Appendix

Enabling Human-In-The-Loop Interpretability Methods of Machine Learning Models: The Case of Bird Species Identification

TU Delft

Wei Zeng

November, 2021

TABLE OF CONTENTS

Appendix A: Summary of Interview Results 4
Appendix B: Raw Data from Mini-Surveys 6
Appendix C: Statement Cards 10
Appendix D: Set Up of Online Questionnaires 12
Appendix E: Questionnaires Results 20
Appendix F: Complete Explanation Prototypes 28
Appendix G: Evaluation Protocol of Explanation Tests 30
Appendix H: Raw data Explanation Test Results 34
Appendix I: Analysis of Explanation Test Results 38
Appendix J: The Online Annotation Test Process 40
Appendix K: Raw Data of the Annotation Test Results 50
Appendix L: Prototypes for Interface Tests 54
Appendix M: Evaluation Questionnaire for Interface Tests 56
Appendix N: Interface Test Results 58
Appendix O: Information on the Classification Model and the Dataset 62
Appendix P: The Project Brief 66

Appendix A: Summary of Interview Results

Questions:
1. How do you get started in identifying birds?
2. What helps you at different stages of learning?
3. Throughout the learning experience, which part did you find most helpful/challenging?
4. Have you used any bird ID app? What do you use it for?

No.	Basic information	How to identify a bird?	Most helpful for learning	Most challenging	Use of bird ID apps	How to verify the correctness of bird apps
P1	• 2.5 years experience of bird watching until now (2021) • Joined an association for bird/insect watching during the University • Painting birds as a hobby (participated in the painting of an illustrated bird guide) • Behaviors: bringing a telescope, taking photos, recording bird sound	• Shape/size, Feather color, Location/habitats (eg. forest or bush?), Behavior (eg. what it eats), Sound • The participant herself mostly goes through the process spontaneously • Some of her friends will follow the following steps • First tell by Order: Passerines or Non-Passerines (90% belongs to Passerines); • If Non-Passerines, identify its family.	• Bird guide: Compare the photos and the illustrations • Taking lectures, guidance from experienced bird watchers: build the knowledge system • Discussion with experienced bird watchers	• Identifying birds that are rare seen in this area (eg. stragglers birds)	• eBird, Bird ID Master, Xeno (sound ID) • Use the bird ID apps as a digital version of the bird guide (input keywords like “duck” and search for specific entries) • Check the Hotspot on eBird before birding to see what species are there • Find the feature of sound identification most helpful	• Check photos of predictions ranked 1st, 2nd, 3rd,… to compare one by one
P2	• Started bird watching 5 yrs ago (2016) by joining a birding association • Have done bird watching in most provinces of China, and in some cities in the UK • Majors in Life Science	• Gizz (the overall impression or appearance), Shape/size, Feather color, Location/habitats (eg. forest or bush?), Behavior (eg. what it eats), Sound • Goes through the process spontaneously or through discussions with others. • He mentioned that years ago when cameras were less advanced, people might try to note down the characteristics in words. And they might do so by first recognizing which family it belongs to.	• Bird guide: prefer illustrated ones, which shows more stable characteristics • Lectures, practice makes perfect • Discussion with friends	• Subtle differences of feather color are hard to distinguish • Some birds are hard to see and can only be identified by sound	• eBird, Bird ID Master • Rarely use the bird ID features, use them mainly as documenting tools and digital guidebooks. • Will check others’ documentation of birds on eBird before birding. • Comment on Bird ID Master: It was not that accurate when it was first published in 2018, but it is becoming more and more accurate now.	• Will refer to bird guide and compare the target to sample pictures of the predicted species
P3	• Started bird watching as a hobby since middle school by joining a birding club, which was around 10 years ago • Majors in Life Science	• Similar to above	• Lecture, field guides, knowledgeable friends	• Subtle differences of different ages of birds	• Rarely use bird ID apps	-

Appendix B: Raw Data from Mini-Survey

Research Questions

- What are people's motivations for using bird ID apps?
- When do they trust the prediction result completely? When do they doubt it? When are they 100% sure that it is wrong?
- What information about the prediction do they care about most?

From Reddit

Question:

While using the bird ID apps like Merlin, when will you doubt the prediction result and when will you trust it completely? And when are you 100% sure that it is wrong? What will you do in each situation?

Any stories of being deceived by the apps are more than welcome. And I'm particularly curious to know the answers from the less experienced birders.

[Reply 1]

For me the answer boils down to decades of experience, but of the errors I've seen Merlin come up with for other people, incorrect answers usually are a bird that's out of range, out of the correct habitat, or has some obvious wrong physical feature (bill shape and size don't match, or it's thrown off by shadows across the bird). Same sort of stuff you'd have to be cautious of when IDing without an app to figure out what you're looking at.

[Reply 2]

I am new, just in my first year. My approach is probably a bit different. I don't use Merlin, but I do use BirdNET for sound id (also a Cornell Lab of Ornithology resource). I also do a lot of studying and browsing, especially on AllAboutBirds and using the eBird bar charts.

When brand new, I only accepted as fact the "almost certain" ids (highest confidence level) and would look to confirm them and also follow up with reading later. I got familiar with the most commonly encountered birds pretty quick, and I fell into sync with the app. I learned a lot this way and I feel that the confident ids are very reliably accurate.

The oddball ids are often kind of obvious for being rare or unexpected (recognizing that aspect takes at least a little experience) The less confident ids that I know are wrong tend to be from a pool of usual suspects (in my case, eastern screech owl, mourning dove, wild turkey, human). Bugs, frogs etc especially bring these on, but I know to ignore these and try again. If I keep getting something else over and over I'll look into it later. Sometimes they do indeed show up for real. Either the app has picked up on something I missed, or a new possibility has joined the pool based on timing and I find it later.

Now that I'm pretty good with a lot of bird songs, I do occasionally disagree with the id. But I generally also know why it thinks different. For instance, I can ignore crickets and sometimes get a better listen on the song. Depending on what it is, I might go with my id or the app's or call it a draw. When the id is inconclusive and I've spent the effort I'm willing to, I just move on. There are a lot more birds out there! If you can get a picture, you can study it later (the camera hassle is not always worth it imo).

No one is 100% accurate, and AI isn't either. I'm sure you can trust Merlin to get you going and it will not steer you terribly wrong. You will end up learning a lot from your mistakes and those of the app too. The app is a super start and hugely valuable tool, but you will end up needing to do follow-up (and prep, ideally) to get the most from birding and learning about the whole thing. But you will find the level of investment that keeps the birding fun for you, which could be any amount.

The Cornell tools are pretty gold-plated in my opinion (though I haven't used others for comparison). You can definitely trust it enough, especially compared to the starting point of knowing... a lot less than them. I don't think you should fret about whether the app will "deceive" you or not. It just one tool. Just give it a try and evaluate how it's going later.

However you proceed, you will start to appreciate how involved birding really is. People spend decades working on their own cumulative personal experience and still find id challenges and exciting sightings and interesting everyday behavior all the time. App functionality, internet resources, reference books, news of ornitholgy discoveries, and viewing equipment are all things to use, but ultimately it's experience that counts the most. Just a newbie myself, but that's how it seems to me.

[Reply 3]

I usually use Merlin, but when I'm not very sure I also check out the Audubon app and check out what other people have seen in the area recently.

I've never been in a situation where I had to be correct about it though. I usually go "hmm I think I saw an x bird today!" and that's it.

From DouBan

Questions:

1. 你们主要用什么方式识别见到的鸟啊（跟人讨论/查图鉴/用app识别，顺序是怎样的，觉得哪一种可信度比较高）？

2. 经常用鸟类识别App（比如懂鸟这种）吗？

3. 用鸟类识别App的主要目的是什么（找到大概的方向/验证自己的猜想/查询更多的资料…）？

4. 在用App时会关注哪些信息？什么情况下会觉得预测结果是可信/不可信的？

1. What methods do you mainly use to identify the birds you see (discuss with people/check the picture book/identify with the app; the order that you use them, which do you think is more credible)?

2. Do you often use bird recognition apps?

3. What is the main purpose of using the bird recognition app (find the general direction/verify your guess/look for more information...)?

4. What information will you pay attention to when using the App? Under what circumstances will the prediction result be considered credible/uncredible for you?

[Reply 4]

如果是和别人一起观鸟的话，遇到不认识的会先和同伴讨论，有时候同伴的意见能帮你少走弯路。没有结论就会查图鉴，最后才会看APP。

还有就查图片一般是知道鸟类的科目所属，比如看见一只猛禽，我知道是鹰形目的，就会查图鉴。反之，碰到一只无法判断是什么科目所属的就会用APP。

If you are bird watching with others, you will discuss with your peers when you meet something you don't know. Sometimes your peers' opinions can help you avoid detours. If there is no conclusion, I will check the guidebook, and finally I will look at the APP.

Plus, you look up the guidebook usually when you know what order it belongs to. For example, if I see a bird of prey, I know it's an eagle, then I will look up the guidebook. On the contrary, if I encounter a bird whose order cannot be identified, I will use the APP.

[Reply 5]

观鸟两个多月吧，认识的鸟很少。

1.遇到不认识的鸟先用app的图片搜索，然后再用图鉴对比验证，结合百科或别人发的内容综合了解。有时也查别人的观鸟报告，看该观测点该季节是不是出现过这种鸟，用来再次验证。但是最后这一点未必准确，比如刚好碰见迁徙中歇歇脚的，比如5月在成都望江楼看到一只金眶鸻

2.常用的有爱鸟国际、懂鸟、观鸟记录中心

3.对，确定大概方向，然后再查图鉴

4.鸟的大致样貌是不是一致，如果不一致会看看是不是雌雄、亚种或亚成鸟。都不是的话会觉得不准确。比如在成都东湖公园门口二环路那里的桥上看到过一只黑头小鸟，图片识别是白头鹎，对比图鉴可能是亚种（不太确定）

I have been bird watching for more than two months. I know very few birds.

1. When encountering a bird that I don't know, first use the picture of the app to search, and then use the guidebook to compare and verify, in combination with the content of the encyclopedia or the content posted by others. Sometimes I check other people's bird watching reports to see if this kind of bird has appeared in this season at the observation point, and use it to verify again. But this last point may not be accurate. For example, some birds just happen to be resting during migration. For example, I saw a golden-eyed plover at Wangjiang Tower in Chengdu in May.

2. Commonly used are Love Bird International, Bird ID Master, and Bird Watching Record Center

3. Yes, determine the general direction, and then check the bird guide.

4. I'll check first whether the general appearance of the birds look the same. If they are not the same, I will see if they are male and female, subspecies or sub-adults. If it is neither, it will feel inaccurate. For example, I saw a black-headed bird on the bridge on the second ring road at the entrance of Donghu Park in Chengdu. The photo is identified as a white-headed bulbul, and by comparing to the guidebook, I decide that it may be a subspecies (not sure)

[Reply 6]

我每次是散步时发现不认识的鸟就百度搜搜，感觉还是用靠谱的书识别更好.准备买一本《中国鸟类图鉴》。感觉书可能更全面些，而且有记录，百度过后我说不定忘了，书还可以回翻看看。

Every time I find a bird I don't know when I'm walking, I search on Baidu. I think it's better to use a reliable guidebook to identify it. I'm going to buy a book of "Chinese Birds". I feel that the guidebook may be more comprehensive, and with records. I might forget it after searching it online, but I can look back at the book.

[Reply 7]

懂鸟基本上是靠谱的 当然要积累经验，自己看书记录

The Bird ID Master is basically reliable. Of course, to gain experience you need practices, reading books and taking records.

[Reply 8]

我用的是merlin bird这个软件，识别我觉得还挺准的。这个app有两种识别方式，第一种是描述，根据你对鸟体型的描述以及你的地理位置推测，第二种就是如果你拍到照片了，可以识别图片。初期查证的话我主要是还是对比照片，把自己拍摄的照片对照鸟类图鉴，根据鸟身上的特征来确认，包括叫声等。

另外就是城市的官网或者是本地鸟类小组，会有发布当地鸟类的图鉴，比如xxx地区常见鸟类这种，我最初的时候就是把这个打印出来，先大致对本地区的常见鸟的外观做一个大致的了解，然后再慢慢观察自己家周边的鸟类...这样一点点先熟悉了自己家的鸟类，再逐渐拓展到公园或者专门的保护区....

I use the merlin bird app, and I think the recognition is quite accurate. This app has two recognition methods. The first is through description, it predicts based on your description of the bird's body shape and your geographic location. The second is that if you take a photo, you can recognize the picture. In the initial verification, I mainly compare the photos. Compare the photos I took with the bird guide, and confirm according to the characteristics of the bird, including its call, etc.

The other resource is the official website of the city or the local bird group, which will publish the local bird pictures, such as the commonly seen birds in the xxx area. I printed this out at the beginning, and first roughly got acquainted with the appearance of the common birds in the area. Then slowly observe the birds around my home...In this way, I get familiar with the birds in my area a little bit, and then gradually expand to parks or special sanctuaries...

[Reply 9]

对于鸟类，声音识别可能比照片还靠谱

For birds, voice recognition may be more reliable than photos

[Reply 10]

https://www.zhihu.com/question/35455024/answer/62924616 可以看看这篇回答，我也是前两天才看到的，确实声音有的时候更准确

https://www.zhihu.com/question/35455024/answer/62924616

You can take a look at this answer, I only saw it two days ago, and indeed the sound is sometimes more accurate

(Supplement from the Interview (P1))

先看图, 然后看名字，用英文或者学名去搜资料。资料关注分布地、行为，以及图示。但是图示比较难辨认，所以我一般会问擅长的朋友而不是自己搜。可信度会看但不太关注。

Look at the picture first, then the name, and search for information use the English or scientific name. When searching for information, I pay attention to its distribution, behavior, and graphics. But the graphics are hard to identify, so I usually ask knowledgeable friends instead of searching by myself. I will look at the confidence number but not very concerned about it.

Appendix C: Statement Cards

I know the wrong IDs are tend to be from a pool of usual suspects.(for audio ID)

Quote:The less confident ids that **I** know are wrong tend to be from a pool of usual suspects. [R2]

Voice recognition maybe more reliable than photos.

