A hybrid approach to invertebrate biomonitoring using computer vision and DNA metabarcoding

Jarrett D. Blair [1,2], Michael D. Weiser [3], Cameron Siler [3,4], Michael Kaspari [3,5], Sierra N. Smith [3,4,6], Jessica F. McLaughlin [3,4], Katie E. Marshall [1]

[1] Department of Zoology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

[2] Department of Ecoscience, Aarhus University, 8000 Aarhus C, Denmark

[3] School of Biological Sciences, University of Oklahoma, Norman, OK 73019-0235, USA

[4] Sam Noble Oklahoma Museum of Natural History, University of Oklahoma, Norman, OK 73072-7029, USA

[5] Conservation Ecology Center, Smithsonian's National Zoo and Conservation

[6] Department of Biology Institute, Front Royal, VA 22630, University of Texas at Arlington, Arlington, TX, 76019, USA.

Corresponding author: Jarrett Blair, Department of Ecoscience, Aarhus University, C.F. Møllers Allé 8, 8000 Aarhus C, Denmark, jarrett@ecos.au.dk

1 **<u>Abstract</u>**

2     1.   Automated invertebrate classification using computer vision has shown significant

3        potential to improve specimen processing efficiency. However, challenges such as

4        invertebrate diversity and morphological similarity among taxa can make it difficult to

5        infer fine-scale taxonomic classifications using computer vision. As a result, many

6        invertebrate computer vision models are forced to make classifications at coarser levels,

7        such as at family or order.

8     2.   Here we propose a novel framework to combine computer vision and bulk DNA

9        metabarcoding specimen processing pipelines to improve the accuracy and taxonomic

10        granularity of individual specimen classifications. To improve specimen classification

11        accuracy, our framework uses multimodal fusion models that combine image data with

12        DNA-based assemblage data. To refine the taxonomic granularity of the model's

13        classifications, our framework cross-references the classifications with DNA

14        metabarcoding detections from bulk samples. We demonstrated this framework using a

15        continental-scale, invertebrate bycatch dataset collected by the National Ecological

16        Observatory Network. The dataset included 17 taxa spanning three phyla (Annelida,

17        Arthropoda, and Mollusca), with the finest starting taxonomic granularity of these taxa

18        being order-level.

19     3.   Using this framework, we reached a classification accuracy of 79.6% across the 17 taxa

20        using real DNA assemblage data, and 83.6% when the assemblage data was "error-free",

21        resulting in a 2.2% and 6.2% increase in accuracy when compared to a model trained

22        using only images. After cross-referencing with the DNA metabarcoding detections, we

23         improved taxonomic granularity in up to 72.2% of classifications, with up to 5.7%

24         reaching species-level.

25    4.  By providing computer vision models with coincident DNA assemblage data, and

26         refining individual classifications using DNA metabarcoding detections, our framework

27         has the potential to greatly expand the capabilities of biological computer vision

28         classifiers. This framework allows computer vision classifiers to infer taxonomically

29         fine-grained classifications when it would otherwise be difficult or impossible due to

30         challenges of morphologic similarity or data scarcity. This framework is not limited to

31         terrestrial invertebrates and could be applied in any instance where image and DNA

32         metabarcoding data are concurrently collected.

## 1. Introduction

Computer vision has the potential to transform invertebrate ecology by automating estimations of invertebrate abundance, biomass, and diversity (Høye *et al.*, 2021; Schneider *et al.*, 2022). However, accurately classifying invertebrate species using computer vision is challenging. This is partly due to the sheer diversity of invertebrates, as there are an estimated 7.5 million (~1.5 million named) terrestrial invertebrates species globally (Stork, 2018). This has led most invertebrate classification models to opt for coarser taxonomic granularity (e.g. order-level instead of species-level classifications) with relatively few unique classification groups (usually <50; (Ärje *et al.*, 2020; Blair *et al.*, 2022; Schneider *et al.*, 2022). However, ecology studies can involve hundreds or thousands of species, which poses a challenge for simpler machine vision techniques.

One way computer vision models have overcome the challenge of handling many thousands or millions of classification labels is by including additional data modalities such as contextual metadata (e.g. collection location) in computer vision models. The mobile app iNaturalist uses this spatiotemporal data in combination with user-submitted photos to classify nearly 80,000 taxa across the tree of life (Leary *et al.*, 2023). Other studies conducted on a smaller scale have also found substantial improvements to classification accuracy with multimodal models that include both metadata and images compared to image-only models (Berg *et al.*, 2014; Terry, Roy and August, 2020; Blair *et al.*, 2022). However, despite the potential improvements in accuracy, there are several pitfalls to consider when including spatiotemporal metadata in a computer vision model. For one, spatiotemporal metadata is a lagging indicator of species habitat occupancy (i.e. the presence or absence of a species at a given place and time), and as such it is susceptible to data drift over time (Friedland, 2024). That is, spatiotemporal distributions of taxa

56    change over time, but computer vision models can only learn from past data. Unless a computer

57    vision model is updated frequently with more recent data, the species range distributions it has

58    learned may quickly become outdated. Finally, when dealing with many machine learning

59    classes, spatiotemporal metadata does not solve the challenge of gathering enough training data

60    to sufficiently train a computer vision model (Beery, Van Horn and Perona, 2018; Beery *et al.*,

61    2020). In short, studies that incorporate spatiotemporal metadata have shown that supplemental,

62    non-visual data can improve ecological computer vision models, but spatiotemporal metadata

63    itself has several potential drawbacks. In this study, we leverage an alternative data stream that

64    does not pose the same challenges associated with spatiotemporal metadata: DNA

65    metabarcoding.

66    DNA metabarcoding is an established tool in ecological research that allows for multiple species

67    to be identified from a single sample using high-throughput sequencing (Deiner et al., 2017; Liu

68    et al., 2020; Taberlet et al., 2012). Using this method, DNA can be collected from the

69    environment (eDNA) or from preservative media (e.g. ethanol in insect bycatch samples),

70    sequenced, and then used to infer ecological metrics such as species richness and community

71    composition (Wood *et al.*, 2017; Marquina *et al.*, 2019; Weiser *et al.*, 2022). Due to its improved

72    cost-effectiveness, DNA metabarcoding is becoming more frequently used in large-scale studies

73    where traditional morphological identification techniques cannot keep up financially or

74    logistically (Liu et al., 2020). However, despite being an excellent tool for detecting occurrence

75    at fine taxonomic granularity (even below species-level; Stewart & Taylor, 2020; Wilcox et al.,

76    2015), DNA metabarcoding cannot be used to reliably estimate species abundance or biomass

77    (Bista *et al.*, 2018; Lamb *et al.*, 2019). Instead, eDNA metabarcoding is more suitable for binary

78    presence/absence detections of species. Additionally, while DNA metabarcoding is generally

79    reliable for taxonomic identifications, it is not exempt from false positive and false negative

80    detections (Guillera-Arroita *et al.*, 2017). Some examples of how this may occur include DNA

81    contamination and primer mis-priming (false-positives), or DNA degradation and insufficient

82    sampling effort (false-negatives) (Guillera-Arroita *et al.*, 2017; Liu *et al.*, 2020). Therefore,

83    while DNA metabarcoding offers considerable advantages for biodiversity assessment (e.g.,

84    species inventories, species richness) its limitations often necessitate the use of complementary

85    indicators such as visual observations for other metrics (e.g., abundance, biomass) (Schneider et

86    al., 2022).

87    Given DNA metabarcoding's ability to produce reliable fine-scale community composition data,

88    and computer vision's ability to measure abundance and biomass at coarse taxonomic

89    granularity, several studies have called for a synergistic classification pipeline that takes

90    advantage of the strengths of each tool (Badirli et al., 2023; Schneider et al., 2022; Sys et al.,

91    2022). In theory, such a pipeline could leverage DNA metabarcoding's fine taxonomic

92    granularity against computer vision's ability to infer specimen-level characteristics (identity,

93    morphology, etc.) to make ecological inferences that would not be possible using either data

94    stream on their own. DNA might also be a favourable alternative to spatiotemporal metadata, as

95    it is a more direct and coincident indicator of species habitat occupancy, likely making it more

96    resistant to data drift over time (Taberlet *et al.*, 2018). Despite the potential benefits of

97    multimodal image-DNA classification models for ecological research, few studies have explored

98    this approach. Additionally, proposed hybrid classification pipelines either leave the DNA and

99    image data streams separate (Sys *et al.*, 2022), or sequence specimens individually, and thus do

100   not take advantage of metabarcoding's ability to process bulk samples (Badirli *et al.*, 2023).

