

INVESTIGATION OF THE COGNITIVE SATURATION THRESHOLD OF THE YOLOV8 MODEL UNDER LIMITED TRAINING DATA

DOI: 10.36724/2072-8735-2026-20-9-68-75

Manuscript received 20 June 2026;
Accepted 27 August 2026

Maria Yu. Karelina,
State University of Management;
Moscow Automobile and Road Construction State Technical University
(MADI), Moscow, Russia, karelinamu@mail.ru

Alexey A. Akulov,
State University of Management;
Moscow Automobile and Road Construction State Technical University
(MADI), Moscow, Russia, a.akulov.98@mail.ru

Dmitriy S. Taldykin,
Moscow Automobile and Road Construction State Technical University (MADI);
State University of Management, Moscow, Russia,
dima.dima.taldykin@mail.ru

This article was prepared as part of the implementation of research work, implemented at the expense of the federal budget (source of financing - the Ministry of Education and Science of the Russian Federation) on the topic: "Development of software for the agro-industrial complex for monitoring the state of the tops and fruits of greenhouse plants using machine vision and artificial intelligence" (code of the scientific theme FSFM-2025-0002)

Keywords: YOLOv8; tomatoes; weed detection; small datasets; model hallucinations

This study examines the training-set size at which YOLOv8m ceases to benefit from adding visually similar frames in weed detection among tomato plants. The experimental design reflects a common precision-agriculture constraint: RGB images are relatively easy to acquire, whereas agronomist-verified annotation requires substantially more time. The experiment used 600 images with a resolution of 640x640 pixels collected in greenhouse tomato rows during a single growing season. Three classes were annotated: tomato, weed vegetation, and soil-substrate background. Seven training series were used to test the hypothesis that generalization peaks near $N=250$ when the number of unique scenes is below 300. Beyond this point, additional frames acquired under nearly unchanged conditions were accompanied by an increase in high-confidence false detections. The effect was characterized by a hallucination coefficient and an oversaturation index comparing the variance of the weights in the final detection layer with the mean validation loss. At $N=600$, Precision was 41.94% below its peak value and the hallucination coefficient reached 1.80. The practical implications concern annotation planning and detector-capacity selection: increasing the number of near-duplicate frames should not be treated as an unconditional means of improving the reliability of an automated system.

Information about authors:

Maria Yu. Karelina, Doctor of Technical Sciences, Doctor of Pedagogical Sciences, Professor, Head of the Department of Logistics and Transport and Technological Systems, State University of Management, Head of the Department of Machine Parts and Theory of Mechanisms at the Moscow Automobile and Road Construction State Technical University (MADI), Moscow, Russia. ORCID ID: 0000-0003-0335-7550

Alexey A. Akulov, Candidate of Technical Sciences, Associate Professor of the Department of Logistics and Transport and Technological Systems of the State University of Management, Associate Professor of the Department of Machine Parts and Theory of Mechanisms of the Moscow Automobile and Road Construction State Technical University (MADI), Moscow, Russia. ORCID ID: 0000-0001-8952-0237

Dmitriy S. Taldykin, Moscow Automobile and Road Construction State Technical University (MADI), State University of Management, Moscow, Russia. ORCID ID: 0009-0008-5071-426X

Для цитирования:

Карелина М.Ю., Акулов А.А., Талдыкин Д.С. Исследование порога когнитивного насыщения модели YOLOv8 в условиях ограниченной обучающей выборки // Т-Комм: Телекоммуникации и транспорт. 2026. Том 20. №9. С. 68-75.

For citation:

M.Yu. Karelina, A.A. Akulov, D.S. Taldykin, "Investigation of the cognitive saturation threshold of the YOLOv8 model under limited training data," T-Comm, 2026, vol. 20, no. 9, pp. 68-75.

Introduction

A detector output acquires technological significance before any mechanical action on the plants begins. The camera evaluates a section of a tomato row, the recognition module separates weed vegetation from the inter-row space and the soil-substrate surroundings, and the result is then transferred to the actuator system. At this point, an error ceases to be merely a classification error. An incorrectly assigned class may alter the command sent to the working tool and, consequently, the course of a specific agromonic operation.

The visual structure of a greenhouse row offers no single separating feature. Soil, substrate, trellis components, drip tape, plant residues, shadows, and structural elements can occur in the same frame. Against this background, tomato leaves have a complex dissected contour: their orientations differ, adjacent blades overlap, and texture varies even within one plant. Weed color is often close to that of the crop. The boundaries among the classes “tomato - weed - background” must therefore be inferred from a combination of morphological and background cues; no stable single criterion based on color or geometry exists in the scenes considered.

The main data shortage did not arise at the image-acquisition stage. Capturing RGB images required comparatively little time, whereas expert preparation of one frame containing several plants and background regions could take up to 15 minutes. Scenes with small weeds, overlapping leaves, and ambiguous object boundaries were especially time-consuming to annotate. Consequently, the volume of raw photographs substantially exceeded the number of images verified by a specialist. Hundreds of reliably validated scenes, rather than thousands of raw frames, were therefore available for training.

For such a dataset, the number of files must be distinguished from the number of newly observed conditions. A frame acquired with a different plant geometry or a different background state contributes new information to the model. A sequence recorded during one season, in the same row type, and under similar illumination is different. Similar leaf shadows, substrate patterns, trellis directions, drip-tape edges, and reflections on wet surfaces recur from frame to frame. The network receives useful object cues together with a repeated noise structure; under these conditions, increasing N is no longer equivalent to increasing diversity.

