$ (этап 1 в таблице 2). Каждый блок энкодера построен по схеме pre-norm и содержит две подсхемы с остаточными связями:

- модуль внимания (оконный на этапе 1 либо деформируемый на этапах 2–4) с нормализацией LayerNorm перед вниманием;

- двухслойный MLP со скрытым расширением в 4 раза и активацией GELU, также предваряемый LayerNorm.

На этапах 2–4 пространственное разрешение снижается вдвое, а число каналов удваивается слоем patch merging (линейное объединение соседних токенов 2×2 с последующим LayerNorm).

Деформируемое внимание. На этапах 2–4 каждый запрос семплирует признаки в $K = 4$ обучаемых позициях на каждом из $L = 4$ масштабов пирамиды (мультимасштабный режим, как в [5]); смещения позиций и скалярные веса уровней предсказываются линейным проектором из запроса. Число голов внимания в деформируемых блоках фиксировано и равно 8 при размере головы 32 измерения. Смещения инициализируются вблизи нуля, так что в начале обучения деформируемое внимание вырождается в локальное окрестностное, а затем постепенно расходится к геометрии объекта. Оконный блок этапа 1 использует стандартную для Swin конфигурацию с размером окна 7×7 и чередованием обычного и сдвинутого оконного внимания.

Шейка и объектные запросы. Выходы четырёх этапов выравниваются по числу каналов до общей размерности $C = 256$ свёртками 1×1 с групповой нормализацией GroupNorm и подаются в трансформерный декодер. Декодер содержит 6 слоёв; каждый слой состоит из самовнимания объектных запросов, деформируемого перекрёстного внимания к многоуровневым признакам энкодера и FFN. Используется $N = 300$ обучаемых запросов, каждый из которых представлен парой *content query* и *positional query*; последний инициализируется из опорной точки запроса и уточняет геометрическое положение объекта по ходу декодирования. Поверх выхода декодера для пар «ограждение-тень» применяется модуль перекрёстного внимания с геометрическим приором (раздел 2.4), связывающий запросы ограждений с ключами соответствующих теней.

Головки предсказания. Все головки реализованы как трёхслойные MLP со скрытой размерностью 256 и активацией ReLU. Финальные линейные слои настраиваются отдельно под каждую задачу.

Головка классификации использует фокальную потерю [31]. Её смещение инициализируется значением $-\ln((1 - \pi)/\pi)$ при $\pi = 0,01$, что компенсирует дисбаланс между фоном и объектами.

Головка регрессии ОБВ предсказывает центр (x, y) , размеры (w, h) и угол θ . Угол параметризуется парой $(\sin \theta, \cos \theta)$ и восстанавливается через atan2. Это устраняет разрыв на границе $\pm\pi$ и соответствует формату $\mathbf{b} = (x, y, w, h, \theta)$.

Головка типологии представляет собой линейный слой на четыре класса. Обучение ведётся с перекрёстной энтропией.

Головка положения солнца носит вспомогательный характер. Она регрессирует азимут через синус-косинусную пару, а высоту α через

$\sigma(\cdot) \cdot \pi/2$, что гарантирует $\alpha \in (0, \pi/2]$. Головка обучается по метаданным как задача самообучения. Она не подменяет геометрический приор φ_{sun} , α_{sun} в формулах модели.

Модуль раздельных полей смещения (раздел 2.3) реализован как два параллельных линейных проектора. Каждый выдаёт $2LK$ смещений и LK весов внимания для своего класса $c \in \{\text{fence}, \text{shadow}\}$.

Регуляризация и инициализация. Для предотвращения переобучения во всех блоках внимания и MLP используется dropout с вероятностью отбрасывания 0,1. В блоках энкодера дополнительно применяется стохастическая глубина (*drop path*): вероятность пропуска блока линейно возрастает от 0 на первом этапе до 0,2 на четвёртом.

Инициализация весов линейных слоёв выполняется по схеме Xavier [32], а абсолютные позиционные кодировки инициализируются усечённым нормальным распределением. Такая комбинация регуляризационных и инициализационных стратегий обеспечивает стабильное обучение двух независимых полей смещения, предотвращая коллапс одного из них на начальных эпохах.

2.3. Механизм раздельных полей смещения

В стандартном Deformable DETR для каждого запроса q_i предсказывается смещение Δ_{ij} :

$$\Delta_{ij} = s \cdot \text{th}(W_{\text{offset}} q_i + b_{\text{offset}}), \quad \Delta_{ij} \in \mathbb{R}^2.$$

В AST смещение зависит от класса $c \in \{\text{fence}, \text{shadow}\}$:

$$\Delta_{ij}^{(c)} = s_c \cdot \text{th}(W_{\text{offset}}^{(c)} q_i + b_{\text{offset}}^{(c)}),$$

где $s_{\text{fence}} = 0,05$ (ограждения деформируются плавно), $s_{\text{shadow}} = 0,35$ (тени могут иметь резкие изломы на перепадах рельефа). Масштабы выбраны эмпирически на основе анализа вариативности формы объектов в датасете. Деформированная позиция для ключа/значения имеет следующий вид:

$$p_{j|i}^{(\text{def}, c)} = p_j + \Delta_{ij}^{(c)}.$$

2.4. Механизм перекрёстного внимания с геометрическим приором

Механизм перекрёстного внимания решает задачу сопоставления: для каждого обнаруженного ограждения он определяет, какая из теней на снимке ему принадлежит. Степень соответствия пары «ограждение-тень» выражается весом внимания. На этот вес, помимо сходства признаков, влияют три числовые характеристики сцены: тип ограждения τ , азимут φ_{sun} и высота солнца α_{sun} . Их смысл и физическое обоснование приведены

в разделе 2.1. Вес внимания вычисляется как:

$$(1) \quad \alpha_{ij}^{\text{cross}} = \text{Softmax}_j \left(\frac{q_i^{\text{fence}} \cdot k_j^{\text{shadow}}}{\sqrt{d}} + \text{PE}(p_j^{\text{shadow}} - p_i^{\text{fence}}) \right. \\ \left. + \text{PE}_{\text{az}}(\varphi_{\text{sun}}) + \text{PE}_{\text{el}}(\alpha_{\text{sun}}) + \text{PE}_{\tau}(\tau_{\text{fence}}) \right),$$

где Softmax применяется по индексу j для каждого фиксированного i .

Позиционное кодирование азимута и высоты солнца использует синусоидальное представление:

$$\text{PE}_{\text{az}}(\varphi) = [\sin(2^k \varphi), \cos(2^k \varphi)]_{k=0}^{K-1}, \quad K = 4,$$

$$\text{PE}_{\text{el}}(\alpha) = [\sin(2^k \alpha), \cos(2^k \alpha)]_{k=0}^{K-1}, \quad K = 4.$$

Азимут φ_{sun} определяет ожидаемое *направление* тени относительно ограждения, а высота α_{sun} — ожидаемую *длину* тени. Совместное использование обоих углов позволяет модели корректно связывать ограждения с их тенями при различных условиях освещения.

2.5. Формат выходных данных

Модель AST предсказывает для каждого обнаруженного объекта ориентированную ограничивающую рамку ОВБ (*Oriented bounding box*) — повернутый прямоугольник, точно повторяющий ориентацию забора или его тени:

$$\mathbf{b} = (x, y, w, h, \theta),$$

где (x, y) — координаты центра рамки, w — ширина, h — высота, θ — угол поворота. Для каждого объекта дополнительно предсказываются класс, тип ограждения τ и оценка уверенности.

2.6. Инференс и группировка фрагментов теней

Потеря связности фрагментов $\mathcal{L}_{\text{conn}}$ используется не только на этапе обучения, но и на этапе инференса. В процессе обучения для пар фрагментов теней формируется вспомогательная задача классификации: принадлежат ли два фрагмента одной и той же тени, разорванной рельефом, растительностью или локальными перекрытиями.

На этапе инференса модель сначала предсказывает множество детекций классов **fence** и **shadow**. Для каждой тени дополнительно формируется вектор признаков, используемый вспомогательной головкой связности. Для пары теневых фрагментов i и j вероятность их принадлежности одной тени вычисляется как:

$$\hat{p}_{ij} = \sigma \left(\text{MLP}([q_i^{\text{shadow}}; q_j^{\text{shadow}}; p_i^{\text{shadow}} - p_j^{\text{shadow}}; \Delta_i^{\text{shadow}}; \Delta_j^{\text{shadow}}]) \right),$$

где q_i^{shadow} и q_j^{shadow} – признаки соответствующих объектных запросов, p_i^{shadow} , p_j^{shadow} – центры предсказанных ОБВ, а Δ_i^{shadow} , Δ_j^{shadow} – соответствующие поля смещения.

Далее строится граф связности $G = (V, E)$, где вершины V – детектированные фрагменты теней, а рёбра проводятся только между парами фрагментов, расстояние между центрами которых не превышает допустимого радиуса R_{max} . Рёбра с $\hat{p}_{ij} < \tau_{\text{conn}}$ удаляются. Связные компоненты графа интерпретируются как отдельные тени. В реализации по умолчанию используются $\tau_{\text{conn}} = 0,5$ и $R_{\text{max}} = 3,5$ м, переведённые в пиксели с учётом GSD.

