



High Recall and Weak Localization in Spatially Weighted Camera Soiling Segmentation

Abdulazizxon Axmedov Ganijon O'g'li

PhD student, Namangan State University, Namangan, Uzbekistan

ORCID: 0009-0004-6784-4010

Email: abdulazizahmedov97@gmail.com, abdulaziz_axmedov@namdu.uz

Abstract

A vehicle camera soiling detector can achieve high recall by labelling most of an image as contaminated. This study re-examines recorded pixel counts from a U-Net baseline and two spatially weighted training runs, using a fixed 3×3 grid and a Gaussian priority map. Each evaluation contains 2,457,600 pixels at a decision threshold of 0.5. Recalculation shows that the Gaussian run detects 99.989% of labelled defect pixels but correctly retains only 0.055% of clean pixels. Its predicted defect area covers 99.972% of the evaluated pixels. Its Dice score of 0.761949 exceeds an all-defect predictor by only 0.00497 percentage points. The baseline Dice score, 0.762570, is slightly below its own all-defect reference; the grid run gives 0.675116. Balanced accuracy and Matthews correlation confirm weak discrimination in all three records. The baseline and weighted evaluations have different foreground totals, so these results are not treated as a controlled comparison of loss functions. An analytical examination also shows that unnormalized spatial weights change the balance between binary cross-entropy and Dice terms. The contribution is a retrospective failure analysis: it identifies extensive oversegmentation, separates this observation from unproven explanations, and specifies the controls needed before spatial weighting can support decisions about vehicle camera availability.

Keywords

vehicle camera; lens soiling; U-Net; spatially weighted loss; oversegmentation; Dice coefficient; camera health monitoring.



1 Introduction

A useful camera soiling mask must identify contaminated pixels while preserving clear parts of the image. Detecting nearly every contaminated pixel is insufficient if the same model rejects almost every clean pixel. For a car using cameras to observe lanes, surrounding vehicles, or parking obstacles, this distinction affects how much visual information remains available after a defect warning.

WoodScape includes fisheye imagery and perception tasks for automotive surround-view systems [1]. SoilingNet addresses contamination detection specifically for automotive cameras [2]. These studies establish the application context, but a particular implementation still requires evaluation of both missed contamination and unnecessary rejection of clear image regions. A segmentation mask is an intermediate measurement; its overlap score does not establish vehicle-level safety.

Spatial weighting offers an intuitive way to emphasize selected image locations. A central region may seem more important than an image corner, and a smooth Gaussian map avoids abrupt changes between regions. However, an image-coordinate preference does not by itself improve defect discrimination. It can change the training objective without resolving a model's tendency to predict contamination everywhere.

The present work revisits a published U-Net baseline [3] and recorded follow-up runs with grid and Gaussian weights. It asks a narrower question than whether spatial weighting is generally effective: what do these particular evaluation counts reveal when clean-pixel specificity, predicted area, and a trivial all-defect reference are examined alongside Dice and recall? The new contribution consists of this recalculation, an audit of comparability, and an analysis of the mixed loss. The baseline measurements are reused with attribution; no new network training is claimed.

2 LITERATURE REVIEW



"KELAJAK TEKNOLOGIYALARI VA SUN'IY INTELLEKT"
nomli respublika ilmiy-amaliy masofaviy konferensiyasi
VOLUME-1, ISSUE-6, 2026

U-Net combines an encoder–decoder structure with skip connections to recover spatial detail [4]. Its architecture permits many training objectives, and those objectives address different problems. Generalised Dice uses class-volume information to handle imbalanced segmentation [5]. Focal loss reduces the relative influence of easy examples during detection training [6]. The Tversky formulation permits different penalties for false positives and false negatives [7]. These mechanisms differ from assigning a fixed multiplier to an image coordinate.

A coordinate weight is applied wherever a pixel lies, regardless of whether that pixel is clean or contaminated. It should therefore not be described as class balancing unless its construction explicitly accounts for class membership. In the present records, contaminated pixels comprise about 62% of the pooled labels. The defect class is not the minority class at this evaluation scale, although individual images may have different proportions.

Evaluation measures also answer different questions. Foreground Dice and intersection over union (IoU) omit true negatives from their formulas. Matthews correlation coefficient (MCC) incorporates all four entries of the confusion matrix and can expose imbalanced prediction behaviour that an isolated F1 score does not adequately describe [8]. Here, MCC is used together with recall, specificity, and balanced accuracy. No single measure substitutes for the full counts.