Most published studies in agricultural machine vision use YOLO to detect fruit, weeds, plant diseases, and technological objects. With large datasets, attention is naturally concentrated on final accuracy. A small dataset raises a different question: what happens when additional images barely expand the space of observed scenes? Of particular interest is the regime in which aggregate metrics have deteriorated only moderately, whereas high-confidence false bounding boxes have already appeared among the errors.

For an agricultural robot, two predictions assigned to the same wrong class may have different technological consequences. A low-confidence result can be removed by thresholding or sent for repeated analysis. A value above 0.9, by contrast, normally falls within the detector's confidence region. If such a bounding box is false, the actuator loop may treat a weed-free area as a target and perform an unnecessary local operation. In this study, this specific type of error is referred to as a model hallucination.

The study was limited to $N \leq 600$ and to a single task: weed detection in tomato plantings using YOLOv8m. The objective was to locate the boundary beyond which enlarging a structurally similar dataset ceases to improve generalization. A dataset with a fixed annotation scheme was prepared, and training series were carried out for different values of N . For each series, mAP@0.5, Precision, Recall, the hallucination coefficient, the L_2 gradient norm, and the oversaturation index were compared; activation maps were additionally examined for the rational and oversaturated regimes.

In agricultural object detection, YOLO is attractive primarily because of its one-stage prediction scheme: object coordinates and class are determined within a single network. The family is therefore used for plant monitoring, fruit detection, robotic harvesting, and weed detection. The requirements of a mobile platform, however, extend beyond accuracy alone. Computational load and robustness to heterogeneous backgrounds become part of the same engineering problem because the detector must operate within a constrained onboard computing loop.

For tomato crops, the literature more often addresses fruit ripeness, developmental stages, and robotic harvesting. Typical sources of error are familiar: leaf occlusion, complex backgrounds, and variable illumination. Attention modules, lightweight blocks, or other YOLOv8 modifications are commonly proposed in response. Such architectural changes may improve detection, but the number of original annotated scenes remains a separate issue. Architectural complexity cannot create plant or environmental states that are absent from the source material.

The high mAP values reported for multiclass weed detection in cotton using YOLOv8 [4, p. 2] were obtained from a different observation structure. A field dataset covers a broader range of scenes, whereas the present experiment was deliberately confined to a greenhouse row and one growing season. Weed species composition, background, and the range of acquisition conditions all differ. The information content of the two datasets is therefore not directly comparable. The required size of the small tomato dataset was established in a dedicated experimental series rather than by transferring a threshold from the field domain.

Additional ambiguity is introduced by the tomato leaf apparatus itself. For cereal crops, recognized difficulties include similarity in crop and weed color, object overlap, and the absence of an unambiguous size cue. The large dissected tomato leaf adds another class of background repetition: its shadows and characteristic texture fragments recur at different locations along the row and often appear near annotated target objects. In a small dataset, such co-occurrence can reinforce an accidental background combination as a weed-related cue.

The concept of double descent is used here only as a framework for discussing non-monotonic error behavior. The conducted series does not reproduce the classical double-descent curve and is not intended to demonstrate its universality. The relevant point is a possibility recognized in modern learning theory: more examples or a more complex model do not improve generalization monotonically in every regime. In the present formulation, the content of each new image is decisive. Dataset expansion is informative while a frame introduces a previously unrepresented scene state. When additional frames are close realizations of an already observed distribution, the increase in N remains quantitative and may not expand the feature space available to the model.

YOLOv8m has substantial parametric capacity for this regime. According to the Ultralytics documentation, the model contains 25.9 million parameters at an input size of 640 pixels. This expressiveness is sufficient for complex object detection; on several hundred closely related scenes, however, the same capacity also creates an opportunity to memorize incidental background dependencies. Along with tomato and weed morphology, the network may interpolate covariations associated with recurrent image structure.

Changes in the network's internal focus were evaluated using Grad-CAM. The method produces a map of image regions that contribute most strongly to the decision of a convolutional model. Under a rational data regime, activations are expected to concentrate on leaves, stems, and weed vegetation. A shift toward background textures in an oversaturated regime, conversely, indicates weakening of the detector's object-specific focus.

The scientific problem addressed in this study is therefore the identification of a boundary beyond which expansion of a small, weakly variable dataset no longer broadens the model's representation of target objects. Past this boundary, the repeated noise component of the empirical distribution receives additional weight and strengthens memorization. The corresponding state is termed cognitive saturation of the model.

Materials and Methods

The source dataset contained 600 RGB images with a resolution of 640x640 pixels, acquired in tomato rows during one growing season. This deliberately narrow domain was selected to isolate saturation from ordinary domain variability. A change of season, cultivar, camera, or cultivation technology would have introduced an additional covariate shift and complicated interpretation of the saturation effect. The experiment therefore did not aim to estimate maximum domain transferability; it sought an internal limit within a closely related distribution of scenes.

Annotation was performed for three classes. The “tomato plant” class included leaves, distinguishable stem fragments, and other visible tomato parts that could be separated from the surroundings. The “weed vegetation” class covered unwanted plants in the inter-row area and near the root zone. Problematic background fragments - soil, substrate, plant residues, drip tape, shadows, and free background regions - were assigned to the “soil-substrate background” class. The third class was retained specifically because such fragments could provoke false detections.

Training subsets were formed for seven values of N : 100, 200, 250, 300, 400, 500, and 600 images. Class proportions were maintained as the dataset size changed. Source scenes included in validation were excluded from training. This restriction is methodologically important for a small dataset because placing nearly identical frames on opposite sides of the split artificially inflates the estimated performance.