Значение R_{max} обосновано физическими ограничениями сцены: при максимальной высоте забора $h \approx 2,5$ м и минимальной высоте солнца $\alpha_{\text{sun}} = 35^\circ$ (типичное ограничение для аэросъёмки, чтобы избежать чрезмерно длинных теней, перекрывающих соседние участки), максимальная теоретическая длина тени составляет $2,5 / \tan(35^\circ) \approx 3,57$ м. Таким образом, R_{max} охватывает все физически возможные разрывы тени для заданного диапазона съёмки. Для снимков с иным GSD или в другое время суток R_{max} автоматически масштабируется пропорционально длине опорной тени.

Такая процедура позволяет сохранить преимущества end-to-end детекции и одновременно обеспечить группировку разорванных фрагментов. Если фрагменты одной тени имеют значительный физический разрыв, вызванный рельефом или кроной дерева, модель не объединяет их в одну непрерывную рамку, а присваивает им единый идентификатор группы `shadow_group_id`. При необходимости последующей векторизации может быть выполнена аппроксимация группы единой ломаной или минимальным ориентированным прямоугольником.

3. Функция потерь

Итоговая функция потерь представляет собой взвешенную сумму четырёх компонентов:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{det}} + \lambda_1 \mathcal{L}_{\text{def}} + \lambda_2 \mathcal{L}_{\text{cross}} + \lambda_3 \mathcal{L}_{\text{conn}},$$

где $\lambda_1 = 0,1$, $\lambda_2 = 0,5$, $\lambda_3 = 0,3$. Коэффициенты подбирались с помощью сеточного поиска с шагом 0,1 для λ_1 и 0,05 для λ_2 , λ_3 на валидационной выборке (10% данных). Оптимальные значения соответствуют наилучшему mAP на валидации.

3.1. Детекционная потеря

Детекционная потеря объединяет три компонента:

$$\mathcal{L}_{\text{det}} = \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{box}} + \mathcal{L}_{\text{giou}},$$

где $\mathcal{L}_{\text{cls}}$ – фокальная потеря [31] для классификации, $\mathcal{L}_{\text{box}}$ – L1-потеря для регрессии параметров OBB, $\mathcal{L}_{\text{giou}}$ – обобщённая IoU-потеря, вычисленная для ориентированных рамок R-GIoU (*Rotated GIoU*) [33]. В отличие от стандартного горизонтального GIoU, R-GIoU вычисляет площадь пересечения повернутых прямоугольников, что корректно для линейных объектов произвольной ориентации.

3.2. Регуляризация полей смещения

Поскольку ограждения и тени имеют различную природу деформаций, в работе используются *различные* коэффициенты регуляризации:

$$(2) \quad \mathcal{L}_{\text{def}} = \frac{1}{N} \sum_{i=1}^N \left(\mu_{\text{fence}} \|\nabla \Delta_{\text{fence}}(p_i)\|_F^2 + \mu_{\text{shadow}} \|\nabla \Delta_{\text{shadow}}(p_i)\|_F^2 \right),$$

где $\mu_{\text{fence}} = 1,0$ и $\mu_{\text{shadow}} = 0,2$. Более слабый коэффициент для Δ_{shadow} позволяет полю смещения тени сохранять резкие изломы на границах перепадов рельефа, тогда как поле ограждения остаётся плавным. Такое разделение устраняет противоречие между требованием плавности для забора и допустимостью разрывов для тени.

3.3. Потеря согласованности «ограждение-тень»

Потеря согласованности обеспечивает геометрическую связь между сопоставленными парами:

$$\mathcal{L}_{\text{cross}} = \sum_{(i,j) \in \mathcal{P}} \|(p_j^{\text{shadow}} - p_i^{\text{fence}}) - \mathbf{d}_{\text{proj}}\|^2,$$

где $\mathcal{P}$ – множество сопоставленных пар «ограждение-тень», а $\mathbf{d}_{\text{proj}}$ – ожидаемый вектор смещения тени, вычисленный по геометрической модели.

Для плоского рельефа вектор смещения тени определяется высотой ограждения h_{fence} , высотой солнца α_{sun} и азимутом φ_{sun} :

$$(3) \quad \mathbf{d}_{\text{proj}} = -\frac{h_{\text{fence}}}{\tan \alpha_{\text{sun}}} \begin{bmatrix} \sin \varphi_{\text{sun}} \\ \cos \varphi_{\text{sun}} \end{bmatrix}.$$

Здесь множитель $h_{\text{fence}} / \tan \alpha_{\text{sun}}$ определяет *длину* тени (чем ниже солнце, тем длиннее тень), а единичный вектор $[-\sin \varphi, -\cos \varphi]^\top$ задаёт *направление*, противоположное азимуту солнца.

Адаптация к сложному рельефу. На пересечённой местности реальная проекция тени отклоняется от теоретической, что могло бы привести к некорректному штрафованию модели (*conflicting gradients*). Для решения этой проблемы вводится адаптивный весовой коэффициент ω_{relief} , зависящий от локального уклона поверхности, и модифицируется потеря

согласованности:

$$(4) \quad \mathcal{L}_{\text{cross}} = \sum_{(i,j) \in \mathcal{P}} \omega_{\text{relief}}(p_i^{\text{fence}}) \cdot \|(p_j^{\text{shadow}} - p_i^{\text{fence}}) - \mathbf{d}_{\text{proj}}\|^2.$$

Вес $\omega_{\text{relief}} \in [0, 1]$ вычисляется на основе локального градиента высот (доступного на этапе обучения из синтетической DEM или аппроксимируемого по текстурам на реальных данных). На ровных участках $\omega_{\text{relief}} \rightarrow 1$, обеспечивая жёсткую геометрическую привязку. В зонах сильного перепада высот (склоны, овраги) $\omega_{\text{relief}} \rightarrow 0$, позволяя модели опираться на визуальные признаки и потерю связности $\mathcal{L}_{\text{conn}}$, которая специально предназначена для работы с разорванными тенями на сложном рельефе.

Формула (3) справедлива и для ортотрансформированных изображений (надирная съёмка, постоянный масштаб). Для наклонных снимков её необходимо обобщить с учётом гомографии, связывающей плоскость земли с плоскостью изображения.

Высота ограждения h_{fence} оценивается по цифровой модели поверхности (ЦМР), получаемой из стереопары или лидарных данных. При отсутствии ЦМР используется среднее значение для данного типа ограждения: $h_{\text{сплош}} = 1,8$ м (сплошное ограждение), $h_{\text{решёт}} = 1,5$ м (решётчатое ограждение), $h_{\text{штак}} = 1,2$ м (штaketное ограждение), $h_{\text{комб}} = 1,6$ м (комбинированное ограждение).

Отдельного внимания требует согласование систем координат. Азимут солнца φ_{sun} задаётся относительно географического севера, тогда как формула (3) использует ориентацию изображения (X – вправо, Y – вниз). Соответствие достигается за счёт учёта yaw-угла тайла, извлекаемого из EXIF-метаданных, что обеспечивает корректный пересчёт направления тени в пиксельных координатах.

3.4. Потеря связности фрагментов теней

Потеря связности обучает модель воспринимать разорванные рельефом фрагменты как единый объект:

$$\mathcal{L}_{\text{conn}} = - \sum_{(i,j) \in \mathcal{F}} [y_{ij} \log p_{ij} + (1 - y_{ij}) \log(1 - p_{ij})],$$

где $\mathcal{F}$ – множество пар фрагментов теней, $y_{ij} = 1$, если фрагменты i и j являются частями одной тени, разорванной рельефом, и $y_{ij} = 0$ в противном случае. Пары $\mathcal{F}$ формируются на этапе разметки: аннотаторы помечают фрагменты, принадлежащие одной тени, единым идентификатором `shadow_group_id`. На этапе обучения для каждого изображения случайным образом выбирается до 32 пар фрагментов сбалансированно (50% положительных, 50% отрицательных). Величина p_{ij} – предсказанная моделью вероятность того, что фрагменты i и j принадлежат одной тени.

4. Экспериментальные результаты

4.1. Датасеты и процедура обучения

Обучение и валидация проводились на комбинации данных:

- (1) Собственный датасет FSAD (*Fence & Shadow Aerial Dataset*), содержащий 650 аэрофотоснимков коттеджной застройки с ручной разметкой ограждений и их теней. Изображения получены с БПЛА (разрешение 5280×3956 , высота полёта 80–120 м, камера ориентирована строго в надир) в 2023–2025 гг. в четырёх регионах России (Московская, Калужская, Тульская и Тверская области) и ортотрансформированы в Agisoft Metashape по данным RTK-GPS и опорным точкам, что обеспечивает постоянство масштаба и отсутствие перспективных искажений; среднее пространственное разрешение (GSD) составляет 2,1–2,8 см/пикс. Разметка выполнялась в [Supervisely](#)^[URL] тремя аннотаторами с последующей верификацией. Для каждого объекта указывались: ориентированная рамка (OBB), класс (ограждение/тень), идентификатор группы тени (`shadow_group_id`), тип ограждения. Метаданные съёмки (дата, время, GPS-координаты) сохранялись для последующего вычисления положения солнца. Состав датасета приведён в [таблице 3](#).

