

Video-Based Criterion Validation of a Proprietary Collar-Mounted IMU Step Counter in Dogs

Marlon Malheiros^{1,*}, Gustavo Pegado^{1,*}, Charles Ferreira¹, Bianca Cifali², Dorian Arnaud A. Dack², Guilherme Alexandre Sant'Ana², Ronaldo Zampolo¹, and Antonio Pereira¹

Abstract—Collar-mounted inertial measurement units (IMUs) offer practical canine activity monitoring, although neck signals only indirectly represent paw-ground contact. We compared a proprietary collar IMU step counter with manual video annotation in 15 walking and trotting/running bouts from 11 dogs. The video reference contained 357 steps and the IMU output 355. Mean absolute error was 1.60 steps/bout, MAPE was 7.96%, WAPE was 6.72%, and Bland–Altman limits were -4.87 to 4.60 steps/bout. Dog-cluster bootstrapping addressed repeated observations. The collar-system and video-reference counts showed close bout-level agreement under the walking and trotting/running conditions examined.

Keywords—canine gait, collar sensor, dogs, inertial measurement unit, pedometry

I. INTRODUCTION

Objective measurement of canine activity can support welfare monitoring, rehabilitation, exercise assessment, and veterinary research. Changes in activity may provide information about mobility, discomfort, health status, and recovery after intervention [1]. Wearable sensors are attractive in this setting because they can be deployed outside instrumented laboratories and provide repeated measurements under owner-managed conditions.

Collar-mounted sensors are practical because they are familiar to owners and impose little additional handling burden. However, collar placement introduces a biomechanical limitation. A collar sensor does not directly measure paw-ground contact; it records motion transmitted through the trunk, neck, collar, and local soft tissues. The signal may therefore vary with head movement, sniffing, leash tension, turning, gait transitions, collar fit, body size, and breed morphology.

Wearable inertial measurement unit (IMU) systems have been used in dogs to characterize gait and activity. Limb-mounted systems record motion mechanically closer to paw contact and gait-phase events and have been evaluated against manually annotated video and pressure-sensor walkway measurements [2], [3]. Collar-mounted sensors pose a different measurement problem because locomotion is observed indirectly through motion transmitted to the trunk and neck. Their

step-count output therefore requires direct validation against an external criterion reference.

Previous studies have evaluated canine activity monitors and collar-mounted accelerometry for locomotor and behavioral monitoring [1], [4], [5]. We therefore tested whether the bout-level output of a proprietary collar-mounted IMU step counter agrees with counts obtained through frame-by-frame video annotation during walking and trotting/running.

We treated the collar system as a black-box instrument and compared its final step-count output with the video reference. Agreement was characterized using absolute and relative error measures, Bland–Altman analysis, concordance coefficients, dog-cluster bootstrap intervals, and sensitivity analyses.

II. METHODS

Figure 1 summarizes the validation workflow. Manual paw-contact annotation produced a bout-level video-reference count, V_i , while the collar system produced the corresponding IMU-derived count, I_i . Agreement between V_i and I_i was evaluated using the procedures described below.

A. Study Design and Evaluated System

The evaluated system was a proprietary collar-mounted IMU step counter. The analyzed collar recordings provided triaxial acceleration sampled at 50 Hz. The device was evaluated as a fixed black-box system by comparing its bout-level step-count output with the corresponding video-reference count; its internal step-counting algorithm was not analyzed.

The validation set consisted of 15 reviewed locomotor bouts from 11 dogs. Some dogs contributed more than one bout. Bouts were included when walking or trotting/running was visible in the video, successive paw-ground contacts of the selected forelimb were clearly visible throughout the analyzed interval, and the recording could be matched to the corresponding collar-system output. The unit of analysis was the locomotor bout. Agreement was assessed between matched video-reference and IMU-derived counts obtained from each reviewed walking or trotting/running segment.

B. Video Reference and Operational Step Definition

Manual video annotation was used as the operational criterion reference. Videos were reviewed using BORIS, an open-source platform for frame-based behavioral annotation [6]. Recordings were obtained at approximately 30 frames/s using high-definition mobile cameras positioned to provide a visible view of locomotion.

