Machine Learning: Earth

Crossmark

RECEIVED
dd Month yyyy

REVISED
dd Month yyyy

JOURNAL ARTICLE

Computer Vision for White-Tailed Deer Age Estimation: A Dual-Modality Approach Using Trail Camera Images and Jawbone Morphology

Aaron J. Pung, Ph.D.

E-mail: aaron.pung@gmail.com

Keywords: machine learning, computer vision, neural network, deer, age, classification, prediction, dental analysis, tooth wear, wildlife management, automation, deep learning, transfer learning

Abstract

Accurate age estimation of white-tailed deer remains a critical challenge for wildlife management, with existing methods limited by low accuracy, high cost, or extensive processing delays. This study presents the first comprehensive computer vision approach addressing both field scenarios (live deer from trail cameras) and post-harvest scenarios (jawbone dental analysis). Using transfer learning with Convolutional Neural Networks, two complementary systems were developed: a ResNet-18 ensemble for trail camera images achieving $76.7\% \pm 5.9\%$ cross-validation accuracy, and an EfficientNet ensemble for jawbone images achieving $90.7\% \pm 2.6\%$ cross-validation accuracy. Both models substantially outperform traditional methods including human visual assessment (58.6%), morphometric models (63%), and manual tooth wear analysis, while exceeding the 70% accuracy threshold required for wildlife management decisions. Attention map analysis confirms both models autonomously discover and utilize biologically relevant features despite never being explicitly programmed with rules from either domain. Combined with dramatic improvements in speed and consistency over traditional methods, the model's biological plausibility positions the complementary computer vision systems as practical tools for transforming deer population monitoring and management across North America.

1 Introduction

1.1 Wildlife Management Context

Accurate age estimation of white-tailed deer (*Odocoileus virginianus*) is fundamental to effective wildlife management, impacting harvest regulations, population modeling, and conservation strategies across North America. Age-structure data inform critical decisions including harvest quotas, antler point restrictions, and habitat management priorities. However, obtaining reliable age estimates remains a persistent challenge since non-invasive field observations of live deer lack accuracy, while methods that achieve higher accuracy (laboratory dental analysis) require harvested specimens and substantial processing time.

Realistically, wildlife management requires age data from *both* live populations and harvested specimens. Pre-season trail camera surveys of live deer enable population monitoring, buck-to-doe ratio estimation, and age-class distribution assessment without requiring harvest. Post-harvest analysis of jawbones from hunter-harvested deer provides validation data and detailed age-structure information for population models. Current methods address these scenarios independently, each with significant limitations. This study presents the first unified computer vision solution to address both modalities, providing complementary tools that match the diverse data collection contexts wildlife managers encounter.

1.2 Current Methods and Their Limitations

For live deer, wildlife professionals and outdoor enthusiasts rely on "aging on the hoof" (AOTH), a set of visual assessment guidelines applied to trail camera images or field observations. First documented in 1978 [1], AOTH attempts to predict deer age based on morphological features including chest depth, stomach sag, neck thickness, and leg proportions [2, 3, 4, 5]. The biological

rationale follows predictable ontogenetic changes: yearling bucks (1.5 years) exhibit proportionally longer legs, narrow chests, and minimal neck development due to incomplete skeletal and muscular maturity. As bucks mature (2.5-4.5 years), chest depth increases relative to body length, neck circumference expands due to muscle hypertrophy during breeding season, and stomach contour changes from flat to increasingly sagging as body mass increases. Aged bucks (≥ 5.5 years) typically show pronounced stomach sag, thick necks, shorter-appearing legs relative to body depth, and graying facial fur.

Despite these predictable biological patterns, human AOTH accuracy remains problematic. Gee et al. [6] found that wildlife enthusiasts and trained professionals achieved only 36% age prediction accuracy in systematic testing, with individual scores ranging between $16 - 56\%$. Recent data from the National Deer Association's (NDA) "Age This!" survey (used as ground truth in this study given the scarcity of publicly available known-age deer data) confirm these challenges. As shown in Figure 1, human AOTH prediction accuracy revealed an average correct prediction rate of 58.64% with an inverse relationship between accuracy and buck age, consistent with previous studies [6, 7]. Incorrect predictions for 2.5-year-old deer split nearly evenly between 1.5 and 3.5 years, whereas incorrect predictions for 3.5- and 4.5-year-old deer overwhelmingly skewed younger. For deer ≥ 5.5 years, incorrect predictions dropped to 4.5 years nearly half the time. No age classes 2.5 years or higher achieved the 70% threshold professionals identify as necessary for management decisions, much less the 80% threshold required for research applications.

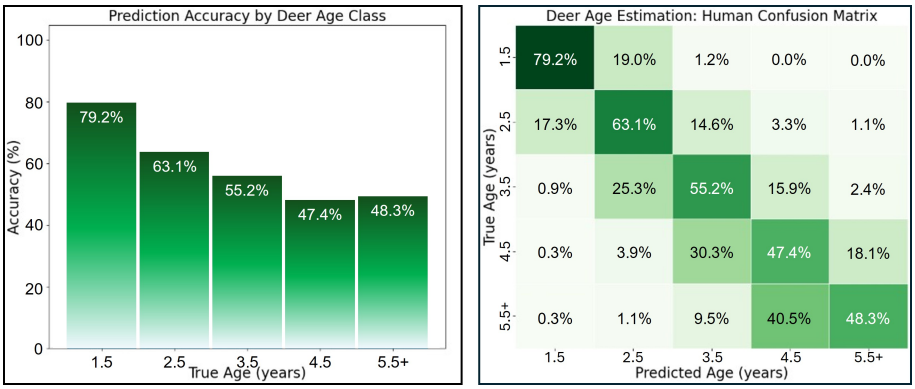

Figure 1. Human prediction accuracy for white-tailed deer aging on the hoof from trail camera images. (Left) Accuracy by age class shows consistent underperformance below management thresholds. (Right) Confusion matrix reveals systematic bias toward predicting younger ages for mature deer (2.5-4.5 years) while yearlings are more reliably identified.

The fundamental limitation of AOTH stems from individual morphological variation. Body condition varies dramatically with local nutrition quality, genetics, population density, and environmental stressors. Morphometric approaches [8] have attempted to identify mophological relationships through quantitative measurements of 64 body part ratios, achieving 63% accuracy during post-breeding periods. While this approach is an improvement over unaided human assessment, its performance still falls short of practical thresholds and requires precise body measurements rarely obtainable from trail camera imagery.

For harvested deer, postmortem dental analysis provides two alternatives: tooth replacement and wear (TRW) and cementum annuli (CA). TRW relies on predictable dental development patterns. White-tailed deer are born with deciduous premolars and develop permanent dentition following a known eruption sequence. By 1.5 years, the third premolar (P_3) exhibits a characteristic three-crested structure. As deer age, molar tooth wear progresses in a predictable sequence based on masticatory mechanics and diet abrasiveness. The dentine-to-enamel ratio (DER) of successive molars (M_1, M_2, M_3) increases systematically with age as enamel wears away, exposing underlying dentine. By ≥ 6.5 years, the first molar (M_1) typically shows extreme wear with flattened occlusal surfaces [9, 10].