Quote: For birds, voice recognition may be more reliable than photos. [R9]

Errors of the prediction can be told by range, habitat and physical features.

Quote: Of the errors **I**'ve seen Merlin come up with for other people, incorrect answers usually are a bird that's out of range, out of the correct habitat, or has some obvious wrong physical feature (bill shape and size don't match, or it's thrown off by shadows across the bird). [R1]

I am certained that the predictions are false when they are rare or unexpected species.

Quote:The oddball ids are often kind of obvious for being rare or unexpected (recognizing that aspect takes at least a little experience) [R2]

I can tell whether the prection is true when getting more experienced.

Quote: For me the answer boils down to decades of experience [R1]

I verify the prediction by comparing appearance and call.

Quote: Compare the photos **I** took with the bird guide, and confirm according to the characteristics of the bird, including its call, etc. [R8]

Look at the picture first, then the habitat and behaviors.

Quote: Look at the picture first, then the name, and search for information use the English or scientific name. When searching for information, **I** pay attention to its distribution, behavior, and graphics.. [R10]

Check the different look within a species(gender, sub-species, sub-adults).

Quote: I'll check first whether the general appearance of the birds look the same. **I**f they are not the same, **I** will see if they are male and female, subspecies or sub-adults. **I**f it is neither, it will feel inaccurate. [R5]

The occurance records of the species in this area is another important clue, but misleading sometimes.

Quote: Sometimes **I** check other people's bird watching reports to see if this kind of bird has appeared in this season at the observation point, and use it to verify again. But this last point may not be accurate. For example, some birds just happen to be resting during migration. [R5]

Use the apps in compination with guidebook and other online resources.

Quote: When encountering a bird that **I** don't know, first use the picture of the app to search, and then use the guidebook to compare and verify, in combination with the content of the encyclopedia or the content posted by others. [R5]

Use the apps in compination with guidebook and other online resources.

Quote: When encountering a bird that **I** don't know, first use the picture of the app to search, and then use the guidebook to compare and verify, in combination with the content of the encyclopedia or the content posted by others. [R5]

The Bird ID Apps are important starting tools.

Quote: You can definitely trust it enough, especially compared to the starting point of knowing... a lot less than them. **I** don't think you should fret about whether the app will "deceive" you or not. **I**t just one tool. Just give it a try and evaluate how it's going later.[R2]

I get familiar with the birds' appearance in the area before outing.

Quote:**I** printed this out at the beginning, and first roughly get acquaintance with the appearance of the common birds in the area. [R8]

Follow-up and prep are needed besides using the apps.

Quote: The app is a super start and hugely valuable tool, but you will end up needing to do follow-up (and prep, ideally) to get the most from birding and learning about the whole thing. [R2]

When I know the order of that bird, I use the book. Otherwise I use the app.

Quote: you look up the guidebook usually when you know what order it belongs to. For example, if **I** see a bird of prey, **I** know it's an eagle, then **I** will look up the guidebook. On the contrary, if **I** encounter a bird whose order cannot be identified, **I** will use the APP. [R4]

I check the APP after asking people and checking books.

Quote: Sometimes your peers' opinions can help you avoid detours. If there is no conclusion, I will check the guidebook, and finally I will look at the APP.[R4]

I use the ID app to tell the general direction.

Quote: (**I** use it to) determine the general direction, and then check the bird guide. [R5]

The correctness of the prediction result is not very important to me.

Quote: **I**'ve never been in a situation where **I** had to be correct about it though. **I** usually go "hmm **I** think **I** saw an x bird today!" and that's it. [R3]

Appendix D: Set Up of Online Questionnaires

Start of Block: Block 1

Q2 Thank you for opening this link. 🦤🦉🦆🦅 You are being invited to participate in a research study on machine learning interpretability and bird species identification. This study is being done by a master student from the TU Delft. The survey will take you approximately 8 minutes to complete. To the best of our ability, your answers in this study will remain confidential.Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any question. Please contact W.Zeng@student.tudelft.nl for any questions you may have regarding this study.

Page Break

Q1 What's your age?

- 0-15
- 16-25
- 26-35
- 36-45
- 46-55
- >55

X→

Q3 Do you have a professional relationship with ornithology?

- Yes, my education or job is closely related to ornithology.
- Yes, my education or job is partly related to ornithology.
- No, my education and job are not related to ornithology.

Q4 Do you have a professional relationship with machine learning?

- Yes, my education or job is closely related to machine learning.
- Yes, my education or job is partly related to machine learning.
- No, my education and job are not related to machine learning at all.

End of Block: Block 1

Start of Block: Birding experience

Q6 Birdwatching, or birding, is a recreational activity. It can be done with the naked eye, through a visual enhancement device like binoculars and telescopes, by listening for bird sounds, or by watching public webcams.

Q7 How often do you do bird watching outdoor (on average)?

- More than once a week
- More than once a month but less than once a week
- Several times a year but less than once a month
- Less than once a year
- I've never done that

Display This Question:
If How often do you do bird watching outdoor (on average)? != I've never done that

Q5 How many years have you been birding(bird watching), approximately?

- Less than 1 year
- 1-3 years
- 3-5 years
- 5-8 years
- More than 8 years

Q8 What level you are at in telling birds apart, compared to people around you?

- Entrance-level
- Intermediate-level
- Advanced-level
- Expert-level

Q9 Which of the following activities do you usually participate in? (multi-choice)

- [ ] Photographing birds
- [ ] Keeping records of birds
- [ ] Participating in bird watching activities organized by local organizations
- [ ] Bird-watching on your own (or with few friends)
- [ ] Feeding birds in back-yard
- [ ] Taking lectures on bird watching
- [ ] None of the above

Q8 What level you are at in telling birds apart, compared to people around you?

- ○ Entrance-level
- ○ Intermediate-level
- ○ Advanced-level
- ○ Expert-level

Q9 Which of the following activities do you usually participate in? (multi-choice)

- ☐ Photographing birds
- ☐ Keeping records of birds
- ☐ Participating in bird watching activities organized by local organizations
- ☐ Bird-watching on your own (or with few friends)
- ☐ Feeding birds in back-yard
- ☐ Taking lectures on bird watching
- ☐ None of the above

Q12 What references do you usually use for learning about birds?
And please rate them based how important they are for your learning. (0=not important/not used, 5=very important)

Lectures of bird watching	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5
Video and DVD guides	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5
Knowledgeable friends/experts around me	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5
Books or fieldguides of birds	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5
Discussions in online forums / interest groups	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5
Other online resources	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5
Birding records of other birders (eg. eBird)	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5
Bird ID apps (eg. Merlin Bird ID, Bird ID Master)	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5

Q13 What do you usually do to tell birds apart? (multi-choice)

- ☐ Ask knowledgeable friends
- ☐ Check the field guide
- ☐ Take a picture of the bird and search in bird ID apps
- ☐ Use searching engines (like google)
- ☐ Check others' birding records in this area
- ☐ Other _____

End of Block: Birding experience

Start of Block: Bird ID apps

Q14 Have you tried any bird ID apps?

- ○ Yes
- ○ Maybe
- ○ No

Skip To: Q24 If Have you tried any bird ID apps? = No

Q15 Which of the following apps do you usually use for bird identification?

- ☐ eBird
- ☐ Bird ID Master (Dong Niao)
- ☐ Merlin Bird ID
- ☐ Other _____

Q16 Which of the following information do you usually use (input) for bird ID in apps? (multi-choice)

- ☐ Photos
- ☐ Audios
- ☐ Characteristics (eg, size, color, habitat, location, etc)
- ☐ Other _____

Q17 Why do you use the bird ID apps? (multi-choice)

- [ ] I can find the answer more quickly than looking into a bird guide
- [ ] Because it's convenient to reach compared to a bird guide
- [ ] Because I can show it to others easily
- [ ] Because I'm curious whether it could identify correctly
- [ ] I'm not a fan of bird ID apps
- [ ] Other _____

Q18 How do you use the bird ID apps? (multi-choice)

- [ ] I use it as a digital bird book to look up for entries
- [ ] I use it as the main tool to help me identify birds
- [ ] I use it at the first step to narrow down the range of possible species
- [ ] I use it for verification after checking other references
- [ ] Other _____

Q19 Have you encountered false predictions provided by the ID apps?

- ○ Yes, very often.
- ○ Yes, few times.
- ○ Some predictions were doubtable but I wasn't sure.
- ○ No, I haven't.

Skip To: Q22 If Have you encountered false predictions provided by the ID apps? = No, I haven't.

Q20 How did you know that a prediction was false according to how you experienced it? (multi-choice)

- [ ] Told by a human
- [ ] The predicted species had a different appearance
- [ ] The predicted species had a different call
- [ ] The predicted species were out of the correct location/habitat
- [ ] The spotted species had different behaviors from the predicted ones
- [ ] Other _____

Display This Question:
If Which of the following information do you usually use (input) for bird ID in apps? (multi-choice) = Photos

Q21 What do you think were the reasons for the false predictions that you encountered? (multi-choice)

- [ ] The captured photos were too vague to be recognized
- [ ] In the captured photos, some features of the spotted bird were masked by something
- [ ] The species themselves are tricky to identify merely with the photos
- [ ] Something wrong with the algorithms/datasets of the apps
- [ ] I don't know what was the reason
- [ ] Other _____

Q22 What do you want to do when seeing a false prediction? (multi-choice)

- [ ] Give it another try
- [ ] Report it to the app developers
- [ ] Check and learn about the predicted species even though it's the false one
- [ ] Use the clues(eg. family name) from the false predictions to find the right answer
- [ ] Try other tools to find the correct answer (friends, bird guide, search online, etc)
- [ ] None of the above
- [ ] Other _____

Q23 What are the biggest challenges for you to tell birds apart with bird ID apps? (multi-choice)

- [ ] Capturing clear enough photos for the apps to recognize
- [ ] The subtle differences between different species are hard to be told apart
- [ ] The various appearance within the same species
- [ ] No challenge for me at all
- [ ] Other _____

Q24 Last question: How do the bird ID apps work, as you imagine?

_____

_____

_____

_____

_____

End of Block: Bird ID apps

Start of Block: Block 3

Q25 Thank you for taking part in this survey.
Meanwhile, we make some creative bird ID interfaces that we really want to show you in a few weeks and hear your comments on them.
Will you be interested in checking them out? If yes, please leave your contact method below for us to get in touch with you later. If no, please just click the "➡" button to submit the response.
Your personal information will not be shared with other individuals or organizations and you will only be contacted for issues regarding this research.

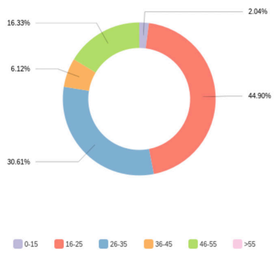
Q28 Here you can leave your contact information (Email/Whatsapp, or any other you prefer), and please state what contact channel it is.

_____

End of Block: Block 3

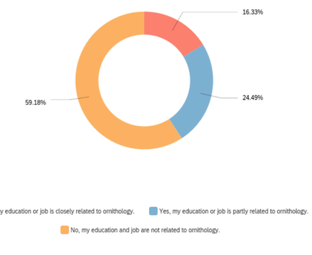
Appendix E: Questionnaires Results

Q1 - What's your age?



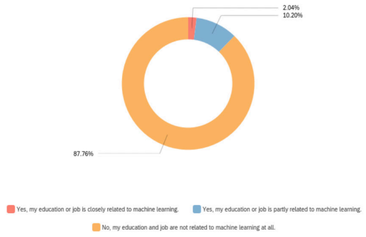
Answer	%	Count
0-15	2.04%	1
16-25	44.90%	22
26-35	30.61%	15
36-45	6.12%	3
46-55	16.33%	8
>55	0.00%	0
Total	100%	49

Q3 - Do you have a professional relationship with ornithology?



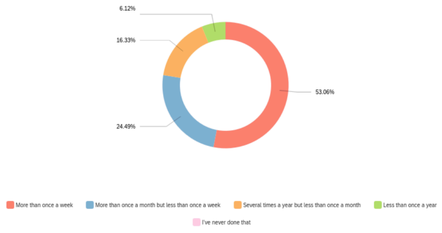
Answer	%	Count
Yes, my education or job is closely related to ornithology.	16.33%	8
Yes, my education or job is partly related to ornithology.	24.49%	12
No, my education and job are not related to ornithology.	59.18%	29
Total	100%	49

Q4 - Do you have a professional relationship with machine learning?



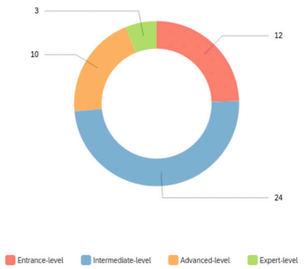
Answer	%	Count
Yes, my education or job is closely related to machine learning.	2.04%	1
Yes, my education or job is partly related to machine learning.	10.20%	5
No, my education and job are not related to machine learning at all.	87.76%	43
Total	100%	49

Q7 - How often do you do bird watching outdoor (on average)?



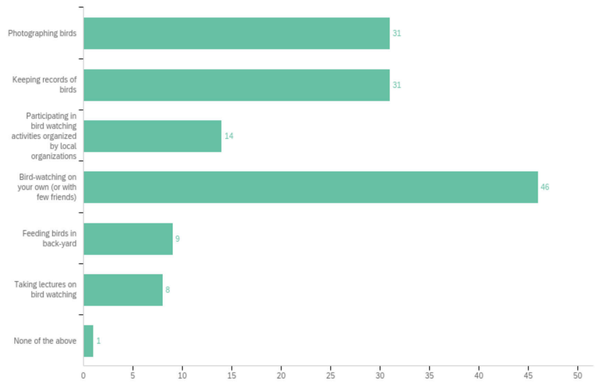
#	Answer	%	Count
1	More than once a week	53.06%	26
2	More than once a month but less than once a week	24.49%	12
3	Several times a year but less than once a month	16.33%	8
4	Less than once a year	6.12%	3
5	I've never done that	0.00%	0
	Total	100%	49

Q8 - What level you are at in telling birds apart, compared to people around you?