101    Here we present a novel approach for identifying and classifying invertebrate taxa which

102    integrates DNA metabarcoding and computer vision. The objective of this hybrid approach is to

103    improve the accuracy and taxonomic granularity of computer vision classifications by adding

104    concurrent community assemblage data derived from DNA metabarcoding into a bulk specimen

105    classification pipeline (Figure 1). The combination of DNA and image data occurs twice

106    throughout the pipeline: first during classification inference in the computer vision model, and

107    then again as a post-processing step for the model's classifications. While developing this

108    approach, we ask two primary questions: (1) How does error in DNA metabarcoding data affect

109    the accuracy of the computer vision classification model? (2) What are the strengths and

110    limitations of different taxonomic granularity refinement methods? In addition to the case study

111    we present here, we have also developed a GitHub repository to allow this framework to easily

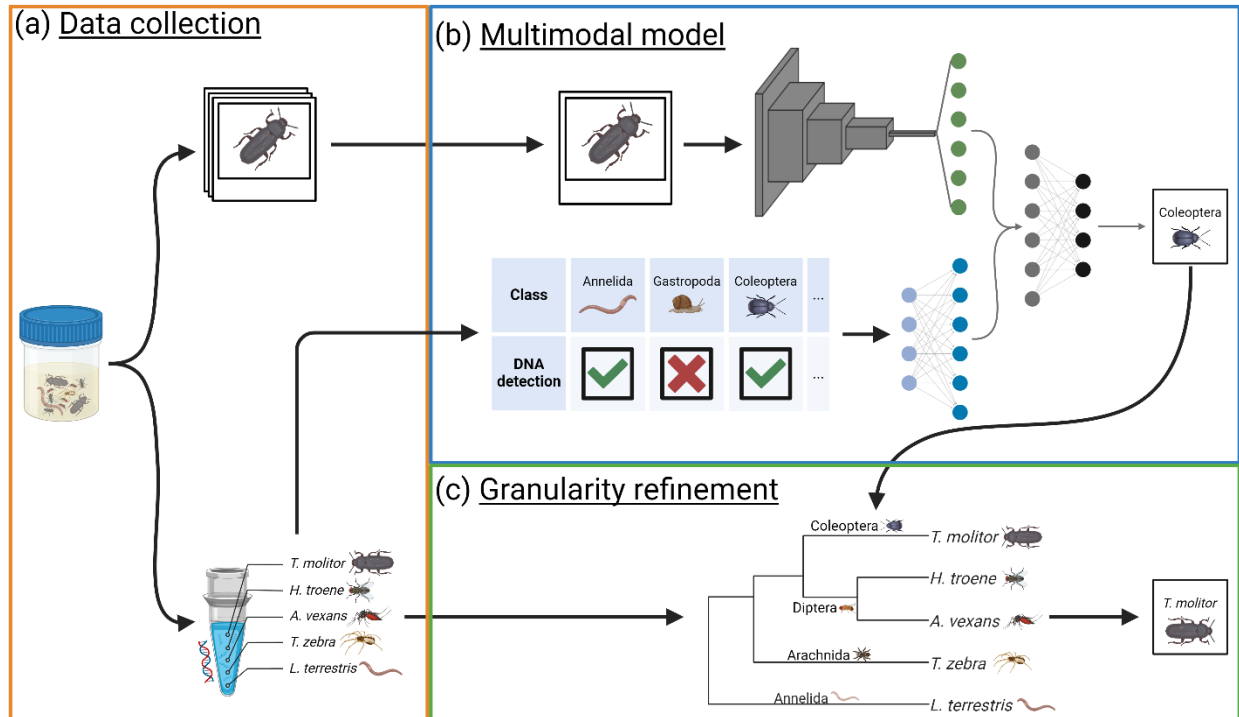112    be adapted to other study systems (Blair, 2024).

Figure 1: Our framework for combining computer vision and DNA metabarcoding to improve the accuracy and taxonomic granularity of classifications. (a) Images and DNA metabarcoding data are collected concurrently from bulk samples. (b) Images and DNA assemblage data are used as input for a multimodal classification model. The features of the image input are extracted using a convolutional neural network. The DNA assemblage data provides presence/absence information for the model's known classes and is input as a binary vector into a dense neural network. The image features and the DNA features are concatenated and processed to produce a final classification. (c) By interpreting the DNA metabarcoding detections hierarchically and cross-referencing them with the model's classifications, the taxonomic granularity of the classifications can be refined.

## 2. Methods

**2.1 Data collection**

2.1.1 Specimen collection

Each year, the National Ecological Observatory Network (NEON) performs standardized pitfall

trap array sampling across the United States, including Alaska, Hawaii, and Puerto Rico

(Hoekman et al. 2017). The focal taxa of the pitfall trap array project are ground beetles

(Coleoptera: Carabidae), which are collected, identified, and counted by NEON staff members

7

130 once every two weeks during the growing season (defined as "the weeks when average minimum

131 temperatures exceed 4 ℃ for 10 days and ending when temperatures remain below 4 ℃ for the

132 same period", Kaspari et al., 2022). The remaining pitfall trap contents are set aside as

133 'Invertebrate Bycatch' and archived in 95% ethanol-filled 50 mL falcon tubes. Hereon, a single

134 collection period from a pitfall trap plot is referred to as a "sampling event".

135 The invertebrate bycatch specimens used in this research were taken from 56 NEON trap plots

136 from 27 sites (usually two plots per site; Figure 2, Figure S.1). Generally, we used three

137 sampling events per plot, selected at the beginning, middle, and end of each site's growing

138 season. This resulted in a total of 150 sampling events. All sampling events used here were

139 collected in 2016 and processed in 2019. The focus of this project was to classify the invertebrate

140 bycatch, so ground beetles and non-invertebrate specimens were not considered.
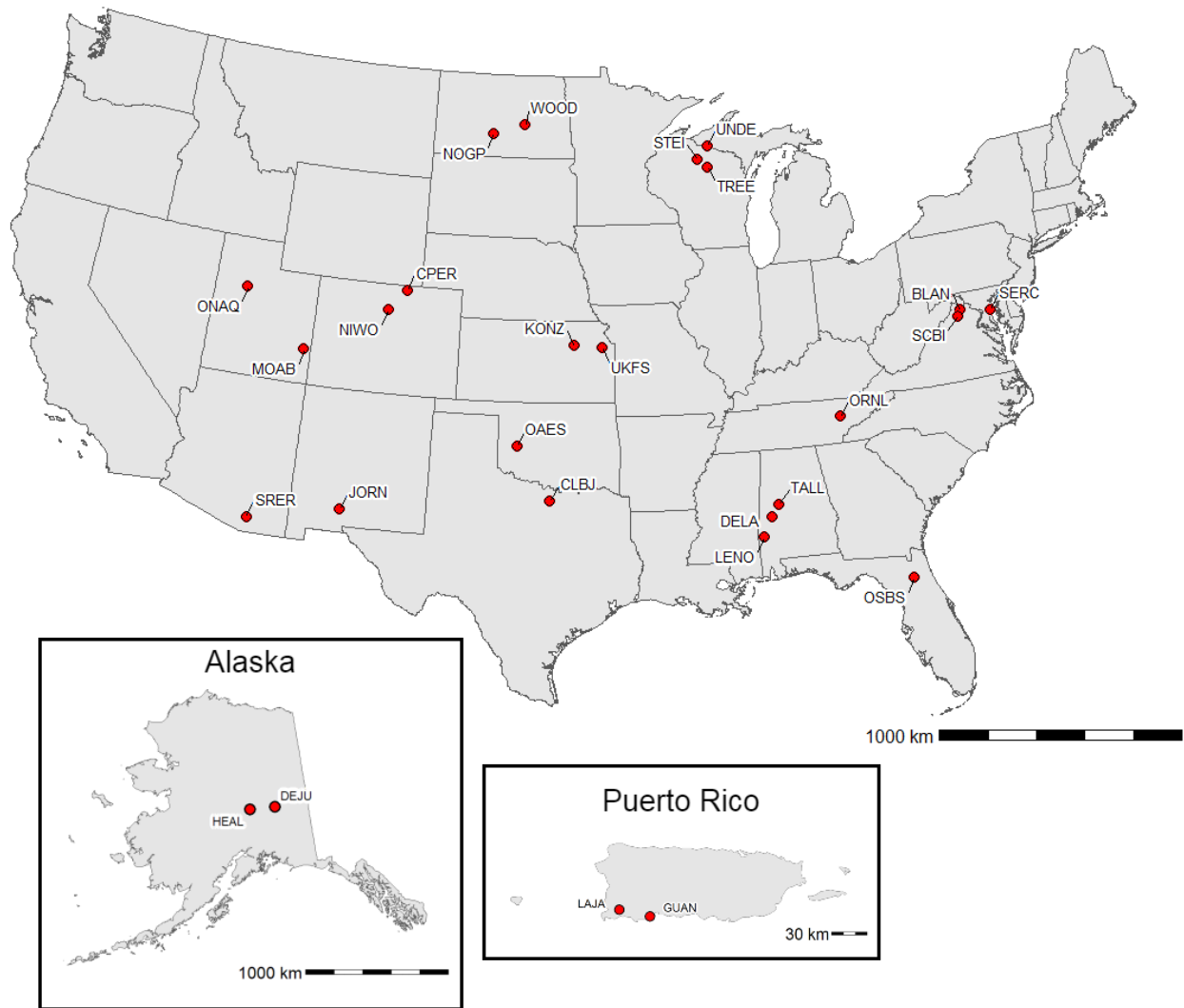
141
142   Figure 2: Map of the 27 NEON sampling sites used in this study. The sites are labelled with their
143   abbreviated names.