CA exploits annual cementum deposition patterns. Cementum layers form seasonally on tooth roots, with light bands depositing during summer due to abundant nutrition and rapid growth. Dark bands form during winter where nutritional stress and slower growth take place. Thin-section microscopy of incisors or premolars enables annuli counting similar to tree ring analysis [9, 11, 12]. While conceptually elegant, both TRW and CA face practical limitations. TRW accuracy varies with soil abrasiveness affecting wear rates, nutritional stress affecting eruption timing, and analyst experience [13, 14, 15]. CA faces technical challenges including indistinct or condensed annulus

patterns [12, 15] and requires specialized laboratory processing with multi-week turnaround times [16].

Critically, studies comparing TRW and CA show inconsistent results. Some report equivalent performance [17], others favor CA [18, 19], while still others demonstrate TRW superiority even with specialized CA laboratories [15]. This lack of consensus reflects genuine biological variability in dental aging markers, suggesting neither method provides a definitive gold standard. The practical consequence for wildlife management is clear: current methods lack the combination of accuracy, speed, and accessibility required for large-scale population monitoring.

1.3 Computer Vision and Transfer Learning

Machine learning (ML) and computer vision (CV) have revolutionized classification tasks across domains by automatically learning hierarchical feature representations from raw data. Convolutional Neural Networks (CNNs) eliminate laborious manual feature engineering, instead discovering optimal features through data-driven optimization. For image classification, CNNs learn progressively complex representations: early layers detect edges and textures, intermediate layers recognize shapes and patterns, and deep layers capture high-level semantic concepts.

Transfer learning further enables high accuracy with limited domain-specific data by leveraging CNNs pre-trained on large datasets like ImageNet (14 million images across 1000 categories). Architectures including ResNet [20], EfficientNet [**?**], and DenseNet [21] achieve exceptional performance through architectural innovations like residual connections and compound scaling. By freezing early layers and fine-tuning deep layers, models are able to preserve general visual features and adapt to domain-specific patterns. Transfer learning's ability to achieve strong performance in data-scarce domains is a critical advantage for wildlife research where obtaining large quantities of professionally verified specimens is prohibitively expensive and time-consuming.

This study presents the first comprehensive application of deep learning to white-tailed deer age estimation, developing complementary models for both trail camera imagery and dental specimens. The dual-modality approach directly addresses the practical reality that wildlife managers collect age data from both live monitoring and post-harvest sampling, requiring tools optimized for each context.

2 Trail Camera Assessment

2.1 Dataset Collection and Preparation

The trail camera dataset contains 197 color images sourced exclusively from National Deer Association (NDA) materials including the weekly "Age This!" survey, educational videos, and published guides. While many images meet strict NDA quality standards (a broadside posture, head-up position, minimal motion blur, good lighting, and entire body visibility), the dataset also includes video screenshots and media captures with more challenging conditions, expanding morphological variation and imaging quality. Geographic coverage spans 14 U.S. states, capturing regional variation in body morphology. Age determinations represent NDA expert panel consensus, the most readily available source of systematically vetted age labels for supervised learning [7].

Importantly, image metadata including capture date, time, and location were stored but deliberately excluded as model inputs. This design choice forces the model to rely exclusively on visual morphology rather than potentially spurious correlations with temporal or geographic patterns. While metadata might improve predictions (e.g., correlating body condition with pre-rut timing), the goal of this study was to develop a generalizable model based purely on biological features that wildlife professionals could interpret and trust.

Age distribution within the dataset reflects natural collection patterns: 30 yearlings (1.5 years), 36 images each of 2.5 and 3.5 year-olds, 52 images of 4.5 year-olds, and 43 images of deer ≥ 5.5 years. Following standard wildlife management practice, all deer ≥ 5.5 years were grouped into a single class, as age-related morphological changes become less distinct in mature animals.

Image standardization followed a minimal preprocessing approach. Original images were cropped to square format capturing maximum deer body coverage while eliminating extraneous background. Cropped images were resized to 224×224 pixels to match ResNet-18 input requirements. Crucially, backgrounds remained unmodified, forcing the model to identify and learn deer morphology while ignoring diverse environmental contexts including forests, fields, feeding stations, and human structures. No artificial borders or digital artifacts were added, ensuring classification accuracy reflects genuine feature recognition rather than dataset-specific markers. Figure 2 illustrates the standardized dataset's diversity in deer pose, lighting, background, and image quality.

Figure 2. Representative sample from the standardized trail camera dataset, demonstrating variation in deer age, pose, lighting conditions, background environments, and image quality

2.2 Data Augmentation Strategy

Deep learning classification typically requires ∼1000 samples per category for optimal performance, nearly 25× the data available in this study. To address the limitation of data scarcity, systematic data augmentation expanded the training set (157 images) 40-fold through transformations simulating natural field variation while preserving biological features critical for age assessment.

Augmentation transformations included rotations ($\pm 10°$) to simulate camera mounting angle variation and deer positioning relative to trail camera, horizontal flipping to mimic deer different deer walking directions, and image brightness ($0.8 - 1.2\times$) to replicate lighting changes across time of day, weather conditions, and seasonal canopy coverage. Other transformations include contrast variation ($0.8 - 1.2\times$) to account for camera sensor differences and exposure settings, and Gaussian noise addition ($\sigma = 0.01$) to mimic sensor noise and image compression artifacts. Critically, augmentation ranges were constrained to avoid distorting age-diagnostic features. For instance, excessive rotation may alter apparent body proportions, extreme brightness changes might obscure facial features and coat color, and aggressive cropping could eliminate key body regions. Instead, each transformation preserves the morphological relationships wildlife biologists rely upon for visual assessment. The test set (40 images) remained non-augmented and held out throughout model development, providing unbiased performance evaluation.

2.3 Model Architecture and Training

Preliminary evaluation of traditional machine learning methods like K-Nearest Neighbor, Random Forest, and others achieved ∼ 57% accuracy, substantially below CNN performance. In addition, over 60 deep learning architectures were explored; of those, ResNet-18 [20] was selected based on superior validation accuracy while maintaining computational efficiency suitable for practical deployment.