3 Materials and retrospective analysis

3.1 Evaluation records and their limits

The experiment records describe a WoodScape soiling subset with 499 development images and 150 test images, resized to 128×128 pixels. The development split is recorded as 399 training and 100 validation images. Masks are converted to a binary target, with nonzero values treated as defect pixels. This operational definition combines the retained nonzero mask categories; it does not distinguish contaminant



"KELAJAK TEKNOLOGIYALARI VA SUN'IY INTELLEKT"
nomli respublika ilmiy-amaliy masofaviy konferensiyasi
VOLUME-1, ISSUE-6, 2026

types or opacity levels. A sigmoid output is thresholded at 0.5. Table 1 reports the retained pooled counts, where a positive prediction denotes a defect.

Table 1. Recorded pixel counts used for all recalculations.

Run	True negative	False positive	False negative	True positive
Baseline	47,520	891,471	33,390	1,485,219
3×3 grid	215,754	729,495	369,982	1,142,369
Gaussian	522	944,727	162	1,512,189

Each row sums to 2,457,600 pixels, equal to $150 \times 128 \times 128$. The baseline contains 1,518,609 positive labels, whereas both weighted records contain 1,512,351. The difference is 6,258 pixels, or 0.255% of the evaluation total. Equal image counts therefore do not establish identical processed masks. Matching totals for the two weighted runs also do not prove identical image and mask identities.

The retained baseline script resizes categorical masks using the image library's default interpolation before binarization; the weighted configuration records specify nearest-neighbour resizing. These procedures can produce different binary boundaries. They provide a plausible reason to investigate the label discrepancy, but the aggregate counts do not establish its cause. A checkpoint-linked list of processed image and mask hashes is needed to resolve it.

There are further configuration uncertainties. The published baseline describes Dice training, while a retained script combines binary cross-entropy with Dice. That script also supplies masks as the label argument of a single image augmentation generator, without an explicit synchronized mask transformation. This is an unresolved alignment concern for geometric augmentation. The surviving records do not establish which script version generated every checkpoint. Accordingly, the analysis does not attribute differences between runs solely to spatial weighting.

3.2 Spatial maps and recorded loss

The two weighted configurations use a U-Net family model and the same stated mixed-loss form. Let $y_i \in \{0,1\}$ be the target, p_i the predicted defect probability, w_i a positive spatial weight, and N the number of pixels over which the loss is reduced. Binary cross-entropy at a pixel is $\ell_i = -y_i \log p_i - (1 - y_i) \log(1 - p_i)$. The recorded mixing coefficient is $\lambda = 0.5$:

$$L = \frac{\lambda}{N} \sum_{i=1}^N w_i \ell_i + (1 - \lambda)(1 - D_w) \quad (1)$$

$$D_w = \frac{2 \sum_i w_i y_i p_i + \varepsilon}{\sum_i w_i y_i + \sum_i w_i p_i + \varepsilon} \quad (2)$$

The smoothing constant is $\varepsilon = 10^{-7}$. The grid assigns weights $[1, 1.5, 1]$ to the top row, $[1.5, 3, 1.5]$ to the middle row, and $[1, 2, 1]$ to the bottom row. The Gaussian map is $w(x,y) = 1 + 2 \exp\{-(x - 64)^2/(2 \times 32^2) + (y - 74.24)^2/(2 \times 24^2)\}$, for integer coordinates from 0 to 127. Its centre lies slightly below the image midpoint. Direct calculation gives a mean weight of 1.554279 and a maximum of 2.999900. Figure 1 depicts these training priorities; it does not show predicted contamination.