Answer	%	Count
Entrance-level	24.49%	12
Intermediate-level	48.98%	24
Advanced-level	20.41%	10
Expert-level	6.12%	3
Total	100%	49

Q9 - Which of the following activities do you usually participate in? (multi-choice)

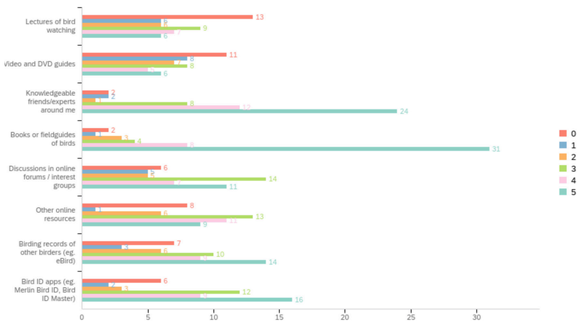


#	Answer	%	Count
4	Bird-watching on your own (or with few friends)	93.88%	46
1	Photographing birds	63.27%	31
2	Keeping records of birds	63.27%	31
3	Participating in bird watching activities organized by local organizations	28.57%	14
5	Feeding birds in back-yard	18.37%	9
6	Taking lectures on bird watching	16.33%	8
7	None of the above	2.04%	1
	Total	100%	49

Q10 - What is your main motivation for bird watching:

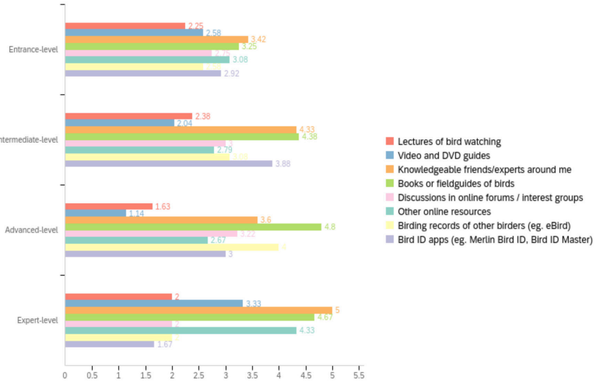
What is your main motivation for bird watching:	EN translation
Being able to identify new species to me and connect with my birder stepmom	Being able to identify new species to me and connect with my birder stepmom
I like it. I am always looking and when I watch them I am happy and sometimes whem I watching birds in my backyard. Iforgot the time and relax. I forget th	I like it. I am always looking and when I watch them I am happy and sometimes whem I watching birds in my backyard. Iforgot the time and relax. I forget th
兴趣	interest
爱好，希望学习鸟类知识	Hobby, hope to learn bird knowledge
I don't do it actively, but when I go walking I try to spot nice birds. So it is more of an extra fun thing when I am outside.	I don't do it actively, but when I go walking I try to spot nice birds. So it is more of an extra fun thing when I am outside.
喜欢大自然，喜欢观察鸟类	Like nature, like watching birds
喜欢鸟类	Love birds
喜欢	I like it
享受自然	Enjoy nature
休闲娱乐	For entertainment
娱乐	Entertainment
看到就開心	It makes me happy
放松	To relax
误打误撞	By accident
喜欢博物	Like natural history

Q12 - What references do you usually use for learning about birds? And please rate them based how important they are for your learning. (0=not important/not used, 5=very important)



#	Question	0		1		2		3		4		5		Total
1	Lectures of bird watching	27.66%	13	12.77%	6	12.77%	6	19.15%	9	14.89%	7	12.77%	6	47
2	Video and DVD guides	24.44%	11	17.78%	8	15.56%	7	17.78%	8	11.11%	5	13.33%	6	45
3	Knowledgeable friends/experts around me	4.08%	2	4.08%	2	2.04%	1	16.33%	8	24.49%	12	48.98%	24	49
4	Books or fieldguides of birds	4.08%	2	2.04%	1	6.12%	3	8.16%	4	16.33%	8	63.27%	31	49
5	Discussions in online forums / interest groups	12.50%	6	10.42%	5	10.42%	5	29.17%	14	14.58%	7	22.92%	11	48
6	Other online resources	16.67%	8	2.08%	1	12.50%	6	27.08%	13	22.92%	11	18.75%	9	48
7	Birding records of other birders (eg. eBird)	14.29%	7	6.12%	3	12.24%	6	20.41%	10	18.37%	9	28.57%	14	49
8	Bird ID apps (eg. Merlin Bird ID, Bird ID Master)	12.50%	6	4.17%	2	6.25%	3	25.00%	12	18.75%	9	33.33%	16	48

Mean broken down by expertice



Entrance-level

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
3	Knowledgeable friends/ experts around me	0.00	5.00	3.42	1.98	3.91	12
4	Books or fieldguides of birds	0.00	5.00	3.25	1.96	3.85	12
6	Other online resources	0.00	5.00	3.08	1.71	2.91	13
8	Bird ID apps (eg. Merlin Bird ID, Bird ID Master)	0.00	5.00	2.92	1.80	3.24	12
5	Discussions in online forums / interest groups	0.00	5.00	2.75	1.96	3.85	12
7	Birding records of other birders (eg. eBird)	0.00	5.00	2.58	1.55	2.41	12
2	Video and DVD guides	0.00	5.00	2.58	1.98	3.91	12
1	Lectures of bird watching	0.00	5.00	2.25	1.88	3.52	12

Intermediate-level

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
4	Books or fieldguides of birds	2.00	5.00	4.38	0.95	0.90	24
3	Knowledgeable friends/ experts around me	3.00	5.00	4.33	0.80	0.64	24
8	Bird ID apps (eg. Merlin Bird ID, Bird ID Master)	0.00	5.00	3.88	1.62	2.61	24
7	Birding records of other birders (eg. eBird)	0.00	5.00	3.08	1.80	3.24	24
5	Discussions in online forums / interest groups	0.00	5.00	3.00	1.47	2.17	24
6	Other online resources	0.00	5.00	2.79	1.63	2.66	24
1	Lectures of bird watching	0.00	5.00	2.38	1.65	2.73	24
2	Video and DVD guides	0.00	5.00	2.04	1.55	2.39	23

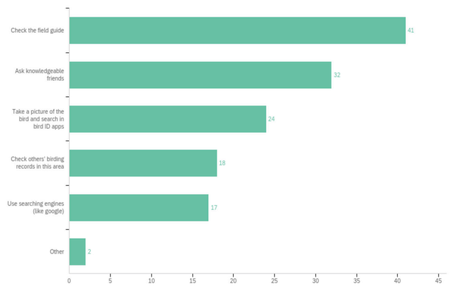
Advanced-level

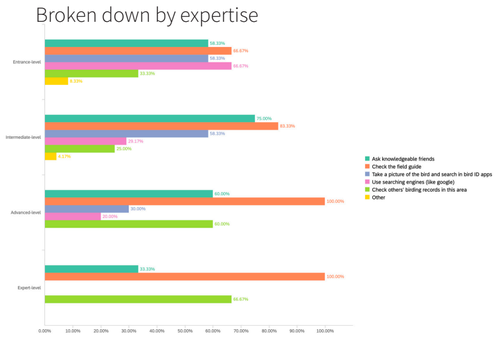
#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
4	Books or fieldguides of birds	4.00	5.00	4.80	0.40	0.16	10
7	Birding records of other birders (eg. eBird)	2.00	5.00	4.00	1.00	1.00	10
3	Knowledgeable friends/ experts around me	1.00	5.00	3.60	1.11	1.24	10
5	Discussions in online forums / interest groups	0.00	5.00	3.22	1.69	2.84	9
6	Other online resources	0.00	4.00	2.67	1.49	2.22	9
1	Lectures of bird watching	0.00	4.00	1.63	1.65	2.73	8
2	Video and DVD guides	0.00	4.00	1.14	1.36	1.84	7
8	Bird ID apps (eg. Merlin Bird ID, Bird ID Master)	1.00	4.00	3.00	0.94	0.89	9

Expert-level

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
3	Knowledgeable friends/ experts around me	5.00	5.00	5.00	0.00	0.00	3
4	Books or fieldguides of birds	4.00	5.00	4.67	0.47	0.22	3
6	Other online resources	3.00	5.00	4.33	0.94	0.89	3
2	Video and DVD guides	2.00	5.00	3.33	1.25	1.56	3
7	Birding records of other birders (eg. eBird)	0.00	5.00	2.00	2.16	4.67	3
8	Bird ID apps (eg. Merlin Bird ID, Bird ID Master)	0.00	3.00	1.67	1.25	1.56	3

Q13 - What do you usually do to tell birds apart? (multi-choice)



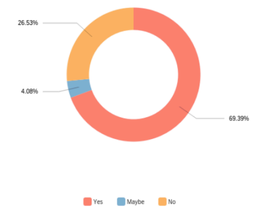


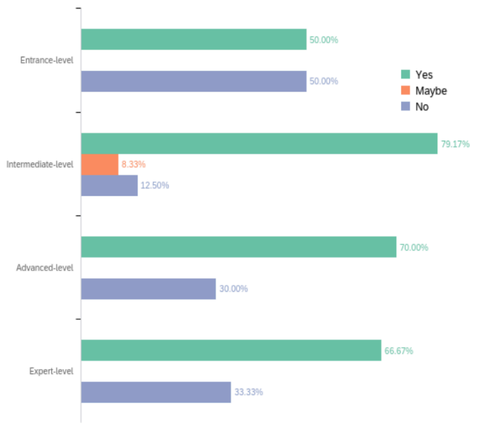
#	Answer	%	Count
1	Ask knowledgeable friends	65.31%	32
2	Check the field guide	83.67%	41
3	Take a picture of the bird and search in bird ID apps	48.98%	24
4	Use searching engines (like google)	34.69%	17
5	Check others' birding records in this area	36.73%	18
6	Other	4.08%	2
	Total	100%	49

Q13_6_TEXT - Other

Other - Text	EN Translation
Ask my dad	Ask my dad
不太会去查，不知道就不知道了，遇到很好看的会拍下来搜图	I would not look it up, just let it be. If I came across something very pretty I would take a photo of it to search

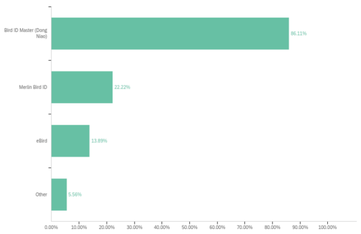
Q14 - Have you tried any bird ID apps?





#	Answer	%	Count
1	Yes	69.39%	34
2	Maybe	4.08%	2
3	No	26.53%	13
	Total	100%	49

Q15 - Which of the following apps do you usually use for bird identification?

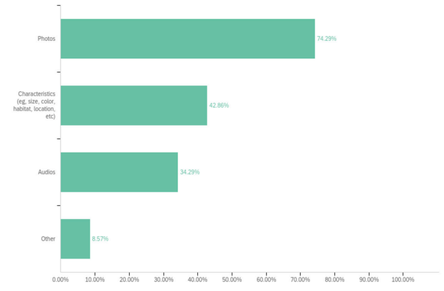


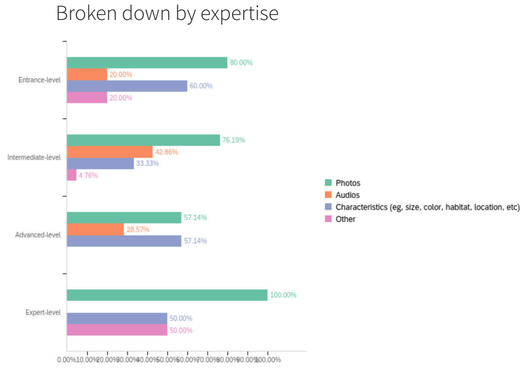
#	Answer	%	Count
2	Bird ID Master (Dong Niao)	86.11%	31
3	Merlin Bird ID	22.22%	8
4	Other	5.56%	2
1	eBird	13.89%	5
	Total	100%	36

Q15_4_TEXT - Other

Other - Text	EN
鸟典	Bird Dictionary (an app)
mockingbird	

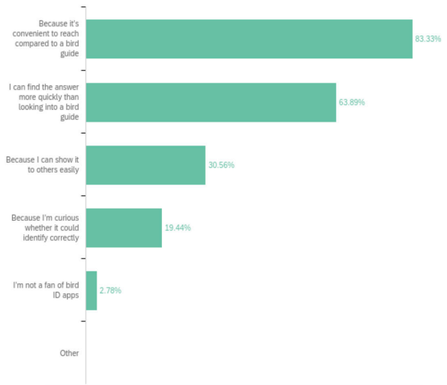
Q16 - Which of the following information do you usually use (input) for bird ID in apps? (multi-choice)

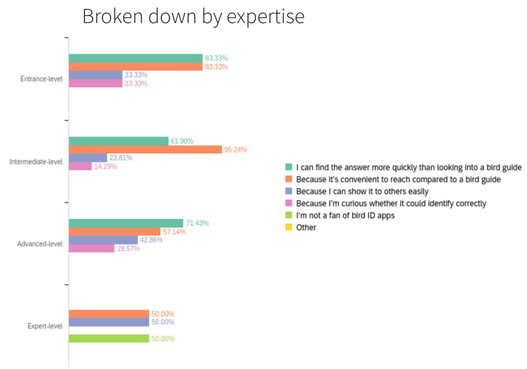




#	Answer	%	Count
1	Photos	74.29%	26
2	Audios	34.29%	12
3	Characteristics (eg, size, color, habitat, location, etc)	42.86%	15
4	Other	8.57%	3
	Total	100%	35

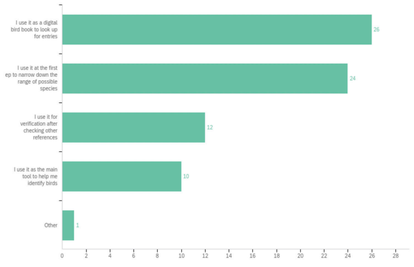
Q17 - Why do you use the bird ID apps? (multi-choice)