144   <u>2.1.2 Imaging</u>

145   The contents of each 50mL falcon tube were spread out across a 20.32 cm ✖ 30.48 cm (8" ✖

146   12") white ceramic tile and photographed at a resolution of 729 pixels per mm$^2$, as described by

147   Weiser et al., 2021 (Figure S.2). Using the FIJI implementation of ImageJ (Schindelin et al.,

148   2012), each specimen was detected and cropped to its bounding box to produce a final image.

149   <u>2.1.3 DNA extraction and metabarcoding</u>

150    The DNA metabarcoding data used in this study was collected for Weiser et al., 2022, which

151    used the same sampling events described in Section 0. In brief, DNA metabarcoding was

152    conducted on a per-tube basis (Figure S.1). Ethanol from each falcon tube was filtered

153    individually (i.e., one filter per tube) and DNA was extracted from the filters using established

154    protocols (Weiser et al., 2022). The cytochrome c oxidase I (COI) barcode region (141-254 base

155    pairs) was then amplified using a two-step polymerase chain reaction (PCR) protocol and

156    sequenced on an Illumina MiSeq. Three COI primers were used: 157, LCO, and Lep (Rennstam

157    Rubbmark et al., 2018; Hajibabaei et al., 2019; Weiser et al., 2022). Sequences were clustered

158    into Operational Taxonomic Units (OTUs) and each OTU was assigned a taxonomic

159    classification using NCBI BLASTn (Altschul *et al.*, 1990) and Integrated Taxonomic

160    Information System (ITIS) (U.S. Geological Survey, 2013). Only sequences with $\geq 97\%$

161    similarity between the OTU consensus sequence and the BLASTn search were used. See Weiser

162    et al., 2022 for the full DNA extraction and metabarcoding methods.

163    In total, across all sampling events, there were 10,212 DNA metabarcoding detections. To align

164    the DNA data with the imaging data, we removed any DNA detections from sampling events not

165    included in the image dataset, as well as duplicate detections (i.e. multiple detections of the same

166    taxon in a single sampling event, for example due to amplification using multiple primers). This

167    yielded a final DNA metabarcoding dataset with 3,361 detections and 1,212 unique taxa,

168    primarily consisting of family (369 detections; 85 unique), genus (468 detections; 183 unique),

169    and species-level (2,471 detections; 922 unique) detections.

**2.2 Data and labelling**

171 <u>2.2.1 Computer vision class labels</u>

172 The taxonomic scope of the image and DNA metabarcoding data spanned three invertebrate

173 phyla: Annelida, Arthropoda, and Mollusca. The specimen images were labelled by a single

174 technician to the best of their ability (as described in Blair *et al.*, 2022). The final labels used for

175 our study ranged from order to phylum-level. Classes with a taxonomic granularity coarser than

176 order-level but with no subtaxa present in the dataset (e.g. Phylum: Annelida) were included.

177 Specimens labelled as nested classes at a taxonomic granularity coarser than order-level

178 (Phylum: Arthropoda, Class: Insecta, and Class: Arachnida) were excluded. Classes with fewer

179 than 100 specimens in the dataset were also excluded. This resulted in a final image dataset with

180 a total of 36,988 specimens across 17 classes (13 orders and one subclass, class, subphylum, and

181 phylum; Figure 3).

1    Annelida
2    Acari
3    Araneae
4    Opilioacarida
5    Opiliones
6    Collembola
7    Blattodea
8    Coleoptera
9    Diptera
10   Hemiptera
11   Hymenoptera
12   Lepidoptera
13   Orthoptera
14   Zygentoma
15   Isopoda
16   Myriapoda
17   Gastropoda

5 mm

182

Figure 3: Collage of photographs of all invertebrate classes used in the training dataset (n = 17). The taxonomic granularity of the classes ranges from order to phylum. Specimens were cropped from their original photographs, and the background was removed. Relative scale of each specimen is conserved.

186   2.2.2 Hierarchical labels

187   Hierarchical labels containing taxonomic information at multiple levels and were assigned to

188   images and sampling events. Image-based hierarchical labels contained information at six levels

189   from phylum to order-level for individual specimens (Table 1). DNA-based hierarchical labels

190   contained information at 13 levels from phylum to species for all DNA detection in each

191   sampling event (Table 2). The levels in the DNA-based hierarchical labels were phylum,

192   subphylum, class, subclass, superorder, order, suborder, infraorder, superfamily, family,

193   subfamily, genus, and species. We used these labels as part of the taxonomic granularity

194   improvement process detailed in Section 2.5.

195   Table 1: Three examples image-based hierarchical labels (a,b,c). (a) The hierarchical label for Blattodea
196   with all taxonomic levels between phylum and order filled. (b) Some taxa do not have names for every
197   taxonomic sub-level. In these cases, the name from the preceding, finer-grain level is used. In this
198   example, Coleoptera does not belong to any superorder, so "Coleoptera" is used as a placeholder
199   superorder name. (c) Classes with a taxonomic granularity above order-level used indeterminate ("indet.")
200   labels at all remaining taxonomic levels. In this example, "Annelida indet." is used for all levels below
201   phylum.

| Example | Phylum | Subphylum | Class | Subclass | Superorder | Order |
|---------|--------|-----------|-------|----------|------------|-------|
| (a) | Arthropoda | Hexapoda | Insecta | Pterygota | Polyneuroptera | Blattodea |
| (b) | Arthropoda | Hexapoda | Insecta | Pterygota | *Coleoptera* | Coleoptera |
| (c) | Annelida | *Annelida indet.* | *Annelida indet.* | *Annelida indet.* | *Annelida indet.* | *Annelida indet.* |

202

Table 2: Simplified example of a DNA multi-class hierarchical label for a sampling event. Our DNA hierarchical labels used 13 taxonomic levels from phylum to species, but for the sake of space only major taxonomic levels are shown here.

| Phylum | Class | Order | Family | Genus | Species |
|--------|-------|-------|--------|-------|---------|
| Arthropoda | Insecta | Diptera | Culicidae | *Aedes* | *A. vexans* |
| | | | | *Culex* | *C. salinarius* |
| | | Orthoptera | Gryllidae | *Gryllus* | *G. veletis* |
| | Arachnida | Araneae | Lycosidae | *Schizocosa* | *S. humilis* |

## 2.2.3 Binary assemblage data

Two sets of binary assemblage data were recorded for each sampling event: one using detections from the image labels and one using detections from the DNA class labels. DNA class labels were based on the same 17 classes used by the computer vision model, and followed the same naming scheme described in Section 0. The assemblage data was multi-hot encoded (Goodfellow, Yoshua and Courville, 2016). That is, each sampling event was assigned a 17 element long binary vector where each element corresponded to a computer vision class. If the class was detected in the sampling event, its element was assigned a score of 1, whereas it was assigned a score of 0 if it was not detected. This data was used as input for some classification models described in Section 0.

## 2.3 Training and testing data split

Quasi-replication occurs in machine learning datasets when the same or very similar data occur in both the training and testing datasets. This violates the assumption of independence between training and testing data and should be avoided to make valid inferences on the test data. The

220 DNA-based assemblage data presented a quasi-replication risk, as specimens from the same

221 sampling event would have the same assemblage data. To avoid quasi-replication, we split the

222 training and testing such that all specimens from a given sampling event were only included in

223 either the training or testing data. We set our target training:testing ratio to 85:15, and we

224 randomly added sampling events to the test dataset until the test dataset contained >15% of the

225 total number of specimens. The final train:test split was 31,381: 5,617 specimens and 122:28

226 sampling events.