The transformation set was intentionally restricted. Only rotations of ± 15 degrees and horizontal flips were applied; intensive color correction, Mosaic, CutMix, and strong geometric distortions were excluded. This regime made it possible to observe the boundary between useful data expansion and memorization of the noise component of the empirical distribution without masking it by strong augmentation-based regularization.

Training used early stopping with patience=50 epochs, and weight checkpoints were saved every 10 epochs. Final weights, validation loss, the confidence-score distribution, and the L_2 norm

of gradients in the final detection layer were recorded for each series. The stochastic parameter-update rate was held constant across the series. Its potential influence on the location of the saturation threshold was therefore excluded from the cross-regime comparison.

The loss function was represented as the sum of a localization component and a focal component. The latter accounted for imbalance between object and background regions. The basic form was:

$$L_{total} = LCIoU + \alpha \cdot LFocal, \alpha = 0.75 \quad (1)$$

In expression (1), $LCIoU$ characterizes the localization quality of bounding boxes. $LFocal$ reduces the contribution of easily classified examples and reallocates weight toward difficult objects; the focal term was weighted by $\alpha=0.75$. This choice is consistent with the foreground-background imbalance known for one-stage detectors [9, p. 319].

High-confidence false responses were described using a separate hallucination coefficient. It was calculated as the mean logarithmically normalized number of false detections with confidence above 0.9:

$$Khal = (1/M) \cdot \sum (j=1..M) \ln(1 + FP_{0.9,j}) \quad (2)$$

Here, M is the number of images in the validation set, and $FP_{0.9,j}$ is the number of false bounding boxes in the j -th image at confidence > 0.9 . Logarithmic normalization converts the indicator into a dimensionless form and limits the influence of a single frame containing an anomalously large number of false objects.

Precision, Recall, and $mAP@0.5$ alone were insufficient to describe changes within the final detection layer. The relationship between parameter dispersion and validation error was therefore compared across the training series. For this purpose, the oversaturation index was introduced:

$$Sover = \sigma^2(W_{last}) / \mu(L_{val}) \quad (3)$$

In expression (3), the numerator $\sigma^2(W_{last})$ characterizes the variance of the weights in the final detection layer, whereas $\mu(L_{val})$ in the denominator is the mean validation loss. The *Sover* index combines two quantities that are usually evaluated separately. Within the present experimental series, its increase was considered together with the appearance of locally confident predictions associated with unstable background cues. This interpretation is specific to the adopted formulation and does not define a universal threshold for other architectures or datasets.

The restructuring of the detection head was also monitored through gradients. Their L_2 norm was calculated for the final layer using:

$$G2 = \sqrt{\sum (i=1..m) (dL / dw_i)^2} \quad (4)$$

Conventional performance assessment was retained through three standard object-detection metrics: Precision, Recall, and $mAP@0.5$. They were calculated as:

$$Precision = TP / (TP + FP) \quad (5)$$

$$Recall = TP / (TP + FN) \quad (6)$$

$$mAP@0.5 = (1/C) \cdot \sum (c=1..C) AP_c(IoU=0.5) \quad (7)$$

The spatial Grad-CAM comparison was supplemented with a numerical measure of the difference between activation distributions. The Kullback-Leibler divergence was calculated between the rational and oversaturated regimes:

$$DKL = \sum_z P_{250}(z) \cdot \ln(P_{250}(z) / P_{600}(z)) \quad (8)$$

Results

The results of the seven training series were grouped by the type of observed characteristic. Metrics directly related to detection performance are presented in Table 1. Training time, L_{val} , the L_2 norm, and $Sover$ are summarized in Table 2. This separation does not alter the data and keeps both tables readable in portrait A4 format.

Table 1

Detection performance metrics for different training-set sizes

N	mAP	$Prec.$	$Rec.$	$Khal$
100	0.67	0.71	0.58	0.05
200	0.83	0.86	0.79	0.12
250	0.91	0.93	0.88	0.18
300	0.88	0.89	0.81	0.45
400	0.80	0.78	0.70	0.89
500	0.71	0.65	0.59	1.35
600	0.62	0.54	0.48	1.80

Table 2

Training time, gradient L_2 norm, and oversaturation index as N changes

N	$Time, h$	L_{val}	L_2	$Sover$
100	0.8	0.91	0.40	0.82
200	1.5	0.52	0.52	1.03
250	2.0	0.31	0.64	1.18
300	2.4	0.39	0.91	2.08
400	3.1	0.58	1.36	2.54
500	3.9	0.75	1.82	3.02
600	4.6	0.94	2.30	3.45

The first series represents a shortage of object examples. With 100 images, $mAP@0.5$ was 0.67, Precision was 0.71, and Recall was 0.58. The detector already responded to large objects but frequently missed small weeds and visually ambiguous parts of tomato plants. $Khal$ was only 0.05. The small number of high-confidence false responses should not be interpreted as robustness: the model remained insufficiently active in this regime and simultaneously exhibited low recall.

The largest positive shift in the three primary metrics occurred not at the end of the experimental range but after the second hundred images were added. At $N=200$, $mAP@0.5$ reached 0.83, Precision 0.86, and Recall 0.79. The dataset now contained new leaf positions, different planting densities and weed levels, and altered background states. $Khal$ increased from 0.05 to 0.12, but this increase was substantially smaller than the gain in detection performance. Over this interval, increasing N was still accompanied by a meaningful broadening of the observed scenes.