Answer	%	Count
Because it's convenient to reach compared to a bird guide	83.33%	30
I can find the answer more quickly than looking into a bird guide	63.89%	23
Because I can show it to others easily	30.56%	11
Because I'm curious whether it could identify correctly	19.44%	7
I'm not a fan of bird ID apps	2.78%	1
Other	0.00%	0
Total	100%	36

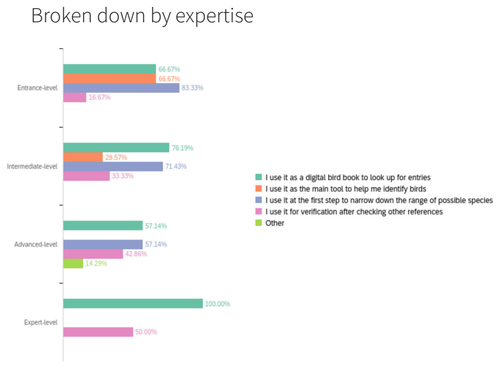
Q18 - How do you use the bird ID apps? (multi-choice)



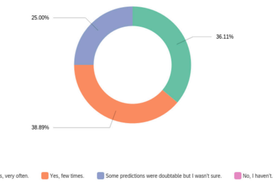
#	Answer	%	Count
1	I use it as a digital bird book to look up for entries	72.22%	26
2	I use it as the main tool to help me identify birds	27.78%	10
3	I use it at the first step to narrow down the range of possible species	66.67%	24
4	I use it for verification after checking other references	33.33%	12
5	Other	2.78%	1
	Total	100%	36

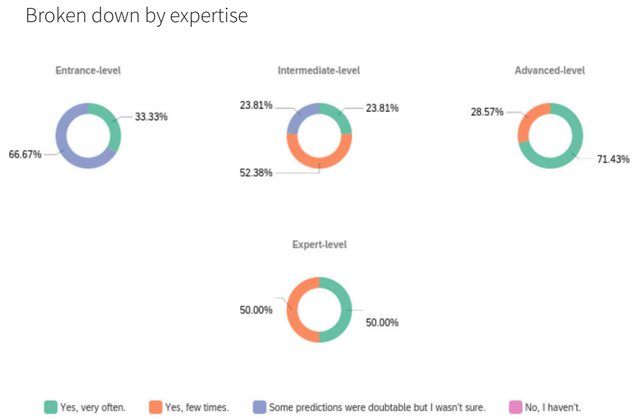
Q18_5_TEXT - Other

Other - Text
At first, I used mainly as a way to ID unknown birds. But after my 4th or 5th outing, I really don't use bird ID apps. However, I do prep in advanced. I check local sightings of birds and study what is in my area.



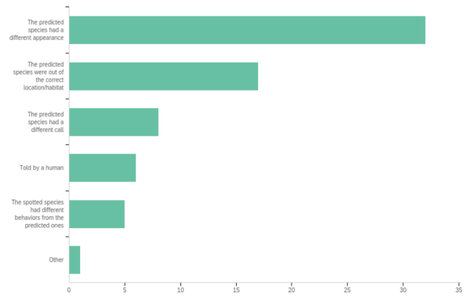
Q19 - Have you encountered false predictions provided by the ID apps?





#	Answer	%	Count
1	Yes, very often.	36.11%	13
2	Yes, few times.	38.89%	14
3	Some predictions were doubtable but I wasn't sure.	25.00%	9
4	No, I haven't.	0.00%	0
	Total	100%	36

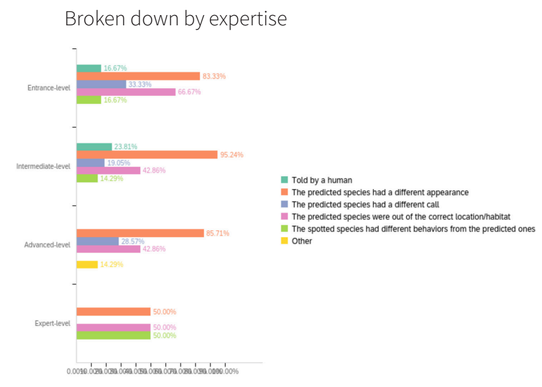
Q20 - How did you know that a prediction was false according to how you experienced it? (multi-choice)



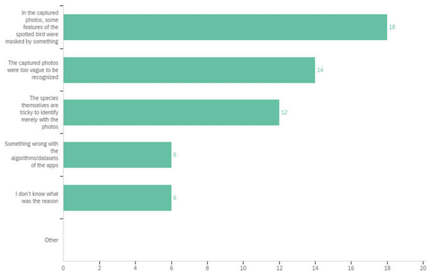
#	Answer	%	Count
1	Told by a human	16.67%	6
2	The predicted species had a different appearance	88.89%	32
3	The predicted species had a different call	22.22%	8
4	The predicted species were out of the correct location/habitat	47.22%	17
5	The spotted species had different behaviors from the predicted	13.89%	5
6	Other	2.78%	1
	Total	100%	36

Q20_6_TEXT - Other

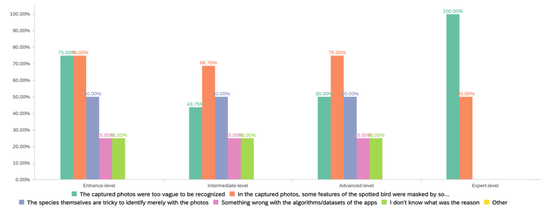
Other - Text
Usually do to my own error, input incorrect size, or even color. So user input needs to be improved on many of these apps



Q21 - What do you think were the reasons for the false predictions that you encountered? (multi-choice)



Broken down by expertise

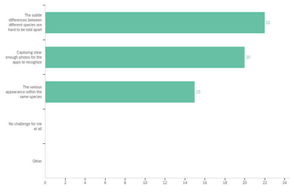


#	Answer	%	Count
1	The captured photos were too vague to be recognized	53.85%	14
2	In the captured photos, some features of the spotted bird were masked by something	69.23%	18
3	The species themselves are tricky to identify merely with the photos	46.15%	12
4	Something wrong with the algorithms/datasets of the apps	23.08%	6
5	I don't know what was the reason	23.08%	6
6	Other	0.00%	0
	Total	100%	26

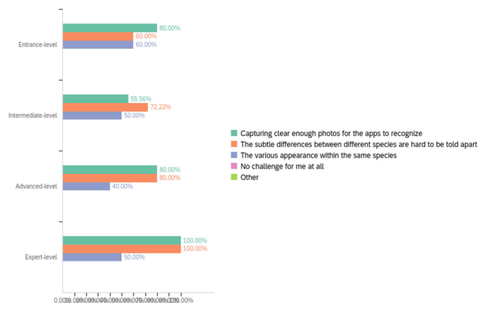
Q22 - What do you want to do when seeing a false prediction? (multi-choice)

#	Answer	%	Count
1	Give it another try	25.00%	9
2	Report it to the app developers	8.33%	3
3	Check and learn about the predicted species even though it's the false one	33.33%	12
4	Use the clues(eg. family name) from the false predictions to find the right answer	33.33%	12
5	Try other tools to find the correct answer (friends, bird guide, search online, etc)	77.78%	28
6	Other	0.00%	0
7	None of the above	8.33%	3
	Total	100%	36

Q23 - What are the biggest challenges for you to tell birds apart with bird ID apps? (multi-choice)



Broken down by expertise



#	Answer	%	Count
1	Capturing clear enough photos for the apps to recognize	66.67%	20
2	The subtle differences between different species are hard to be told apart	73.33%	22
3	The various appearance within the same species	50.00%	15
4	No challenge for me at all	0.00%	0
5	Other	0.00%	0
	Total	100%	30

Last question: How do the bird ID apps work, as you imagine? EN

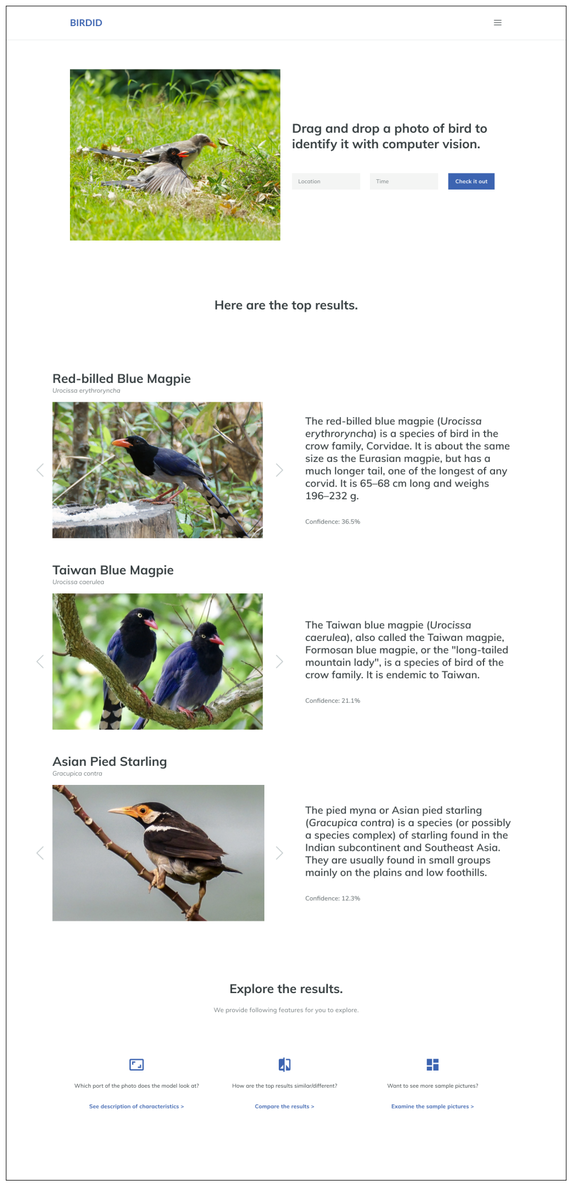
I'm not very knowledgeable in this category, but I'm sure they use some sort of algorithm to combine you location, season, and description to narrow the species down to give you the birds that are the highest probability.
When identifying, also indicate how often this type of bird has been to the sprayed place.
Don't know, search for similar pictures after recognizing contours?
Learning from large amounts of data
they recognize it with AI (I only guess this, because before you asked if I was affiliated with AI in my profession)
Recognize through the appearance and sound characteristics of birds
Machine learning training
Some more vivid action characteristics beyond taxonomy, you know that the classification criteria in the application are too objective, and it is difficult to resonate with readers
Train directly with pictures
Color styling
Image color block matching?
A list of questions similar to how Merlin already works. Questions in Order: Your location, Size of bird, Color of Wings, Color of the Head, Color or the breast/belly, what it was seen doing.... after all questions are answered it shows the 10 most likely birds, if the bird is not there you can hit "bird not shown" and the next 10 most likely will come up. To note, the color questions would have multiple choice. Just in case the bird has multiple colors on its breast or head.
Gallery
Compare with a variety of standard pictures and give the closest option
Graphic recognition? I don't quite understand
do not know
Identify birds from the background-identify various body parts of birds, extract color, shape and other information-compare with the database, calculate similarity, find candidate species, and return to the previous section for intensive comparison based on the characteristics of the candidate species -The final result is verified with the bird distribution, habitat, behavior and other data, and some unreasonable results are excluded

Picture and sound
Automatically detect the current geographic location to delimit a rough range, narrow the range through the surrounding environment, then further narrow the range through behavioral patterns/ appearance characteristics, and finally determine the category and outline through the shooting details
First, find the bird in the picture and get its outline; secondly, extract the color distribution in the outline and compare it with the information in the database
do not know
Machine learning
Learn a lot of photos
shape
It should be through some characteristics of birds
Body shape and body color distribution
Train AI to recognize the color patches and combinations of parts of different birds to reach the level of identifying birds.
Perform feature extraction and match with the database
According to appearance or humming sound.
I don't know... Maybe it's through some of the bird's appearance and the overall color?
Similarity of image detail features and arrangement pattern
Body line ratio, feather distribution, color, etc.

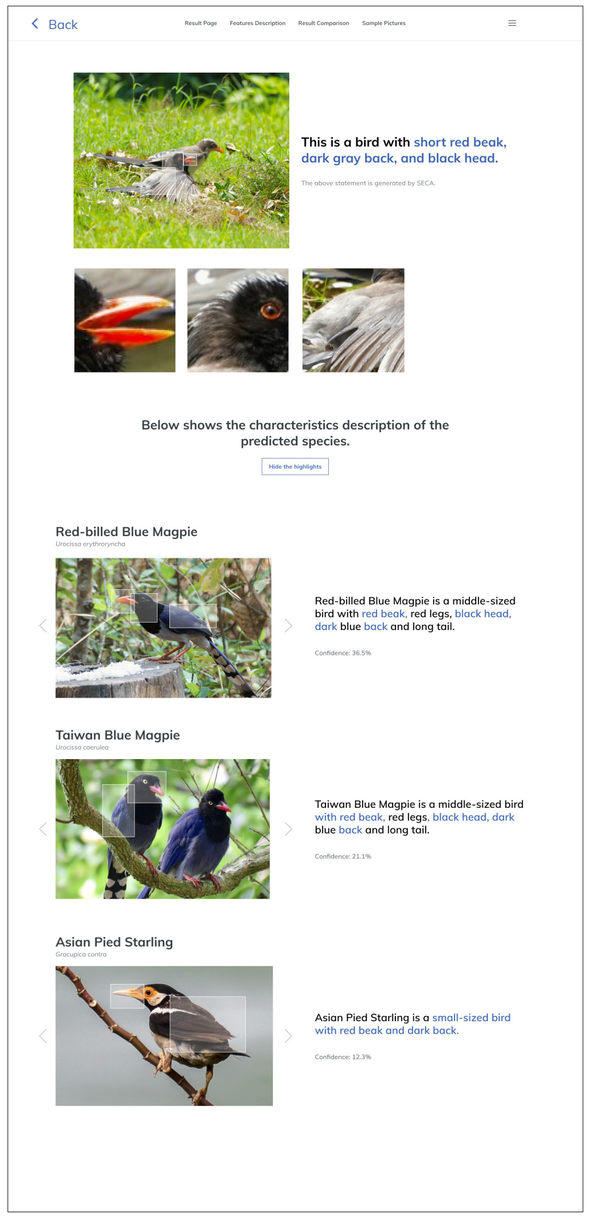
Appendix F: Complete Explanation Prototypes