¹Marlon Malheiros, Gustavo Pegado, Charles Ferreira, Ronaldo Zampolo, and Antonio Pereira are with the Signal Processing Laboratory, Institute of Technology, Federal University of Pará, 66075-110 Belém, Brazil. E-mails: marlon.malheiros1@gmail.com, gustavo.pegado@itec.ufpa.br, charleslf.ceng@ieee.org, zampolo@ufpa.br, apereira@ufpa.br. ²Bianca Cifali, Dorian Arnaud A. Dack, and Guilherme Alexandre Sant'Ana are with ZOOMIES S.A., 01310-925 São Paulo, Brazil. E-mails: biancacifali@zoomies.com.br, dorian.dack@zoomies.com.br, guilherme.santana@zoomies.com.br. ^{*}Marlon Malheiros and Gustavo Pegado contributed equally to this work.

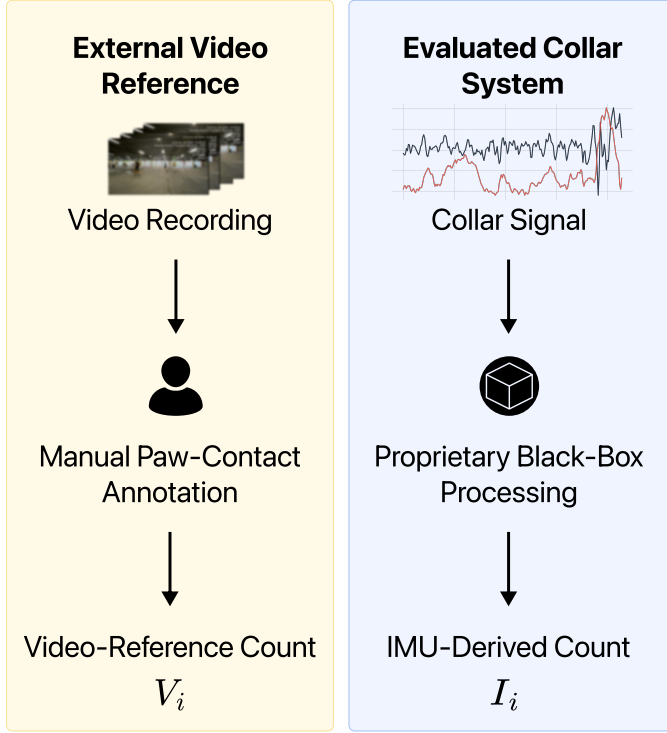


Fig. 1. Overview of the validation workflow. An external video reference was manually annotated to obtain the video-reference count V_i , while the evaluated collar system processed the corresponding IMU signal to produce the IMU-derived count I_i .

For each bout, the annotator selected the forelimb that was most clearly visible; the tracked side could therefore differ between bouts. A point event was marked at each visible paw-ground contact of that forelimb. A video-reference step was defined as the interval between two consecutive point events. Thus, M valid contact events yielded up to $M - 1$ step intervals, and the bout-level video-reference count V_i was the number of intervals retained for analysis.

For each reviewed bout, the video-reference and IMU-derived counts were computed over identical temporal boundaries. The corresponding IMU segment was extracted using the same start and end timestamps defined for the video-reference analysis.

C. Agreement Analysis

For bout i , let V_i be the video-reference step count and I_i the IMU-derived step count. Signed error was defined as

$$e_i = I_i - V_i,$$

so that positive values indicate IMU overcounting and negative values indicate undercounting relative to the video reference.

Agreement was characterized using complementary absolute, relative, and concordance measures. Mean signed bias quantified the average systematic difference between the two methods, whereas mean absolute error (MAE) quantified the average magnitude of the bout-level error in steps:

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N e_i, \quad \text{MAE} = \frac{1}{N} \sum_{i=1}^N |e_i|.$$

Mean absolute percentage error (MAPE) expressed the absolute error in each bout relative to its video-reference count:

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \frac{|e_i|}{V_i}.$$

Weighted absolute percentage error (WAPE) expressed the total absolute error relative to the total number of video-reference steps:

$$\text{WAPE} = 100 \frac{\sum_{i=1}^N |e_i|}{\sum_{i=1}^N V_i}.$$

MAPE gives equal weight to the percentage error of each bout, whereas WAPE weights bouts according to their contribution to the total video-reference count.