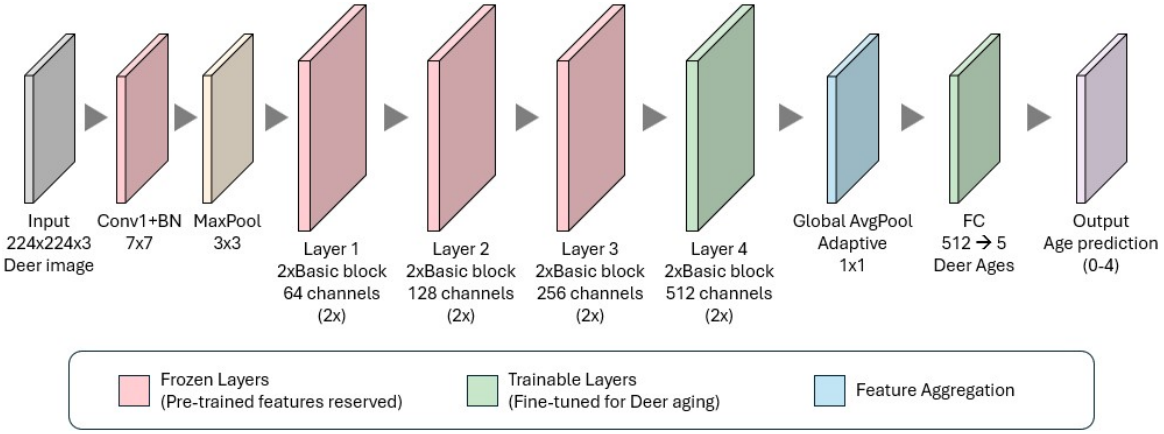

Figure 3. ResNet-18 architecture adapted for deer age classification.

The ResNet-18 architecture (Figure 3) comprises an initial convolution layer (Conv1) and batch normalization (BN), followed by four residual blocks (Layers 1-4) and a fully connected classifier. Transfer learning was implemented by freezing the initial convolution, batch normalization, and first three residual blocks, preserving ImageNet-learned features for edges, textures, and shape primitives. These low- to mid-level features transfer effectively across visual domains. The fourth residual block remained trainable, adapting high-level semantic features to deer-specific morphology including body proportion patterns, coat texture changes with age, and muscle conformation. The original 1000-class ImageNet classifier was replaced with a 5-class output layer for deer age categories (1.5, 2.5, 3.5, 4.5, ≥ 5.5 years).

To maximize utilization of the limited sample size, a 5-fold stratified cross-validation ensemble approach was employed by partitioning training data into five folds with proportional age class representation. Each fold was subdivided into training (125 images) and validation (32 images) components. Training images underwent 40× augmentation, yielding 5000 balanced samples per fold. Five independent ResNet-18 models were trained, one per fold, with each model training exclusively on its fold's augmented data and validating on the corresponding non-augmented validation set.

Training utilized AdamW optimization with differential learning rates: 0.0003 for frozen backbone layers (minimal adjustment was used to preserve ImageNet features) and 0.001 for trainable layers, enabling deer-specific adaptation. Learning rates followed exponential decay

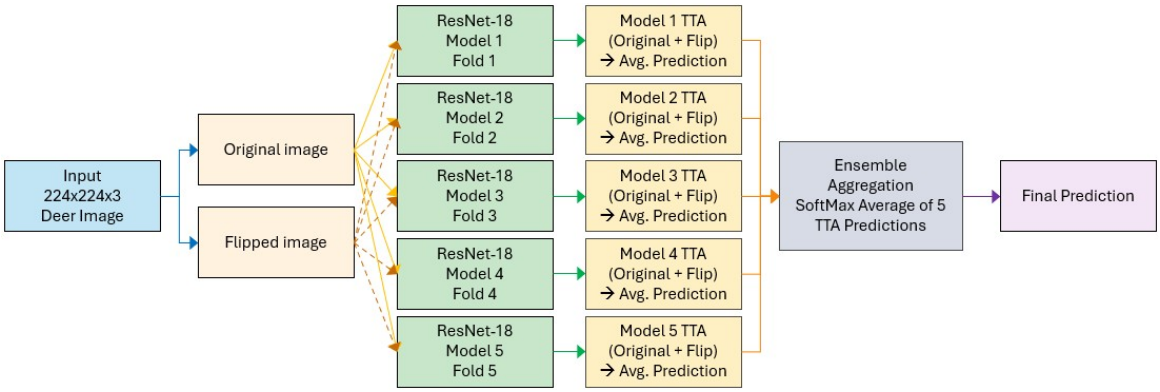

Figure 4. ResNet-18 ensemble inference process employing test-time augmentation (TTA).

($\gamma = 0.95$) to gradually reduce step sizes as models converged. Label smoothing ($\alpha = 0.1$) provided regularization by softening one-hot encoded labels, preventing overconfident predictions on the limited dataset. Cross-entropy loss served as the objective function. Early stopping with 20-epoch patience prevented overfitting, terminating training when validation loss ceased improving. Models typically converged after approximately 40 epochs per fold, requiring 45 minutes total training time on NVIDIA RTX 2060 hardware.

Inference utilized test-time augmentation (TTA) with ensemble averaging (Figure 4). For each test image, all five models generated predictions on both the original image and its horizontal flip. Within each model, predictions from original and flipped versions were averaged. The five TTA-averaged predictions were then ensemble-averaged using softmax normalization to produce final classification probabilities. The multi-stage process was chosen to reduce prediction variance and improves robustness to minor pose variations.

2.4 Trail Camera Results

The ResNet-18 ensemble achieved $76.7\% \pm 5.9\%$ mean cross-validation accuracy across five folds and 97.5% test accuracy. The substantial test performance likely reflects overfitting to the small dataset, making cross-validation the more reliable metric for expected real-world performance. Nonetheless, the cross-validation accuracy exceeds human expert assessment (58.6%), morphometric models (63%), and the 70% threshold required for management decisions.

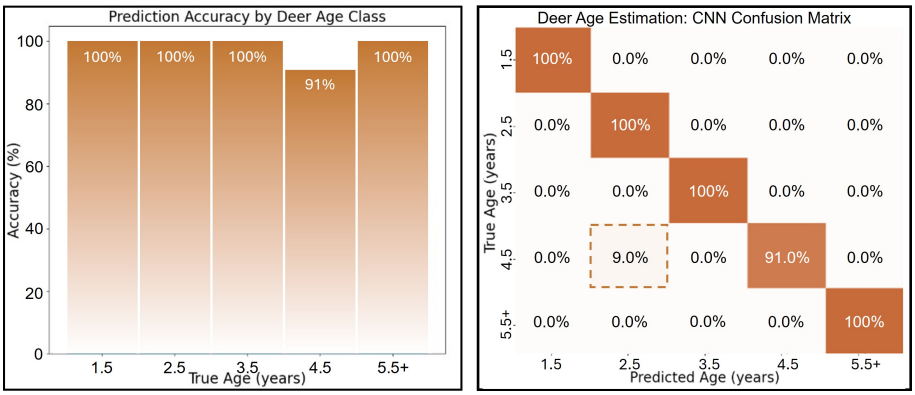

Figure 5. Trail camera model performance showing (left) a bar chart for per-class accuracy, and (right) a confusion matrix illustrating incorrect predicitons.

shows strong performance across all age categories except 4.5 years, where one test specimen was misclassified as 2.5 years—the same age class most challenging for human assessors. (Right) Confusion matrix demonstrates near-perfect classification with minimal off-diagonal errors. The model achieves 100% accuracy for yearlings (1.5 years), 2.5 years, 3.5 years, and mature deer (≥ 5.5 years) on the test set.