[Preview of live prototypes (EN)](#) [(CH)](#)

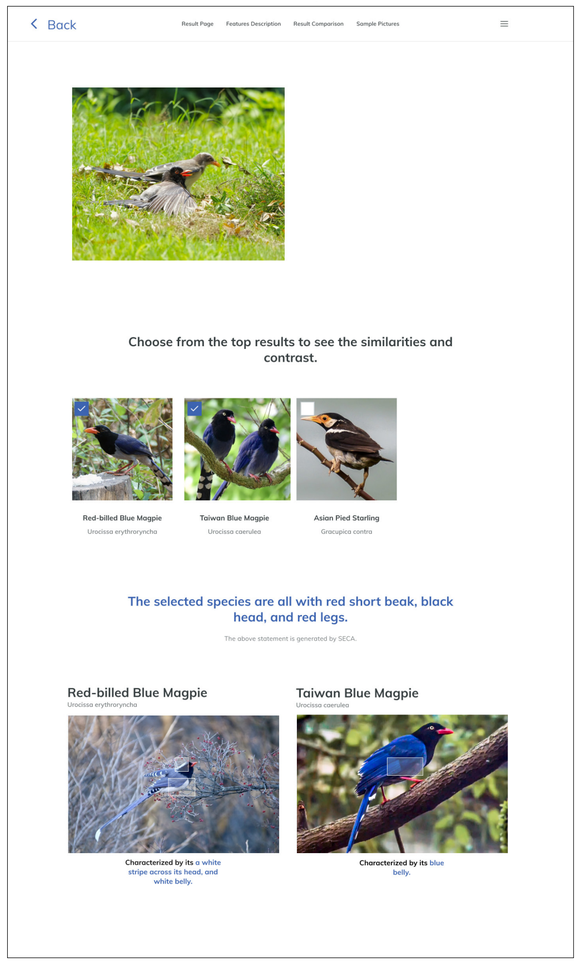
Normal Prediction Page



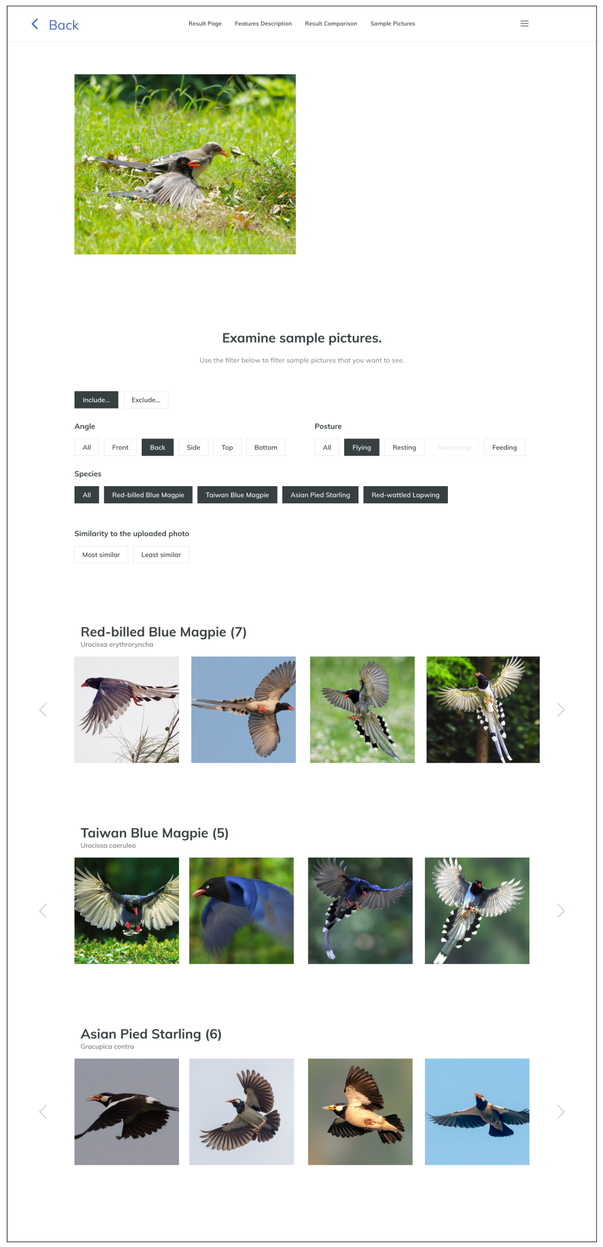
Feature Description



Result Comparison



Showing samples



Appendix G: Evaluation Protocol of Explanation Tests

Mock-up Evaluation Protocol

Research goal

- What are the goals they are pursuing while using the prototypes?
- Which aspects of the prototypes help them reach their goals? ->qualities of explanation needed by birders
- What else do they expect from the explanation?

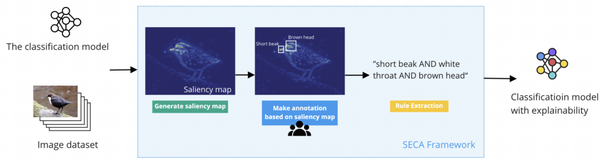
Informed Consent Form

You are being invited to participate in a research study on machine learning interpretability and bird species identification. This study is being done by a master student from the TU Delft.

To the best of our ability, your answers in this study will remain confidential. Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any question.

The process of this research will be recorded, for only reviewing and documenting and will not be shared with any third party.

Intro to SECA (SEmantic Concept Extraction and Analysis)



Opening Questions

1. What level you are at in telling birds apart?

   ☐Entrance-level
   ☐Intermediate-level
   ☐Advanced-level
   ☐Expert-level

2. Have you tried any bird apps?

3. Do you recognize this bird?

https://www.figma.com/proto/Kp6J17U9QchTpyVJYrISRf/GP-Mock-ups?page-id=147%3A4048&node-id=248%3A9164&viewport=1113%2C-1613%2C0.3879758417606354&scaling=min-zoom

Prototype NO.1

Rating

1=totally disagree, 5=totally agree, please rate the following statement:

In comparison to a normal bird ID app that you have tried…

1. I think this function is easy to understand and easy to use.
   ☐1 ☐2 ☐3 ☐4 ☐5
2. I think this function helps me understand why the predictions were made (and the reasons why certain predictions were false, if applicable).
   ☐1 ☐2 ☐3 ☐4 ☐5
3. I think this function brings me more certainty on whether the prediction is true or false.
   ☐1 ☐2 ☐3 ☐4 ☐5
4. I think this function enables me to learn more about birds.
   ☐1 ☐2 ☐3 ☐4 ☐5
5. I like this function.
   ☐1 ☐2 ☐3 ☐4 4.5 ☐5

Other questions

- Is there any other information you would like to see on the interface?
- Do you want it to be interactive? How?

Prototype NO.2

Rating

1=totally disagree, 5=totally agree, please rate the following statement:

In comparison to a normal bird ID app that you have tried…

6. I think this function is easy to understand and easy to use.
   ☐1 ☐2 ☐3 ☐4 ☐5

7. I think this function helps me understand why the predictions were made (and the reasons why certain predictions were false, if applicable).
   ☐1 ☐2 ☐3 ☐4 ☐5

8. I think this function brings me more certainty on whether the prediction is true or false.
   ☐1 ☐2 ☐3 ☐4 ☐5

9. I think this function enables me to learn more about birds.
   ☐1 ☐2 ☐3 ☐4 ☐5

10. I like this function.
    ☐1 ☐2 ☐3 ☐4 ☐5

Other questions

- Are there any other information you would like to see on the interface?
- Would you want the presentation of information be different?
- Do you want it to be interactive? How?

Prototype NO.3

Rating

1=totally disagree, 5=totally agree, please rate the following statement:

In comparison to a normal bird ID app that you have tried…

11. I think this function is easy to understand and easy to use.
    ☐1 ☐2 ☐3 ☐4 ☐5

12. I think this function helps me understand why the predictions were made (and the reasons why certain prediction was false, if applicable).
    ☐1 ☐2 ☐3 ☐4 ☐5

13. I think this function brings me more certainty on whether the prediction is true or false.
    ☐1 ☐2 ☐3 ☐4 ☐5

14. I think this function enables me to learn more about birds.
    ☐1 ☐2 ☐3 ☐4 ☐5

15. I like this function.
    ☐1 ☐2 ☐3 ☐4 ☐5

Other questions

- Is there any other information you would like to see on the interface? (Do you expect different filters?)
- Would you like to have a feature that enables you to describe the characteristics in text? How would you like to use it?
- Do you think it is important whether you are presented with photos from the training dataset? Does it make a difference to you?

Closing Questions

- What are the goals that you are pursuing when checking these prototypes? Which one is more important to you? What are the relationships between these goals? (Sequential, containment, Parallel)
- Are there any other ideas on how the web app can be designed? (It can be either with or without the SECA tech)
- Would you be interested in contributing your own annotation when using this website? Why?

Appendix H: Raw data Explanation Test Results

Opening Questions

1. **What level you are at in telling birds apart?**
   ☐**Entrance-level**
   ☐**Intermediate-level**
   ☐**Advanced-level**
   ☐**Expert-level**

👧-1: Entrance
👩-2: Intermediate
👦-3: Entrance
🧓-4: Intermediate
👩-5: Entrance
👨-6: Advanced/Expert (Master in ecology)

2. **Have you tried any bird apps?**

Entrance:
👧-1: Nope
👦-3: Tried some image search apps for both plants and animals, not specifically for birds.
👩-5: Yes, the Bird ID Master, Bird Fans International (a Chinese app for bird fans with a bird ID feature)

Intermediate:
👩-2: Yes, mainly Bird ID Master, and image search engine.
🧓-4: Yes, tried Bird ID Master, eBird

Advanced/Expert:
👨-6: Bird ID Master, and some digital bird guidebooks.

3. **Do you recognize this bird?**

👧-1: Nope
👦-3: No.
👩-5: No.
🧓-4: I think it is a red-billed or yellow-billed blue magpie.
👩-2: The closer one looks like a red-billed blue magpie, not sure about the further one.
👨-6: Yes, this is a red-billed blue magpie.

The Test Prototype

https://www.figma.com/proto/Kp6J17U9QchTpvVJYrlSRf/GP-Mock-ups?page-id=257%3A10489&node-id=147%3A4049&viewport=664%2C421%2C0.11633702367544174&scaling=min-zoom

Prototype NO.1

Rating

1=totally disagree, 5=totally agree, please rate the following statement:

In comparison to a normal bird ID app that you have tried…

1. **I think this function is easy to understand and easy to use.**
   ☐**1** ☐**2** ☐**3** ☐**4** ☐**5**

👧-1: **4**
💡I wish the photo uploaded and the predicted photos were on the same screen so I can compare them directly. Now they are too far away from each other.
💡For example, at the long press of the bounding boxes, show a close-up view of that area is shown.
👦-3: **4**
The information can be structured according to their importance hierarchy. I don't know where to check first the way you present it now.
👩-5: **4**

👩-2: **5**
Very straightforward and easy to use.
🧓-4: **5**

👨-6: **5**

2. **I think this function helps me understand why the predictions were made (and the reasons why a certain prediction was false, if applicable).**
   ☐**1** ☐**2** ☐**3** ☐**4** ☐**5**

👧-1: **5**
Based on my shallow understanding, I would think the model ONLY looks at these few areas that were marked out. But there must be more to the recognition than that.
💡Some subtle differences were not described, for example, the nuance of the bill shape. (👦-3 thinks so too ⬆️)
But the confidence value serves as a clue to know about the correctness.
Would be better if the time and location info is extracted automatically from the uploaded photo, and be taken into consideration.
prediction)
👦-3: **4**
It shows exactly why the predictions were made but maybe lacking some details.
👩-5: -
(The participants said she didn't understand the question, but gave a score anyway.)
I don't understand this question. For me, I only care about which one is the right answer and never bother thinking about how the computer works.

But I can tell the third prediction is incorrect because what's shown in the photo is different from the textual description. I think this is due to the different color perceptions of machines and of humans. Or maybe the data used to train the machine shows different colors under the influence of ambiance.

👩-2: **5**
If the prediction was false, by comparing my photo and the textual description, I can check if it is because certain features in my uploaded photo are not obvious. (understand the false
🧓-4: **4**
It is mostly clear why the machine made such predictions. But I was a little confused where the prediction of Pied Starling and Red-wattled Lapwing came from. Then I guess from this page that the machine makes its prediction based on separate characteristics instead of an overall impression.

But as humans the reasoning process works differently, I know it is very likely to be a blue magpie (genus), as the top 2 results are both blue magpies, so I would choose from the 4 kinds of blue magpies. So I think the weight of different factors might be different for humans and machines.

👨-6: **5**

3. **I think this function brings me more certainty on whether the prediction is true or false.**
   ☐**1** ☐**2** ☐**3** ☐**4** ☐**5**

👧-1: **4**
It really confused me as there are different answers and the highest confidence is just 36.5%. It doesn't tell me firmly which is the right one. I want it to tell me directly which is the right one, instead of leaving me uncertain. (stress out the most likely result)
If you show me the result first, then present photos with the bounding boxes, it will seem more trustworthy to me. (information presentation)
👦-3: **4**
It helps to compare the characteristics in the result pictures and the target pictures.
👩-5: **5**
✅The photos are clear and the highlight of the main characteristics helps.

👩-2: **4**
I would also need to check distribution, posture, and behavior information to be certain. Appearance is not the only clue involved.
🧓-4: **4**
I am very sure about the correctness of the prediction when combining all these 3 pages. But if it is only for this page, I will give it a 4.
I can not make a decision solely based on the 3 characteristics provided above (of the target photo), but by comparing the characteristics in the prediction photos and that in the target one, I can reach a conclusion.

If I didn't know this bird before, I will stay unsure after checking this page, cuz it only describes the characteristics, and the predicted species share some similar characteristics.

