agriculture

MDPI

Reply

Reply to Mahler, S. Comment on "Jia et al. Exploratory Study of Sex Identification for Chicken Embryos Based on Blood Vessel Images and Deep Learning. *Agriculture* 2023, *13*, 1480"

Nan Jia [1], Bin Li [1,*], Yuliang Zhao [1], Shijie Fan [2], Jun Zhu [1], Haifeng Wang [1] and Wenwen Zhao [1]

[1] Intelligent Equipment Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; jianan@nercita.org.cn (N.J.); zhaoyl@nercita.org.cn (Y.Z.); zhujun@nercita.org.cn (J.Z.); wanghf@nercita.org.cn (H.W.); zhaoww@nercita.org.cn (W.Z.)
[2] Beijing Huadu Yukou Poultry Co., Ltd., Beijing 101206, China; fansj23@126.com
* Correspondence: lib@nercita.org.cn

Abstract

We respond to the Comment regarding the reproducibility and generalizability of in ovo sex identification from extraembryonic vessel images. Our 2023 exploratory study evaluated Jingfen No. 6 white-shell eggs using green-light transmission images and a YOLO-based object detector. Mahler et al. evaluated LSCI images from a different egg population using whole-image classifiers and reported per-egg validation accuracies below 60%. To assess what could still be verified, we reconstructed the detection workflow from retained transmission images using a new implementation based on the published methods. The reconstruction yielded validation-set mAP@0.5 values of 84.90%, 91.74%, and 80.56% for days 3, 4, and 5, respectively. These are internal validation estimates because the same partitions were used for checkpoint selection and performance reporting. Substantial differences between the studies in imaging modality, egg population, prediction task, reference labeling, and evaluation procedures preclude direct comparison of their results. The reconstructed results are specific to the retained transmission-image dataset and do not establish external validity or commercial readiness.

Keywords: chicken embryo; in ovo sexing; extraembryonic vasculature; green-light transmission imaging; reproducibility; generalizability

1. Scope of the Original Study

We welcome the study by Mahler et al. and agree that early, non-invasive sex identification requires robust external validation before commercial use [1–3]. Our 2023 exploratory study evaluated Jingfen No. 6 white-shell eggs imaged with a defined green-light transmission system at a single hatchery; the test eggs came from the same acquisition domain as the training eggs [4]. It did not assess generalization across genetic lines, shell colors, devices, sites, or imaging modalities. The original study should therefore be regarded as a proof of concept under the reported acquisition conditions, not as evidence of external validity. We acknowledge that the original discussion of practical potential did not make this limitation sufficiently clear.

Mahler et al. reported validation accuracies below 60% for several models applied to their LSCI dataset [2]. Assessing the implications of these results for our original findings requires consideration of differences in image formation, biological interpretation, study population, reference labeling, and model development and evaluation procedures. The

Check for updates

Academic Editor: Les Copeland

Received: 15 August 2026
Revised: 15 October 2026
Accepted: 22 October 2026
Published: 2 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

following sections examine these differences and their implications for interpreting the reported results.

2. Green-Light Transmission Imaging and LSCI

Green-light transmission imaging and LSCI can both reveal extraembryonic vessels, but their image contrast arises from different physical processes. In green-light transmission images, intensity is governed by absorption and scattering through the eggshell and internal structures. LSCI, by contrast, derives contrast from motion-induced changes in coherent speckle patterns and is sensitive to blood-flow dynamics [2,4–7]. The visibility of vessels in both modalities therefore does not mean that they provide equivalent inputs for sex prediction.

The studies also differed in model inputs and prediction tasks. Our study used red–green–blue (RGB) transmission images with bounding-box annotations to train an object detector [4], whereas Mahler et al. used monochrome LSCI-derived images to train whole-image binary classifiers [2].

The Comment also notes that the green channel may contain most of the visible vessel information in our RGB images [1]. This observation concerns channel-specific contrast within the transmission images and does not alter the physical distinction between transmission imaging and LSCI.

The evaluation endpoints also differed. Mahler et al. reported per-egg classification accuracy, whereas our original study reported object-detection mAP [2,4]; the specific day-4 values are discussed in Section 4.3. The egg-level accuracy estimates from the reconstruction permit a descriptive comparison but not a matched quantitative assessment. Accordingly, the present cross-study comparison does not isolate the effect of imaging modality.