Per-class performance (Figure 5) remained strong across all age categories with one notable exception: a single 4.5-year-old buck was misclassified as 2.5 years old on the test set. Interestingly, 4.5 years is also the age class most challenging for human assessors [6], suggesting this may represent a genuinely difficult specimen exhibiting atypical morphology for its age rather than a

systematic model failure. All other age classes achieved 100% test accuracy, though this should be interpreted cautiously given the small test set size.

2.5 Biological Interpretation via Attention Maps

Beyond classification accuracy, understanding *which* image features drive predictions is critical for biological validation and professional acceptance. Attention map analysis visualizes spatial regions the model deems most important for classification decisions. Figure 6 shows attention maps for representative specimens from each age class. 5

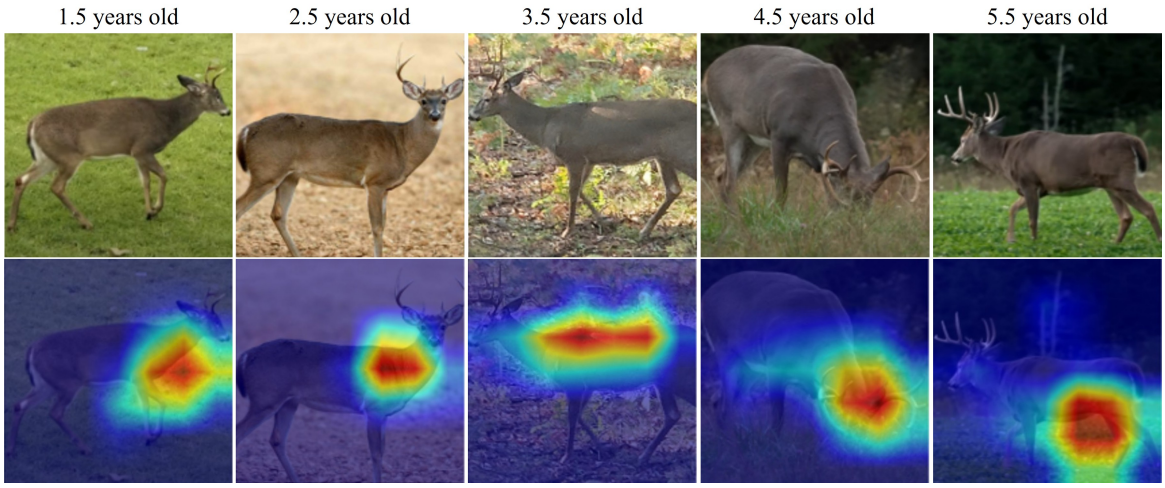

Figure 6. Attention map analysis for trail camera images. (Top row) Original images labeled with NDA-determined age. (Bottom row) Attention heatmaps overlaid on original images, with red/yellow indicating high attention and blue indicating low attention.

Despite variation in background, pose, and lighting, attention maps consistently focus on biologically relevant morphological features. In 1.5 year old bucks (yearlings), attention concentrates on neck and chest regions, key indicators of immature body conformation. Young bucks exhibit narrow chests and thin necks relative to body size. For 2.5 year old bucks, focus distributes across the neck, chest, and body, reflecting the transitional growth phase where proportions shift toward mature conformation but remain incomplete. In mature deer, $\geq$ 5.5 years, attention strongly emphasizes the stomach region, a primary characteristic experts use for identifying aged bucks. Pronounced stomach sag results from accumulated body mass and age-related changes in body composition.

Critically, attention maps largely ignore antlers across all age classes, aligning with expert guidance that antler size is an unreliable age indicator due to high variation with genetics, nutrition, and injury history [3, 4]. Despite never being explicitly instructed to do so, the model's spontaneous focus on body morphology rather than antlers provides compelling evidence that learned features reflect genuine age-related biological patterns rather than spurious correlations.

3 Dental Assessment

3.1 Biological Foundations of Dental Aging

White-tailed deer dental aging exploits two fundamental biological processes: predictable tooth eruption sequences during development and systematic tooth wear progression throughout life. Fawns are born with four deciduous premolars, with permanent premolars and molars erupting following a genetically determined schedule. By 1.5 years, the full adult dentition has erupted, with the third premolar (P_3) exhibiting a diagnostic three-crested structure. Deer $\geq$ 2.5 years show two-crested P_3, providing clear separation between yearlings and older animals [9, 10].

For deer $\geq$ 2.5 years, aging relies on tooth wear patterns driven by masticatory mechanics. White-tailed deer are browsers and grazers, consuming vegetation with varying abrasiveness. Chewing mechanics produce predictable wear progression: the first molar (M_1) erupts earliest and experiences the longest wear period, followed sequentially by M_2 and M_3. As enamel wears away on occlusal surfaces, the underlying dentine becomes increasingly exposed. Therefore, the dentine-to-enamel ratio (DER) of each molar increases systematically with age, providing quantitative aging criteria.

Traditional tooth replacement and wear (TRW) protocols follow a sequential decision tree based on tooth count (fawns vs. older deer), P_3 crest structure (1.5 years vs. $\geq$ 2.5 years), and progressive DER thresholds for M_1, M_2, and M_3 to separate older age classes. However, work by Meares et al. [22] demonstrated that individual variation in tooth wear rates undermines DER reliability for separating 2.5-4.5 year classes, suggesting this age range represents a fundamental biological challenge for dental aging.

3.2 Dataset Collection and Preparation

The jawbone dataset comprises 243 color images collected from 17 independent sources including Quality Deer Management Association, National Deer Association, state wildlife agencies, and university wildlife programs via educational videos, training tutorials, and published guides. Unlike trail camera images, no restrictions were placed on geographic region or deer sex, as dental aging principles apply broadly across white-tailed deer populations. Age distribution included 39 fawns (0.5 years), 62 yearlings (1.5 years), 33 images of 2.5 year-olds, 29 images of 3.5 year-olds, 20 images of 4.5 year-olds, 22 images of 5.5 year-olds, and 38 images from deer aged 6.5-16.5 years (grouped as 5.5+ years). Specimens confirmed as $\geq$ 9.5 years were exclusively sourced from NDA documentation.

Jawbone image standardization followed similar principles to trail camera processing but was adapted for dental specimens. Raw images often contained annotations, age labels, or measurement markings added during original educational material creation. These were digitally removed using editing tools on a Samsung Galaxy S25 Ultra to prevent the model from exploiting text information rather than learning dental features. Images were cropped to 2:1 aspect ratio ensuring all visible teeth remained in frame. Original backgrounds and lighting were preserved, and fingertips holding jawbones were retained in some images to simulate realistic field submission conditions where enthusiasts or agency personnel would photograph specimens in-hand. Figure 7 illustrates the standardized dataset's variation in jawbone orientation, lighting, background, and presentation.