👨-6: **3**
Such explanation only helps to eliminate prediction with obvious mistakes, for example, taking one body part as another. But the subtle differences in their color, pattern, and vibe are still hard to distinguish.

4. **I think this function enables me to learn more about birds.**
   ☐**1** ☐**2** ☐**3** ☐**4** ☐**5**

👧-1: **2**
I only learn from the text, but it only describes the pictures. Something beyond the pictures themsevles is missing here. And seeing more sample pictures of different angles can also help to recognize.
👩-5: **2**
I would want to know more about the habits, habitats, behaviors of these birds.
👦-3: **3**
I would want to check more information other than seeing only the appearance of birds.

👩-2: **3**
✅This function is interesting. It gives me insights on which part to look at when observing these birds, though it might be universal for most birds.
I would like to know more about the appearance of birds.
🧓-4: **3**
It only provides information on birds' appearance, which I think is still different from so-called "bird knowledge". For me, I would like to see more detailed information about those birds. Introduction to their habitat, location, etc.
✅Mainly in that it helps me recognize what bird it is, so I can look up further info in the bird guide.
👨-6: **2**
⛔Provides limited knowledge for myself. ✅But it provides beginners tricks for identifying birds.

5. **I like this function.**
   ☐**1** ☐**2** ☐**3** ☐**4** ☐**5**

👧-1: **4**
The hierarchy is clearer here compared to the other two pages, as it put the most likely result on top and shows confidence.
👦-3: **5**
I always like trying out new features, so I would definitely like to use this function.
👩-5: **3**
This function is too basic to be liked.

👩-2: **3.5**
✅At first sight, it's very clear. And some birds are characterized by specific patterns in their body parts, so this will be helpful sometimes.

⛔But for me, the appearance doesn't take up too much weight in terms of justification.
🧓-4: **4.5**
✅Very nice that it can highlight the characteristics.
I really like this function. The only reason I don't give it a 5 is that I like the second one more.

👨-6: **5**

Supplementary questions

- **Is there any other information you would like to see on the interface?**
- **Do you want it to be interactive? How?**

👧-1: The text beside the picture shows exactly what I see in the picture, it's somewhat redundant. I'll expect it to show more detailed information than that, the intro of the birds with more pictures or videos, for example.
👦-3: The information structure can be improved based on users' aims and be catered for professionals and non-professionals. And provide entrance to more detailed information.
👩-5: Could show more information about these bird at the click of a "more" button.

👩-2: 💡 For species like Night Heron, which look very differently during sub-adult and adult periods, how would you present it?
🧓-4: 💡 There are only 4 kinds of blue magpies in total. So it would be of great help if you also link to the other 2 kinds of blue magpie here, which helps to determine a potential range. (provide information of species in a likely genus as reference)
💡 I want a function that allows me to pick an individual on the photo when there are multiple birds in it. For example in this photo, these two birds are of different ages, and I might want to check the bird behind. (👨-6 wants this function too)
👨-6: **5**
(about different mentalities)
With different expertise, you might distinguish the birds based on different clues. When you are beginer, you might tell birds apart based on color blocks. As you get advanced, you tell birds apart based on over all impression. And when you become expert, you are albe to tell the bird based on a very small part on its body. But all of these methods make sense.

For example, I am certained that this is a red-billed blue magpie mainly because of the small white patches on its tail, and the white patch on its head. So it surprise me that the machine actually looks at its back and head and bill.

Prototype NO.2

Rating

1=totally disagree, 5=totally agree, please rate the following statement:

In comparison to a normal bird ID app that you have tried…

6. **I think this function is easy to understand and easy to use.**
   ☐**1** ☐**2** ☐**3** ☐**4** ☐**5**

👧-1: **3**
I feel confused about why I need to select different species. It's like you raise a new question when my last question (what is this bird) still unsolved. I wish you could bring the conclusion (of what bird it is) to me first. Lack of guidance for beginners
💡I want the information's presentation to be different: after choosing the species, the comparison appears on where the photos are. Now it's a bit hard to match photos above to photos below.
👦-3: **2**
⛔I completely don't understand why I need this function and how to use it. As a beginner, I would choose to trust the prediction result completely, and if I don't trust it I will just leave. So I don't know what these functions are here for. Lack of guidance for beginners
👩-5: **4**
💡I want the presentation of information to be briefer and more to the point. For example, show the description on the photo or link them with lines. Then they will be easier to read and understand.

👩-2: **4**
✅It's fine overall. Just not as straightforward as the previous one. I like the textual description.
🧓-4: **5**
Very good.

👨-6: **4**
I didn't know at first that I could scroll down this page.

7. **I think this function helps me understand why the predictions were made (and the reasons why a certain prediction was false, if applicable).**
   ☐**1** ☐**2** ☐**3** ☐**4** ☐**5**

👧-1: **3**
I don't understand why the predictions were made. And I cannot see the confidence value here.
👦-3: **4**
I think the predictions were made based on more specific visual information than what is shown now.
👩-5: -
Still don't know how the model functions after seeing this page.

👩-2: **3**
It mainly compares different predictions. (Doesn't provide information on how they were reached)
🧓-4: **5**
The description of their similarities shows me the reasons for the predictions. And I also know from the description why it made the prediction of the Pied Starlings. The Pied Starlings don't have red bills and red legs, so I think it might be because the machines have a different perception of colors. -> correct annotations?
👨-6: **2**
⛔ It is not obviously revealed in this function as it doesn't show what the machine has recognized in the target photo. For example, you highlight the characteristics of "white belly" in the comparison, while it is not recognizable in the uploaded photo. So I think this doesn't explain its decision but provides extra information.

8. **I think this function brings me more certainty on whether the prediction is true or false.**
   ☐**1** ☐**2** ☐**3** ☐**4** ☐**5**

👧-1: **4**
⛔I'm still unsure which bird my bird is. I'm overwhelmed by the 4 birds presented here.
👦-3: **4**
I need more information than the only appearance.
👩-5: **5**
✅I like that it marks out the contrast of these two birds, especially some subtle differences that I couldn't notice myself.

👩-2: **5**
Very nice that it highlights the difference between the 2 similar birds. Even my uploaded photo doesn't include those characteristics, I might have other photos or have seen the bird from different angles (so I can check).
🧓-4: **5**
With the highlight of different characteristics of these two blue magpies, I can easily tell which prediction is correct. It is very accurate and helpful.
👨-6: **2**
Same as the previous question.
✅ But this(providing extra info other than what's in the photo) helps in some way, maybe the characteristics were not captured by photo but seen by the users in the flesh.

9. **I think this function enables me to learn more about birds.**
   ☐**1** ☐**2** ☐**3** ☐**4** ☐**5**

👧-1: **4**
✅It leads me to compare those different birds, and I learn what are their differences and similarities. The website is leading me to acquire knowledge from comparison. I'm viewing it with a clearer aim in mind.
✅And it shows pictures of different angles.

👦-3: **3**
⛔I expect to see more info.
👩-5: **5**
✅I think it teaches me hands-on knowledge in telling birds apart.

👩-2: **4**
✅It teaches me how to distinguish two similar birds.
🧓-4: **4**
✅This function can teach me how to distinguish two similar birds, which is very helpful.
👦-6: **3**
✅ Compared to other resources that only list out all the features of the birds, making comparison is helpful.

10. I like this function.
☐1 ☐2 ☐3 ☐4 ☐5
👧-1: **2**
✅ I like the information provided below.
⛔ But I feel completely lost when I was asked to choose from those species.
👦-3: **3**
⛔I think it shows only limited information.
👩-5: **4**
✅It is useful, ⛔but hard to read.

👩-2: **5**
✅Because it is very novel in the field of bird ID.
🧓-4: **5**
✅I like it very much! It's even better than the first function in that it provides a comparison after describing the features.
👦-6: **2**
I would prefer using bird guide to compare dozen birds at once.

Supplementary questions

- **Is there any other information you would like to see on the interface?**
- **Would you want the presentation of information to be different?**
- **Do you want it to be interactive? How?**

👧-1: 💡 I want it to tell me what bird this is first. Show the conclusion first and then show the justification. Now there's no focus in terms of the information hierarchy.
💡 Instead of showing a whole picture, show close-ups of certain body parts that are similar/different, making it easier to compare.
👦-3: information structure, to be more focused, more guidance
💡When showing the description of color, take into account the influence of sunlight.
👩-5: I want it to show more basic information, for example, the distribution of this bird, whether it is a migrant bird, when in a year does it show up around this area. These are important information for me when it comes to justification.
(Do you recognize this bird based on the information so far?)
I think it is most likely to be a red-billed blue magpie, but am not sure. Why not sure: the color is hard to identify as it is easily influenced by the ambient light, and the characteristics highlighted are masked on the target photo (the white stripe on the head and the white belly). And I noticed the bird in the target photo has red pupils in its eyes, while the pupils are black in the sample pictures.

👩-2: (no comments)
🧓-4: 💡Show some reference birds (eg. of the same genus) to compare.
💡Or even a feature that could compare the differences in the appearance of any two birds.
💡Train the model based on genus instead of species. Then it tells how likely this belongs to a certain genus. If I were the user, if I cannot know for sure which specific species it is, I would like to know how likely it is a blue magpie or a starling (on genus level). (already exist on iNaturalist)
But it could possibly be a problem that there is quite a diversity within some genus and the species look pretty different from one another, which might be a challenge for the machine to learn. So this idea might be more suitable for the genus which has more consistency within.

👦-6: 💡Comparison of their behaviors, moves, habitats, etc.

Prototype NO.3

Rating

1=totally disagree, 5=totally agree, please rate the following statement:

In comparison to a normal bird ID app that you have tried…

11. I think this function is easy to understand and easy to use.
☐1 ☐2 ☐3 ☐4 ☐5
👧-1: **4**
✅The similarity filter is most useful to me, so I want it to show first.
I will not need the species filters as I'm not sure what species it is so far, and will just choose all.
👦-3: **5**
It looks like those online shopping websites, so it's easy to understand.
👩-5: **5**
This is quite comprehensive, enabling me to see the birds from a tricky angle.

👩-2: **3**
⛔The filter section takes too much space for me, not very necessary. I would prefer to see photos of some default angles and posture directly.
🧓-4: **5**
👦-6: **5**

12. I think this function helps me understand why the predictions were made (and the reasons why a certain prediction was false, if applicable).
☐1 ☐2 ☐3 ☐4 ☐5

👧-1: **3.5**
✅ I also expect the boxes and description to be combined with the photos here. But it's nice that there are some textual filters I can choose from.
👦-3: **4**
Same as question 7, I think there must be more to it than what is shown here.
👩-5: -
I still can not make a connection between this page and the mechanism of the machines.

👩-2: **3**
I think it's not explaining the mechanism but providing information for me to make my own guess.
🧓-4: **4**
I can conclude from the dataset what are the rules behind the machines' predictions. And the rules I assumed are pretty much consistent with what I saw in function 2.
✅The advantage is that in these photos, I can see more details that are hard to describe, compared to the descriptions provided by the machines.
👦-6: **1**
⛔I think this is not related to how the model functions.

13. I think this function brings me more certainty on whether the prediction is true or false.
☐1 ☐2 ☐3 ☐4 ☐5
(The sample pictures I put here might be problematic and doesn't provide useful information)
👧-1: **2**
I still don't know what bird it is, especially when the photos below are completely different from the photo I uploaded.
👦-3: **3**
⛔Would be great if this can be combined with the previous 2 functions, showing the links between the target photo and the sample pictures.
👩-5: **5**
For example, if I want to identify a bird with a white patch on its wings, then by examining flying photos of the birds, I am more likely to tell what is this bird.

👩-2: **4**
✅Similar to the second function, by seeing the comparison of similar species from different angles, I can know the difference between them.
🧓-4: **5**
✅This function is in line with what I used to do to confirm the prediction: looking for more pictures in bird books.
👦-6: **1**
⛔ I think the certainty will rely more on how the model makes its prediction, so what characteristics it looks at.
✅ The similarity filter is very helpful for justification.

14. I think this function enables me to learn more about birds.
☐1 ☐2 ☐3 ☐4 ☐5
👧-1: **3**
At least I know furthermore how these birds look like.
👦-3: **3**
I would be interested in knowing information other than the appearance of birds.
👩-5: **5**
✅ With this function, I can easily compare the same posture of 2 similar-looking birds, the contrast will be more obvious. It teaches me more knowledge of birds' appearances.

👩-2: **3**
⛔ This page shows only pictures, lacking more in-depth information.
🧓-4: **4**
⛔ Extra info other than pictures is needed. ✅ But it is helpful indeed.
👦-6: **5**
✅ Checking such photo databases will be very helpful in improving bird ID skills.
This function is similar to what the oriental bird club does but more user-friendly.

15. I like this function.
☐1 ☐2 ☐3 ☐4 ☐5

👧-1: **3**

👦-3: **4**
This function is intuitive, helpful, and interesting. I will be happy to try it out.
👩-5: **5**
✅Sometimes when I search for more bird photos, I am shown all the pictures of a certain bird, and can not filter what I want to see. So the filter function will be very helpful to me.
✅And I like that there is an option "exclude", so I can exclude what I don't want from the result.

👩-2: **3**
🧓-4: **5**
👦-6: **5**
✅ Is very needed by advanced users.

Supplementary questions

- **Is there any other information you would like to see on the interface? (Do you expect different filters?)**

Expected: Filters of age, state of the feather, gender

👧-1: (no comment)
👦-3: I hope the filters not to be too professional, which will be a bit overwhelmed for me. I think the angle and posture options will be enough.
👩-5: I want to see videos showing how the birds move. 💡For example, White Wagtails are characterized by their flying in a sine curve. These kinds of dynamic characteristics also help to recognize the birds.

👩-2: 💡Filter for location, ages(different life stages), genders of the birds.
💡Would be better to show photos of multiple default postures of the predicted species, and move the filters section to the place below the sample pictures. (presentation of information)
🧓-4: 💡Filters of age, state of the feather, gender
👦-6: 💡As for filters, see what oriental bird database uses.