227 **2.4 Classification models**

228 The objective of all the classification models was to accurately classify the class labels of

229 individual specimen photos. Classification masks (Section 0) and data fusion approaches

230 (Section 0) were used to assess how classification accuracy changes when DNA-based

231 assemblage data was added to the specimen classification pipeline. Performance of each

232 classification model was assessed using the original image labels as a ground truth. All code

233 required for running these models can be found in the "Model_Scripts" subdirectory of our

234 GitHub repository (Blair, 2024).

235 <u>2.4.1 Baseline model</u>

236 To evaluate model performance in the absence of DNA-based assemblage data, we trained a

237 ResNet-50 (He et al., 2016) as a baseline model using only image data. The model was pre-

238 trained using the ImageNet weights from He et al. (2016), and then fine-tuned using the NEON

239 invertebrate bycatch image data. The ImageNet classification layer was removed and replaced

240 with a new classification layer for our 17 classes.

241    <u>2.4.2 Classification masks</u>

242    *2.4.2.1 Naïve mask*

243    For our first experiment in combining the image and DNA-based assemblage data, we used what

244    we call a "naïve mask" on the baseline model's test data softmax classification layer outputs.

245    Using the naïve mask, the softmax layer values for each test data specimen were multiplied by

246    their sampling event's binary DNA-based assemblage data. Thus, any classes not detected by the

247    DNA metabarcoding in a given sampling event had their respective softmax values set to 0,

248    whereas the remaining classes were unaffected. After applying the mask, the class label with the

249    highest softmax value was used as the classification for a given specimen.

250    *2.4.2.2 Weighted mask*

251    "Hard" masks like the naïve mask, which set the softmax values of undetected classes to 0,

252    assume the DNA metabarcoding data is error-free. However, in reality, DNA metabarcoding can

253    have false positive and/or false negative detections (Taberlet *et al.*, 2018). To create a "softer"

254    version of the naïve mask, we also created a "weighted mask", which allows classes not detected

255    by the DNA metabarcoding to still be classified.

256    The weights for the weighted mask were calculated using the DNA's true positive rate

257    (precision) and false negative rate (1 - recall) for each class. The DNA metabarcoding precision

258    and recall were calculated by comparing the DNA-based assemblage labels to the image-based

259    assemblage labels, using the image-based assemblage labels as a ground truth. To create the

260    weighted mask, we assigned a class's precision for positive DNA detections, and the class's false

261    negative rate for negative detections. The application of the weighted mask was the same as the

262 naïve mask: The softmax values for each test data specimen were multiplied by their sampling

263 event's weighted mask values, and the new top-1 class was used as the classification.

264 *2.4.2.3 Image-based mask*

265 To simulate a scenario where the DNA detections were in perfect alignment with the image-

266 based detections, we applied the image-based binary assemblage data to the softmax values of

267 the baseline model. We did this to provide an upper-bound for the classification mask accuracy,

268 and to understand how DNA detection accuracy impacts specimen classification accuracy when

269 using classification masks.

270 2.4.3 Multimodal fusion

271 To allow the classification models to learn patterns from the DNA data, we trained multimodal

272 models that combined images with binary DNA-based assemblage data using intermediate fusion

273 (Boulahia *et al.*, 2021) (Figure 4). In these multimodal models, the image and assemblage data

274 for a given specimen were fed to the model separately. The images were fed through the ResNet-

275 50 architecture, ultimately producing a flat feature layer. DNA-based assemblage data was paired

276 with individual specimens based on their sampling event (the same approach as the classification

277 masks described in Section 2.4.2). This data is fed through a single fully-connected layer. The

278 flattened image layer and fully-connected DNA-based assemblage layer were then concatenated

279 (i.e. fused) and passed through another fully-connected layer before reaching the final

280 classification layer.

281 To understand the effect that DNA detection accuracy has on classification performance, we ran

282 three versions of the multimodal model using different types of assemblage data as input: (1)

283 using the DNA-based assemblage data, (2) using the image-based assemblage data, and (3) using

284 'zero-filled' assemblage data (all values in the assemblage data are set to zero). In all three

285 experiments the training and testing datasets used the same assemblage data type (i.e. DNA-

286 based, image-based, or zero-filled). All three experiments used the same overall model

287 architecture as described in Figure 4. The purpose of the image-based assemblage experiment

288 was to simulate the results of a model where the DNA detections perfectly aligned with the

289 ground truth labels. The purpose of the zero-filled assemblage experiment was to control for

290 differences in model architecture when comparing the multimodal models to models trained

291 without DNA-based assemblage data, as the zero-filled data would provide no informative value

292 to the model. The zero-filled assemblage data had the same dimensions as the other assemblage

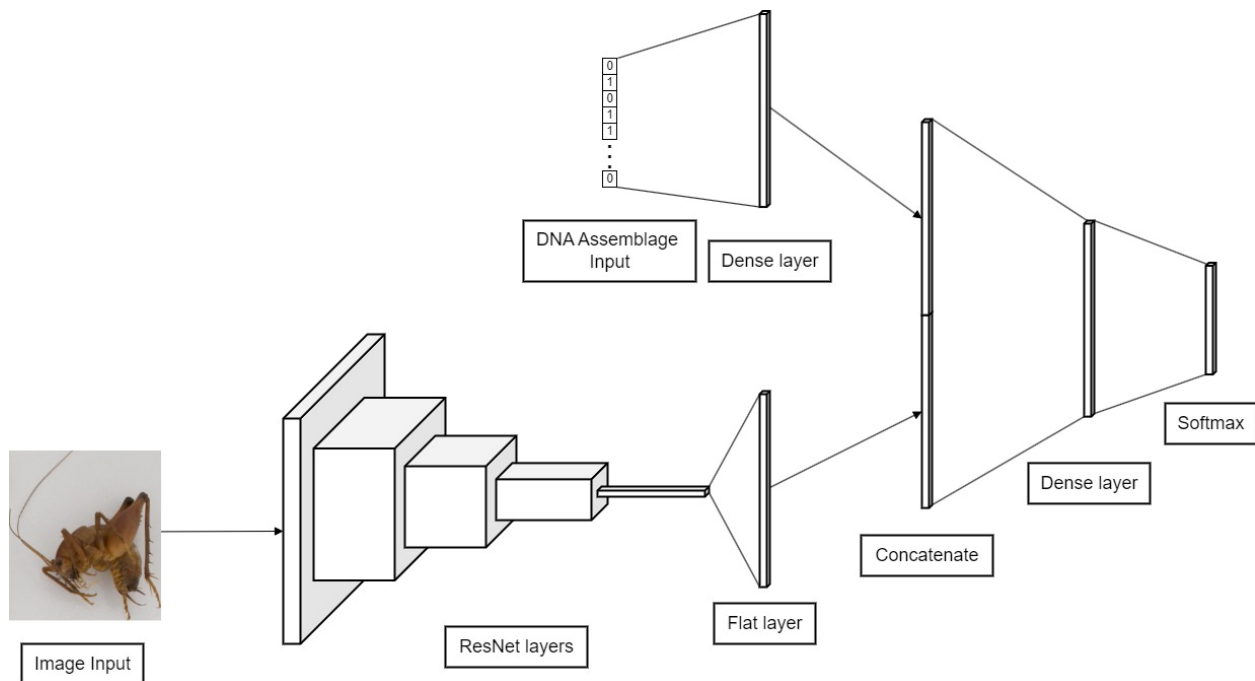293 data (17 values for each sampling event).



295 Figure 4: The described classification model architecture that combines image data with DNA-based
296 binary assemblage data. Images are fed through ResNet convolutional layers to produce a flat feature
297 layer, and DNA assemblage data produces a fully connected dense layer. The flat layer and dense layer
298 are then concatenated and passed through one more dense layer before final classification in the softmax

299     layer. The visual proportions of each layer have been simplified to ease interpretation and are not meant
300     to be interpreted as 1:1 representations of the exact layer sizes.

**2.5 Refining taxonomic granularity using DNA-based assemblage data**

302     To take advantage of the relatively fine taxonomic granularity of the raw DNA detections

303     compared to the computer vision classes, we refined the taxonomic granularity of the DNA

304     fusion model classifications by cross-referencing them with the DNA detections. In cases where

305     the model classifications and DNA detections agreed on the presence of a class, the granularity

306     of the classification improved until the number of subtaxa detected by the DNA metabarcoding

307     was greater than 1 or the granularity reached species-level, as that was the finest granularity

308     reported by the DNA (Figure 5a,b). In cases where the model classifications and DNA detections

309     disagreed on the presence of the class classified by the model, we implemented two methods

310     which we call "model-biased" and "DNA-biased". The model-biased method is simple. In cases

311     where the model and DNA metabarcoding disagreed, the model classification remained

312     unchanged (Figure 5c). When using the DNA-biased method, we compared the hierarchical label

313     of the classified specimen to the hierarchical labels of the sampling event as determined by the

314     DNA metabarcoding (Figure 5d, Table 1,

315    Table **2**). Starting from the original classification level, we coarsened the granularity of the label

316    until an agreement between the model and DNA metabarcoding was reached. The taxonomic

317    name at this level became an intermediate label, and the taxonomic granularity of the label was

318    refined until the number of subtaxa detected by the DNA metabarcoding was greater than 1 or

319    the granularity reached species-level.