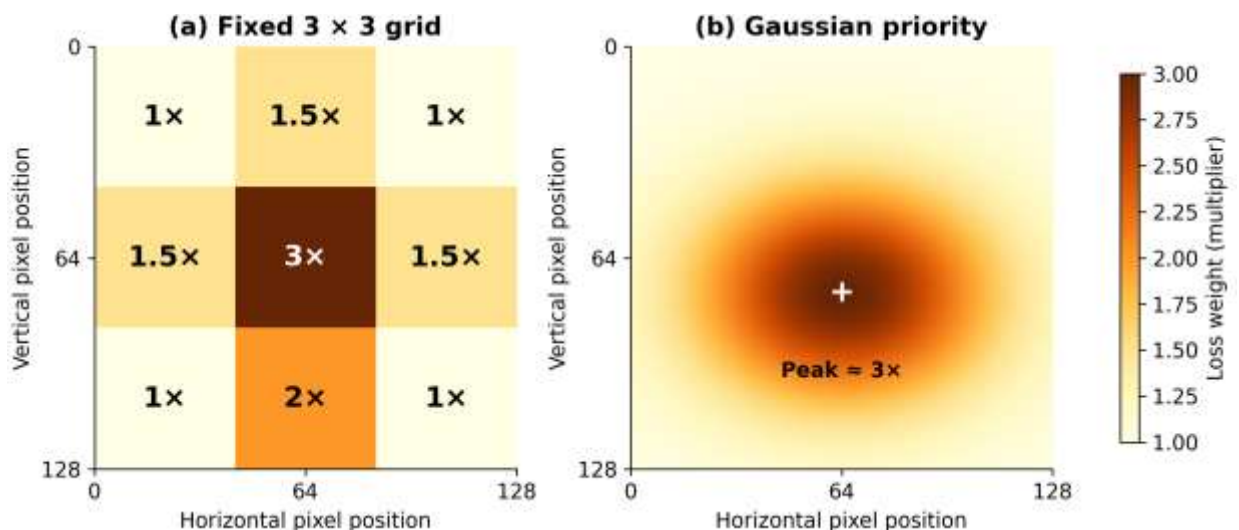


Figure 1. Spatial loss multipliers. Larger values increase the contribution of a location to the loss. The maps encode fixed image priorities, not measured driving risk.

3.3 Recalculated measures and a null reference

All evaluation measures are unweighted and calculated from the pooled counts in Table 1. Recall is $TP/(TP + FN)$, specificity is $TN/(TN + FP)$, and balanced accuracy is their arithmetic mean. Foreground IoU is $TP/(TP + FP + FN)$, and Dice is $2TP/(2TP + FP + FN)$. MCC is $(TP \cdot TN - FP \cdot FN)$ divided by the square root of $(TP + FP)(TP + FN)(TN + FP)(TN + FN)$. These pooled values are not averages of per-image scores.

A useful reference predicts every pixel as defective. If $\pi = (TP + FN)/N$ is the observed label prevalence, this predictor has recall 1, specificity 0, balanced accuracy 0.5, and the following foreground scores:

$$IoU_{all} = \pi, \quad Dice_{all} = \frac{2\pi}{1 + \pi} \quad (3)$$

The reference is calculated separately for each record because their label totals differ. Its MCC is undefined owing to a zero denominator. It is an evaluation check, not a candidate deployment method. A model close to this reference may receive a substantial foreground overlap score while contributing little useful separation of clean and contaminated pixels.

4 Results

4.1 High recall conceals rejection of clean pixels

Table 2. Metrics recalculated from pooled counts. Higher is better for all listed measures; MCC can be negative.

Measure	Baseline	3 × 3 grid	Gaussian
Precision	0.624911	0.610284	0.615483
Recall	0.978013	0.755360	0.999893
Specificity	0.050608	0.228251	0.000552

"KELAJAK TEKNOLOGIYALARI VA SUN'IY INTELLEKT"
nomli respublika ilmiy-amaliy masofaviy konferensiyasi
VOLUME-1, ISSUE-6, 2026

Foreground IoU	0.616253	0.509566	0.615442
Foreground Dice	0.762570	0.675116	0.761949
Balanced accuracy	0.514310	0.491805	0.500223
MCC	0.077936	-0.018714	0.012982

The baseline detects 97.801% of labelled defect pixels, but preserves only 5.061% of clean pixels. The Gaussian record is more extreme: recall reaches 99.989%, while specificity is 0.055%. Of 945,249 clean pixels in that record, 944,727 are classified as defective. Only 684 pixels in the entire evaluation are predicted clean: 522 true negatives and 162 false negatives.