Bland–Altman analysis was used to assess agreement across the observed range of bout counts [7]. For each bout, the signed difference e_i was plotted against the mean of the two measurements,

$$m_i = \frac{V_i + I_i}{2}.$$

The mean signed difference estimates systematic bias between the two methods. The 95% limits of agreement were estimated as

$$\text{Bias} \pm 1.96 \text{SD}(e_i),$$

and describe the range within which approximately 95% of individual bout-level differences are expected to lie, assuming the differences are approximately normally distributed. The plot was also inspected for changes in disagreement across bout sizes.

Lin's concordance correlation coefficient and an absolute-agreement intraclass correlation coefficient were calculated as complementary summaries of concordance [8], [9]. Primary interpretation was based on the signed and absolute errors and the Bland–Altman analysis rather than on concordance coefficients alone.

D. Clustered Uncertainty and Sensitivity Analysis

Because some dogs contributed multiple bouts, 95% confidence intervals were estimated using dog-cluster bootstrap resampling. Dogs were sampled with replacement, with all bouts from each sampled dog retained together. This approach accounts for the clustering of repeated bouts within dogs.

Confidence intervals for bias, MAE, MAPE, WAPE, and the Bland–Altman limits of agreement were obtained from 10,000 bootstrap replicates using percentile intervals.

Robustness was evaluated through leave-one-bout-out and leave-one-dog-out analyses, exclusion of the longest bout, and descriptive comparison of walking and trotting/running bouts.

TABLE I

AGREEMENT BETWEEN VIDEO-REFERENCE AND COLLAR-SYSTEM
BOUT-LEVEL STEP COUNTS. CONFIDENCE INTERVALS ARE DOG-CLUSTER
BOOTSTRAP 95% INTERVALS.

Metric	Estimate	95% CI
Video-reference total	357	–
Collar-system total	355	–
Bias, steps/bout	-0.13	-1.36 to 1.13
MAE, steps/bout	1.60	0.88 to 2.64
Median absolute error, steps	1.00	–
MAPE	7.96%	4.57 to 12.23
WAPE	6.72%	3.26 to 12.57
Lower LoA, steps/bout	-4.87	-7.03 to -2.35
Upper LoA, steps/bout	4.60	1.38 to 7.43
Lin CCC	0.994	–
Absolute-agreement ICC	0.995	–

III. RESULTS

A. Primary Agreement

B. Primary Agreement

Across the 15 reviewed bouts, the video reference yielded 357 steps, whereas the collar system yielded 355 steps. Mean signed bias was -0.13 steps per bout (95% CI: -1.36 to 1.13), indicating little average over- or undercounting relative to the video reference.

Mean absolute error was 1.60 steps per bout (95% CI: 0.88 to 2.64), and the median absolute error was 1.00 step. MAPE was 7.96% (95% CI: 4.57% to 12.23%), and WAPE was 6.72% (95% CI: 3.26% to 12.57%).

The Bland–Altman limits of agreement ranged from -4.87 to 4.60 steps per bout. Bootstrap 95% intervals were -7.03 to -2.35 for the lower limit and 1.38 to 7.43 for the upper limit. Lin’s concordance correlation coefficient was 0.994, and the absolute-agreement ICC was 0.995.

Figure 2 shows the distribution of bout-level differences across the observed count range. Eleven of the 15 bouts had an absolute error of at most one step, and 14 were within five steps. Individual signed errors ranged from -5 to $+6$ steps (Fig. 3).

C. Sensitivity Analyses

Sensitivity analyses produced estimates similar to the primary results. The proportional-bias slope was 0.001 steps of error per one-step increase in mean count (95% CI: -0.063 to 0.064), providing no evidence of proportional bias across the observed bout sizes.

After excluding the longest bout, bias was -0.07 steps per bout, MAE was 1.64 steps per bout, and the limits of agreement were $[-4.96, 4.82]$ steps, all similar to the primary estimates. WAPE increased from 6.72% to 9.02%.

Descriptively, walking and trotting/running bouts yielded WAPE values of 5.13% and 9.76%, respectively. Leave-one-bout and leave-one-dog analyses yielded agreement estimates similar to those from the full dataset.