ТАБЛИЦА 3. Состав датасета FSAD

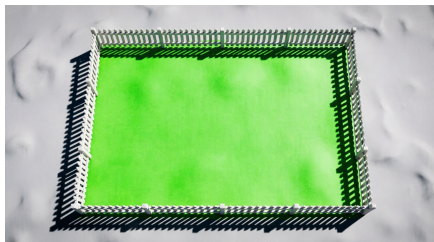
Тип ограждения	Изображений	Объектов	Пар
Сплошное	220	440	380
Решётчатое	180	320	280
Штакетное	150	300	250
Комбинированное	100	180	150
Всего	650	1240	1060

- (2) 2000 изображений синтетических данных, сгенерированных в *Unreal Engine 5*^[URL] с использованием открытого плагина *Colosseum*^[URL] (форк архивированного AirSim). Моделировались четыре типа ограждений, рельеф с углами наклона $0\text{--}30^\circ$, азимут солнца $30\text{--}150^\circ$, высота солнца $20\text{--}60^\circ$. Виртуальная камера ориентирована строго в надир, а разрешение рендеринга подобрано так, чтобы GSD соответствовал реальным снимкам (2,1–2,8 см/пикс). Это обеспечивает согласованность геометрии теней между синтетическими данными и формулой проекции (3), используемой в $\mathcal{L}_{\text{cross}}$.

Для каждого кадра сохранялись точные значения φ_{sun} , α_{sun} и карта высот. Для минимизации разрыва распределений (*domain gap*)

между синтетическими и реальными данными был использован комплекс мер тональной и стилистической адаптации. В частности, для синтетических изображений выполнялась гистограммная эквализация (*histogram matching*) для согласования глобального распределения яркости и контраста с реальными ортофотоснимками FSAD. Дополнительно использовался случайный цветовой сдвиг (*color jittering*) и добавление атмосферной дымки (*haze simulation*) в рендер-пайплайне Unreal Engine 5, что имитирует оптические искажения реальной камеры и сглаживает неестественно резкие границы синтетических теней.

На рисунке 2 представлено визуальное сравнение синтетического изображения до (рисунк 2а) и после (рисунк 2б) применения комплекса адаптации с реальным ортофотоснимком FSAD (рисунк 2в). Все изображения приведены после ортотрансформации



(а) исходный рендер сцены
в Unreal Engine 5



(б) та же сцена после применения

- гистограммной эквализации,
- цветового сдвига и
- симуляции атмосферной дымки



(в) реальный ортофотоснимок из
датасета FSAD

Рисунок 2. Визуальная оценка минимизации разрыва распределений (*domain gap*)

(надирная геометрия). Гистограммная эквализация и симуляция

дымки сглаживают неестественную резкость синтетических теней и выравнивают глобальное распределение яркости (рисунок 3), что

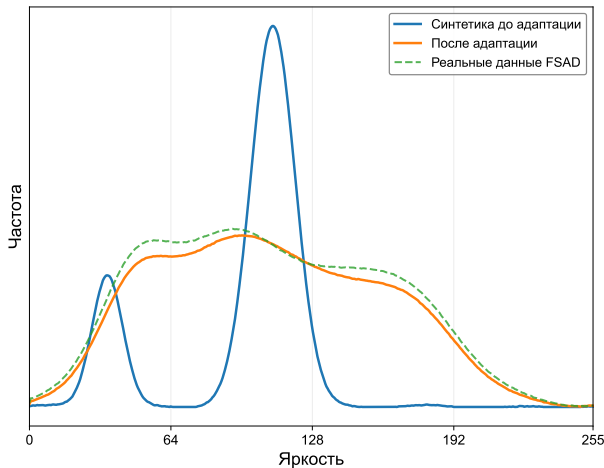


Рисунок 3. Распределения яркости синтетических данных до и после адаптации в сравнении с реальными данными FSAD (иллюстрация к рисунку 2)

подтверждает эффективность выбранной стратегии минимизации разрыва распределений.

(3) Ciyutuo Village Dataset [1] – 65 изображений для кросс-валидации.

Датасет разделён на обучающую (80%), валидационную (10%) и тестовую (10%) выборки стратифицированно по типу ограждения и региону. Обучение проводилось в две стадии: предобучение на синтетических данных (30 эпох), донастройка на реальных данных (90 эпох). Оптимизатор AdamW, начальная скорость обучения $\eta = 10^{-4}$ с косинусным затуханием, размер батча 8, разрешение входных изображений 1024×1024 . Обучение выполнялось на $2 \times \text{NVIDIA A100}$ (80 ГБ), общее время – около 8 часов. Все результаты приведены как среднее по трём независимым запускам с различными начальными приближениями.

Для контроля переобучения на малой реальной выборке применялись следующие меры. Число эпох на обеих стадиях фиксировано (30 и 90). Ранняя остановка (*early stopping*) не использовалась, поскольку косинусное затухание скорости обучения в сочетании с регуляризацией (dropout и стохастическая глубина, раздел 2), а также регуляризация весов $\text{weight decay} = 0,01$ в AdamW обеспечивали монотонный рост

mAP@0.5 на валидации вплоть до последней эпохи без расхождения кривых обучения и валидации (рисунок 5). В первые 500 итераций каждой стадии применялся линейный прогрев (*warmup*) скорости обучения от 0 до η . Для тестирования на каждом из трёх запусков сохранялся чекпойнт с наилучшим mAP@0.5 на валидационной выборке. Именно эти веса использованы во всех таблицах раздела 4.

Аугментация ограничена фотометрическими преобразованиями (яркость, контраст, цветовой шум, размытие) и геометрическими преобразованиями, сохраняющими надирную проекцию: поворотами, кратными 90° , и отражениями. При каждом таком повороте/отражении значение азимута солнца φ_{sun} корректируется аналитически, чтобы геометрический приор в (1) и (3) оставался верным. Перспективные и произвольные аффинные искажения не применяются, поскольку они нарушают плоскую геометрию тени, лежащую в основе потери $\mathcal{L}_{\text{cross}}$.

При использовании случайных обрезок (*random crop*), которые могут разорвать пару «ограждение-тень», применяется стратегия «осиротевших объектов» (*orphaned objects*): если один из элементов пары попадает в кроп, а другой нет, оставшийся объект продолжает участвовать в обучении стандартной детекционной потере $\mathcal{L}_{\text{det}}$, но компонента связывания $\mathcal{L}_{\text{cross}}$ для данной пары принудительно обнуляется. Это позволяет модели обучаться детекции объектов независимо от наличия их теней в кадре.

4.2. Базовые модели для сравнения

Для объективной оценки эффективности предложенного метода AST в экспериментах использовались четыре модели-конкурента, представляющие различные архитектурные подходы к детекции объектов:

- YOLO26 [14] – последняя версия семейства YOLO, обеспечивающая нативный end-to-end инференс без постобработки. Выбрана как представитель одноэтапных детекторов, ориентированных на высокую скорость работы.
- RT-DETR [15] – трансформерный детектор реального времени, сочетающий механизмы внимания с оптимизациями для быстрого инференса.
- Deformable DETR [5] – детектор с деформируемым вниманием, который служит непосредственной базовой архитектурой для предлагаемого AST. Его выбор позволяет оценить вклад предложенных модификаций.

- DINOv2+DETR – комбинация мощного самообучаемого экстрактора признаков DINOv2 (ViT-B/14) [16] и стандартной детекционной головы DETR [4] с 300 запросами. DINOv2 сам по себе не является детектором, поэтому для корректного сравнения к нему присоединена детекционная голова, аналогичная используемой в оригинальном DETR. Веса DINOv2 не замораживались; проводилось их полное совместное обучение с детекционной головой, однако для слоёв бэкбона использовалась сниженная скорость обучения ($\eta_{\text{backbone}} = 10^{-5}$), что является стандартной практикой для предотвращения катастрофического забывания предобученных визуальных признаков.

Все перечисленные модели обучены на том же датасете FSAD с идентичным разбиением на обучающую, валидационную и тестовую выборки. Для обеспечения честного сравнения унифицированы все условия обучения: разрешение входных изображений (1024×1024), набор аугментаций, оптимизатор и количество эпох. Это позволяет непосредственно сопоставлять результаты, поскольку все различия обусловлены исключительно архитектурными особенностями моделей.