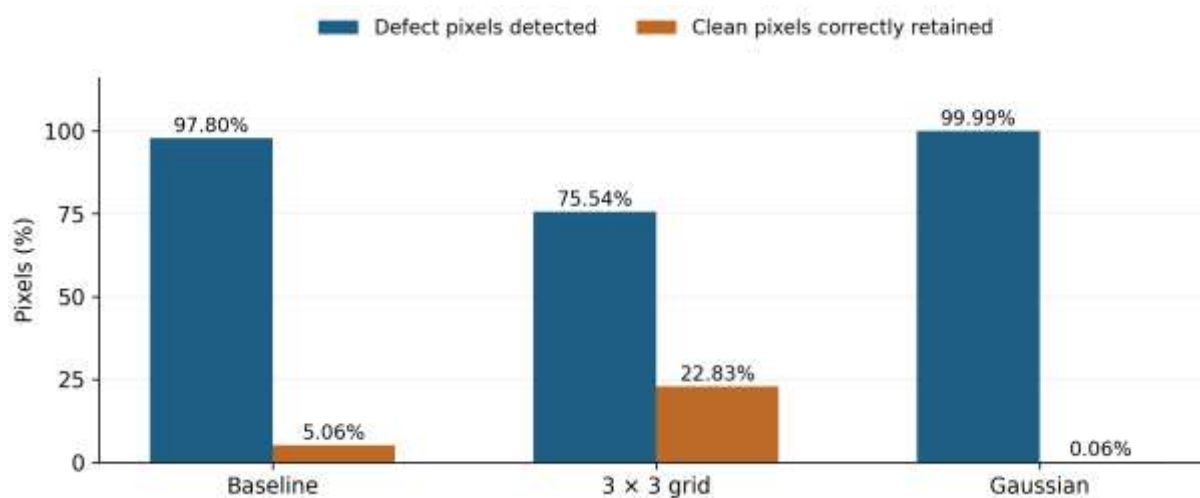


Figure 2. Defect detection and clean-pixel preservation at threshold 0.5. The Gaussian run's almost complete defect recall accompanies almost complete rejection of clean pixels.

The grid record preserves a larger fraction of clean pixels, 22.825%, but its defect recall is 75.536%. Its balanced accuracy is 0.491805 and MCC is -0.018714 . This describes weak pooled discrimination at the recorded threshold. Without matched evaluation inputs and repeated training runs, it is not evidence that the grid weighting itself caused the lower scores.

4.2 Predicted area and the all-defect comparison

The predicted foreground proportions are 96.708% for the baseline, 76.166% for the grid, and 99.972% for the Gaussian run. Each exceeds its corresponding labelled foreground proportion. Figure 3 makes this excess visible. It supports a conclusion of oversegmentation at the pooled pixel level, but cannot show where the excess occurs or whether individual masks are connected, fragmented, or displaced.

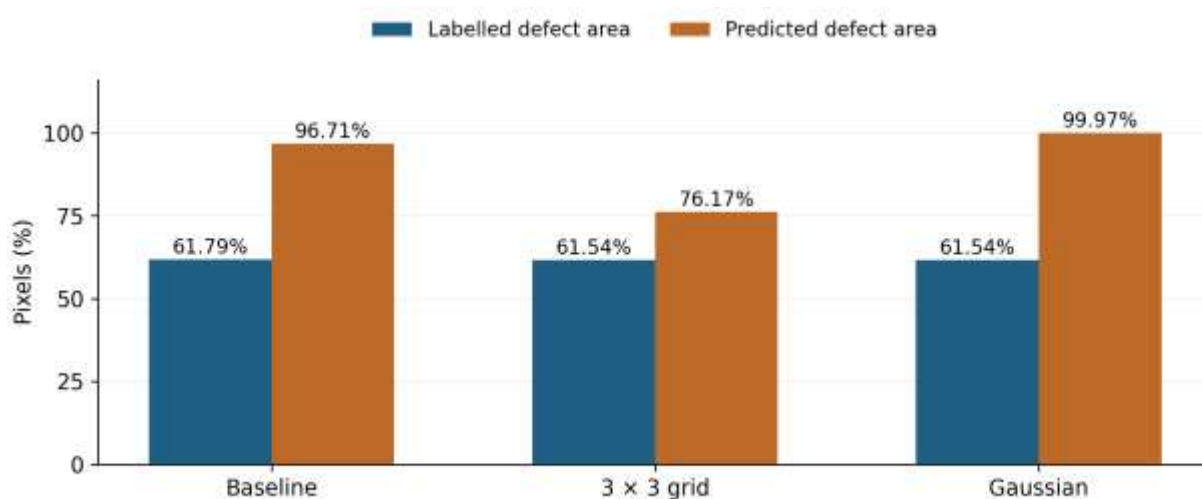


Figure 3. Labelled and predicted defect coverage. Bars summarize pooled pixels in each record; they do not imply that the three runs used identical processed masks.