The highest combined performance was obtained with 250 images. $mAP@0.5$, Precision, and Recall were 0.91, 0.93, and 0.88, respectively. For the dataset under study, this regime was adopted as the rational saturation level. The number of scenes was already

sufficient for class discrimination, whereas recurrent background details - substrate patterns, drip-tape edges, and leaf shadows - had not yet begun to dominate object decisions according to the results of the series.

The next experiment interrupted the upward trend. An additional 50 images reduced $mAP@0.5$ to 0.88; Precision and Recall fell to 0.89 and 0.81. The decline in aggregate metrics was still modest. $Khal$ changed much more strongly: 0.18 at $N=250$ became 0.45 at $N=300$. The first visible restructuring therefore appeared not as an abrupt collapse of mAP but as an accumulation of false bounding boxes to which the network assigned high confidence.

Inspection of predictions from the 400-image series revealed a pronounced spatial localization of errors. False boxes appeared predominantly on furrows, substrate boundaries, elongated stem shadows, and regions containing plant residues. Repeated linear geometry was a common property of these areas. At the same time, $mAP@0.5$ decreased to 0.80, Precision to 0.78, Recall to 0.70, and $Khal$ increased to 0.89. Thus, degradation in this series was not spatially uniform and was associated with specific configurations of the background scene.

The final two regimes differed from the preceding points not by an isolated deviation in one metric but by a general deterioration of detection. At $N=500$, $mAP@0.5=0.71$, $Precision=0.65$, $Recall=0.59$, and $Khal=1.35$ were obtained. With 600 images, these values changed to 0.62, 0.54, 0.48, and 1.80. The nature of the error also changed: some false objects had high confidence. Such predictions are more likely to be accepted by the actuator loop as valid. For targeted inter-row treatment, the technologically dangerous transition is specifically the shift from frequent misses to confident false detection of a background region.

To quantify the difference between the regime with maximum Precision and the final experimental point, the relative decrease in this metric was calculated:

$$\Delta P = ((P_{250} - P_{600}) / P_{250}) \cdot 100\% = 41.94\% \quad (9)$$

Substitution of the $N=250$ and $N=600$ values into expression (9) gives 41.94%, or approximately 42%. Over the same range, $Khal$ reaches 1.80. These quantities describe different aspects of the same practical risk: positive predictions become less accurate, while the share of high-confidence errors among false decisions increases. A background region may therefore be misinterpreted as a weed more frequently and transferred to the control loop as a basis for actuation.

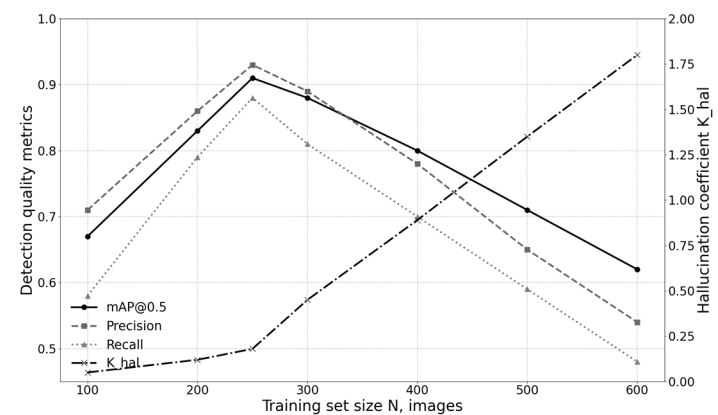


Fig. 1. Changes in detection metrics and hallucination coefficient with increasing training-set size

The four curves in Figure 1 move in the same direction only up to $N=250$. At this point, $mAP@0.5$, *Precision*, and *Recall* reach their maxima; further enlargement of the dataset is accompanied by a decline in all three. *Khal* follows the opposite trajectory, increasing over almost the entire experimental range and more rapidly after $N=300$. The interval between 250 and 300 images therefore cannot be treated as an ordinary plateau. Mean metrics remain close to their maxima, yet the structure of false decisions is already shifting toward high-confidence detections.

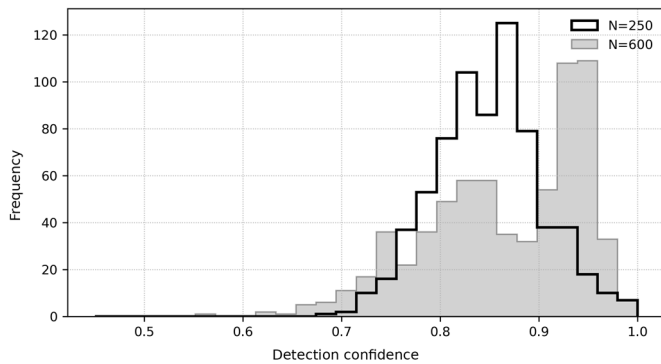


Fig. 2. Detection-confidence distributions for $N=250$ and $N=600$

Figure 2 exposes the difference without reducing it to a single averaged metric. At $N=250$, correct predictions are concentrated mainly around a confidence of 0.85. In the $N=600$ series, the distribution contains two pronounced regions; the second corresponds to false responses at approximately 0.92-0.96. Simply raising the confidence threshold does not resolve the problem. The erroneous objects lie precisely in the part of the scale that is usually considered the most reliable.