320    All code required for running these methods can be found in the "Granularity_Refinement"

321    subdirectory of our GitHub repository (Blair, 2024).
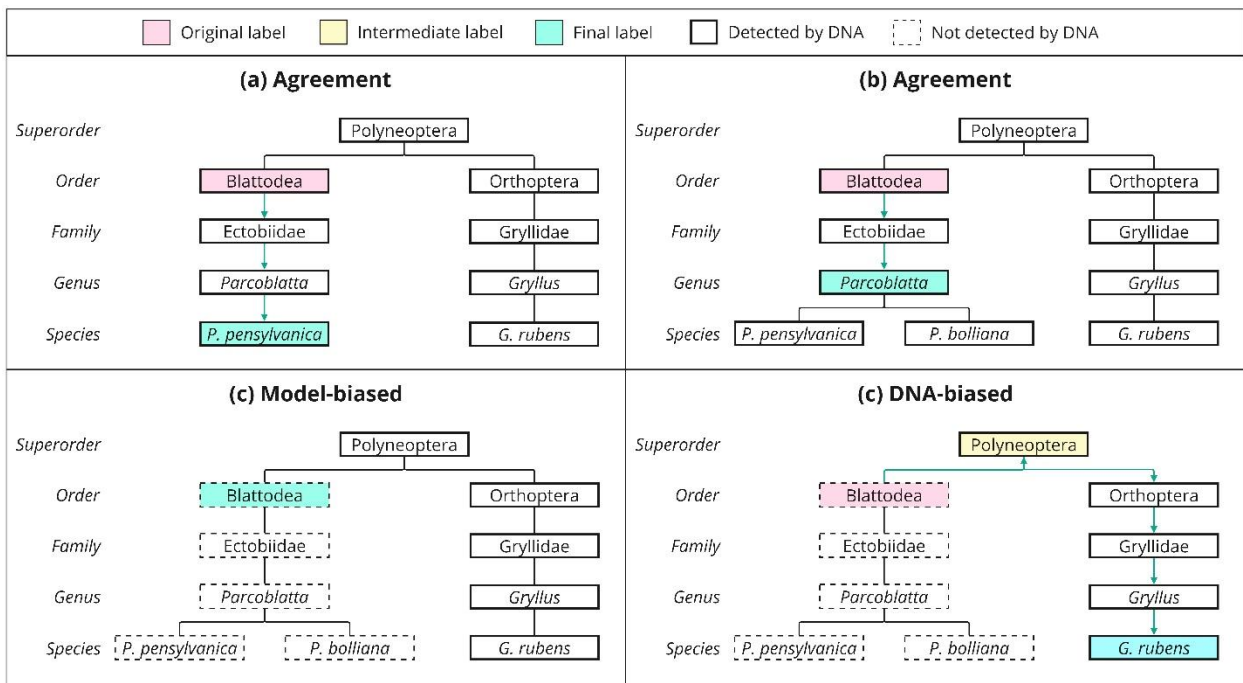


322
323    Figure 5: Four methods of changing a label's granularity using DNA detections. (a,b) When the
324    classification label and DNA detections are in agreement about the presence of a class, granularity will be
325    refined until the number of subtaxa detected by the DNA metabarcoding is > 1 or the classification
326    reaches species level. (c) Under the model-biased approach, when the classification label and DNA
327    metabarcoding do not agree on the presence of a class, the classification label remains unchanged. (d)
328    Under the DNA-biased approach, when the classification label and DNA metabarcoding do not agree on
329    the presence of a class, the granularity of the classification label is coarsened until a DNA detection for
330    the class is found ("intermediate label"). The granularity of the intermediate label is then refined using the
331    same rules as (a,b).

## 3. Results

### 3.1 DNA metabarcoding precision and recall

To generate the weights for the weighted mask, we calculated the DNA-based assemblage data's precision and recall for each class using the image-based assemblage data as the ground truth. This calculation only included sampling events from the training dataset. Across the 17 predicted classes, recall ranged from 0.905 to 0.000, with an average recall of 0.570. The weight for a negative detection was 1 - recall, so negative detection weights ranged from 1.000 to 0.095, with an average of 0.430. Two classes (Opilioacarida and Zygentoma) were never detected by the DNA. Across the 15 classes detected by the DNA, the average precision was 0.761. The precision and recall values per class are reported in Table S.1.

### 3.2 Classification accuracy

Compared to the baseline model, the DNA assemblage fusion model improved accuracy by 2.2% (79.6% vs 77.4%; Table 3). However, the DNA assemblage fusion model's accuracy was 1.0% lower than the zero-filled model's accuracy (80.6%). This suggests that some of the performance improvements seen in the fusion models could come from changes to the model architecture. The naïve classification mask recorded the lowest classification metrics, with an accuracy 22.6% below the baseline model (54.8%), and a top-3 accuracy of only 64.0%.

Experiments with image-based assemblage data performed better, with the fusion model reaching an average accuracy of 83.6% and a balanced accuracy (i.e. macro-averaged recall) of 0.713 (Table 4). The image-based naive mask accuracy was 25.8% better than the DNA-based

21

352   naive mask, and 3.2% better than the baseline model. It also had the highest top-3 accuracy

353   across all experiments at 93.2%.

354   Table 3: Performance metrics for experiments trained using DNA-based assemblage data, other than the
355   baseline model, which was trained only using images, and the 'zero-filled' experiment, which replaced all
356   assemblage data values with zero to control for the impact of model architecture. Underlined scores
357   indicate they are the highest for a given metric.

| Experiment | Accuracy | Balanced Accuracy | Top-3 Accuracy |
|---|---|---|---|
| Baseline | 0.774 | 0.674 | 0.952 |
| Naive mask | 0.548 | 0.509 | 0.640 |
| Weighted mask | 0.764 | 0.666 | 0.950 |
| Fusion (DNA) | 0.796 | <u>0.680</u> | <u>0.957</u> |
| Fusion (zero-filled) | <u>0.806</u> | 0.670 | 0.952 |

358   Table 4: Performance metrics for the image-based assemblage data experiments. These experiments used
359   binary assemblage data taken from the ground truth specimen labels. Underlined scores indicate they are
360   the highest for a given metric.

| Experiment | Accuracy | Balanced Accuracy | Top-3 Accuracy |
|---|---|---|---|
| Mask | 0.806 | 0.690 | 0.959 |
| Fusion | <u>0.836</u> | <u>0.713</u> | <u>0.963</u> |

361   **3.3 Taxonomic granularity**

362   Of the DNA fusion model classifications, 68.2% (3833/5617) were present in their

363   corresponding sampling event's DNA assemblage data. Nonetheless, both approaches for

364   handling disagreements between the DNA and classification model (model-biased or DNA-

365   biased) improved the average taxonomic granularity of the classifications (Figure 6). Despite

366   starting with no classifications finer than order-level, the DNA-biased approach resulted in 5.7%

367   of classifications improving to species-level and the model-biased approach resulted in 5.1% of

368   the classifications improving to species-level. The DNA-biased approach was more effective

369    overall at refining the granularity of classifications, with 72.2% of classifications becoming finer

370    than their original classification, and 43.1% reaching at least family-level. In the model-biased

371    approach, 51.7% of classifications improved their granularity, and 31.0% reached family-level or

372    lower. When looking exclusively at classifications where the DNA assemblage data and fusion

373    model classifications agreed on the presence of a class, 7.5% reached species level, 45.4%

374    reached family-level or lower, and 75.7% became finer than their original classifications. Both