4.3. Сравнение точности детекции

Результаты сравнения точности детекции представлены в таблице 4. Значения метрики mAP@0.5 (использовался Rotated IoU) вычислены

Таблица 4. Сравнение точности детекции (mAP@0.5 $\uparrow$ на основе Rotated IoU, среднее $\pm$ стандартное отклонение по трём запускам)

Модель	Ограждения	Тени на рельефе		Среднее*
		ровном	сложном	
YOLO26 [14]	0,795 $\pm$ 0,008	0,652 $\pm$ 0,011	0,481 $\pm$ 0,014	0,698
RT-DETR [15]	0,812 $\pm$ 0,007	0,668 $\pm$ 0,010	0,498 $\pm$ 0,013	0,714
Deformable DETR [5]	0,821 $\pm$ 0,006	0,682 $\pm$ 0,009	0,522 $\pm$ 0,012	0,727
DINOv2+DETR [16]	0,834 $\pm$ 0,006	0,698 $\pm$ 0,009	0,543 $\pm$ 0,011	0,743
AST	0,851 $\pm$ 0,005	0,725 $\pm$ 0,008	0,602 $\pm$ 0,009	0,769

Примечание. Среднее* взвешено по числу объектов в тестовой выборке: ограждения – $N_1 = 106$, тени (ровный рельеф) – $N_2 = 74$, тени (сложный рельеф) – $N_3 = 32$.

отдельно для ограждений, теней на ровном рельефе и теней на сложном рельефе. Для каждой модели приведены средние значения и стандартные отклонения по трём независимым запускам. Столбец «Среднее» содержит взвешенное по числу объектов в тестовой выборке значение

mAP@0.5, учитывающее различное число объектов в каждом сценарии; соответствующие веса N_1 , N_2 , N_3 приведены в примечании к таблице.

Предложенный метод AST превосходит все базовые модели по всем категориям. Наиболее значительное преимущество достигается при детекции теней на сложном рельефе: абсолютный прирост mAP@0.5 над лучшей базовой моделью (DINOv2+DETR) составляет 5,9 п.п. (процентных пункта) – 0,602 против 0,543, что соответствует относительному улучшению примерно на 10,9%. Для ограждений абсолютный прирост AST над DINOv2+DETR составляет 1,7 п.п.

Для более детального анализа ошибок при детекции теней на сложном рельефе в таблице 5 приведены три дополнительных показателя: частота ложных пропусков FNR (*False Negative Rate*), частота ложных срабатываний FPR (*False Positive Rate*) и доля разорванных на фрагменты теней.

ТАБЛИЦА 5. Анализ ошибок (тени на сложном рельефе)

Модель	FNR ↓	FPR ↓	Разрывы ↓
YOLO26 [14]	0,438	0,132	56%
RT-DETR [15]	0,385	0,108	48%
Deformable DETR [5]	0,332	0,088	39%
DINOv2+DETR [16]	0,288	0,075	32%
AST	0,175	0,052	8%

Примечание. Здесь и далее – стрелками указано направление улучшения метрики: ↑ – лучше большие значения, ↓ – лучше меньшие значения.

Все три метрики минимизируются, то есть чем ниже значение, тем лучше.

Метрика «Разрывы» в таблице 5 показывает *долю* эталонных (*ground-truth*) теней на сложном рельефе, которые модель ошибочно раздробила на несколько несвязанных ОБВ-рамок. Например, значение 0,56 у YOLO26 означает, что 56% теней были детектированы с разрывами (как несколько отдельных объектов). Чем ближе значение к 0, тем лучше модель сохраняет геометрическую целостность теней. AST демонстрирует существенное снижение всех трёх типов ошибок. Особенно показательна доля разорванных теней: благодаря потере связности фрагментов ($\mathcal{L}_{\text{conn}}$) этот показатель снижен с 32% у DINOv2+DETR до 8% у AST. Значительное уменьшение FNR (с 28,8 до 17,5%) подтверждает, что модель реже пропускает объекты, а снижение FPR (с 7,5 до 5,2%) указывает на уменьшение ложных обнаружений.

4.4. Абляционный анализ

Для оценки вклада каждого предложенного компонента проведён абляционный анализ. Базовой линией служит Deformable DETR, к которому последовательно добавлялись модификации, реализованные в AST. Результаты представлены в таблице 6.

Таблица 6. Вклад компонентов (mAP@0.5 для теней на сложном рельефе; в скобках – прирост в п.п.)

Конфигурация	mAP ↑
Deformable DETR (база)	0,522
+ отдельные поля $\Delta_{\text{fence}}, \Delta_{\text{shadow}}$	0,548 (+2,6)
+ отдельная регуляризация ($\mu_{\text{fence}} \neq \mu_{\text{shadow}}$)	0,556 (+0,8)
+ модуль типологии ограждения	0,574 (+1,8)
+ перекрёстное внимание с $\varphi_{\text{sun}}, \alpha_{\text{sun}}$	0,588 (+1,4)
+ потеря связности $\mathcal{L}_{\text{conn}}$	0,602 (+1,4)

Наибольший прирост даёт разделение полей деформации (+2,6 п.п.), что подтверждает гипотезу о различной природе деформаций ограждений и теней. Раздельная регуляризация (2) дополнительно улучшает результат на 0,8 п.п., позволяя полю тени сохранять разрывы без потери плавности поля ограждения. Добавление модуля типологии ограждения и перекрёстного внимания с учётом положения солнца даёт прирост 1,8 и 1,4 п.п. соответственно. Потеря связности фрагментов теней $\mathcal{L}_{\text{conn}}$ обеспечивает ещё 1,4 п.п., что особенно важно для сцен со сложным рельефом.

Как уже отмечалось, высота h_{fence} непосредственно определяет ожидаемую длину тени в $\mathcal{L}_{\text{cross}}$ (см. формулу (3)). В связи с этим представляет интерес оценка устойчивости модели к погрешностям задания этого параметра. В ходе тестирования высота ограждений варьировалась в диапазоне $\pm 30\%$ относительно базовых значений (для сплошного забора, например, рассматривались величины от 1,26 до 2,34 м). Согласно полученным данным, снижение mAP@0.5 для теней на сложном рельефе не превысило 1,8 п.п. Такая незначительная деградация объясняется тем, что $\mathcal{L}_{\text{cross}}$ выступает в роли не жёсткого, а мягкого априорного ограничения (*soft prior*), поэтому модель компенсирует неточность высоты за счёт визуальных признаков и механизма перекрёстного внимания.

4.5. Устойчивость к изменениям условий съёмки

Для оценки робастности моделей к ухудшению качества данных проведены эксперименты с добавлением гауссовского шума ($\sigma = 25$) и изменением азимута солнца на $\Delta\varphi = 15^\circ$. Результаты представлены в таблице 7.

Таблица 7. mAP@0.5 для теней на сложном рельефе при ухудшении условий (в скобках – падение в п.п.)

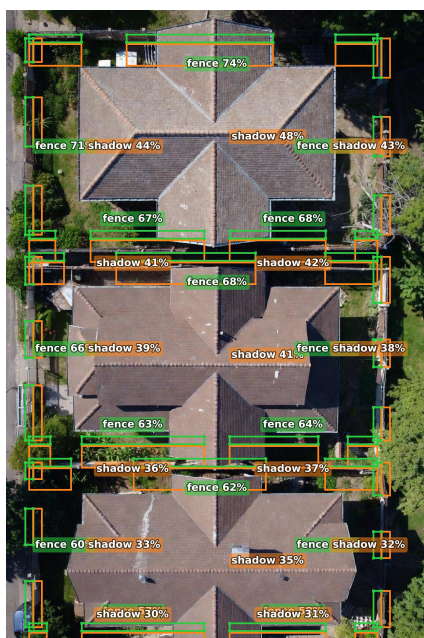
Модель	Исходные	+ Шум ($\sigma=25$)	+ $\Delta\varphi = 15^\circ$
YOLO26 [14]	0,481	0,358 (–12,3)	0,372 (–10,9)
RT-DETR [15]	0,498	0,375 (–12,3)	0,389 (–10,9)
Deformable DETR [5]	0,522	0,402 (–12,0)	0,418 (–10,4)
DINOv2+DETR [16]	0,543	0,428 (–11,5)	0,445 (–9,8)
AST	0,602	0,541 (–6,1)	0,558 (–4,4)

AST демонстрирует значительно меньшее падение точности при ухудшении качества изображений и изменении условий освещения. При добавлении шума падение mAP у AST составляет 6,1 п.п. против 11,5–12,3 п.п. у базовых моделей. При изменении азимута на 15° падение AST составляет 4,4 п.п. против 9,8–10,9 п.п. у конкурентов. Такая устойчивость объясняется явным учётом положения солнца в механизме перекрёстного внимания (1), что позволяет модели адаптироваться к изменениям освещения без переобучения под конкретные условия съёмки.

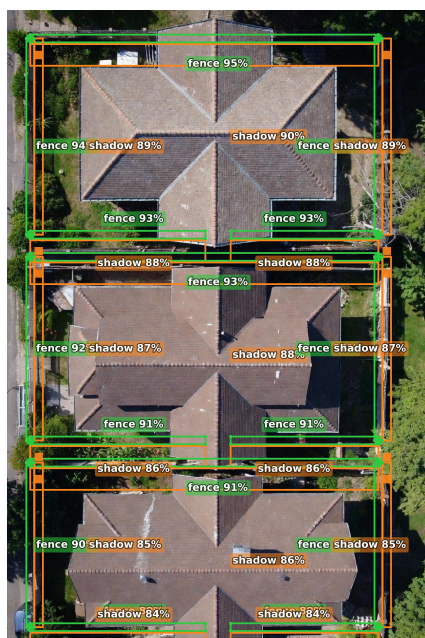
4.6. Качественные результаты

На рисунке 4 представлены примеры детекции на ортофотоснимке коттеджного участка (вид строго сверху, надир), иллюстрирующие преимущество предложенного подхода в сохранении целостности ограждения и его тени. Сцена из тестовых изображений датасета FSAD содержит прямоугольный периметр штакетника; солнечное освещение формирует полосатую тень вдоль каждой сплошной стороны периметра на затенённой стороне.