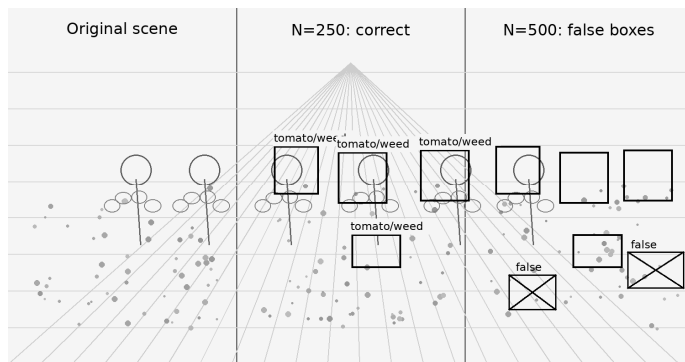


Fig. 3. Comparison of YOLOv8m prediction localization on the tomato scene and background regions

The schematic in Figure 3 separates two fundamentally different types of wrong decision. In the $N=250$ series, bounding boxes correspond to tomato plants and weed vegetation that are actually present. At $N=500$, additional bounding boxes appear on background regions; they are marked with crosses in the schematic. These cases no longer concern an inaccurate object boundary. The target object itself is absent, yet the detector still produces a positive prediction.

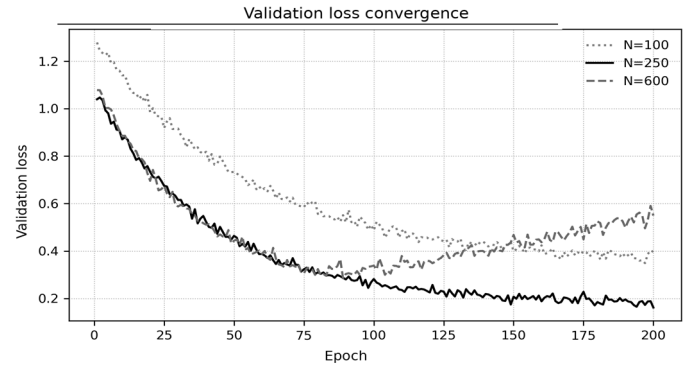


Fig. 4. Validation-loss trajectories for three training regimes

Comparison of the curves in Figure 4 reveals different causes of elevated error. At $N=100$, validation loss remains at a high residual level, which is consistent with a shortage of training examples. The $N=250$ regime behaves differently: the value decreases steadily and then stabilizes. For $N=600$, the initial downward trend is also present, but the direction of the curve later reverses and a secondary increase appears. This is also the series in which memorization of the noise component of the empirical distribution becomes stronger.

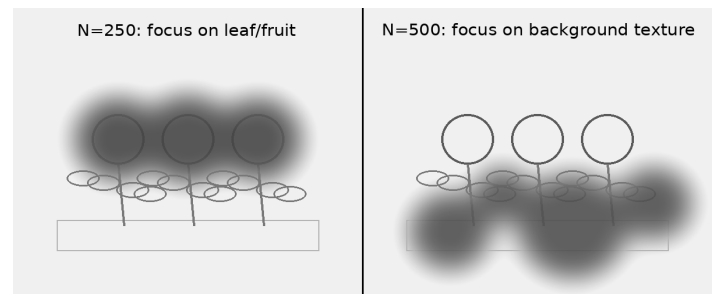


Fig. 5. Comparison of Grad-CAM regions in the rational and oversaturated regimes

The change in the internal basis of the decision can be traced in the Grad-CAM maps in Figure 5. At $N=250$, the strongest contribution is associated with the leaf blade, fruit contour, and fragments immediately adjacent to the object. After the transition to $N=500$, part of the active regions appears on the background texture. The difference between the regimes therefore concerns more than the final magnitude of the error: the spatial set of cues used by the network to form a bounding box also changes.

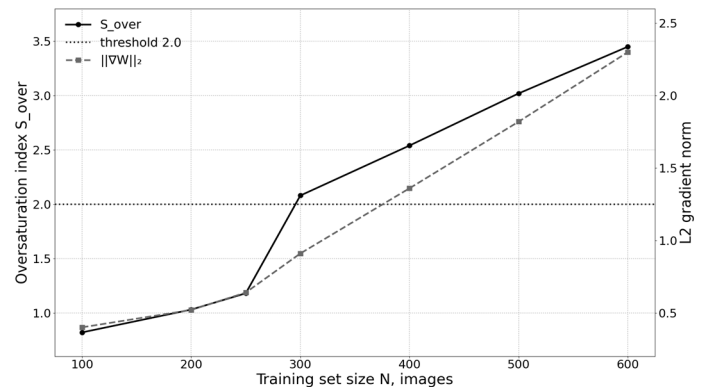


Fig. 6. Dynamics of *Sover* and the L_2 norm of the final detection-layer gradient

Sover exceeds 2.0 after $N=300$. In parallel, the L_2 norm of the final-layer gradient increases from 0.40 to 2.30 across the experimental series. No attenuation of the detection head response is observed. On the contrary, the final-layer parameters continue to undergo active restructuring in the presence of contradictory micro-patterns, and the number of high-confidence false detections increases in the same training regimes.

Discussion

Between 100 and 250 images, enlargement of the dataset coincided with a genuine expansion of the observed scene states. Leaf position, planting density, weed level, soil appearance, and substrate appearance all changed. These differences supplied the network with new combinations of object-related cues; the improvement in mAP@0.5, Precision, and Recall cannot therefore be attributed solely to the number of files. In the initial part of the experimental range, both N and the meaningful variability of the training distribution increased.