- **Would you like to have a feature that enables you to describe the characteristics in text? How would you like to use it?**

👧-1: I would like the feature of describing characteristics to be shown after the sample pictures. So I can examine other species if I don't find my answer.
👩-5: ✅It will be very helpful if I don't capture a clear picture of the bird or with no picture at all. It will be hard for beginners like me just to look into the guidebooks to find out what it is. so this sort of function will help a lot.

👦-3: ✅I'd like to use the feature input function as a supplement to the image information. Some information is hard to recognize on the image. I would also like to explore other birds with this function.

👩-2: ⛔I think textual information is less accurate. For example, it's hard to describe the back color of the winter kingfishers, it could be either blue or green. Input of color information

🧓-4: ✅It sounds good. And I hope I could set parallel conditions with this function, for example, red bill and red legs and dark back. And maybe also include the filters of orders and locations (beyond visual info). Then it will be very helpful.

I would expect something like the identification key in biology (which scores 5 in terms of learning), which is very accurate and takes the mutations into account. Though I know it might be too demanding for image identification. It is a difficult topic even in the academy.

👦-6: ✅ Yes I think this will be very helpful for the prediction. It will be very similar to looking into bird guides.
⛔Bird could be problematic for beginners as they might tell the characteristics incorrectly (taking blue as gray, blue as yellow, etc).

- **Do you think it is important whether you are presented with photos from the training dataset? Does it make a difference to you?**

👧-1: It makes no difference to me.
👩-5: don't care.
👦-3: I don't mind. As long as they are with clear backgrounds and the birds look good in the photos.
👩-2: (wasn't asked)
🧓-4: Don't mind. But the photos have to be accurate, not labeling one bird as another.
👦-6: (wasn't asked, but could tell from previous answer that he doesn't mind)

Closing Questions

- **What are the goals that you are pursuing when checking these prototypes? Which one is more important to you? What are the relationships between these goals? (Sequential, containment, Parallel)**

👧-1: **Learning** (knowing) > justification>transparency
Knowing what the bird is is the most important goal. And then the justification so I trust the result.
👦-3: **Learning** (knowing)
I care more about knowing what bird it is. And want to learn something along the way. And I don't care about how the machine works.

👩-5: I need to make sure the it is correct (**justification**), so that I can **learn**
Justification is the most important goal. I don't trust the prediction result blindly, it has happened to every app I used when the predictions are doubtable, especially when the picture I uploaded was not very clear. Usually, I doubt the prediction mainly because I notice differences in their appearances, or they are out of regular areas. But some have different appearances during molting, so it's hard to tell whether the prediction is correct in this case.

👩-2: Learning>Transparency>Justification
When watching birds I first want to learn bird knowledge. Then I want to know what characteristics do the machines focus on, which lets me know what to pay attention to when watching birds. Meanwhile, the accuracy of the prediction result is less important to me, I would ask around to get a good answer instead of relying on photo identification.

🧓-4: Justification > Transparency > Learning
For me, the most important goal is to confirm which species it is, and knowing the mechanism of the models helps in reaching that goal. And the learning part will follow naturally after reaching those two goals, it's not that important but will add to it.

👦-6: Justification
I would feel more comfortable using this app as it provides me with clues to justify its predictions, so I don't feel confused or probably being deceived by it.
Transparency is not important to me.
For learning, there are lots of competitors. ✅ I still believe in the potential of your product as it's one in a kind. ⛔ It's just that I already have much knowledge of birds so it won't be helpful to me.

- **Are there any other ideas on how the web app can be designed? (It can be either with or without the SECA tech)**

Dealing with unusual angle, ambient light.
👦-3: 💡A function that helps to deal with the influence of ambient light. For example, by showing sample pictures with similar ambiance, or by providing reminders on the interface to the users.
👩-5: I want to know what species are spotted by other birders near this area, and how can I spot them too. But it also worries me that this might be used by people who catch birds.

👩-2: 💡 It will be great if the technology can modify the ambient light and angle of the photos.
💡Include location and call as conditions to filter prediction results.
On the primary prediction result page, I would like to see the information shown in a fixed format, revealing the name, scientific name, habitat, distribution of that species. That will be helpful for me to make decisions.
💡The confidence value is important to me. For example, when I see the confidence of the most likely result is only 36.5%, I would think it is possibly due to the low quality of my photo.

🧓-4: 💡Provide links to other online photo databases, even they are not used for training.

- **Would you be interested in contributing your own annotation when using this website? Why?**

Not to be too open, reasonable reward (point, knowledge, or fun)
👧-1: What worries me is that the request for making annotations will be too open-ended and I don't know what to fill in. But I'm willing to help by putting the bounding boxes.
👦-3: 💡 I might be willing to do the annotation once or twice, but not that interested in doing it consistently.
👩-5: 💡While I am browsing information about such birds and notice some mistakes in the annotation, I would like to correct it. But I don't want to put too much extra effort into it, cause I am easily addicted to those easy and effortless tasks, and fear that I can't stop myself.

👩-2: I have participated in the construction of Wiki pages, so I think with an appealing bonus system or point reward system, I would be willing to do it.
🧓-4: Yes. I would like to make a contribution to the community. And I think I can also consolidate my knowledge throughout this process.
👦-6: The advanced birding hobbyists will be not only willing to but very passionate about taking part in such jobs. The key is to make it interesting, provide people with a sense of achievement, also help them in learning things at the same time.

Appendix I: Analysis of Explanation Test Results

Clusters of properties	Properties	Prototype NO.	Feedback	Relevant goal	Who holds a similar opinion?
Highlights on image and the textual description	Highlights on the target photo	1	💡"It will be more user-friendly if the highlights in the photo are presented in relation to the textual description."	(usability)	3(ABC)
	Description of the target photo	1	✅ "The description enables me to identify whether the prediction is trustworthy, or whether it fails because my uploaded photo is not good enough, for example, certain features not obvious."	✅Transparency, Justification	3(DEF)
	Highlights on the prediction photos	1	✅ "It taught me where to look at when seeing these birds."	✅Learning	3(CDE)
		1	⛔"I think it shows exactly how the computer makes its prediction, but I doubt there's more to it for identifying birds. I want to see more detailed descriptions of their features instead of general ones."	⛔ Transparency, Learning	2(AB)
	Description of the predicted photos	1, 2	✅"By providing info of features not shown in the uploaded photo, users are enabled to compare the info to what they have seen in the wild (but not captured by the camera)"	✅Transparency	4 (ACDF)
Comparison	Comparison between different prediction results	2	✅ This feature is useful for learning to distinguish birds"	✅Learning	5 (ACDEF)
		2	⛔"I want to see the textual description more connected to the features highlighted in the photo."	(usability)	3 (ABC)
		2	⛔"I want the description to tell me directly the differences between them instead of the way it is now, preferably with a close-up of the features."	(usability)	2 (AC)
	Comparison between target and prediction results	2,3	"I want to see the link between the target and predicted species, the fact that it's missing from the current prototype makes it hard to read."	(usability)	4 (ABCF)
Showing training data	Showing sample photos	3	✅"Checking sample pictures is helpful in identifying and learning birds."	✅Learning	4 (ACEF)
		3	"Showing a large number of sample pictures is a bit overwhelming for the beginners, but very helpful for advanced users."	✅Learning (usability)	5 (ABCDEF)
	Filters for sample photos	3	✅"The filters really help, compared to showing all photos of the birds."	(usability)	4 (BCDF)
		3	💡"I'd love to have filters for age, gender, or different features, such as feather color."	Learning	3 (DEF)

	Quote	Who hold a similar opinion?
Annotation collection	"I hope the queries will not be too open, because otherwise, I don't know what to fill in."	1 (A)
	"I will not be interested in doing it consistently."	3 (ABC)
	"I would love to do it with an appealing reward system."	1 (D)
	"Doing this will provide me with a sense of achievement and contribute to the birder community and enable myself to learn at the same time."	3 (DEF)
	"Be careful that not all the people are qualified to make the annotations, so you'd better test them first."	1 (F)

Other comments and Ideas:

Usability
(Now the explanation is a bit overwhelming to be useful to beginners.)
・Provides more cues on how to use the explanation for entrance-level users. For example, when it means inconsistency between the sample pictures and target pictures, what are the reasons behind the untruthful results.
・A hierarchical information structure is recommended to be easily understood by beginners.
・Show connections and contrast between target photo, prediction sample pictures, and textual description.

General
・Compared to how it is now in the mock-ups, the participants want to learn more detailed characteristics of the birds (like pattern, shape)

Input
・Input features manually as a supplement to the input of photos. (caution: beginners might interpret some features incorrectly)
・Modification of the influence of ambient light.

Prediction
・Present species of the same genus as reference.

Explanation
・Show comparison between target photo and prediction results
・Show comparison between prediction results
・Show sample pictures of the predicted species of different ages, gender, postures, molting stages, to present various appearances.

Annotation
・Verifying the qualification of the annotators
・Ways that are intriguing, enabling learning
・Request not to be too open-ended (for entrance-level annotators)

Appendix J: The Online Annotation Test Process

Test Plan

Research questions:

1. Are the end-users able to make annotations correctly on the photos with bounding boxes?
    a. Are the end-users able to make annotations for photos that the model found hard to classify?
2. Does making the annotation enable the end-users to be better / more confident at telling birds apart?

Hypothesis:

1. The participants can make annotations on the testing bird photos with bounding boxes (made according to the saliency maps of the classification model)
    a. The participants can make annotations even on the photos that are difficult to classify for the model.
2. The participants can do better in telling apart the birds after completing the annotation tasks.
    a. The participants will be more confident in telling birds apart after completing the annotation tasks.

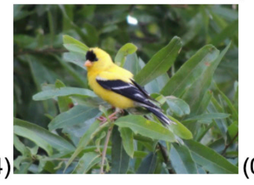
Participants:
8~10
I posted the online test in the subreddit /SurveyExchange and /SampleSize (surveys without limitation for participants), and a Chinese online forum for students studying abroad.

Materials:
A=American Goldfinch
B=Lesser Goldfinch

-Identification task I & II (each)

- For practice: 4 bird photos (2A+2B)

(0.988)

(0.999)

(0.544)

(0.999)

- Real task: 10 bird photos (5A+5B), among which, 4A and 3B are with normal classification confidence (confidence>0.8) by the model, 1A and 2B are with lower one (confidence<0.61).

task I:

(0.609)

(0.497)

(0.492)

(0.997)

(0.988)

(0.859)

(0.999)

(0.612)

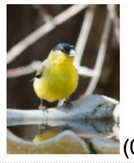
(0.939)

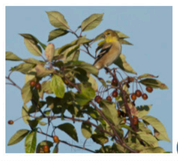
(0.999)

task II:

(0.029)

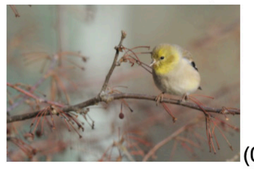
(0.278)

(0.270)

(0.992)

(0.993)

(0.748)

(0.990)

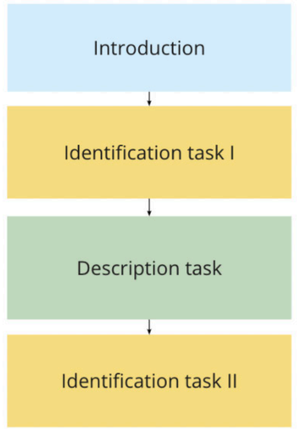
(0.858)

(0.990)

(0.999)

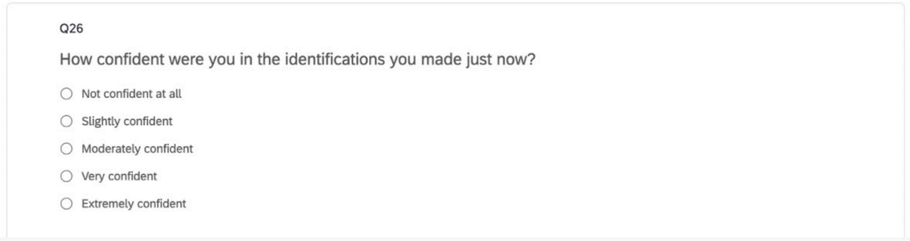
-Description task

- For practice: 2 bird photos with bounding boxes (1A+1B), both are with normal confidence (>0.8) by the classification model
- Real task: 4 bird photos (2A+2B), 2 with normal confidence(1A+1B), 2 with low confidence (confidence<0.6) (1A+1B).

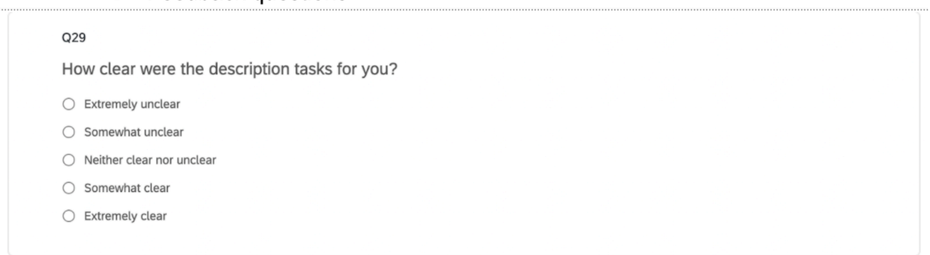
Steps:

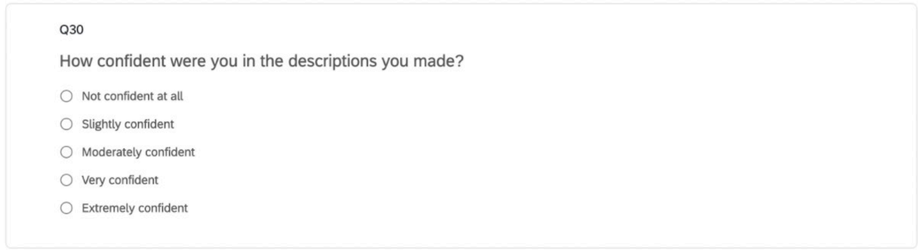


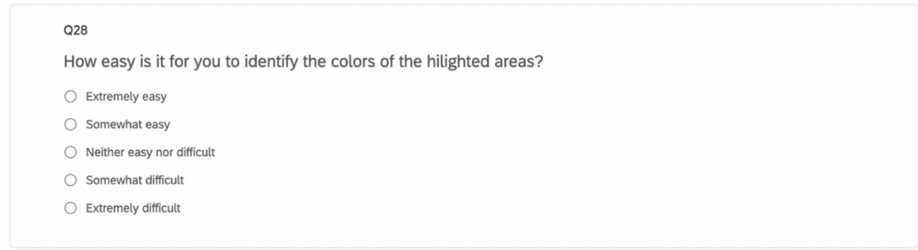
1. Introduction
    - An introductory diagram showing the main differences between A&B
    - Show sample photos (4A+4B)
2. Identification task I (pre-test)
    - Exercises:
        - Identification exercise with 4 bird photos (2A+2B). Show the correct answer after submission.
    - 10 bird photos (5A+5B)
    - Feedback question



3. Descriptions task:
    - Instructions and examples
    - Exercise*2:
        - Description exercise on 2 photos(1A+1B) with bounding boxes. Show the correct answer after submission.
    - Description task on a photo of A*2
    - Description task on a photo of B*2
    - Feedback questions







4. Identification task II(post-test)
    - 10 bird photos (5A+5B)
    - Feedback questions

Pilot test

Though the participant's score increased a bit (from 7/10 to 8/10), she said she wasn't consciously learning during the description tasks.