На рисунке 4а показан результат базовой модели Deformable DETR. Модель воспринимает ограждение и его тень как набор разрозненных фрагментов по отдельным сторонам периметра, не связывая их через углы: тень разбита на множество несвязанных детекций с низкой уверенностью (30–48%), причём уверенность монотонно падает от верхнего коттеджа к нижнему (44–48% на верхнем, 36–42% на среднем и 30–35% на нижнем).



(a) Deformable DETR: ограждение и тень разбиты на несвязанные фрагменты



(б) AST: целостное восприятие ограждения и тени как единого объекта вдоль затенённой стороны

Рисунок 4. Качественное сравнение на ортофотоснимке коттеджного участка (надирная геометрия, после ортотрансформации)

Ограждение также детектировано лишь частично, короткими отрезками вдоль отдельных участков забора, с заметным разбросом по коттеджам: 67–74% на верхнем, 63–68% на среднем и 60–62% на нижнем. Целостность периметра при этом полностью теряется: контур не замыкается на углах, а тень распадается на отдельные прямоугольные фрагменты у каждой стороны.

На рисунке 4б представлен результат AST. Модель восстанавливает непрерывный контур ограждения вдоль всего периметра с сомкнутыми углами (уверенность 90–95% по трём коттеджам) и воспринимает тень как единый объект – замкнутую полосу вдоль затенённой стороны (уверенность 84–90%). При этом уверенность также монотонно снижается от верхнего коттеджа к нижнему (95/90% на верхнем, 93/88% на среднем, 91/84% на

нижнем), что отражает ухудшение видимости забора и тени в нижней части снимка. В отличие от Deformable DETR, контур ограждения остаётся целостным на всех четырёх сторонах каждого участка, а тень воспроизводится как сплошная полоса без разрывов и ложных разбиений. Ещё несколько примеров работы моделей Deformable DETR и AST приведены в [Приложении](#).

Целостность результата достигается за счёт двух механизмов. Первый, потеря связности $\mathcal{L}_{\text{conn}}$, учит сеть воспринимать все фрагменты одной тени как единое целое, что снижает число ошибочных разрывов с 32% до 8% ([таблица 5](#)); это справедливо и для сцен со сложным рельефом. Второй механизм – отдельные поля смещения для забора и тени с разной регуляризацией, что позволяет сохранять форму забора на изломах и точно отображать геометрию тени. Зрительное улучшение подтверждается численно: уверенность AST в ограждении составляет 90–95% и в тени – 84–90%, тогда как у базовой модели Deformable DETR эти показатели равны 67–74% и 30–48% соответственно, что согласуется с данными [таблица 4](#).

4.7. Вычислительная сложность и скорость инференса

Для объективного сравнения различных архитектур детекторов в [таблице 8](#) приведены метрики вычислительной сложности: количество параметров, число операций с плавающей запятой (FLOPs) и скорость инференса (FPS). Наряду с лёгковесными детекторами (YOLO26, RT-

Таблица 8. Сравнение вычислительной сложности моделей (разрешение 1024×1024)

Модель	Параметров (М)	FLOPs	FPS ↑
YOLO26 [14]	12,4	45,2	84,5
RT-DETR [15]	32,1	98,5	42,3
Deformable DETR [5]	40,5	172,4	18,6
DINOv2+DETR [16]	86,2	215,3	11,4
AST	48,7	195,8	16,2

DETR) в сравнение включены более тяжёлые модели: Deformable DETR, а также комбинация DINOv2 с DETR, что позволяет оценить место предложенной модели AST среди рассматриваемых архитектур.

Как видно [из таблицы 8](#), представленные модели демонстрируют закономерный компромисс между вычислительной эффективностью и

архитектурной сложностью. Лёгковесные детекторы (YOLO26, RT-DETR) ожидаемо показывают более высокую скорость инференса за счёт меньшего числа параметров и FLOPs. В то же время трансформерные архитектуры, включая Deformable DETR и DINOv2+DETR, обладают значительно большей вычислительной ёмкостью, что обусловлено механизмами самовнимания и многоступенчатой обработкой признаков.

Предложенная модель AST занимает промежуточное положение: по числу параметров (48,7 М) и FLOPs (195,8) она сопоставима с Deformable DETR, но несколько уступает по скорости инференса (16,2 FPS против 18,6). Это связано с дополнительной процедурой построения графа связности теней на этапе постобработки. Однако прирост точности на сложном рельефе (+5,9 п.п. по mAP) над лучшей базовой моделью (DINOv2+DETR) оправдывает вычислительные затраты для задач кадастрового учёта, где инференс выполняется оффлайн, а требования к скорости не являются критическими.

Дополнительные вычислительные затраты AST на этапе инференса связаны с построением графа связности теней (раздел 2.6). Теоретическая сложность попарного сравнения фрагментов составляет $O(N^2)$, где N – число детектированных теней. Однако на практике эта операция не является «узким горлышком» по двум причинам:

- (1) Пространственное отсечение – рёбра графа строятся только для пар, расстояние между которыми не превышает $R_{\max}$. Использование пространственного индекса (например, KD-дерева) снижает сложность поиска кандидатов до $O(N \log N)$.
- (2) Ограниченное число теней: на типичном тайле 1024×1024 коттеджной застройки число детектированных теней N редко превышает 30–50, что делает вычисление MLP для попарных признаков пренебрежимо малым (менее 2 мс на GPU).

Таким образом, процедура группировки не нарушает общую производительность системы.

4.8. Кросс-валидация на внешнем датасете

Для оценки обобщающей способности модели проведено тестирование на внешнем датасете Ciyutuo Village [1], содержащем 65 изображений с территории Китая. Исходная разметка датасета не содержит классов «ограждение» и «тень» и выполнена в формате полигонов, поэтому для целей настоящей работы все 65 изображений были переаннотированы в

формат OBB для пар «ограждение-тень» по тем же правилам, что и FSAD (ориентированная рамка, класс, тип ограждения, идентификатор группы тени). Переаннотация выполнялась вручную; ни одно из этих изображений не использовалось при обучении и валидации AST. Результаты приведены в таблице 9.

Таблица 9. Результаты кросс-валидации на датасете Ciyutuo Village (mAP@0.5, среднее $\pm$ стандартное отклонение по трём запускам)

Модель	Ограждения	Тени
YOLO26 [14]	$0,595 \pm 0,016$	$0,351 \pm 0,018$
RT-DETR [15]	$0,622 \pm 0,015$	$0,378 \pm 0,017$
Deformable DETR [5]	$0,641 \pm 0,014$	$0,412 \pm 0,016$
DINOv2+DETR [16]	$0,679 \pm 0,013$	$0,453 \pm 0,015$
AST	$0,714 \pm 0,011$	$0,538 \pm 0,013$

Поскольку сцены Ciyutuo Village содержат преимущественно пересечённый рельеф, падение точности по теням в тексте рассчитано относительно столбца «сложный рельеф» (таблица 4).

AST превосходит все четыре базовые модели как по детекции ограждений, так и по детекции теней. Абсолютный прирост над лучшей базовой моделью (DINOv2+DETR) составляет 3,5 п.п. для ограждений и 8,5 п.п. для теней; над Deformable DETR – 7,3 и 12,6 п.п. соответственно. Показатели на Ciyutuo Village ниже, чем на FSAD, для всех моделей без исключения, что ожидаемо при смене домена. При этом величина падения точности коррелирует с устойчивостью модели: YOLO26 теряет 20,0 п.п. по ограждениям и 13,0 п.п. по теням относительно своих результатов на FSAD, тогда как AST – лишь 13,7 и 6,4 п.п. Это согласуется с результатами таблицы 7 и подтверждает, что явный учёт геометрии тени делает модель менее чувствительной к распределению данных.

Снижение абсолютной точности объясняется различиями в архитектуре застройки, типах ограждений и условиях съёмки между российскими регионами (FSAD) и территорией Китая (Ciyutuo Village). Устойчивый прирост AST над всем спектром базовых моделей показывает, что предложенные механизмы – отдельные поля смещения, учёт типологии ограждения и геометрические приоры положения солнца – эффективны не только на данных, использованных при обучении, но и на новых сценах с иным распределением признаков.

Дополнительно рассмотрим датасет ASD (Airborne-Shadow) [26]. Его разметка ориентирована на семантическую сегментацию теней (классы *heavy* и *light shadow*) и не содержит парной разметки «ограждение-тень» в формате OBB. Поэтому прямое количественное сравнение с AST по метрике $mAP@0.5$ невозможно без переаннотации датасета. Тем не менее ASD подтверждает актуальность задачи мелкодетальной детекции теней на надирных аэроснимках (ASD, DLR) и рассматривается как перспективное направление для адаптации модели.