Figure 7. Representative sample from the standardized jawbone dataset, demonstrating variation in jawbone orientation, tooth condition, lighting, background, and imaging context.

Data splitting followed 80/20 train/test stratification with proportional age representation in each. Training data (80%, 194 images) underwent balanced augmentation to 1200 samples per class through rotations ($\pm 10°$), horizontal flipping (simulating imaging either side of the jaw), brightness adjustments ($0.8 - 1.2\times$), and contrast variation ($0.8 - 1.2\times$). Unlike trail camera augmentation, no Gaussian noise was added given that jawbone images are typically captured under more controlled conditions. The test set (20%, 49 images) remained non-augmented.

3.3 Model Architecture and Training

EfficientNet was selected for dental assessment based on its compound scaling approach balancing network depth, width, and resolution for computational efficiency. The architecture employs seven

Mobile Inverted Bottleneck Convolution (MBConv) blocks with squeeze-and-excitation optimization, enabling efficient feature extraction with fewer parameters than comparably accurate architectures.

Transfer learning froze the stem convolution and first three MBConv blocks (blocks 0-2), preserving low- to mid-level features learned from ImageNet, including edges, textures, and basic shapes. Similar to trail camera images, general visual features transfer effectively to dental imaging. Subsequent blocks (blocks 3-6) remained trainable, adapting to dental-specific features including tooth outline shapes, wear surface textures, and eruption patterns. The classification head was replaced with a 6-class output to include fawns (0.5 years) in addition to the five age classes used for trail camera analysis.

A 5-fold nested cross-validation ensemble was employed with per-fold architecture selection. Training data were stratified into five folds (155 training, 39 validation per fold). For each fold, EfficientNet-B0, B1, and B2 variants were evaluated, selecting the best-performing architecture based on validation accuracy. The final ensemble comprised two EfficientNet-B2 models (folds 1 and 4) and three EfficientNet-B0 models (folds 2, 3, 5), as illustrated in Figure 8. This mixed-architecture approach captures complementary feature hierarchies: B0 emphasizes computational efficiency while B2 captures finer details through additional parameters.

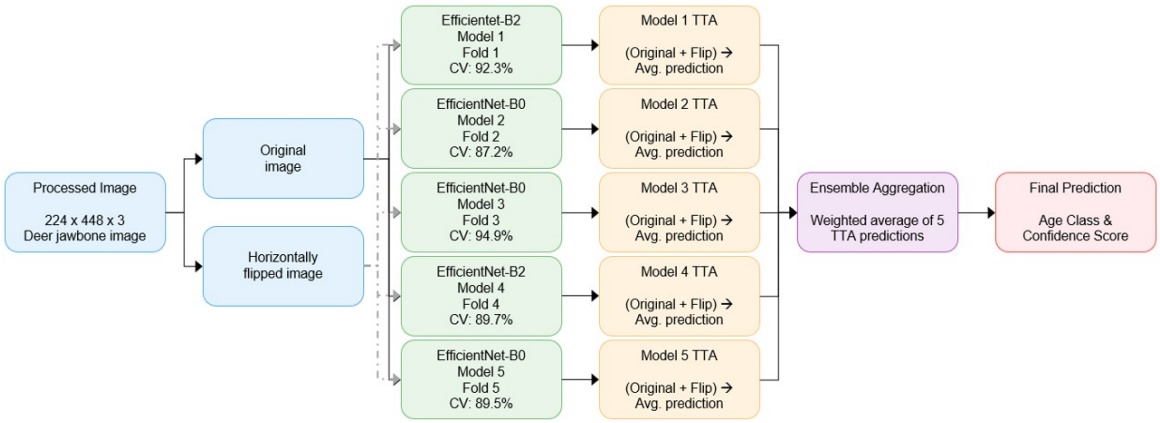

Figure 8. EfficientNet ensemble structure for dental age assessment.

Training utilized AdamW optimization with differential learning rates (0.0003 for frozen layers, 0.001 for trainable layers). Learning rates followed cosine annealing scheduling with $T_{max} = 80$ and $\eta_{min} = 1 \times 10^{-6}$, providing smooth decay for stable convergence. Label smoothing ($\alpha = 0.1$) and dropout ($p = 0.3$) provided regularization. Cross-entropy loss with mixed precision training accelerated convergence. Early stopping with 20-epoch patience was used to improve efficiency. Training converged after approximately 40-60 epochs per fold, requiring 536 minutes total on NVIDIA RTX 2060 hardware. Inference utilized TTA with cross-validation score weighting, where predictions from each model were weighted by its validation accuracy before ensemble averaging.

3.4 Dental Assessment Results

The EfficientNet ensemble achieved $90.7\% \pm 2.6\%$ mean cross-validation accuracy and 77.6% test accuracy. The 13.1% discrepancy between cross-validation and test performance likely reflects specimen-level data leakage (multiple images from individual specimens appearing in both training and test sets) combined with limited test set size. The more conservative test accuracy (77.6%) still substantially exceeds traditional TRW performance and surpasses the 70% management threshold, while the cross-validation accuracy (90.7%) well exceeds the 80% research threshold.

Per-class performance (Figure 9) reveals strong accuracy for fawns (0.5 years, 100%), yearlings (1.5 years, 91.7%), 4.5 years (100%), and aged deer ($\geq$ 5.5 years, 83.3%). However, 2.5 and 3.5 year classes showed reduced accuracy. For specimens with a true age of 2.5 years, the model split predictions evenly between 1.5 and 2.5 years. For specimens with a true age of 3.5 years, the model predicted 2.5 years half the time, correctly identifying 3.5 years only one-third of the time. Although the results cannot be directly compared to DER, the difficulty of separating 2.5-4.5 year classes is similar to Meares et al.'s finding [22]. The model's confusion in this age range likely reflects genuine biological ambiguity rather than a correctable model limitation.

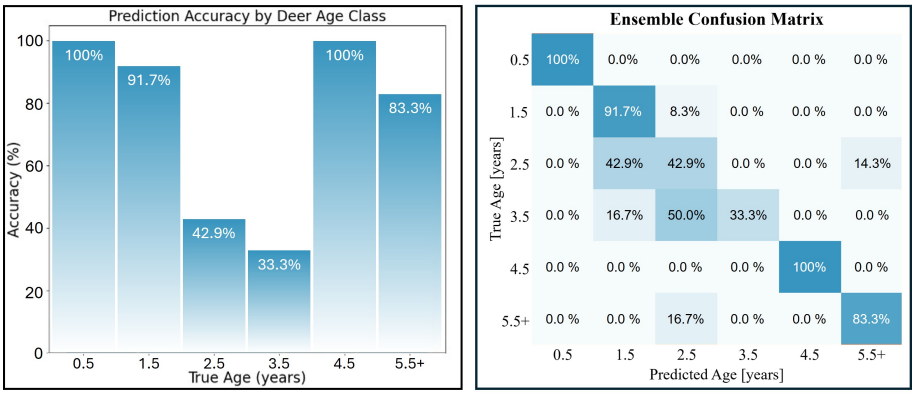

Figure 9. Jawbone model performance is illustrated in (left) a bar chart showing per-class F1 scores and (right) a confusion matrix.