The situation changed after the peak regime. New frames continued to enlarge the dataset but increasingly reproduced conditions already represented in the same technological zone. For a highly parameterized YOLOv8m model, this provides additional realizations of local visual dependencies, some of which may belong to the background rather than to stable target morphology. The number of examples then grows faster than the number of independent scene states. This divergence between dataset size and informational breadth forms the basis of our subsequent analysis of oversaturation.

An early sign of the transition is visible when $N=250$ and $N=300$ are compared. mAP@0.5 decreases by only 0.03, from 0.91 to 0.88. Over the same step, *Khal* rises from 0.18 to 0.45, a 2.5-fold increase. If only aggregate accuracy is examined, the change may resemble a moderate fluctuation near the optimum. High-confidence false predictions give a different picture: their accumulation begins before the decline in mAP becomes pronounced.

For a system that controls a working tool, the earlier response of *Khal* relative to mAP is of separate importance. An averaged metric describes overall detection performance but does not indicate how dangerous a particular error is for the subsequent action. A false bounding box with high confidence may pass the trust threshold and enter an actuator command without repeated verification. The hallucination coefficient is therefore used to characterize a technologically significant class of errors rather than as an alternative name for mAP.

We associate the observed behavior with repeated background combinations inside one technological zone. A shadow of a particular shape, the edge of drip tape, or an accumulation of plant residues occurred in many closely related scenes. The parametric capacity of YOLOv8m permits memorization of a statistical association between such local fragments and the target class. This explanation is restricted to the experiment conducted here: with broader domain diversity, the same architecture may distribute its capacity across a larger number of stable visual regularities.

The errors at the two extremes of the experimental range have different origins. At $N=100$, the model lacks examples of object morphology. At $N=600$, the number of files is much larger, but the informational dimensionality of the scene grows disproportionately slowly relative to N . In the latter case, the network repeatedly

receives close realizations of a latent noise structure instead of new independent states of the agrobiocenosis. Data scarcity and an excess of similar data should therefore not be regarded as the same training regime.

The transition in the internal characteristic of the final layer is visible directly in the numerical values: *Sover* is 1.18 at $N=250$ and reaches 2.08 at $N=300$; the index remains above 2.0 thereafter. *Khal* increases over the same range, while Grad-CAM shifts part of the activations toward background regions. For the obtained series, the coincidence of these three features describes the same transition to an oversaturated regime. The present data do not justify treating *Sover* as a universal standalone criterion, but the index can serve as an additional diagnostic signal when deciding whether to continue dataset expansion.

Augmentation can increase the number of variants derived from a single image, but it does not necessarily alter the informational diversity of the scene. Rotation, geometric transformation, or correction of visual characteristics operate on an existing frame. A new plant growth stage or a new environmental condition cannot be introduced in this way. If repeated background structures and random covariations are already present many times in the source set, their transformed versions remain content-related to the same scenes. An increase in the number of augmented examples therefore cannot automatically be equated with the appearance of independent observations.

For the activation distributions, $DKL(P250||P600)=0.37$ and $DKL(P600||P250)=0.52$ were obtained. The asymmetry is particularly important in the reverse direction: the oversaturated regime contains activation states that were not represented at $N=250$. Independently, a second pronounced mode of false responses appears in the confidence-score distribution for $N=600$ at 0.92-0.96. Within the present experiment, part of the additional internal states coincides with the emergence of high-confidence erroneous predictions.

This regularity should be considered before annotation of the next image series begins. After approximately 250 high-quality annotated frames acquired during one season and under similar conditions, the automatic addition of several hundred nearly identical scenes does not guarantee a performance gain. More meaningful expansion requires new acquisition conditions: a different time of day and lighting scheme, another substrate type, or another cultivation area. Growth stage, cultivar, and planting density also alter the observed distribution. Under such expansion, the diversity of states increases rather than N alone.

The influence of architectural capacity can also be tested as a means of reducing the tendency of a small dataset to encourage local texture memorization. YOLOv8n contains fewer trainable parameters than YOLOv8m and may therefore be less able to adapt to idiosyncratic features of individual scenes; for a mobile platform, a smaller model simultaneously reduces computational requirements. This remains a separate experimental hypothesis. The two detectors should be compared not only by final accuracy but also by their robustness to the background structures on which high-confidence false detections occur.

In the transition region $N=250-300$, mAP detects the degradation later than *Khal*. The next stage of the study should therefore address early diagnosis of oversaturation using *Khal*, *Sover*, the dynamics of the gradient L_2 norm, and changes in activation maps jointly. Conventional early stopping is oriented mainly toward validation loss.

For autonomous inter-row treatment, the diagnostic loop should additionally register the moment at which high-confidence false objects begin to appear, because these detections are directly associated with the risk of an erroneous control action.

Conclusions

The maximum performance in the present series occurred at $N=250$: $mAP@0.5=0.91$, $Precision=0.93$, and $Recall=0.88$. Up to this point, dataset growth was accompanied by improved generalization; beyond it, the dependence became non-monotonic. For the scene set considered here, $N=250$ corresponds to the rational cognitive saturation threshold of YOLOv8m.

The final $N=600$ regime differs substantially from the peak regime. Precision decreases to 0.54, $Khal$ increases to 1.80, and the relative loss of Precision is 41.94%. Correct annotation of additional frames alone therefore does not guarantee improved practical reliability: when scene variability is weak, increasing the number of images may be accompanied by deterioration of detector performance.