3. Gonadal Differentiation and Early Image Associations

The Comment notes that gonadal differentiation occurs over approximately embryonic days 4.5 to 6.5, with several established features of female differentiation becoming evident around days 6 to 6.5, after the day-5 endpoint of our study [1,8]. The association observed in images from days 3 to 5 therefore cannot be attributed to established gonadal differentiation. Because the study was designed to test for sex-associated information in transmission images rather than to investigate mechanism, the result should be regarded as a predictive association within the dataset, not as evidence of sex-specific vascular development.

Sex-related molecular differences have been reported before morphological gonadal differentiation. Chickens have a ZZ/ZW sex-determination system, and cell-autonomous sexual identity and sexually dimorphic gene expression have been observed as early as the blastoderm stage [9,10]. These findings provide biological context for investigating early sex-associated signals, but they do not link such signals to extraembryonic vascular morphology or transmission-image appearance during days 3 to 5.

4. Egg Population, Reference Labeling, and Evaluation

4.1. Egg Population and Reference Labeling

Our study used Jingfen No. 6 white-shell eggs, whereas Mahler et al. used brown-shell eggs [2,4]. This difference is relevant because shell pigmentation and optical transmittance can affect the appearance of transmission images [5,6], not because the studies imply different patterns of embryonic sex development. Because egg population and imaging modality varied simultaneously, their individual effects cannot be separated in a cross-study comparison.

Reference labels were obtained using different procedures. In our study, sex was determined at hatch by three experienced hatchery workers using the line-specific feather-color characteristics of Jingfen No. 6, following the routine commercial sexing procedure at the cooperating hatchery [4]. Each worker assessed a separate group of chicks and was unaware of the image-based predictions. The research team recorded the final labels. The procedure did not involve three independent assessments of each chick, and no repeat-assessment or inter-observer agreement records were collected. Mahler et al. used gonadal examination combined with polymerase chain reaction (PCR) [2]. Molecular confirmation was not part of the routine hatchery workflow in which our dataset was collected. Because the labeling procedures were not evaluated under otherwise matched conditions, their contribution to the difference in reported performance cannot be isolated.

4.2. Prior Work on Jing-Series Eggs

Previous studies examined sex-associated optical features in Jingfen No. 1 eggs using transmission images, transmission spectra, combined image and spectral features, and hyperspectral measurements [11–15]. These studies provided methodological background for evaluating whether sex-associated optical information could also be detected in the Jingfen No. 6 population.

4.3. Dataset Size, Model Architecture, and Evaluation Metrics

The Comment notes that differences in dataset size and model architecture alone are unlikely to fully explain the difference in reported performance [1]. Because neither factor was evaluated under otherwise matched conditions, the present studies cannot quantify the contribution of either factor.

The original object detector was evaluated using average precision (AP) and mean average precision (mAP). The 88.79% day-4 value cited in the Comment was mAP, not per-egg classification accuracy [1,4]. The original article also reported a mean day-4 accuracy (Acc) of 89.25% in Table 7 [4]. Because the aggregation unit for Acc was not specified, this value should be interpreted descriptively. In the reconstruction, mAP@0.5 is the mean average precision across the two sex classes at a bounding-box intersection-over-union (IoU) threshold of 0.5, whereas egg-level accuracy is the proportion of evaluated eggs assigned the correct sex label.

For the retrospective egg-level analysis, each image contributed the class of its highest-confidence detection after confidence filtering at 0.25 and class-aware non-maximum suppression (NMS) at an IoU threshold of 0.45. Egg-level labels were obtained by majority voting, with tied votes resolved by summed confidences. Supplementary Methods S3 specifies the rules for multiple detections, missing detections and unresolved ties, and Table S1 gives the class counts and confusion matrices. Re-evaluation of the existing checkpoints at these thresholds yielded validation-set accuracies of 89.57%, 88.26%, and 87.98% for days 3, 4, and 5, respectively. Eggs without a resolved prediction were counted as incorrect. Because Mahler et al. also reported egg-level results, the values may be contrasted descriptively; however, the design differences described above preclude a matched quantitative comparison or attribution of the numerical difference to a single factor.

4.4. Egg-Level Partitioning and the Role of the Evaluation Set

The original article stated that 2000 eggs were used for training and that the remaining 854 eggs formed the test set [4]. However, Table 2 lists 418 female and 426 male test eggs, totaling 844 eggs, consistent with 2844 total eggs minus 2000 training eggs. Thus, 854 was a numerical error, and the correct total was 844 eggs. Because the original split manifest and

run records are unavailable, the original partition and the role of the reported test set in model selection cannot now be verified.