375    the DNA-biased and model-biased approaches produced 89 unique final labels (Table S2).
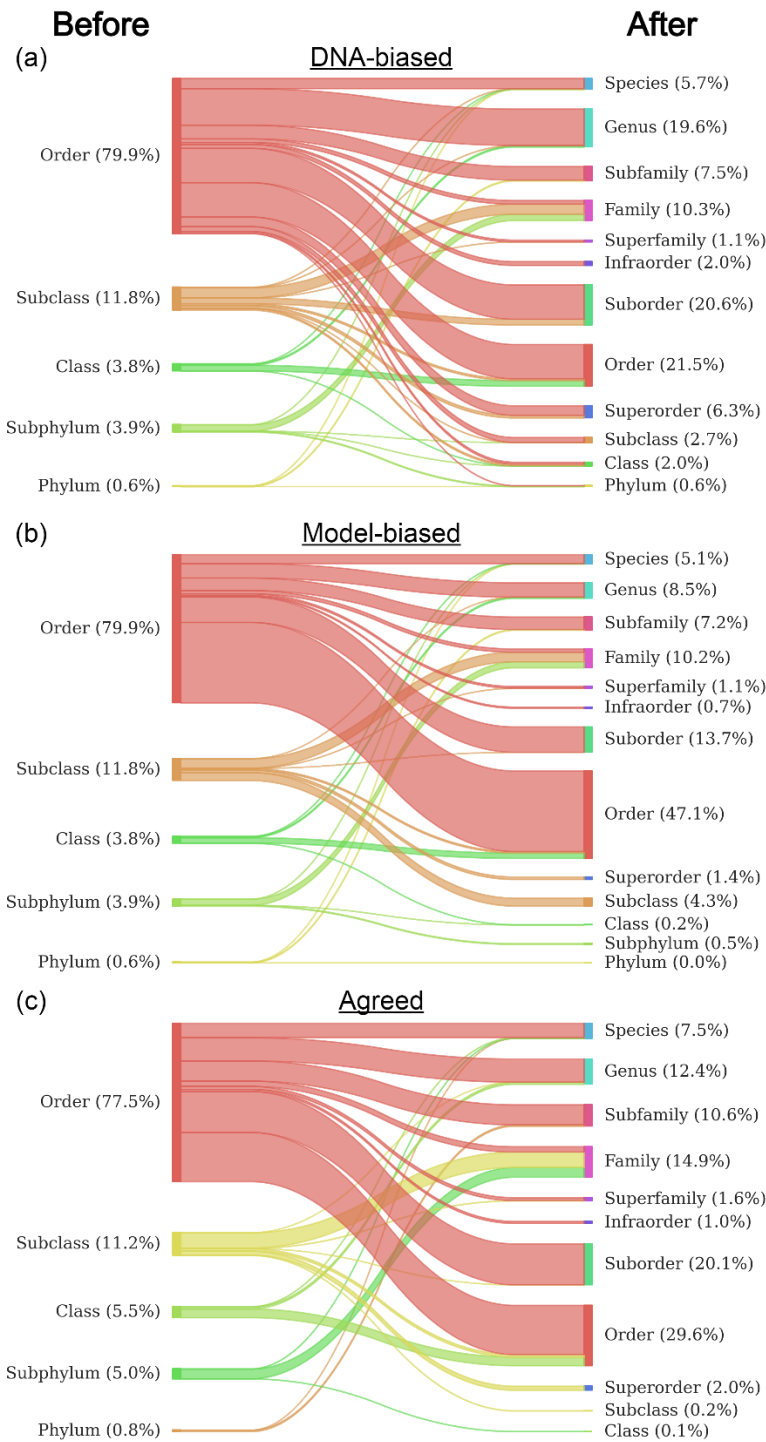
Figure 6: Sankey diagrams showing the change in taxonomic granularity before (left) and after (right) cross-referencing labels with the DNA detections. (a) DNA-biased approach. (b) Model-biased approach. (c) Results when the model classification and DNA detections agree on the presence of the labelled class.

380 **4 Discussion**

381 Here we show that combining concurrent DNA-based assemblage data with computer vision can

382 improve the accuracy and taxonomic granularity of computer vision classifications. Unlike most

383 classification pipeline enhancements which only focus on improving the ability to classify

384 classes on which it was trained ("known classes"), this approach adds the ability for the pipeline

385 to infer classifications beyond the model's usual taxonomic scope ("unknown classes"). To

386 thoroughly explore the benefits and implications of this hybrid approach, we focused on the

387 following two research questions.

388 **4.1 How does DNA metabarcoding accuracy affect specimen classification accuracy?**

389 The effect of metabarcoding accuracy on classification accuracy differed between the

390 classification masks and the multimodal fusion models, illuminating a key difference between

391 them: classification masks (as used in this study) do not take class co-occurrence into

392 consideration, whereas the multimodal models do. Put another way, in a classification mask the

393 only factor that directly influences the weight given to a class is the presence or absence of the

394 class itself. Conversely, the neural networks of multimodal fusion models—with their fully-

395 connected structure—allow the presence or absence of all classes to holistically influence each

396 class's classification probability. This allows the model to use patterns of class co-occurrence to

397 inform its classification decisions.

398 The different mechanisms used by classification masks and multimodal models are best

399 demonstrated in Table 4, where the assemblage data was derived from the ground-truth image

400 labels. In the classification mask, the model's original classifications were exclusively based on

401 image data, and any classifications that did not match their respective assemblage data were

402     reclassified as the class with the highest softmax score that was present in the specimen's

403     assemblage. Given that the assemblage data was derived from the ground truth labels, this mask

404     acted as a sieve that filtered out any classes that could not possibly be the correct classification

405     based on the assemblage data. As a result, it could only have a positive impact on accuracy.

406     However, despite this, the fusion model still scored higher on all three metrics measured (top-1

407     accuracy, balanced accuracy, and top-3 accuracy). This implies that the fusion model was not

408     just using the assemblage data as a filter, but that it provided additional contextual information

409     (such as class co-occurrence or class exclusion) that further improved accuracy. Thus, due to the

410     fusion model's ability to holistically evaluate occurrence data, and as illustrated through its

411     superior performance compared to classification masks even under ideal conditions, multimodal

412     fusion models are likely to be preferable in most use cases. This conclusion is reinforced by the

413     results of Table 3, where both naïve and weighted masks showed negative effects on all

414     classification performance metrics when the DNA-based assemblage data contained substantial

415     amounts of error.

416     **4.2 What are the strengths and limitations of each granularity refinement method?**

417     Here we proposed two approaches for cross-referencing DNA metabarcoding data to achieve the

418     novel ability of refining the taxonomic granularity of computer vision classifications. The two

419     approaches differ in how they resolve disagreements between the detections of the DNA

420     metabarcoding and computer vision classifications, with the model-biased approach favouring

421     the computer vision classifications, and the DNA-biased approach favouring the DNA

422     detections. As such, each approach has its own set of advantages and limitations.

423    Through the ability to coarsen granularity before refining it, the DNA-biased method can make

424    classifications outside of the taxonomy of the original classification model (Figure S.3).

425    Explained another way, the model-biased approach and traditional hierarchical classifiers (e.g.

426    Badirli *et al.*, 2023) can only adjust classifications "vertically" (i.e. to supertaxa or subtaxa of the

427    original classifications), but the DNA-biased approach can also adjust classifications "laterally"

428    to out-of-distribution taxa through a combination of classification coarsening and refining.

429    Classification of out-of-distribution taxa is usually only possible using feature embedding

430    learning methods such as zero-shot learning (Badirli *et al.*, 2021). An illustrative example of this

431    comes from the DNA-biased approach's detection of taxa within the insect order Psocodea (e.g.

432    *Valenzuela flavidus*; Table S.2). As Psocodea was not included as a class in our model, and our

433    model's finest taxonomic granularity was order-level, Psocodea's branch of the taxonomy was

434    only accessible through a "lateral" taxonomic adjustment (Figure S.3). As such, it was only

435    detected by the DNA-biased approach, and not the model-biased approach (Table S.2). In theory,

436    this extends the range of possible classifications to the full taxonomic scope of the genetic

437    reference database being used (e.g. GenBank, Barcode of Life, etc.) (Ratnasingham and Hebert,

438    2007; Sayers *et al.*, 2024). In practice, it is likely best to self-impose limits on how much the

439    DNA-biased method can coarsen granularity. In our case we limited ourselves to phylum, as we

440    were only interested in classifications within our three focal phyla.

441    While it does not have the same potential taxonomic scope of the DNA-biased approach, an

442    advantage of the model-biased approach is that the taxonomic granularity of the final

443    classification cannot be coarser than the original classification. Applied to the DNA fusion model

444    classifications, 9.9% of all classifications became coarser when the DNA-biased method was

445    used (Figure 6). While the DNA-biased method classified more specimens at family-level or

446    finer (43.1% vs 31.0%), the ability to coarsen granularity resulted in more classifications above

447    order-level (11.7% vs 6.4%).

448    Even when granularity does not reach species, classifications that match with the DNA-based

449    assemblage data still provide more information than what is typically output from a classification

450    model. This is because we can also see the number and identity of subtaxa that the specimen

451    could be according to the DNA metabarcoding detections. For example, if the DNA

452    metabarcoding detected three species of the cricket genus *Gryllus* in a sample, we could say the

453    label of a specimen that would otherwise be classified simply as "*Gryllus* indet." is actually one

454    of three possible species of *Gryllus*, as detected by the DNA metabarcoding (e.g. *G.*

455    *pennsylvanicus*, *G. rubens*, or *G. veletis*). This might also be useful for future developments to

456    these methods, as the number of subtaxa detected by the DNA metabarcoding could be used to

457    inform clustering algorithms that separate the specimens into morphotaxa.