Table 3. Dice relative to predicting every pixel as defective. pp denotes percentage points.

Run	Observed Dice	All-defect Dice	Difference (pp)
Baseline	0.762570	0.763848	-0.12777
3 × 3 grid	0.675116	0.761899	-8.67830
Gaussian	0.761949	0.761899	+0.00497

The Gaussian Dice score exceeds its all-defect reference by 0.00497 percentage points. Its predictions recover 522 clean pixels while losing 162 defect pixels, giving



"KELAJAK TEXNOLOGIYALARI VA SUN'IY INTELLEKT"
nomli respublika ilmiy-amaliy masofaviy konferensiyasi
VOLUME-1, ISSUE-6, 2026

only 360 additional correct classifications compared with the all-defect rule. The baseline Dice score is 0.12777 percentage points below its own reference. The grid score is 8.67830 percentage points below its reference. These comparisons expose how little the reported foreground scores establish about localization in these records.

This does not mean that an all-defect mask is preferable for a vehicle. It would discard the complete camera view. The comparison instead shows why a Dice value near 0.76, interpreted without label prevalence and specificity, can give an unduly favourable impression. Balanced accuracy near 0.5 and MCC near zero reinforce the same diagnosis using different summaries of the counts.

5 Interpretation and implications for vehicle cameras

5.1 Weight geometry is not the only changed quantity

In Equation (1), weighted cross-entropy is divided by N rather than by the sum of weights. Multiplying every weight by a constant therefore multiplies the cross-entropy term by that constant. The weighted Dice ratio is unchanged when ε is zero and approximately unchanged when the smoothing term is negligible. Consequently, increasing the mean weight changes the balance of the two loss components even when their nominal coefficients remain 0.5 and 0.5.

For the recorded Gaussian map, the mean multiplier is about 1.554. This does not imply a 55.4% increase in the realized cross-entropy loss, because large errors may occupy particular locations. It means that the weighting changes both spatial emphasis and overall scaling. A controlled test should compare the original map with a unit-mean version, or divide weighted cross-entropy by the sum of weights. Such normalization would isolate one confound; no improved segmentation result from that change is claimed here.

5.2 Why foreground overlap needs a background check



"KELAJAK TEKNOLOGIYALARI VA SUN'IY INTELLEKT"
nomli respublika ilmiy-amaliy masofaviy konferensiyasi
VOLUME-1, ISSUE-6, 2026

A restricted constant-output example helps explain the attraction of high foreground probabilities. Suppose every pixel has the same probability c , and let π_w be the weighted foreground proportion. Ignoring the negligible smoothing term for a nonempty foreground, Equation (2) becomes

$$D_w(c) = \frac{2c\pi_w}{c + \pi_w}, \quad \pi_w = \frac{\sum_i w_i y_i}{\sum_i w_i} \quad (4)$$

For positive π_w , this expression increases with c . Thus, within the constant-output family, foreground Dice alone rewards increasing the shared probability. This observation is not a proof that the trained network minimizes the combined loss by predicting all defects. Cross-entropy penalizes confident false positives and diverges as c approaches 1 when clean pixels have positive weight. The recorded behaviour may involve optimization, preprocessing, target alignment, or checkpoint selection. The present evidence does not separate those causes.

5.3 Consequences for camera availability decisions

A contamination detector may feed a cleaning request, a regional validity mask, or a decision to limit use of a vehicle camera. Excess false positives could remove clear image regions or produce unnecessary cleaning requests. Excess false negatives could leave contaminated regions in use. These are plausible consequences of the error types, not measured outcomes of this study: no object detector, driving controller, cleaning mechanism, or road trial was evaluated.

The location of an error also matters, but its importance depends on the camera and driving task. The centre of a forward-facing view and the centre of a side fisheye view do not necessarily carry the same information. A fixed central priority map is therefore a modelling assumption. The pooled counts cannot establish improved detection in a road corridor, better visibility of nearby cars, or a gain in parking performance. Those claims require spatially resolved predictions and a defined downstream task.