For the reconstruction, a new egg-level partition was created from the retained images, with 2000 eggs assigned to training and 844 eggs assigned to validation. This reconstructed partition followed the corrected sample counts derived from the original article but did not reproduce the original test set. All images from the same egg were assigned to the same partition. The day-specific validation partitions contained 844, 843, and 840 eggs for days 3, 4, and 5, respectively, reflecting the retained images available for each day. No egg identifier or filename was shared between the reconstructed training and validation partitions, and no duplicate image hash was detected across them. Checkpoints were selected by the maximum validation fitness, calculated as $0.1 \times \text{mAP@0.5} + 0.9 \times \text{mAP@0.5:0.95}$. Checkpoint-selection details are provided in Supplementary Method S2. The same validation data had also been used in earlier exploratory comparisons of accuracy definitions and bounding-box matching thresholds. Because the validation partitions were used both for checkpoint selection and performance reporting, the reconstructed metrics should be interpreted as internal validation estimates rather than independent test-set estimates.

5. Data and Code Availability

We sincerely apologize for any inconvenience related to previous requests for the original data, source code, and additional methodological information. The original 2023 source code, trained weights, split manifests, and contemporaneous run records are no longer available, and the complete original preprocessing and training workflow cannot now be recovered. An exact rerun of the 2023 analysis is therefore not possible. The current reconstruction used the retained images and the methods reported in the original article. The models were trained using PyTorch 2.4.1 + cu124 with CUDA 12.4 and cuDNN 9.1.0. The reconstruction settings, checkpoint-selection criterion and egg-level aggregation procedure are described in Supplementary Methods S1–S3. The retained reconstruction code is not a complete training-time snapshot because some files were modified after training and their earlier versions were not retained. The accompanying code, retained model weights and result data have been submitted to ScienceDB as two records and are currently under repository review. The submitted code is a new implementation, not the training-time source of the retained model weights. Access to the image dataset remains restricted, as described in the Data Availability Statement.

6. Conclusions

The Comment identifies important limitations in the reproducibility and generalizability of early in ovo sex identification [1]. Our 2023 study reported internal object-detection performance for Jingfen No. 6 white-shell eggs under a defined green-light transmission protocol and did not assess external generalization [4]. The reconstructed workflow yielded validation-set mAP@0.5 values of 84.90%, 91.74%, and 80.56% for days 3, 4, and 5, respectively, and egg-level validation-set accuracies ranging from 87.98% to 89.57% across days 3–5. These values are internal validation estimates obtained using the same partitions for checkpoint selection and performance reporting. The egg-level accuracies may be contrasted descriptively with the per-egg validation accuracies below 60% reported by Mahler et al. [2]; however, the studies do not constitute a matched comparison because acquisition, reference-labeling, prediction, and evaluation conditions were not harmonized. The available evidence is therefore limited to the studied transmission-image dataset and does not establish external validity or commercial readiness. We consider the differences between the two studies scientifically interesting and believe that they may inform future research. We welcome continued interest in and constructive discussion of this topic.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agriculture16212396/s1. Supplementary Methods: S1. Reconstruction settings; S2. Checkpoint selection and evaluation; S3. Egg-level aggregation; Table S1. Egg-level confusion counts and accuracy for internal validation.

Author Contributions: Conceptualization, N.J. and B.L.; writing—original draft preparation, N.J.; writing—review and editing, N.J., B.L., Y.Z., S.F., J.Z., H.W. and W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: No additional external funding was received specifically for the preparation of this Reply. Funding for the original study is reported in reference [4].

Data Availability Statement: The accompanying code and retained model weights, together with the result data and supporting records for the 2026 reconstruction and re-evaluation, have been submitted to ScienceDB as two separate records: code and model weights (https://doi.org/10.57760/sciencedb.015k1) and result data (https://doi.org/10.57760/sciencedb.015n1). Both records are currently under repository review. The retained image dataset is not publicly available because confidentiality agreements for ongoing related projects restrict its release.

Conflicts of Interest: One author is employed by Beijing Huadu Yukou Poultry Co., Ltd., which participated in data collection for the original study. The authors declare no other conflicts of interest.