458    Of course, the granularity of classifications matters little if they are not accurate. A caveat of our

459    study is that we cannot verify the accuracy of granularity-refined classifications, as they are at

460    lower taxonomic levels than our ground truth (human-classified) labels. However, we know that

461    our classification models were more accurate than the DNA assemblage data when compared to

462    our ground-truth labels. Thus, the DNA-biased method of refining granularity will likely add

463    more error to the classifications than the model-biased approach. When deciding between the

464    two methods, this is likely to be a determining factor: does the computer vision model or DNA

465    metabarcoding contain more error?

## 4.3 Caveats and areas for future exploration

In an applied context, we cannot definitively conclude that image-DNA fusion models as we present here improve specimen classification accuracy. This is primarily due to the high rates of disagreement (or "error") between our DNA metabarcoding detections and image-based detections. When comparing our three fusion experiments, the zero-filled experiment had a top-1 accuracy 1.0% higher than the DNA-based assemblage data experiment, but 3.0% less than the image-based assemblage data experiment. This suggests that in an ideal situation where the DNA-based assemblage data has low amounts of error (i.e. it is more similar to the image-based assemblage data), image-DNA fusion models will positively impact classification accuracy. However, when the DNA-based assemblage data contains substantial error, differences in performance between the baseline and fusion models likely arise from changes in the model's architecture.

Reconciling genetic-based and morphology-based data—the two chief methods for invertebrate biodiversity monitoring—is a pressing need as previous studies have shown that assemblages determined by visual classification usually differ from assemblages determined using DNA metabarcoding. For example, Emmons *et al.* (2023) found that NEON benthic macroinvertebrate samples classified by taxonomists only shared 59% of order-level detections with DNA metabarcoding data derived from homogenized blends of the same samples. Marquina *et al.* (2019) also found that different DNA sampling protocols can produce inconsistent assemblage data, as DNA metabarcoded from ethanol vs homogenized blends of the same samples yielded significantly different assemblage data, with both methods detecting taxa not detected by the other. For the image-DNA fusion methods we propose here to be maximally effective, advances

488 will need to be made in DNA metabarcoding methodology to limit false positive and false

489 negative detections.

490 Beyond DNA metabarcoding accuracy, there are likely other factors that can impact the efficacy

491 of our methods, such as assemblage homogeneity and assemblage specificity. Assemblage

492 homogeneity refers to the variation among assemblages. For example, the zero-filled assemblage

493 experiment used data that was completely homogenous, as every sampling event had the same

494 assemblage data. Assemblage specificity refers to how many unique classes are detected within a

495 sample. A maximally specific assemblage would only detect one class as present, while a

496 minimally specific assemblage would detect every class as present. Reducing assemblage

497 homogeneity and increasing assemblage specificity should yield models with greater

498 classification performance. This is because heterogeneity is required for learnable patterns to

499 emerge in the data, and increased specificity allows more classes to be filtered out by the model.

500 This is partially demonstrated by comparing the results of Blair *et al.* (2020) to the results we

501 present here. In their study, which built classification models for NEON's carabid beetles, the

502 authors applied classification masks to their models based on the detected ground beetle

503 assemblages at each sampling site. On average, 2.93 out of 25 potential species (11.7%) were

504 detected per site, resulting in an accuracy improvement of 10.9% (84.7% → 95.6%) after

505 applying the classification masks. Comparatively, our image-based assemblages detected an

506 average of 9.09 out of 17 potential classes (53.5%) per sampling event. When compared to the

507 baseline model, this resulted in an accuracy improvement of 3.2% when using the classification

508 mask and 6.2% in the fusion model (Table 3, Table 4). Thus, methods to reduce assemblage

509 homogeneity (e.g. finer-grain class labels) and increase assemblage specificity (e.g. fewer

510 specimens per sample) will likely increase the efficacy of image-DNA metabarcoding fusion

511 classification pipelines.

512 **4.4 Broader applications and implications**

513 In this study, we used assemblage data derived from DNA metabarcoding to improve computer

514 vision classifications of terrestrial invertebrates. However, our general framework could be

515 applied to any study system where images and DNA metabarcoding data are collected

516 concurrently. Given that computer vision and DNA metabarcoding are emerging technologies in

517 ecological research, the number of research systems that include one or both is increasing

518 (Pichler and Hartig, 2023; Shea *et al.*, 2023). Projects that use both computer vision and DNA

519 metabarcoding also span a wide range of fauna and ecosystems, including freshwater, marine,

520 and terrestrial vertebrates (Takeuchi *et al.*, 2021; Mas-Carrió *et al.*, 2022; Holm *et al.*, 2023). By

521 leveraging the strengths of computer vision and DNA metabarcoding, our framework could

522 enhance the capabilities of projects like these.

523 Our framework is composed of two primary modules—a multimodal classification model and a

524 classification granularity refinement method—which can be used and modified independently of

525 each other. Our multimodal model requires image data and assemblage data, but the assemblage

526 data does not need to be derived from DNA. In this study we conducted experiments where the

527 DNA-derived assemblage data was substituted with image-derived assemblage data. Likewise,

528 assemblages derived from other sensor data (acoustic, lidar, etc.) could also be used (Gasc *et al.*,

529 2015; Kaplan *et al.*, 2015; Wedding *et al.*, 2019). Conversely, our classification granularity

530 refinement method requires DNA metabarcoding data and individual specimen classifications,

531 but the specimen classifications do not need to be derived from a computer vision model. We

532 encourage future studies to explore different variations and combinations of the modules we

533 present here.

534 Our framework's ability to refine classification granularity, which is typically not possible in

535 computer vision, could improve the feasibility of building broad-scope, fine-grain classification

536 models (e.g. models spanning entire classes or phyla and capable of producing species-level

537 classifications). This typically requires vast amounts of training data, as training examples need

538 to be provided for every species. Using the approach that we present here, classifiers could be

539 trained at coarser taxonomic levels such as order or family and still have the potential to produce

540 species-level classifications. This would decrease the number of classes in the model, and thus

541 data needed to train it, by orders of magnitude. Hence, the synergy between DNA metabarcoding

542 and computer vision outlined in this study paves the way for new possibilities in computer vision

543 classification of taxa, with the potential for improved accuracy and granularity with far less data

544 dependency.

**References**

Altschul, S.F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Ärje, J. *et al.* (2020) 'Automatic image-based identification and biomass estimation of invertebrates', *Methods in Ecology and Evolution*, 11(8), 922–931. https://doi.org/10.1111/2041-210X.13428.

Badirli, S. *et al.* (2021) 'Fine-Grained Zero-Shot Learning with DNA as Side Information', in *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 19352–19362. https://proceedings.neurips.cc/paper_files/paper/2021/hash/a18630ab1c3b9f14454cf70dc711483 4-Abstract.html (Accessed: 24 July 2024).

Badirli, S. *et al.* (2023) 'Classifying the unknown: Insect identification with deep hierarchical Bayesian learning', *Methods in Ecology and Evolution*, 14(6), 1515–1530. https://doi.org/10.1111/2041-210X.14104.

Beery, S. *et al.* (2020) 'Synthetic examples improve generalization for rare classes', in *Proceedings - 2020 IEEE Winter Conference on Applications of Computer Vision, WACV 2020*. https://doi.org/10.1109/WACV45572.2020.9093570.

Beery, S., Van Horn, G. and Perona, P. (2018) 'Recognition in Terra Incognita', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-01270-0_28.

Berg, T. *et al.* (2014) 'Birdsnap: Large-scale fine-grained visual categorization of birds', in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011–2018. https://doi.org/10.1109/CVPR.2014.259.

Bista, I. *et al.* (2018) 'Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples', *Molecular Ecology Resources*, 18(5), 1020–1034. https://doi.org/10.1111/1755-0998.12888.

Blair, J.D. *et al.* (2020) 'Robust and simplified machine learning identification of pitfall trap-collected ground beetles at the continental scale', *Ecology and Evolution*, 10(23), 13143–13153. https://doi.org/10.1002/ece3.6905.

Blair, J.D. *et al.* (2022) 'Embracing imperfection: Machine-assisted invertebrate classification in real-world datasets', *Ecological Informatics*, 72, p. 101896. https://doi.org/10.1016/j.ecoinf.2022.101896.

Blair, J.D. (2024) 'CV-DNA-Hybrid'. Zenodo. https://doi.org/10.5281/ZENODO.13621526.

Boulahia, S.Y. *et al.* (2021) 'Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition', *Machine Vision and Applications*, 32(6), p. 121. https://doi.org/10.1007/s00138-021-01249-8.

580     Deiner, K. *et al.* (2017) 'Environmental DNA metabarcoding: Transforming how we survey
581     animal and plant communities', *Molecular Ecology*, 26(21), 5872–5895.
582     https://doi.org/10.1111/mec.14350.