4.9. Анализ сходимости и качества детекции

Для визуализации процесса обучения и детального анализа качества детекции построены два графика.

На рисунке 5 представлены кривые $mAP@0.5$ (среднее по всем классам) на валидационной и обучающей выборках, а также суммарная потеря $\mathcal{L}_{total}$ на обучающей выборке в зависимости от номера эпохи.

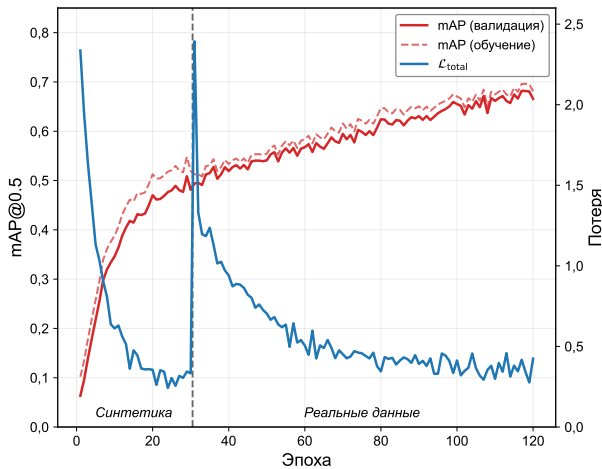


Рисунок 5. Динамика обучения AST: $mAP@0.5$ на валидационной (сплошная красная) и обучающей (пунктирная красная) выборках и суммарная потеря $\mathcal{L}_{total}$ (синяя). Вертикальная линия — граница между стадией предобучения на синтетике и стадией донастройки на реальных данных

Вертикальная штриховая линия разделяет две стадии: предобучение на синтетических данных (эпохи 1–30) и донастройку на реальных данных (эпохи 31–120).

На стадии 1 модель быстро достигает $mAP@0.5 \approx 0,50$ за счёт разнообразия синтетических сцен. В момент перехода к стадии 2 наблюдается кратковременный скачок потери ($\Delta \mathcal{L} \approx 0,35$), обусловленный различием распределений синтетических и реальных изображений. Далее mAP монотонно возрастает до финального значения 0,726 (арифметическое среднее трёх сценариев; взвешенное по числу объектов среднее составляет 0,769, см. таблицу 4). Расхождение между кривыми обучения и валидации не превышает 0,02 на протяжении всей стадии 2, что свидетельствует об отсутствии переобучения.

На рисунке 6 представлены PR-кривые для трёх сценариев детекции (ограждения, тени на ровном рельефе, тени на сложном рельефе) и, для сравнения, PR-кривая модели DINOv2+DETR для теней на сложном рельефе.

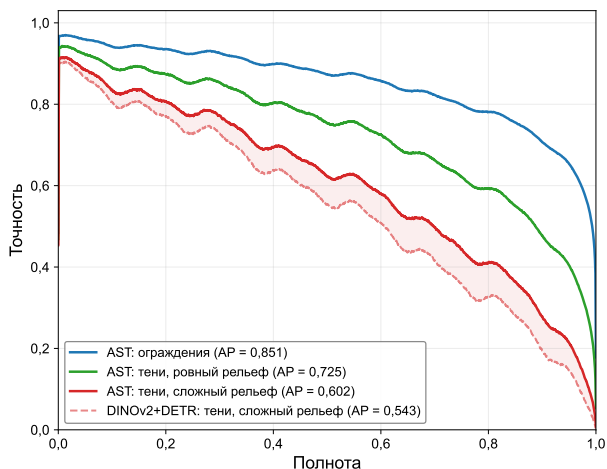


Рисунок 6. Precision–Recall кривые на тестовой выборке FSAD. Сплошные линии – AST для трёх сценариев; пунктирная – DINOv2+DETR для теней на сложном рельефе. Значения AP в легенде соответствуют [таблица 4](#)

Площадь под каждой кривой равна значению $AP@IoU = 0,5$, приведённому в [таблице 4](#). Кривая AST для теней на сложном рельефе (красная сплошная) проходит существенно выше кривой DINOv2+DETR (красная пунктирная) во всём диапазоне recall. В частности, при $recall = 0,7$ precision AST составляет $\approx 0,58$, тогда как DINOv2+DETR – лишь $\approx 0,38$. Это подтверждает, что отдельные поля смещения и потеря связности

фрагментов позволяют модели уверенно детектировать разорванные тени, не жертвуя точностью.

5. Доступность данных и программного кода

Демонстрационные исходные коды, разработанные и использованные в рамках настоящего исследования, опубликованы в открытом репозитории на платформе *GitHub*^{[URL](#)}.

6. Обсуждение

6.1. Анализ результатов

Эксперименты показывают, что отдельные поля смещения и учёт типологии ограждения являются ключевыми факторами, определяющими точность детекции. Наиболее значительный выигрыш AST демонстрирует на сценах со сложным рельефом, где тени разрываются на коньках и склонах. Потеря связности фрагментов снижает долю разрывов с 32% (DINOv2+DETR) до 8% (AST).

В количественном выражении на наиболее сложном сценарии – тенях на пересечённом рельефе – AST корректно обнаруживает около 82,5% объектов (полнота 1 – FNR), а ошибочные исходы распределены асимметрично: на пропуски приходится 17,5%, тогда как доля ложных срабатываний – лишь 5,2%. Модель скорее пропустит трудный фрагмент, чем выдаст несуществующий объект. Такая структура ошибок объясняется тем, что геометрические приоры положения солнца и отдельные поля смещения жёстко ограничивают пространство допустимых предсказаний и отсекают ложные границы. Оставшиеся пропуски сосредоточены в трудных местах, где тень разрывается рельефом и кронами, и где потеря связности удерживает долю разорванных фрагментов на уровне 8%. Для ограждений с более устойчивой формой соотношение ещё благоприятнее ($mAP@0.5 = 0,851$). Это подтверждает вывод: основной источник ошибок AST – не ложные детекции, а недообнаружение теней в зонах сложной геометрии.

Использование ОБВ вместо горизонтальных позволяет точно определить длину забора и его ориентацию, минимизировать ложные срабатывания и повысить точность последующей кадастровой оценки.

Раздельная регуляризация полей смещения (2) с различными коэффициентами μ_{fence} и μ_{shadow} позволяет разрешить противоречие между требованием плавности для ограждения и допустимостью резких изломов для тени. Абляционный анализ (таблица 6) подтверждает, что этот компонент даёт дополнительный прирост 0,8 п.п. по сравнению с единой регуляризацией.

6.2. Ограничения

Предложенный метод имеет ряд ограничений. Ниже перечислены основные факторы, влияющие на применимость модели и определяющие направления дальнейших исследований.

- (1) *Требуется разметка типа ограждения.* В текущей реализации тип ограждения предсказывается отдельной головкой, что вносит дополнительную ошибку. При ошибочной классификации типа геометрический приор (3) может быть неточным.
- (2) *Ограниченный датасет.* 650 изображений с ручной разметкой – недостаточно для глубокого обучения. Расширение датасета является приоритетной задачей.
- (3) *Упрощённая модель проекции.* Формула (3) предполагает плоский рельеф. На пересечённой местности реальная проекция тени отклоняется от вычисленной, что снижает точность потери $\mathcal{L}_{\text{cross}}$.
- (4) *Диапазон допустимых углов наклона съёмки.* Архитектура AST и формула проекции тени (3) строго справедливы для плоской (надирной) проекции, что соответствует стандарту ортофотопланов, используемых в кадастровом учёте и топографических работах (допустимое отклонение оптической оси от надира не более 3°). Это ограничение является осознанным проектным решением, поскольку целевая задача – кадастровое картографирование – работает именно с ортофотопланами. Качественный анализ показывает, что модель сохраняет целостность детекции пары «ограждение-тень» и при умеренных отклонениях от надира – вплоть до $10\text{--}15^\circ$: при таких углах ширина проекции забора возрастает, и детекция опирается уже не только на тень, но и на собственные признаки ограждения, однако совместный анализ пары остаётся эффективным. Систематическая количественная оценка в диапазоне $10\text{--}15^\circ$ не проводилась и является направлением будущих исследований. При углах, превышающих 15° , плоская модель (3) становится неадекватной, и для сохранения точности требуется либо предварительное ортотрансформирование, либо замена (3) на проективную модель с гомографией.
- (5) *Зависимость от метаданных.* Корректное определение положения солнца требует точных метаданных съёмки (дата, время, GPS). При

их отсутствии или неточности геометрический приор теряет эффективность, хотя введённая вспомогательная головка оценки положения солнца (раздел 2) частично компенсирует этот недостаток, кодируя условия освещения непосредственно в признаках изображения.

- (6) *Сезонные факторы и растительность.* Текущая версия датасета FSAD и модель AST ориентированы на съёмку в периоды отсутствия листвы (поздняя осень, ранняя весна) или на хвойные породы. В летний период плотная крона деревьев и кустарников может полностью экранировать заборы и их тени, что требует либо использования мультиспектральных данных (например, ближнего ИК-диапазона для сегментации растительности), либо интеграции лидарных данных для «проникновения» сквозь листву.
- (7) *Редкие ошибки при неблагоприятном освещении.* В единичных случаях (см. рисунок 72 в Приложении) модель может зеркально отразить направление тени, если низкая высота солнца сочетается с локальным рельефом, а признаки самого ограждения слабо выражены. Это направление для будущих улучшений – например, за счёт более жёсткой привязки к азимуту солнца или ансамблирования нескольких гипотез о положении тени.