D. Representative Signal–Reference Alignment

Figure 4 illustrates the temporal relationship between the video-reference annotations and the collar acceleration signal

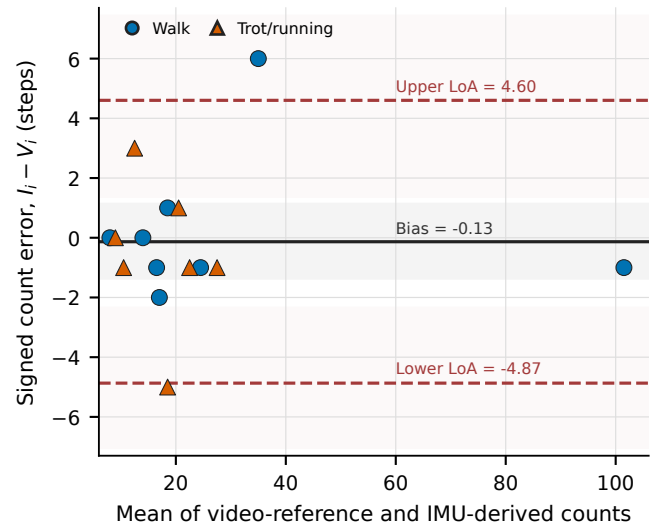


Fig. 2. Bland–Altman plot of bout-level agreement between the video-reference and IMU-derived counts. The solid black line indicates mean signed bias, and the dashed red lines indicate the limits of agreement (LoA). Gray and light-red shaded regions denote dog-cluster bootstrap 95% confidence intervals for the bias and LoA, respectively. Circles indicate walking bouts and triangles indicate trotting/running bouts. Positive values indicate IMU overcounting relative to the video reference.

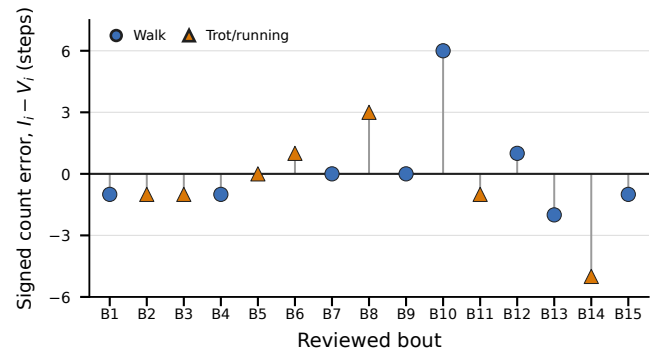


Fig. 3. Signed count error for each reviewed bout. Circles indicate walking bouts, and triangles indicate trotting/running bouts. Positive values indicate IMU overcounting, whereas negative values indicate undercounting relative to the video reference. Bout identifiers were anonymized.

in one reviewed locomotor segment. The manually annotated paw-contact events occur within the recurring signal structure associated with locomotion, providing a visual check of the synchronized video–sensor representation used in the validation.

IV. DISCUSSION

This study found close bout-level agreement between a proprietary collar-mounted IMU step counter and video-reference counts during walking and trotting/running bouts. Across 357 video-reference steps, the collar-system output differed by two steps in aggregate. Mean signed bias was close to zero, and the median absolute error was one step per bout.

The primary interpretation rests on the observed error distribution and Bland–Altman limits of agreement. Dog-cluster bootstrap intervals accounted for repeated bouts from the

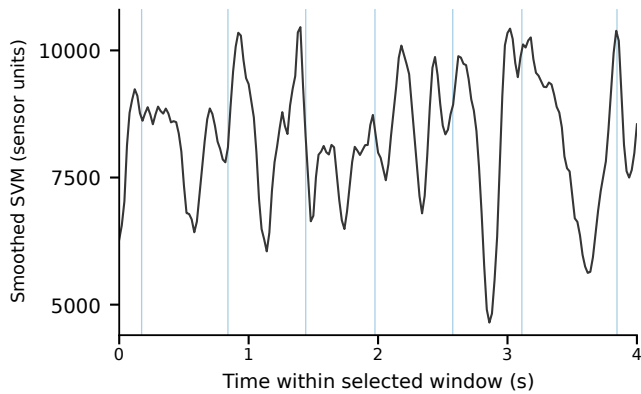


Fig. 4. Representative temporal alignment between the smoothed collar-signal vector magnitude and manually annotated video-reference paw-contact events during a reviewed locomotor segment. Vertical blue lines indicate BORIS annotations. Smoothing was applied only for visualization; no proprietary detector outputs, thresholds, or decision rules are shown.

same dogs. Lin's concordance coefficient and the absolute-agreement ICC were also high and were used as complementary measures of agreement.