583     Emmons, S.C. *et al.* (2023) 'DNA metabarcoding captures different macroinvertebrate
584     biodiversity than morphological identification approaches across a continental scale',
585     *Environmental DNA*, p. edn3.453. https://doi.org/10.1002/edn3.453.

586     Friedland, G. (2024) *Information-Driven Machine Learning: Data Science as an Engineering
587     Discipline*. 1st ed. Cham: Springer International Publishing AG.

588     Gasc, A. *et al.* (2015) 'Acoustic indices for biodiversity assessments: Analyses of bias based on
589     simulated bird assemblages and recommendations for field surveys', *Biological Conservation*,
590     191, 306–312. https://doi.org/10.1016/j.biocon.2015.06.018.

591     Goodfellow, I., Yoshua, B. and Courville, A. (2016) *Deep Learning*. Cambridge: The MIT Press.

592     Guillera-Arroita, G. *et al.* (2017) 'Dealing with false-positive and false-negative errors about
593     species occurrence at multiple levels', *Methods in Ecology and Evolution*. Edited by R. McCrea,
594     8(9), 1081–1091. https://doi.org/10.1111/2041-210X.12743.

595     Hajibabaei, M. *et al.* (2019) 'COI metabarcoding primer choice affects richness and recovery of
596     indicator taxa in freshwater systems', *PLOS ONE*. Edited by A. Ramaiah, 14(9), p. e0220953.
597     https://doi.org/10.1371/journal.pone.0220953.

598     He, K. *et al.* (2016) 'Deep residual learning for image recognition', in *Proceedings of the IEEE
599     Computer Society Conference on Computer Vision and Pattern Recognition*.
600     https://doi.org/10.1109/CVPR.2016.90.

601     Holm, A.M.R. *et al.* (2023) 'Holistic monitoring of freshwater and terrestrial vertebrates by
602     camera trapping and environmental DNA', *Environmental DNA*, 5(6), 1608–1622.
603     https://doi.org/10.1002/edn3.481.

604     Høye, T.T. *et al.* (2021) 'Deep learning and computer vision will transform entomology',
605     *Proceedings of the National Academy of Sciences of the United States of America*, 118(2).
606     https://doi.org/10.1073/PNAS.2002545117.

607     Kaplan, M.B. *et al.* (2015) 'Coral reef species assemblages are associated with ambient
608     soundscapes', *Marine Ecology Progress Series*, 533, 93–107.
609     https://doi.org/10.3354/meps11382.

610     Kaspari, M. *et al.* (2022) 'Activity density at a continental scale: What drives invertebrate
611     biomass moving across the soil surface?', *Ecology*, 103(1), p. e03542.
612     https://doi.org/10.1002/ecy.3542.

613     Lamb, P.D. *et al.* (2019) 'How quantitative is metabarcoding: A meta-analytical approach',
614     *Molecular Ecology*, 28(2), 420–430. https://doi.org/10.1111/mec.14920.

615 Leary, P. *et al.* (2023) 'inatVisionAPI'. https://github.com/inaturalist/inatVisionAPI (Accessed:
616 26 September 2023).

617 Liu, M. *et al.* (2020) 'A practical guide to DNA metabarcoding for entomological ecologists',
618 *Ecological Entomology*, 45(3), 373–385. https://doi.org/10.1111/een.12831.

619 Marquina, D. *et al.* (2019) 'Establishing arthropod community composition using
620 metabarcoding: Surprising inconsistencies between soil samples and preservative ethanol and
621 homogenate from Malaise trap catches', *Molecular Ecology Resources*, 19(6), 1516–1530.
622 https://doi.org/10.1111/1755-0998.13071.

623 Mas-Carrió, E. *et al.* (2022) 'Assessing environmental DNA metabarcoding and camera trap
624 surveys as complementary tools for biomonitoring of remote desert water bodies',
625 *Environmental DNA*, 4(3), 580–595. https://doi.org/10.1002/edn3.274.

626 Pichler, M. and Hartig, F. (2023) 'Machine learning and deep learning—A review for
627 ecologists', *Methods in Ecology and Evolution*, 14(4), 994–1016. https://doi.org/10.1111/2041-
628 210X.14061.

629 Ratnasingham, S. and Hebert, P.D.N. (2007) 'BOLD : The Barcode of Life Data System
630 (http://www.barcodinglife.org)', *Molecular Ecology Notes*, 7(3), 355–364.
631 https://doi.org/10.1111/j.1471-8286.2007.01678.x.

632 Rennstam Rubbmark, O. *et al.* (2018) 'A broadly applicable COI primer pair and an efficient
633 single-tube amplicon library preparation protocol for metabarcoding', *Ecology and Evolution*,
634 8(24), 12335–12350. https://doi.org/10.1002/ece3.4520.

635 Sayers, E.W. *et al.* (2024) 'GenBank 2024 Update', *Nucleic Acids Research*, 52(D1), D134–
636 D137. https://doi.org/10.1093/nar/gkad903.

637 Schneider, S. *et al.* (2022) 'Bulk arthropod abundance, biomass and diversity estimation using
638 deep learning for computer vision', *Methods in Ecology and Evolution*, 13(2), 346–357.
639 https://doi.org/10.1111/2041-210X.13769.

640 Shea, M.M. *et al.* (2023) 'Systematic review of marine environmental DNA metabarcoding
641 studies: toward best practices for data usability and accessibility', *PeerJ*, 11, p. e14993.
642 https://doi.org/10.7717/peerj.14993.

643 Stewart, K.A. and Taylor, S.A. (2020) 'Leveraging eDNA to expand the study of hybrid zones',
644 *Molecular Ecology*, 29(15), 2768–2776. https://doi.org/10.1111/mec.15514.

645 Stork, N.E. (2018) 'How Many Species of Insects and Other Terrestrial Arthropods Are There on
646 Earth?', *Annual Review of Entomology*, 63(1), 31–45. https://doi.org/10.1146/annurev-ento-
647 020117-043348.

648 Sys, S. *et al.* (2022) 'COLLEMBOLAI , a macrophotography and computer vision workflow to
649 digitize and characterize samples of soil invertebrate communities preserved in fluid', *Methods*
650 *in Ecology and Evolution*, 13(12), 2729–2742. https://doi.org/10.1111/2041-210X.14001.

651     Taberlet, P. *et al.* (2012) 'Towards next-generation biodiversity assessment using DNA
652     metabarcoding', *Molecular Ecology*, 21(8), 2045–2050. https://doi.org/10.1111/j.1365-
653     294X.2012.05470.x.

654     Taberlet, P. *et al.* (2018) *Environmental DNA: for biodiversity research and monitoring*. First
655     edition. Oxford, United Kingdom: Oxford University Press.

656     Takeuchi, Y. *et al.* (2021) 'The Asia-Pacific Biodiversity Observation Network: 10-year
657     achievements and new strategies to 2030', *Ecological Research*, 36(2), 232–257.
658     https://doi.org/10.1111/1440-1703.12212.

659     Terry, J.C.D., Roy, H.E. and August, T.A. (2020) 'Thinking like a naturalist: Enhancing
660     computer vision of citizen science images by harnessing contextual data', *Methods in Ecology*
661     *and Evolution*, 11(2), 303–315. https://doi.org/10.1111/2041-210X.13335.

662     U.S. Geological Survey (2013) 'Integrated Taxonomic Information System (ITIS)'.
663     https://doi.org/10.5066/F7KH0KBK.

664     Wedding, L.M. *et al.* (2019) 'Remote sensing of three-dimensional coral reef structure enhances
665     predictive modeling of fish assemblages', *Remote Sensing in Ecology and Conservation*, 5(2),
666     150–159. https://doi.org/10.1002/rse2.115.

667     Weiser, M.D. *et al.* (2021) *Batch extraction of morphological and color metrics from*
668     *invertebrate samples*, *protocols.io*. https://doi.org/dx.doi.org/10.17504/protocols.io.byt4pwqw.

669     Weiser, M.D. *et al.* (2022) 'Robust metagenomic evidence that local assemblage richness
670     increases with latitude in ground-active invertebrates of North America', *Oikos*, 2022(8), p.
671     e08791. https://doi.org/10.1111/oik.08791.

672     Wilcox, T.M. *et al.* (2015) 'The Dual Challenges of Generality and Specificity When
673     Developing Environmental DNA Markers for Species and Subspecies of Oncorhynchus', *PLOS*
674     *ONE*. Edited by A.R. Mahon, 10(11), p. e0142008.
675     https://doi.org/10.1371/journal.pone.0142008.

676     Wood, J.R. *et al.* (2017) 'No single driver of biodiversity: divergent responses of multiple taxa
677     across land use types', *Ecosphere*, 8(11), p. e01997. https://doi.org/10.1002/ecs2.1997.