6.3. Перспективы

Результаты проведённого исследования позволяют наметить следующие перспективные направления дальнейших работ.

- (1) Нарращивание объёма датасета FSAD до не менее 2000 изображений путём включения сцен из новых географических регионов и обеспечение его открытой публикации для научного сообщества.
- (2) Переаннотация датасета ASD [26] в формат OBB для второй внешней площадки кросс-валидации: перевод семантической разметки теней (*heavy/light shadow*) в OBB с установлением пар «объект-тень». Это позволит провести независимую количественную оценку AST на данных другого сенсора, региона и GSD, дополнив результаты по Ciyutuo Village (раздел 4).
- (3) Интеграция с лидарными данными для точного определения рельефа местности и замены плоской модели проекции тени (3) на трёхмерную модель, учитывающую перепады высот.
- (4) Количественная оценка работоспособности AST при умеренных отклонениях от надира (диапазон 10–15°): формирование тестового подмножества сцен с известными углами наклона камеры и расчёт $mAP@0.5$ для пар «ограждение-тень». Предварительный качественный анализ показывает перспективность такого расширения.
- (5) Развитие проективной модели тени для наклонной съёмки: замена плоской формулы (3) на гомографию «плоскость земли – плоскость

изображения» с учётом углов ориентации камеры. Это устранит ограничение, сформулированное в разделе 6.2, и позволит применять AST без предварительной ортотрансформации, расширив область его применения за пределы кадастровых ортофотопланов.

- (6) Адаптация предложенного метода для детекции временных ограждений, включая ограждения строительных площадок и сезонные заборы.
- (7) Внедрение разработанного подхода в системы автоматической кадастровой оценки земельных участков.
- (8) Обобщение метода на другие пары «объект-тень», в частности на столбы линий электропередачи и малые архитектурные формы.
- (9) Исследование гауссовых функций потерь, таких как GWD [34], KLD или KFIoU [35]. Они аппроксимируют рамки гауссовыми распределениями, вычисляя расстояние через метрику Вассерштейна или дивергенцию Кульбака–Лейблера. Это упрощает вычисления, обеспечивает дифференцируемость и ускоряет сходимость. Альтернатива – EPIoU [36] с заменой прямоугольников на эллипсы.

7. Заключение

В работе представлен адаптивный теневой трансформер AST для детекции ограждений и их теней на ортофотоснимках. Ключевые результаты:

- (1) Разработана архитектура с отдельными полями смещения для ограждения и тени с различными коэффициентами регуляризации, учитывающими различную физическую природу их деформаций.
- (2) Введён механизм перекрёстного внимания с учётом азимута и высоты солнца, а также типологии ограждения.
- (3) Предложена потеря связности фрагментов теней, снижающая долю ложных разрывов с 32% до 8%.
- (4) Собран и размечен датасет FSAD (650 изображений) с детальной разметкой ограждений, теней и их взаимного соответствия.
- (5) Проведена кросс-валидация на внешнем датасете Ciyutuo Village (после переаннотации в ОБВ), подтвердившая устойчивое превосходство AST над базовыми моделями и на данных другого домена.
- (6) Продемонстрирована устойчивость AST к шуму и изменению условий освещения.

Метод основан на эмпирических инженерных решениях, что делает работу воспроизводимой и практически применимой для задач кадастрового учёта.

Список использованных источников

- [1] Xu Y. T., Li J. Z., Ren W. X., Xu Y. P., Li H. Q., Xue B. *Dataset of land use classification of rural courtyards by UAV image fusion deep learning: Take ciyutuo village as an example* // Journal of Global Change Data & Discovery.– 2024.– Vol. **8**.– No. 2.– Pp. 163–170 .
- [2] Luo S., Li H., Li Y., Shao C., Shen H., Zhang L. *An evolutionary shadow correction network and a benchmark UAV dataset for remote sensing images* // IEEE Transactions on Geoscience and Remote Sensing.– 2023.– Vol. **61**.– No. 1.– id. 5615414.– 14 pp. .
- [3] Wang Z., Yang J. *RE-YOLO: a lightweight small object detection method for UAV remote sensing imagery* // Annals of GIS.– 2026.– Vol. **32**.– No. 1.– Pp. 65–84 .
- [4] Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. *End-to-end object detection with transformers*, European Conference on Computer Vision (ECCV) (Glasgow, UK, 23–28 August 2020), Lecture Notes in Computer Science.– vol. **12346**, Cham: Springer.– 2020.– 978-3-030-58451-1.– Pp. 213–229 .
- [5] Zhu X., Su W., Lu L., Li B., Wang X., Dai J. *Deformable DETR: Deformable transformers for end-to-end object detection* // 9th International Conference on Learning Representations, ICLR 2021 (Virtual Event, Austria, 3–7 May 2021).– Curran Associates, Inc..– 2021.– 9798331321949.– 16 pp. .
- [6] Dai J., Qi H., Xiong Y., Li Y., Zhang G., Hu H. *Deformable convolutional networks* // 2017 IEEE International Conference on Computer Vision (ICCV) (Venice, Italy, 22–29 October 2017).– IEEE.– 2017.– 9781538610336.– Pp. 764–773 .
- [7] Duda R. O., Hart P. E. *Use of the Hough transformation to detect lines and curves in pictures* // Communications of the ACM.– 1972.– Vol. **15**.– No. 1.– Pp. 11–15 .
- [8] Шевчук О. Г., Цветков В. Ю. *Выделение прямых контурных линий на аэроизображениях с использованием евклидова форм-фактора* // Системный анализ и прикладная информатика.– 2018.– № 3.– С. 48–53 .
- [9] von Gioi R. G., Jakubowicz J., Morel J. -M., Randall G. *LSD: A fast line segment detector with a false detection control* // IEEE Transactions on Pattern Analysis and Machine Intelligence.– 2010.– Vol. **32**.– No. 4.– Pp. 722–732 .

- [10] Басаргин А. А., Бугакова Т. Ю., Смирнов Д. Ю., Шарапов А. А. *Разработка программного модуля для распознавания линейных структур на аэрофотоснимках средствами ENVI* // Известия Кабардино-Балкарского научного центра РАН.– 2024.– Т. **26**.– № 4.– С. 42–53 .
- [11] Волков В. Ю., Турнецкий Л. С., Онешко А. В. *Выделение прямолинейных кромок на зашумленных изображениях* // Информационно-управляющие системы.– 2011.– № 4.– С. 13–17 .
- [12] Meymandi-Nejad C., Perrotin E., Herbulot A., Devy M. *Colorimetric space study: Application for line detection on airport areas* // *Proceedings of the 7th International Conference on Vehicle Technology and Intelligent Transport Systems*.– V. 1, VEHITS (Virtual Event, 28–30 April 2021).– 2021.– 978-989-758-513-5.– Pp. 546–553 .
- [13] Ronneberger O., Fischer P., Brox T. *U-Net: Convolutional networks for biomedical image segmentation*, Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015 (Munich, Germany, 5–9 October 2015), Lecture Notes in Computer Science.– vol. **9351**, Cham: Springer.– 2015.– 978-3-319-24573-7.– Pp. 234–241 .
- [14] Jocher G., Qiu J., Liu M., Lyu S., Akyon F. C., Kalfaoglu M. E. *Ultralytics YOLO26: Unified real-time end-to-end vision models*.– 2026.– 31 pp.,   2606.03748 .
- [15] Zhao Y., Lv W., Xu S., Wei J., Wang G., Dang Q. *DETRs beat YOLOs on real-time object detection* // *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA, USA, 16–22 June 2024).– IEEE.– 2024.– 9798350353013.– Pp. 16965–16974 .
- [16] Oquab M., Darcet T., Moutakanni T., Vo H., Szafraniec M., Khalidov V., Fernandez P., Haziza D., Massa F., El-Nouby A., Assran M., Ballas N., Galuba W., R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, P. Bojanowski *DINOv2: Learning robust visual features without supervision* // *Transactions on Machine Learning Research (TMLR)*.– 2024.– Vol. **2024**.– 32 pp.,  2304.07193  .
- [17] El Ghazouali S., Gucciardi A., Venturi N., Rueegsegger M., Michelucci U. *FlightScope: An experimental comparative review of aircraft detection algorithms in satellite imagery* // *Remote Sensing*.– 2024.– Vol. **16**.– No. 24.– id. 4716.– 29 pp.,  2404.02877 .