3.5 Biological Interpretation via Attention Maps

Attention map analysis (Figure 10) provides critical validation that the model learns dental-specific features matching TRW criteria despite never being explicitly programmed with these rules.

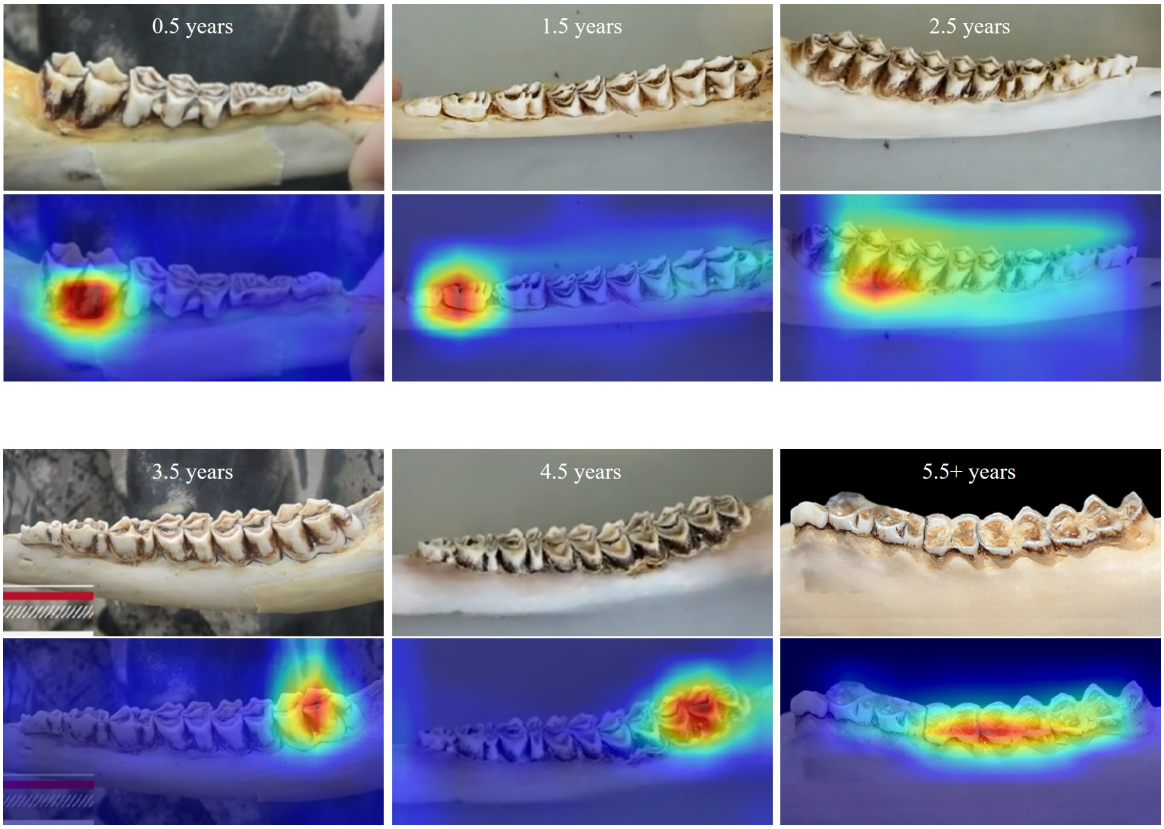

Figure 10. Attention map analysis for jawbone images. Each image pair contains (top) the original image and (bottom) an overlayed attention map. Ages are provided for each image pair.

Attention patterns across age classes demonstrate strong alignment with traditional TRW decision trees, despite never being explicitly programmed with TRW or CA guidelines . Attention in fawn jawbones concentrates on molars, consistent with tooth count assessment, while focus in yearling jawbones shifts to premolars, matching crest-counting criteria. The three-crested P_3 structure characteristic of 1.5-year-olds is the primary diagnostic feature. For jawbones of 2.5 year old deer, attention is distributed across the molar region, particularly M_1, corresponding to DER evaluation for the earliest-erupted molar showing measurable wear. For 3.5 year old deer, focus targets posterior molars (M_2-M_3), consistent with sequential DER assessment as wear progresses to later-erupting teeth. In 4.5 year old jawbones, attention remains focused on the molar region, reflecting continued DER evaluation across all three molars. In mature deer ($\geq$ 5.5 years), the model highlights extensive wear characteristics and flattening across all teeth, a hallmark

characteristic of extreme age.

In each case, attention maps focus exclusively on dental features despite variable backgrounds, finger presence, and jawbone orientations. The model learned to identify and ignore imaging artifacts while extracting genuine biological aging signatures. This spontaneous discovery of the TRW decision tree (tooth count, premolar structure, and sequential molar wear) provides compelling evidence that the model captures authentic biological patterns rather than spurious dataset correlations.

4 Comparative Analysis and Discussion

4.1 Performance Comparison Across Methods

Table 1 summarizes performance across aging methods for both live and harvested deer assessment. The computer vision approaches substantially outperform traditional techniques while offering practical advantages in speed, consistency, and accessibility.

Table 1. Comparison of white-tailed deer aging methods across accuracy, processing time, and practical constraints.
.

Method	Accuracy	Processing Time	Specimen Type	Key Limitation
		Traditional Methods		
Human AOTH	36-58.6%	Immediate	Live	High inter-observer variation
Morphometric	63%	Minutes-hours	Live	Requires precise measurements
Manual TRW	Variable	5-15 minutes	Harvested	Analyst experience dependent
Cementum Annuli	Variable	2-6 weeks	Harvested	Laboratory processing required
		Computer Vision (This Study)		
Trail Camera CNN	76.7%† (97.5%‡)	< 20 seconds	Live	Morphological variation
Jawbone CNN	90.7%† (77.6%‡)	< 20 seconds	Harvested	2.5-3.5 year confusion

†Cross-validation accuracy; ‡Test accuracy

For live deer assessment, the ResNet-18 ensemble (76.7%) exceeds human AOTH (58.6%) and morphometric models (63%) while meeting the 70% management threshold. Inference requires less than twenty seconds per image on modest GPU hardware, enabling rapid processing of large trail camera datasets. The primary limitation remains AOTH's fundamental constraint: morphological variation with nutrition, genetics, and environmental conditions prevents perfect classification. However, the model's ability to consistently exceed human performance while learning biologically interpretable features demonstrates practical value for population monitoring applications.