The dynamics of the final detection layer refine the location of the identified boundary. After $N=300$, $Sover$ exceeds 2.0, while the gradient L_2 norm reaches 2.30 within the experimental range. At the same time, false bounding boxes appear on the greenhouse-row background, soil-substrate surface, shadows, and plant residues. For this series, the combination of these features is associated with an unstable response of the detection head to recurrent background micropatterns.

The obtained value $N=250$ should be regarded only as an experimental optimum within the domain studied. It refers to robotic tomato-care scenes recorded during one growing season and in one technological zone; the result does not establish a universal standard for an arbitrary dataset. Under these conditions, 250 ± 20 annotated frames can be used as a practical reference. Once this range is reached, it is more rational to introduce new observation states – different illumination, growth stage, cultivar, substrate type, planting density, or another acquisition site – rather than to continue adding closely related frames. Such a domain change may also shift the saturation threshold itself.

If heterogeneous scenes cannot be collected, subsequent work should separately test alternative methods for limiting memorization. These include reducing parametric capacity, strengthening regularization, and jointly monitoring $Khal$ and $Sover$. A lighter architecture should not be assumed effective in advance; it must be compared with YOLOv8m in a dedicated series. For an autonomous system, mAP should likewise not remain the sole metric because average accuracy does not express the risk of a confident false response in a region without weeds. In the adopted formulation, $Khal$ and $Sover$ complement the evaluation specifically from this perspective.

Acknowledgements

This article was prepared as part of the implementation of research work, implemented at the expense of the federal budget (source of financing – the Ministry of Education and Science of the Russian Federation) on the topic: “Development of software for the agro-industrial complex for monitoring the state of the tops and fruits of greenhouse plants using machine vision and artificial intelligence” (code of the scientific theme FFSM-2025-0002).

References

1. C. M. Badgujar, A. Poulse, H. Gan, “Agricultural object detection with You Only Look Once (YOLO) Algorithm: a bibliometric and systematic literature review,” *Computers and Electronics in Agriculture*. 2024. Vol. 223. Article 109090. DOI: 10.1016/j.compag.2024.109090.
2. Y. Fu, J. Chen, X. Li, Y. Zhang, “Multi-stage tomato fruit recognition method based on improved YOLOv8,” *Frontiers in Plant Science*. 2024. Vol. 15. Article 1447263. DOI: 10.3389/fpls.2024.1447263.
3. Z. Yang, H. Zhang, J. Li, X. Wang, “A method for tomato ripeness recognition and detection based on an improved YOLOv8 model,” *Horticulturae*. 2025. Vol. 11. No. 1. Article 15.
4. A. T. Khan, S. M. Jensen, A. R. Khan, “Advancing precision agriculture: a comparative analysis of YOLOv8 for multi-class weed detection in cotton cultivation,” *Artificial Intelligence in Agriculture*. 2025. Article 100598. DOI: 10.1016/j.aiaa.2025.01.013.
5. Y. Liu, F. Zeng, H. Diao, J. Zhu, D. Ji, X. Liao, Z. Zhao, “YOLOv8 model for weed detection in wheat fields based on a visual converter and multi-scale feature fusion,” *Sensors*. 2024. Vol. 24. No. 13. Article 4379. DOI: 10.3390/s24134379.
6. P. Nakkiran, G. Kaplan, Y. Bansal, T. Yang, B. Barak, I. Sutskever, “Deep double descent: where bigger models and more data hurt,” *Journal of Statistical Mechanics: Theory and Experiment*. 2021. No. 12. Article 124003. DOI: 10.1088/1742-5468/ac3a74.
7. Ultralytics. YOLOv8 models: performance metrics and model variants. 2023. URL: <https://docs.ultralytics.com/models/yolov8/> (accessed: 03 July 2026).
8. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, “Grad-CAM: visual explanations from deep networks via gradient-based localization,” *International Journal of Computer Vision*. 2020. Vol. 128, pp. 336-359. DOI: 10.1007/s11263-019-01228-7.
9. T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, “Focal loss for dense object detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020. Vol. 42. No. 2, pp. 318-327. DOI: 10.1109/TPAMI.2018.2858826.
10. M. Belkin, D. Hsu, S. Ma, S. Mandal, “Reconciling modern machine-learning practice and the classical bias-variance trade-off,” *Proceedings of the National Academy of Sciences*. 2019. Vol. 116. No. 32. P. 15849-15854. DOI: 10.1073/pnas.1903070116.
11. H. Sun, Y. Zhou, L. Zhao, Q. Wang, “An improved YOLOv8 model for detecting four stages of tomato ripening and its application deployment in a greenhouse environment,” *Agriculture*. 2025. Vol. 15. No. 9. Article 936.
12. J. Liu, K. Wang, L. Chen, Y. Zhao, “Optimization of a multi-environmental detection model for tomato growing points and flower buds based on improved YOLOv8,” *Scientific Reports*. 2025. Vol. 15. Article 6692.