6 Limitations and requirements for a controlled follow-up

The analysis uses one aggregate evaluation record per configuration. It cannot estimate variation across training seeds, recover per-image scores, assess threshold sensitivity, or inspect defect boundaries. The large number of pixels must not be mistaken for millions of independent observations: neighbouring pixels share image content and prediction errors. Confidence intervals based on independent pixel trials would not be justified.

A follow-up should first freeze the test-image list and processed masks, record their hashes, and verify synchronized image–mask transformations. Categorical masks should use a documented interpolation and class mapping. The baseline and weighted models should then share architecture, data split, training budget, and checkpoint selection. Comparing repeated seeds on the same images would permit paired analysis at image or sequence level. Grouping related frames and camera views during splitting would also reduce the risk of information leaking between development and evaluation sets.

The next comparison should report both foreground and background behaviour, including per-image Dice, specificity, predicted area, and performance within predefined image regions. Weight normalization should be a separate experimental factor. Thresholds should be selected using validation data before test evaluation. Only after these checks would it be meaningful to ask whether a priority map improves a particular vehicle perception or camera-maintenance task.

7 Conclusion

The recorded runs demonstrate a failure of interpretation more clearly than a failure of spatial weighting as a general method. Near-perfect recall in the Gaussian run accompanies a mask that marks 99.972% of pixels as defective. Its Dice score is almost identical to an all-defect reference, while the baseline score falls slightly below its



"KELAJAK TEKNOLOGIYALARI VA SUN'IY INTELLEKT"
nomli respublika ilmiy-amaliy masofaviy konferensiyasi
VOLUME-1, ISSUE-6, 2026

corresponding reference. Neither result provides convincing evidence of useful clean–defect separation. These findings support a practical requirement for vehicle camera research: report clean-pixel preservation and an all-defect reference whenever foreground overlap is used to evaluate soiling masks. They also identify two controls for future loss comparisons—identical processed labels and explicit normalization of spatial weights. The available counts justify the diagnosis of extensive oversegmentation; they do not establish its unique cause or rule out spatial weighting under a better controlled implementation.

Data and analysis statement

The baseline measurements originate from the author's published study [3]. The grid and Gaussian counts are retained records of follow-up experiments, reproduced in Table 1. All numerical results added in this article are deterministic recalculations from those counts or from the stated weight maps. The tables and equations provide the inputs needed to reproduce that analysis. No new trained model, additional camera dataset, or vehicle test result is introduced.

References

1. Yogamani, S., Hughes, C., Horgan, J., Sistu, G., Chennupati, S., Uříčář, M., Milz, S., Simon, M., Amende, K., Witt, C., Rashed, H., Nayak, S., Mansoor, S., Varley, P., Perrotton, X., O'Dea, D., & Pérez, P. (2019). WoodScape: A multi-task, multi-camera fisheye dataset for autonomous driving. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9307–9317). <https://doi.org/10.1109/ICCV.2019.00940>
2. Uříčář, M., Křížek, P., Sistu, G., & Yogamani, S. (2019). SoilingNet: Soiling detection on automotive surround-view cameras. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC) (pp. 67–72). <https://doi.org/10.1109/ITSC.2019.8917178>





"KELAJAK TEKNOLOGIYALARI VA SUN'IY INTELLEKT"
nomli respublika ilmiy-amaliy masofaviy konferensiyasi
VOLUME-1, ISSUE-6, 2026

3. Axmedov, A. G. (2025). AI-based image quality and lens defect analysis in autonomous driving: A framework with U-Net-based soiling detection. *American Journal of Mechanics and Applications*, 12(4), 93–101. <https://doi.org/10.11648/j.ajma.20251204.14>

4. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

5. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M. J. (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (pp. 240–248). Springer. https://doi.org/10.1007/978-3-319-67558-9_28

6. Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 2999–3007). <https://doi.org/10.1109/ICCV.2017.324>

7. Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Tversky loss function for image segmentation using 3D fully convolutional deep networks. In *Machine Learning in Medical Imaging* (pp. 379–387). Springer. https://doi.org/10.1007/978-3-319-67389-9_44

8. Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21, Article 6. <https://doi.org/10.1186/s12864-019-6413-7>