For harvested deer, the EfficientNet ensemble achieves 90.7% cross-validation accuracy, substantially exceeding the 80% research threshold. Even the more conservative test accuracy (77.6%) surpasses traditional TRW performance. Processing time (¡20 seconds) represents a dramatic improvement over manual TRW analysis (5-15 minutes) and laboratory CA processing (2-6 weeks). The automated approach eliminates inter-observer variation, a persistent challenge for traditional methods. The model's difficulty with 2.5-3.5 year separation reflects genuine biological ambiguity rather than a correctable failure, a limitation also shared with traditional TRW.

4.2 Biological Plausibility and Professional Acceptance

Attention map validation provides critical evidence that both models learn biologically meaningful features that align with expert knowledge. This biological plausibility distinguishes the approach from opaque "black box" machine learning and could facilitate professional acceptance by wildlife management agencies.

For trail camera assessment, the model's focus on neck, chest, and stomach regions while ignoring antlers precisely matches the anatomical features wildlife professionals emphasize in AOTH training [3, 4]. The model independently discovers these features despite never being explicitly programmed with AOTH guidelines. The convergence between learned features and expert knowledge further suggests the CNN identifies genuine ontogenetic changes like skeletal maturity, muscle development, and body composition shifts rather than spurious correlations.

For dental assessment, the progression from tooth count (fawns) to premolar structure (yearlings) to sequential molar wear (mature deer) perfectly follows the TRW decision tree [9, 10]. Again, these rules were never programmed; instead, the model autonomously discovered them through data-driven optimization. The alignment demonstrates that CNN feature hierarchies can capture domain-specific biological knowledge that experts have codified through decades of field experience.

This biological interpretability addresses a common concern about deploying machine learning in wildlife management: that "black box" predictions lack scientific justification and cannot be

trusted for critical decisions. By demonstrating that learned features match established biological principles, attention map validation builds confidence that predictions reflect genuine aging signals rather than dataset artifacts. The analytic transparency facilitates adoption by wildlife agencies accustomed to traditional methods with clear biological rationales.

4.3 Complementary Deployment Scenarios

The dual-modality approach used in this study directly addresses practical realities of wildlife management data collection. Trail camera analysis enables non-invasive population monitoring of age-structure assessment, buck-to-doe ratio estimation, and cohort tracking over multiple years without requiring harvest. Wildlife agencies conduct pre-season trail camera surveys to inform harvest regulations, and hunters use trail cameras for selective harvest decisions. The 76.7% accuracy meets management thresholds while low processing speeds enable thousands of images to be analyzed each season.

Jawbone analysis provides higher accuracy (90.7% cross-validation) for harvested specimens, meeting research standards for detailed population models and survival analyses. State agencies routinely collect jawbones from check stations during hunting season, often accumulating hundreds to thousands of samples. Traditional processing through manual TRW or laboratory CA creates bottlenecks limiting timely analysis. Automated CV assessment, on the other hand, enables rapid screening allowing wildlife officials to process samples immediately, flag low-confidence cases for expert review, and provide age-structure data to managers within days rather than months.

The complementary nature of both tools addresses the full data collection pipeline wildlife managers actually encounter. Trail cameras monitor live populations throughout the year. Harvest-season jawbone collection validates pre-season estimates and provides high-accuracy age data for population models. Together, the systems transform both field monitoring and post-harvest analysis.

4.4 Limitations and Future Directions

Despite strong performance, several limitations warrant consideration. First, both datasets (197 and 243 images, respectively) are modest in size compared to typical deep learning applications. The data scarcity reflects the real-world wildlife research challenge that obtaining professionally verified known-age specimens is expensive and time-consuming. Transfer learning and data augmentation mitigate this limitation, but larger datasets from institutional collections would likely improve performance. Collaboration with state agencies and research institutions accessing archive collections could substantially expand and improve training data.

Second, specimen-level data leakage may occur in each dataset, where multiple images from individual specimens likely appear in both training and test sets. While attention map analysis suggests the model learns biological features rather than specimen-specific artifacts, future work should implement specimen-level splitting to eliminate this confound. The 13.1% gap between cross-validation and test accuracy of the jawbone model suggests some overfitting, though the conservative test accuracy (77.6%) still exceeds traditional methods.

Third, both models show reduced performance for 2.5-3.5 year classes, reflecting genuine biological challenges. For trail cameras, this age range represents a transitional growth phase where morphology shifts from immature to mature conformation but individual variation increases with nutrition and genetics. For jawbone analysis, Meares et al. [22] demonstrated that individual tooth wear variation undermines DER-based age separation for this range. Both human experts and automated systems struggle with these transitional ages, suggesting an inherent ceiling on classification accuracy absent additional information.

Fourth, geographic generalization requires systematic validation. Both datasets span multiple U.S. states, but body morphology and tooth wear patterns may vary with regional differences in nutrition, genetics (subspecies boundaries), habitat quality, and population density. Validation studies across management units would strengthen confidence in broad deployment. Regional model variants trained on local data might improve performance in specific contexts.

Fifth, ground truth uncertainty deserves acknowledgment. NDA expert consensus represents the best available age labels given data scarcity, but even experts disagree on challenging specimens. In many cases, the true "correct" age remains unknown without independently verified known-age deer (e.g., from captive populations or longitudinal field studies). Model performance metrics inherently reflect agreement with expert consensus rather than absolute accuracy. Obtaining known-age validation sets from research facilities would enable more rigorous evaluation.

For these reasons, future work should prioritize collaboration with wildlife agencies to access larger specimen collections, implement specimen-level data splitting, conduct geographic validation

studies across management units, and develop uncertainty quantification methods to flag ambiguous cases for expert review. Extension to other cervid species (mule deer, elk, taruca, moose, etc.) where similar aging principles apply would demonstrate generalizability. Integration with existing wildlife management databases and workflows would also facilitate practical deployment.

4.5 Broader Implications for Wildlife Research

The success of transfer learning with limited datasets demonstrates important implications beyond deer aging. Wildlife research consistently faces data scarcity since obtaining large quantities of verified specimens is prohibitively expensive and field data collection is inherently challenging. Traditional machine learning struggled in these contexts, requiring extensive domain expertise for manual feature engineering and large sample sizes for reliable performance.

Transfer learning circumvents both limitations. Pre-trained CNNs capture generalizable visual features applicable across domains, enabling strong performance with modest domain-specific data. Automatic feature learning eliminates manual engineering, allowing wildlife biologists to leverage computer vision without extensive machine learning expertise. The approach demonstrated here (collecting limited labeled data from existing educational resources, applying transfer learning with CNN ensembles, and validating biological plausibility through attention analysis) provides a template for developing specialized tools across wildlife biology.

Potential applications of similar pipelines include automated species identification from camera traps, body condition scoring, injury detection, individual identification from coat patterns or other features, habitat quality assessment from vegetation imagery, and behavior classification from video. In each case, transfer learning enables domain experts to build effective tools with achievable data collection efforts. As wildlife agencies increasingly adopt trail cameras and digital documentation, computer vision tools can transform how biological data are extracted from imagery.