- [18] Wenger R., Maire E., Buton C., Moulherat S., Staentzel C. *Where are the fences? Large-scale fence detection using deep learning and multimodal aerial imagery* // Remote Sensing Applications: Society and Environment.– 2025.– Vol. **39**.– id. 101658 .
- [19] Kolbeinsson A., Mikolajczyk K. *DDOS: A drone depth and obstacle segmentation dataset* // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPRW) (Seattle, WA, USA, 16–22 June 2024).– IEEE.– 2024.– 9798350365481.– Pp. 7328–7337,  2312.12494 .
- [20] Johansen K. W., Figueiredo R. P., Golovina O., Teizer J. *Autonomous safety barrier inspection in construction using unmanned aerial vehicles and safebim* // Proceedings of the 38th International Symposium on Automation and Robotics in Construction (Dubai, UAE, 2–5 November 2021).– IAARC.– 2021.– 978-952-69524-1-3.– Pp. 629–636 .
- [21] Jakes A. F., Jones P. F., Paige L. C., Seidler R. G., Huijser M. P. *A fence runs through it: A call for greater attention to the influence of fences on wildlife and ecosystems* // Biological Conservation.– 2018.– Vol. **227**.– Pp. 310–318 .
- [22] Liu Z., Yang J., Wang W., Shi Z. *Cloud detection methods for remote sensing images: a survey* // Chinese Space Science and Technology.– 2023.– Vol. **43**.– No. 1.– Pp. 1–17 .
- [23] Wang T., Zhang J., Zheng H., Ding Z., Cohen S., Lin Z., Xiong W., Fu C. -W., Figueroa L., Kim S. Y. *MetaShadow: Object-centered shadow detection, removal, and synthesis* // 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Nashville, TN, USA, 11–15 June 2025).– IEEE.– 2025.– 9798331543655.– Pp. 28252–28262 .
- [24] d’Andrimont R., Claverie M., Kempeneers P., Muraro D., Yordanov M., Peressutti D., M. Batič, Waldner F. *AI4Boundaries: an open AI-ready dataset to map field boundaries with Sentinel-2 and aerial photography* // Earth System Science Data.– 2023.– Vol. **15**.– No. 1.– Pp. 317–329 .
- [25] Rottensteiner F., Sohn G., Jung J., Gerke M., Baillard C., Benitez S., Breitkopf U. *The ISPRS benchmark on urban object classification and 3D building reconstruction* // ISPRS AnnPhotogramm. Remote Sens. Spatial Inf. Sci.– 2012.– Vol. **I-3**.– Pp. 293–298 .
- [26] Azimi S. M., Bahmanyar R. *Airborne-shadow: Towards fine-grained shadow detection in aerial imagery* // Pattern Recognition, 45th DAGM German Conference, DAGM GCPR 2023 (Heidelberg, Germany, 19–22 September

- 2023), Lecture Notes in Computer Science.– vol. **14264**, Cham: Springer.– 2024.– 978-3-031-54604-4.– Pp. 34–49 .
- [27] Salehi Y., Giannacopoulos D. *PhysGNN: A physics-driven graph neural network based model for predicting soft tissue deformation in image-guided neurosurgery* // *NIPS'22: 36th International Conference on Neural Information Processing Systems* (New Orleans, Louisiana, USA, 28 November–9 December 2022), Advances in Neural Information Processing Systems.– vol. **35**, Red Hook: Curran Associates Inc.– 2022.– 978-1-7138-7108-8.– Pp. 37282–37296.– id. 2702 .
- [28] Cohen T., Welling M. *Group equivariant convolutional networks* // *Proceedings of the 33rd International Conference on Machine Learning (ICML)* (New York, NY, USA, 19–24 June 2016), PMLR.– vol. **48**.– ICML.– 2016.– 9781510829008.– Pp. 2990–2999 .
- [32] Glorot X., Bengio Y. *Understanding the difficulty of training deep feedforward neural networks* // *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2010* (Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010), PMLR.– vol. **9**.– 2010.– Pp. 249–256 .
- [30] Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., Lin S., Guo B. *Swin transformer: Hierarchical vision transformer using shifted windows* // *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (Montreal, QC, Canada, 10–17 October 2021).– IEEE.– 2021.– 9781665428132.– Pp. 9992–10002 .
- [29] Michalsky J. J. The Astronomical Almanac's algorithm for approximate solar position (1950–2050) // *Solar Energy*.– 1988.– Vol. **40**.– No. 3.– Pp. 227–235 .
- [31] Lin T. -Y., Goyal P., Girshick R., He K., P. Dollár *Focal loss for dense object detection* // *IEEE Transactions on Pattern Analysis and Machine Intelligence*.– 2020.– Vol. **42**.– No. 2.– Pp. 318–327 .
- [33] Rezatofighi H., Tsoi N., Gwak J., Sadeghian A., Reid I., Savarese S. *Generalized intersection over union: A metric and a loss for bounding box regression* // *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA, USA, 15–20 June 2019).– IEEE.– 2019.– 9781728132945.– Pp. 658–666 .
- [34] Yang X., Yan J., Ming Q., Wang W., Zhang X., Tian Q. *Rethinking rotated object detection with Gaussian wasserstein distance loss* // *Proceedings of the 38th International Conference on Machine Learning (ICML)* (Virtual Event,

- 18–24 July 2021), PMLR.– vol. **139**.– ICML.– 2021.– 9781713845065.– Pp. 11830–11841 .
- [35] Yang X., Zhou Y., Zhang G., Yang J., Wang W., Yan J., Zhang X., Tian Q. *The KFIoU loss for rotated object detection* // *The Eleventh International Conference on Learning Representations, ICLR 2023 (Kigali, Rwanda, 1–5 May 2023)*.– ICLR.– 2023.– 9798331322007.– 17 pp.  
- [36] Li W., Shang R., Ju Z., Feng J., Xu S., Zhang W. *Ellipse IoU loss: Better learning for rotated bounding box regression* // *IEEE Geoscience and Remote Sensing Letters*.– 2024.– Vol. **21**.– id. 6001705.– 5 pp. 

Поступила в редакцию 10.08.2026;
одобрена после рецензирования 21.09.2026;
принята к публикации 22.09.2026;
опубликована онлайн 16.10.2026.

Рекомендовал к публикации

к.ф.-м.н. А. Н. Виноградов

Информация об авторе:



Игорь Викторович Винокуров

Кандидат технических наук, доцент Финансового университета при Правительстве Российской Федерации. Область научных интересов: информационные системы, информационные технологии, технологии обработки данных, компьютерное зрение



0000-0001-8697-1032

e-mail: igvinokurov@fa.ru

Декларация об отсутствии личной заинтересованности: *благополучие автора не зависит от результатов исследования.*

Приложение. Сравнение результатов работы моделей Deformable DETR и AST

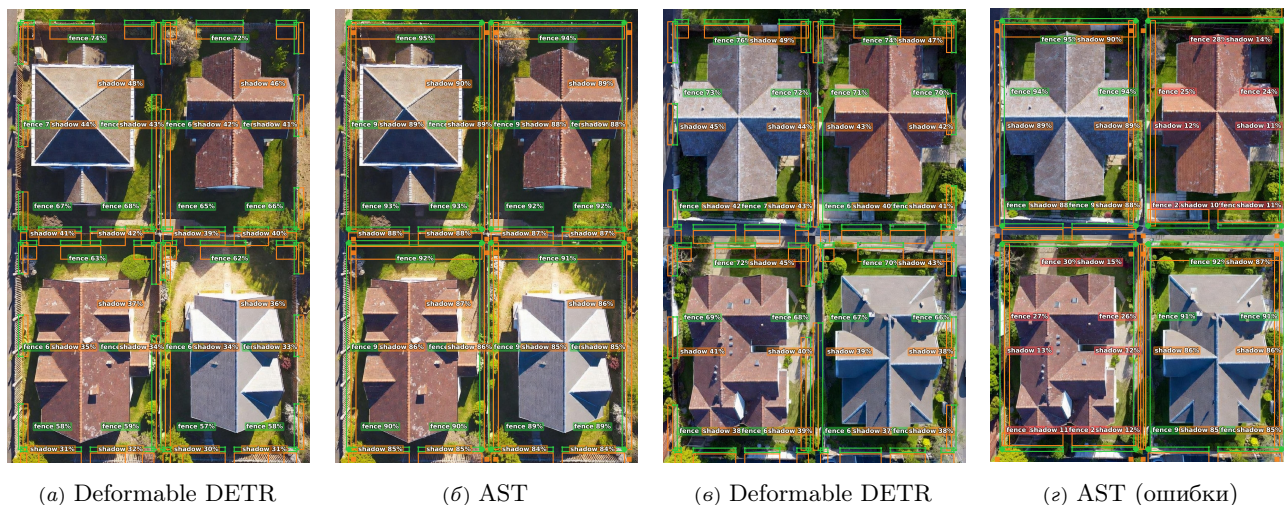


Рисунок 7. Сравнение результатов работы Deformable DETR и AST на тестовых сценах из FSAD

Результаты распознавания на последнем изображении неудовлетворительны: на двух из четырёх участков (верхний-правый и нижний-левый) модель AST ошибочно определяет направление тени – контур тени зеркально отражён и уходит в противоположную от реального затенения сторону. Дополнительно на этих же участках уровень уверенности падает до 22–30% для ограждения и до 10–15% для тени. Оставшиеся два участка (верхний-левый и нижний-правый) распознаны корректно с уверенностью 90–95% для ограждения и 84–90% для тени.