These results support the accuracy of the collar-system output under the structured locomotor conditions examined. The present analysis was designed to assess short walking and trotting/running bouts; longer and less controlled activity remains an important target for subsequent validation.

A. Black-Box Validation

The study evaluated the final output of the proprietary system rather than its internal algorithm. The resulting inference is therefore at the device level: the analysis assesses whether the reported step count agrees with the external video reference.

Observed discrepancies may reflect neck-signal transmission, collar movement, gait transitions, non-step head motion, post-processing, or other device-specific factors. These mechanisms were not examined separately in the present study.

B. Future Validation

Future studies can extend the present results by including more dogs and bouts, broader variation in body size and morphology, and longer recording periods. Evaluation under a wider range of leash, substrate, gait, collar-fit, and behavioral conditions would further characterize device performance outside the structured bouts considered here.

C. Limitations

The main limitations are the number of animals and bouts and the relatively short duration of the analyzed locomotor segments. These factors limit the precision with which performance can be characterized across the broader canine population and during prolonged unrestricted activity.

The video reference was based on repeated visible contacts of a selected forelimb, following the operational step definition used throughout the study. This approach provided a consistent

external reference for the present comparison but does not capture all aspects of full-limb gait dynamics.

V. CONCLUSION

A proprietary collar-mounted IMU step counter was compared with video-reference counts in 15 walking or trotting/running bouts from 11 dogs. The video reference contained 357 steps, and the IMU output contained 355 steps. Mean absolute error was 1.60 steps per bout, MAPE was 7.96%, WAPE was 6.72%, and mean signed bias was -0.13 steps. Bland-Altman limits of agreement were -4.87 to 4.60 steps per bout.

The results demonstrate close bout-level agreement under the walking and trotting/running conditions examined and support the use of the evaluated system for step counting in structured locomotor bouts. Further validation can extend these findings to longer recordings and less controlled activity.

ACKNOWLEDGMENT

The authors thank the participating dogs and handlers.

CONFLICT OF INTEREST

Some authors are affiliated with ZOOMIES S.A., the company responsible for the evaluated system. This affiliation constitutes a potential conflict of interest.

ETHICS STATEMENT

The study involved privately owned dogs undergoing non-invasive collar-mounted sensor recording and video documentation during ordinary locomotion. Owners or legal guardians provided informed consent for participation and use of anonymized derived data for research and publication. The study was conducted by ZOOMIES S.A. as observational product-validation research in accordance with applicable animal welfare and data-protection procedures.

REFERENCES

- [1] J. Colpoys and D. DeCock, "Evaluation of the FitBark activity monitor for measuring physical activity in dogs," *Animals*, vol. 11, no. 3, p. 781, 2021.
- [2] G. J. Jenkins, C. H. Hakim, N. N. Yang, G. Yao, and D. Duan, "Automatic characterization of stride parameters in canines with a single wearable inertial sensor," *PLOS ONE*, vol. 13, no. 6, p. e0198893, 2018.
- [3] X. Zhang, G. J. Jenkins, C. H. Hakim, D. Duan, and G. Yao, "Four-limb wireless IMU sensor system for automatic gait detection in canines," *Scientific Reports*, vol. 12, p. 4788, 2022.
- [4] C. Ladha, Z. Belshaw, J. O'Sullivan, and L. Asher, "A step in the right direction: An open-design pedometer algorithm for dogs," *BMC Veterinary Research*, vol. 14, no. 1, p. 107, 2018.
- [5] I. den Uijl, C. B. Gómez Álvarez, D. Bartram, Y. Dror, R. Holland, and A. Cook, "External validation of a collar-mounted triaxial accelerometer for second-by-second monitoring of eight behavioural states in dogs," *PLOS ONE*, vol. 12, no. 11, p. e0188481, 2017.
- [6] O. Friard and M. Gamba, "BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations," *Methods in Ecology and Evolution*, vol. 7, no. 11, pp. 1325–1330, 2016.
- [7] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.
- [8] L. I. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, vol. 45, no. 1, pp. 255–268, 1989.
- [9] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of Chiropractic Medicine*, vol. 15, no. 2, pp. 155–163, 2016.