The biological interpretability demonstrated through attention analysis addresses a critical barrier to adoption because professional acceptance requires understanding *why* predictions are made, not just achieving high accuracy. By confirming that learned features match established domain knowledge, attention validation builds confidence that models capture genuine biological signals. In turn, the model's transparency facilitates trust in automated systems for informing management decisions.

5 Conclusions

This study presents the first comprehensive computer vision approach to white-tailed deer age estimation, demonstrating that transfer learning with CNN ensembles achieves breakthrough performance for both field (trail camera) and post-harvest (dental) scenarios. The complementary systems address the practical reality that wildlife managers collect age data from both live monitoring and harvested specimens, requiring tools optimized for each context.

The ResNet-18 ensemble for trail camera imagery achieves $76.7\% \pm 5.9\%$ cross-validation accuracy, exceeding human AOTH assessment (58.6%), morphometric models (63%), and meeting the 70% threshold for management decisions. The EfficientNet ensemble for jawbone imagery achieves $90.7\% \pm 2.6\%$ cross-validation accuracy and 77.6% test accuracy, substantially outperforming traditional TRW and CA methods while dramatically reducing processing time from weeks to seconds and eliminating inter-observer variation.

Critically, attention map validation confirms that both models autonomously discover and utilize biologically relevant features that precisely match the anatomical characteristics wildlife biologists rely upon. Body morphology of the neck, chest, and stomach are identified in trail camera images and dental patterns like tooth eruption and sequential wear are identified in jawbone images, despite never being explicitly programmed with domain rules. This biological plausibility distinguishes the approach from opaque machine learning and facilitates professional acceptance by demonstrating that predictions reflect genuine biological aging signals rather than spurious correlations.

Together, these complementary systems offer wildlife managers practical tools for transforming population monitoring and herd management. Trail camera analysis enables rapid, non-invasive age-structure assessment for pre-season population surveys and real-time harvest decisions. Jawbone analysis provides high-accuracy validation from harvested specimens with immediate turnaround, eliminating processing bottlenecks that delay management actions. Both systems maintain or exceed accuracy standards for their respective applications while offering dramatic improvements in speed, consistency, and accessibility.

The success of transfer learning with limited datasets demonstrates broader implications for wildlife research, where data scarcity persistently constrains method development. By leveraging pre-trained CNNs and validating biological plausibility through attention analysis, domain experts can develop specialized computer vision tools without massive data collection efforts or extensive machine learning expertise. As wildlife agencies increasingly adopt digital documentation, automated image analysis can transform how biological information is extracted from visual data, advancing conservation and management across species and contexts.

References

[1] Frederick F Knowlton, Marshall White, and John G Kie. Weight patterns of wild white-tailed deer in southern texas. *Proceedings of the First Welder Wildlife Foundation Symposium*, 1978.

[2] James C. Kroll and Mike Biggs. *Aging and judging trophy whitetails.* Center for Applied Studies in Forestry, College of Forestry, Stephen F. Austin State University, 1996.

[3] S. Demarais, D. Stewart, and R. N. Griffin. A hunter's guide to aging and judging live white-tailed deer in the southeast., 1999.

[4] Dave Richards and Al Brothers. *Observing and evaluating whitetails.* D. Richards, 2003.

[5] Mickey W Hellickson, Karl V Miller, Charles A DeYoung, R. Larry Marchinton, Stuart W Stedman, and Robert E Hall. Physical characteristics for age estimation of male white-tailed deer in southern texas. page 40–45, 2008.

[6] Kenneth L. Gee, Stephen L. Webb, and John H. Holman. Accuracy and implications of visually estimating age of male white-tailed deer using physical characteristics from photographs. *Wildlife Society Bulletin*, 38(1):96–102, Oct 2013.

[7] Aaron J. Pung. Age classification of white-tailed deer via computer vision and deep learning. *bioRxiv*, July 2025.

[8] Jeremy J. Flinn. Accuracy of estimating age and antler size of photographed deer. Master's thesis, Mississippi State University, 2010.

[9] C. W. Severinghaus. Tooth development and wear as criteria of age in white-tailed deer. *The Journal of Wildlife Management*, 13(2):195, Apr 1949.

[10] James S. Larson and Richard D. Taber. Criteria of sex and age. In Sanford D. Schemnitz, editor, *Wildlife Management Techniques Manual*, pages 143–202. The Wildlife Society, Washington, D.C., 4th edition, 1980.

[11] William A. Low and I. McT. Cowan. Age determination of deer by annular structure of dental cementum. *The Journal of Wildlife Management*, 27(3):466, July 1963.

[12] A. Brian Ransom. Determining age of white-tailed deer from layers in cementum of molars. *The Journal of Wildlife Management*, 30(1):197, Jan 1966.

[13] Frederick F. Gilbert. Aging white-tailed deer by annuli in the cementum of the first incisor. *The Journal of Wildlife Management*, 30(1):200, Jan 1966.

[14] John Ludwig. Comparison of age determination techniques for the white-tailed deer of southern illinois. Master's thesis, Southern Illinois University, 1967.

[15] Robert Cook and Raymond Hart. *Ages Assigned Known-Age Texas White-Tailed Deer: Tooth Wear Versus Cementum Analysis*, volume 33, page 195–201. 1979.

[16] National Deer Association. Estimating deer age with cementum annuli. https://deerassociation.com/estimating-deer-age-cementum-annuli/, Oct 2012.

[17] Harry Jacobson and Richard Reiner. Estimating age of white-tailed deer tooth wear vs cementum annuli. In *Proc. Ann. Conf. S.E. Assoc. Fish and Wildl. Agencies*, volume 43, pages 286–291, Jan 1989.

[18] Kenneth L. Hamlin, David F. Pac, Carolyn A. Sime, Richard M. DeSimone, and Gary L. Dusek. Evaluating the accuracy of ages obtained by two methods for montana ungulates. *The Journal of Wildlife Management*, 64(2):441, Apr 2000.

[19] Susan M. Cooper, Shane S. Sieckenius, and Andrea L. Silva. Dentine method: Aging white-tailed deer by tooth measurements. *Wildlife Society Bulletin*, 37(2):451–457, Apr 2013.

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[21] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2018.

[22] Jeremy Meares, Brian Murphy, Charles Ruth, David Osborn, Robert Warren, and Karl Miller. *A Quantitative Evaluation of the Severinghaus Technique for Estimating Age of White-tailed Deer*, volume 60, pages 89–93. 2006.

Acknowledgments
The author acknowledges the wildlife management organizations, state agencies, and educational institutions that provided trail camera and post-mortem dental analysis training materials used in dataset construction. Particular appreciation is extended to the wildlife professionals who developed these educational resources, enabling this interdisciplinary application of computer vision to wildlife biology. Special thanks to the National Deer Association for their "Age This!" survey data and educational content that formed the foundation of both datasets. Appreciation is also extended to the open-source community for the deep learning frameworks and tools that enabled this work.

Funding
This research received no external funding.