ИССЛЕДОВАНИЕ ПОРОГА КОГНИТИВНОГО НАСЫЩЕНИЯ МОДЕЛИ YOLOV8 В УСЛОВИЯХ ОГРАНИЧЕННОЙ ОБУЧАЮЩЕЙ ВЫБОРКИ

Карелина Мария Юрьевна, Государственный университет управления; Московский автомобильно-дорожный технический университет (МАДИ), Москва, Россия, karelinamu@mail.ru

Акулов Алексей Андреевич, Государственный университет управления; Московский автомобильно-дорожный технический университет (МАДИ), Москва, Россия, a.akulov.98@mail.ru

Талдыкин Дмитрий Сергеевич, Московский автомобильно-дорожный технический университет (МАДИ), Государственный университет управления, Москва, Россия, dima.dima.taldykin@mail.ru

Аннотация

В работе проверялось, при каком объеме малой размеченной выборки YOLOv8m перестает получать выигрыш от добавления близких по содержанию кадров при детекции сорной растительности в посадках томатов. Постановка связана с обычным для точного земледелия ограничением: RGB-изображения получить сравнительно легко, тогда как проверка разметки агрономом занимает заметно больше времени. Для эксперимента использованы 600 изображений 640x640 пикселей, снятых в тепличных рядах за один вегетационный период; выделялись томат, сорная растительность и фон почвенно-субстратной поверхности. Семь серий обучения позволили проверить предположение о максимуме обобщения в окрестности $N=250$ при числе уникальных сцен менее 300. После этой точки новые кадры, если условия съемки почти не менялись, сопровождалась ростом ложных детекций с высокой уверенностью. Для описания эффекта использованы коэффициент галлюцинаций и индекс перенасыщения, в котором сопоставлены дисперсия весов последнего детектирующего слоя и средняя валидационная потеря. В серии $N=600$ Precision оказалась на 41,94% ниже пикового значения, коэффициент галлюцинаций достиг 1,80. Практический результат относится прежде всего к планированию разметки и выбору емкости детектора: увеличение числа похожих кадров нельзя рассматривать как безусловный способ повышения надежности автоматизированной системы.

Ключевые слова: YOLOv8; томаты; детекция сорняков; малые выборки; галлюцинации модели

Литература

1. Badgujar C. M., Poulose A., Gan H. Agricultural object detection with You Only Look Once (YOLO) Algorithm: a bibliometric and systematic literature review // Computers and Electronics in Agriculture. 2024. Vol. 223. Article 109090. DOI: 10.1016/j.compag.2024.109090.
2. Fu Y., Chen J., Li X., Zhang Y. Multi-stage tomato fruit recognition method based on improved YOLOv8 // Frontiers in Plant Science. 2024. Vol. 15. Article 1447263. DOI: 10.3389/fpls.2024.1447263.
3. Yang Z., Zhang H., Li J., Wang X. A method for tomato ripeness recognition and detection based on an improved YOLOv8 model // Horticulturae. 2025. Vol. 11. No. 1. Article 15.
4. Khan A. T., Jensen S. M., Khan A. R. Advancing precision agriculture: a comparative analysis of YOLOv8 for multi-class weed detection in cotton cultivation // Artificial Intelligence in Agriculture. 2025. Article 100598. DOI: 10.1016/j.iaia.2025.01.013.
5. Liu Y., Zeng F., Diao H., Zhu J., Ji D., Liao X., Zhao Z. YOLOv8 model for weed detection in wheat fields based on a visual converter and multi-scale feature fusion // Sensors. 2024. Vol. 24. No. 13. Article 4379. DOI: 10.3390/s24134379.
6. Nakkiran P., Kaplan G., Bansal Y., Yang T., Barak B., Sutskever I. Deep double descent: where bigger models and more data hurt // Journal of Statistical Mechanics: Theory and Experiment. 2021. No. 12. Article 124003. DOI: 10.1088/1742-5468/ac3a74.
7. Ultralytics. YOLOv8 models: performance metrics and model variants. 2023. URL: <https://docs.ultralytics.com/models/yolov8/> (accessed: 03 July 2026).
8. Selvaraju R. R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization // International Journal of Computer Vision. 2020. Vol. 128. P. 336-359. DOI: 10.1007/s11263-019-01228-7.
9. Lin T.-Y., Goyal P., Girshick R., He K., Dollar P. Focal loss for dense object detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020. Vol. 42. No. 2. P. 318-327. DOI: 10.1109/TPAMI.2018.2858826.
10. Belkin M., Hsu D., Ma S., Mandal S. Reconciling modern machine-learning practice and the classical bias-variance trade-off // Proceedings of the National Academy of Sciences. 2019. Vol. 116. No. 32. P. 15849-15854. DOI: 10.1073/pnas.1903070116.
11. Sun H., Zhou Y., Zhao L., Wang Q. An improved YOLOv8 model for detecting four stages of tomato ripening and its application deployment in a greenhouse environment // Agriculture. 2025. Vol. 15. No. 9. Article 936.
12. Liu J., Wang K., Chen L., Zhao Y. Optimization of a multi-environmental detection model for tomato growing points and flower buds based on improved YOLOv8 // Scientific Reports. 2025. Vol. 15. Article 6692.

Информация об авторах:

Карелина Мария Юрьевна, д.т.н., д.п.н., профессор, заведующий кафедрой логистики и транспортно-технологических систем, Государственный университет управления, Заведующий кафедрой "Детали машин и теория механизмов" Московский автомобильно-дорожный технический университет (МАДИ), Москва, Россия. ORCID ID: 0000-0003-0335-7550

Акулов Алексей Андреевич, к.т.н., доцент кафедры логистики и транспортно-технологических систем, Государственный университет управления, Доцент кафедры "Детали машин и теория механизмов" Московский автомобильно-дорожный технический университет (МАДИ), Москва, Россия. ORCID ID: 0000-0001-8952-0237

Талдыкин Дмитрий Сергеевич, аспирант, ассистент кафедры "Детали машин и теория механизмов", Московский автомобильно-дорожный государственный технический университет (МАДИ), младший научный сотрудник, Государственный университет управления, Москва, Россия. ORCID ID: 0009-0008-5071-426X