References

1. Mahler, S. Comment on Jia et al. Exploratory Study of Sex Identification for Chicken Embryos Based on Blood Vessel Images and Deep Learning. *Agriculture* 2023, *13*, 1480. *Agriculture* **2026**, *16*, 2395. [CrossRef]
2. Mahler, S.; Arora, A.; Readhead, C.; Yin, S.; Hari, S.N.; Wang, E.; Moxley, C.I.; Adeboye, A.A.; Dong, Z.; Zhou, H.; et al. Detection of Non-Invasive Sexing of Early Chick Embryos in Intact Eggs Using Laser Speckle Contrast Imaging and Deep Neural Networks. *PLoS ONE* **2026**, *21*, e0323847. [CrossRef] [PubMed]
3. Ho, S.Y.; Phua, K.; Wong, L.; Goh, W.W.B. Extensions of the External Validation for Checking Learned Model Interpretability and Generalizability. *Patterns* **2020**, *1*, 100129. [CrossRef] [PubMed]
4. Jia, N.; Li, B.; Zhao, Y.; Fan, S.; Zhu, J.; Wang, H.; Zhao, W. Exploratory Study of Sex Identification for Chicken Embryos Based on Blood Vessel Images and Deep Learning. *Agriculture* **2023**, *13*, 1480. [CrossRef]
5. Dong, Z.; Mahler, S.; Readhead, C.; Chen, X.; Dickson, M.; Bronner, M.; Yang, C. Non-Invasive Laser Speckle Contrast Imaging (LSCI) of Extra-Embryonic Blood Vessels in Intact Avian Eggs at Early Developmental Stages. *Biomed. Opt. Express* **2024**, *15*, 4605–4624. [CrossRef] [PubMed]
6. Ostertag, E.; Scholz, M.; Klein, J.; Rebner, K.; Oelkrug, D. Pigmentation of White, Brown, and Green Chicken Eggshells Analyzed by Reflectance, Transmittance, and Fluorescence Spectroscopy. *ChemistryOpen* **2019**, *8*, 1084–1093. [CrossRef] [PubMed]
7. Boas, D.A.; Dunn, A.K. Laser Speckle Contrast Imaging in Biomedical Optics. *J. Biomed. Opt.* **2010**, *15*, 011109. [CrossRef] [PubMed]
8. Luo, X.; Guo, J.; Zhang, J.; Ma, Z.; Li, H. Overview of Chicken Embryo Genes Related to Sex Differentiation. *PeerJ* **2024**, *12*, e17072. [CrossRef] [PubMed]
9. Zhao, D.; McBride, D.; Nandi, S.; McQueen, H.A.; McGrew, M.J.; Hocking, P.M.; Lewis, P.D.; Sang, H.M.; Clinton, M. Somatic Sex Identity Is Cell Autonomous in the Chicken. *Nature* **2010**, *464*, 237–242. [CrossRef] [PubMed]
10. Ayers, K.L.; Davidson, N.M.; Demiyah, D.; Roeszler, K.N.; Grützner, F.; Sinclair, A.H.; Oshlack, A.; Smith, C.A. RNA Sequencing Reveals Sexually Dimorphic Gene Expression before Gonadal Differentiation in Chicken and Allows Comprehensive Annotation of the W-Chromosome. *Genome Biol.* **2013**, *14*, R26. [CrossRef] [PubMed]
11. Tang, Y.; Hong, Q.; Wang, Q.; Zhu, Z. Sex Identification of Chicken Eggs Based on Blood Line Texture Features and GA-BP Neural Network. *J. Huazhong Agric. Univ.* **2018**, *37*, 130–135. (In Chinese)
12. Zhu, Z.; Tang, Y.; Hong, Q.; Huang, P.; Wang, Q.; Ma, M. Female and Male Identification of Early Chicken Embryo Based on Blood Line Features of Hatching Egg Images and Deep Belief Networks. *Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 197–203. (In Chinese) [CrossRef]
13. Zhu, Z.; Hong, Q.; Wu, L.; Wang, Q.; Ma, M. Early Identification of Male and Female Embryos Based on UV/Vis Transmission Spectroscopy and Extreme Learning Machine. *Spectrosc. Spectr. Anal.* **2019**, *39*, 2780–2787. (In Chinese)

14. Zhu, Z.; Ye, Z.; He, Y.; Yang, K.; Wang, Q.; Ma, M. Gender Identification of Early Chicken Embryo Based on RF-DS Information Fusion of Spectroscopy and Machine Vision. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 308–315. (In Chinese) [CrossRef]
15. Yan, Q.; Tan, Z.; Dong, J.; Wang, Y.; Wang, Q. Non-Destructive Sex Identification of Hatching Eggs during Incubation Based on Improved ResNet-18 and Hyperspectral Features. *Trans. Chin. Soc. Agric. Mach.* **2026**, *57*, 386–394, 426. (In Chinese) [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.