-Problems:

General:

- The participant had to scroll through the long page to check information (table of differences in identification task I and the topography/color diagram)

Identification task:

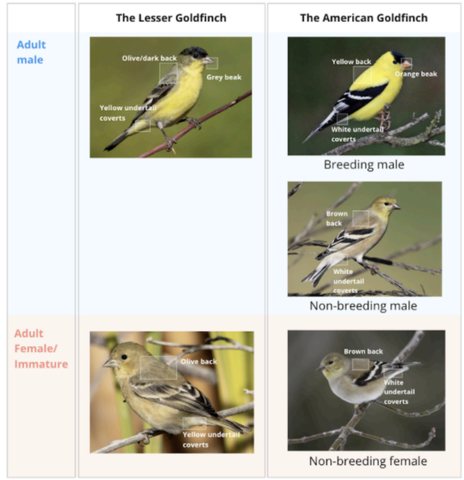
- The differences revealed in the diagrams weren't straightforward enough to memorize.
- Had relied too much on the difference table (took a picture of it)
- Overlooked the sample photos (due to the inconvenience of scrolling)

Description task:

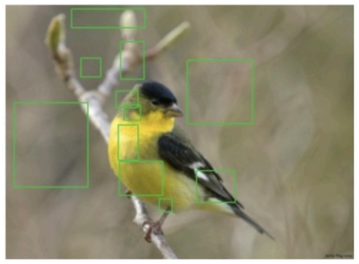
- Didn't learn much about identification because didn't pay attention to the bird names (partly due to inconvenience of use and the workload, and the participant wants to get it done as soon as possible)
- Overlooked the instruction (choose only one color)
- Got tired when doing the last description task

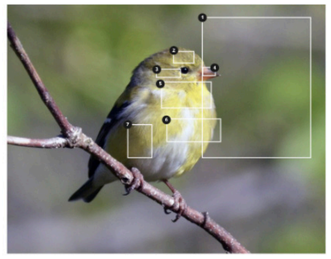
-Improvements:

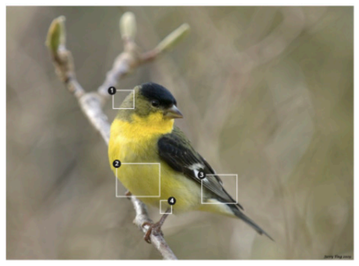
- Simplify the differences diagram (left: before; right: after)

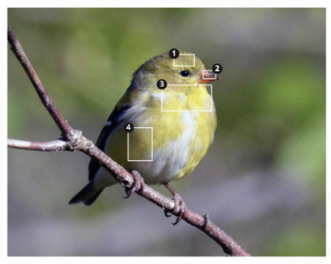


- Make the layout of the sample photos more readable
- In the instructions, emphasize the teaching function of the description task
- Reduce the scale of the description task, by removing some boxes from the easy tasks, for the participants to be more focused on learning
- Remove some bounding boxes on **easy** description photos (above: before; below: after)









Screenshots of Qualtrics Screens

Thank you for opening this link. You are now being invited to a test as part of a scientific research, studying how AI can help people in learning about birds.

During the test, you will be taught to identify two different species of birds.

The whole process will go as follows:

Firstly you will complete an identification task after going through some basic infomation on these birds' appearance.

Then, you are asked to describe the characteristics of the birds to get more familiar with their appearance.

At the end, you will complete another identification task to examine your knowledge.

It will take you about **20** minutes to complete the whole test.

Do you have any previous knowledge in identifying birds?

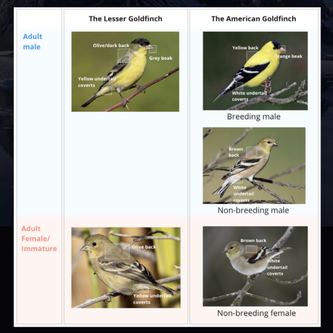
- ○ None at all
- ○ A little
- ○ A moderate amount
- ○ A lot
- ○ I'm an expert in birds

→

Pow

The birds you will be asked to identify are: **the American Goldfinch** and **the Lesser Goldfinch**.

First of all, we will like you to go through the following diagram, which shows their differences in appearance:



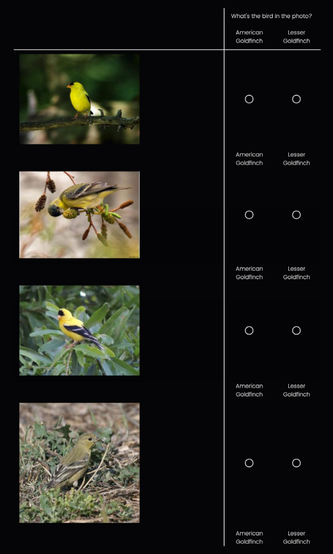
Then, please have a look at these sample photos of these two species, and see if you can notice what makes one species different from the other.

The following four photos are of **the American Goldfinch**.

The following four photos are of **the Lesser Goldfinch**.

With their differences in appearance in mind, now let's see if you can pass this identification challenge!

The following photos show either **an American Goldfinch** or **a Lesser Goldfinch**. Please identify the birds in the photos.

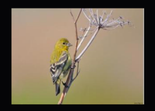


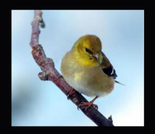
← →

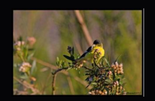
Powered by Qualtrics

You've got a grip! Now let's try some more.

The following photos have either **an American Goldfinch** or **a Lesser Goldfinch**. Please identify what's the bird in each photo.







What's the bird in the photo?	American Goldfinch	Lesser Goldfinch
	○	○
	○	○
	○	○
	○	○
	○	○
	○	○

How confident were you in the identifications you made just now?

- ○ Not confident at all
- ○ Slightly confident
- ○ Moderately confident
- ○ Very confident
- ○ Extremely confident

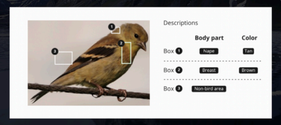
←

→

To help distinguish one species from another, it's useful to identify a particular part of the bird and its characteristics.

For this task, you are asked to describe the body part and color in the highlighted areas of the bird photos.

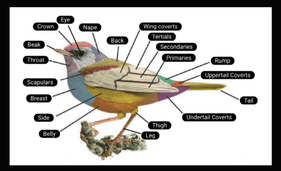
Here's an example of an American Goldfinch photo with highlighted parts and descriptions.

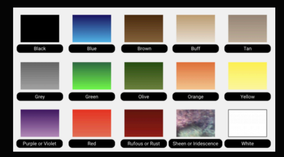


Notice:
1. Please describe **only one dominant part/color** in the highlighting box.

2. If there is no bird part in the highlighting box, choose "**Non-bird area**" and **leave the color blank**.

These two diagrams could help you to decide the words to describe the body parts and colors:





Let's give it a try!
Please describe this photo of an **American Goldfinch**.

	Name of the body part	Color of that area (please choose only the most prominent one. If there are more than one, either will be recognized as correct)
Box 1		
Box 2		
Box 3		
Box 4		

←

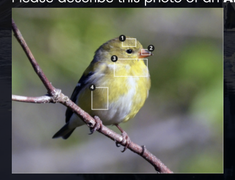
→

Powered by Qualtrics

Well done! Now let's try some more.

Please be aware of what's the species that you're describing during this task, as it will help you with the later identification task.

Please describe this photo of an **American Goldfinch**.



	Name of the body part	Color of that area (please choose only the most prominent one)
Box 1		
Box 2		
Box 3		
Box 4		

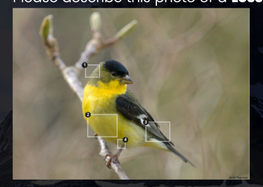
Please describe this photo of an **American Goldfinch**.

	Name of the body part	Color of that area (please choose only the most prominent one)
Box 1		
Box 2		
Box 3		
Box 4		
Box 5		
Box 6		
Box 7		

←

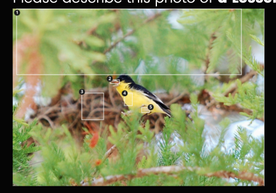
→

Please describe this photo of a **Lesser Goldfinch**.



	Name of the body part	Color of that area (please choose only the most prominent one)
Box 1		
Box 2		
Box 3		
Box 4		

Please describe this photo of **a Lesser Goldfinch**.



	Name of the body part	Color of that area (please choose only the most prominent one)
Box 1		
Box 2		
Box 3		
Box 4		
Box 5		

←

→

How clear were the description tasks for you?

- ○ Extremely unclear
- ○ Somewhat unclear
- ○ Neither clear nor unclear
- ○ Somewhat clear
- ○ Extremely clear

How confident were you in the descriptions you made?

- ○ Not confident at all
- ○ Slightly confident
- ○ Moderately confident
- ○ Very confident
- ○ Extremely confident

How easy is it for you to identify the **body parts** of the birds?

- ○ Extremely easy
- ○ Somewhat easy
- ○ Neither easy nor difficult
- ○ Somewhat difficult
- ○ Extremely difficult

How easy is it for you to identify the **colors** of the hilighted areas?

- ○ Extremely easy
- ○ Somewhat easy
- ○ Neither easy nor difficult
- ○ Somewhat difficult
- ○ Extremely difficult

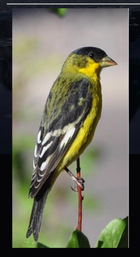
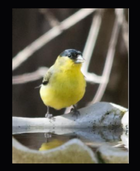
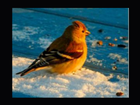
What did you find difficult about the **description task**?

←

→

Last one, let's see how well can you distinguish these birds now!

The following photos have either **an American Goldfinch** or **a Lesser Goldfinch**. Please identify what's the bird in each photo.

	What's the bird in the photo?	
	American Goldfinch	Lesser Goldfinch
	○	○
	○	○
	○	○
	○	○
	○	○
	○	○
	○	○
	○	○
	○	○
	○	○

How confident were you in the identifications you made just now?

- ○ Not confident at all
- ○ Slightly confident
- ○ Moderately confident
- ○ Very confident
- ○ Extremely confident

←

→

Powered by Qualtrics

Here's your score for the first identification task:
5/10

Here's your score for the second identification task:
6/10

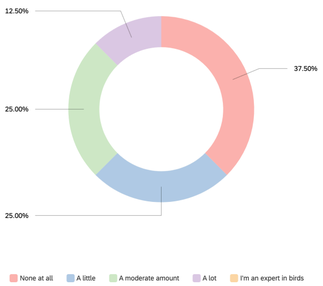
←

→

Appendix K: Raw Data of the Annotation Test Results

Note: The questions and results are presented here in the same sequence as that was shown to the participants. The sequence numbers are not natural numbers that increase constantly because the sequence altered and some questions were added or deleted during the editing process.

Q2 - Do you have any previous knowledge in identifying birds?



#	Answer	%	Count
1	None at all	37.50%	6
2	A little	25.00%	4
3	A moderate amount	25.00%	4
4	A lot	12.50%	2
5	I'm an expert in birds	0.00%	0
	Total	100%	16

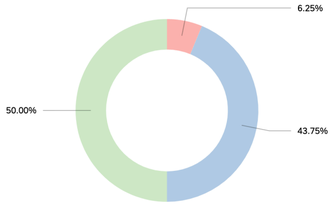
Score of identification task I

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Identification1	3.00	9.00	6.63	1.65	2.73	16

Scores broken down by people's knowledge

Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
None at all	5.00	8.00	6.33	1.25	1.56	6
A little	3.00	9.00	6.00	2.55	6.50	4
A moderate amount	6.00	8.00	7.25	0.83	0.69	4
A lot	7.00	8.00	7.50	0.50	0.25	2
I'm an expert in birds	0.00	0.00	0.00	0.00	0.00	0

Q25 - How confident were you in the identifications you made just now?



#	Answer	%	Count
1	Not confident at all	6.25%	1
2	Slightly confident	43.75%	7
3	Moderately confident	50.00%	8
4	Very confident	0.00%	0
5	Extremely confident	0.00%	0
	Total	100%	16

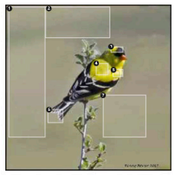
Results of description tasks

Collected annotations on body parts

Average correctness: 3.81/4 (0.95)

Box❶ Crown(16)
Box❷ Beak(16)
Box❸ Throat(8), Breast(6), Scapulars(2)
Box❹ Side(13), Belly(2), Breast(1)

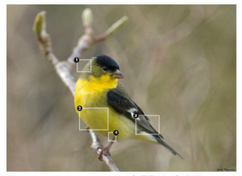
Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
dsc1A	5.00	8.00	6.94	0.90	0.81	16



Average correctness: 6.44/7 (0.92)

Box❶ Non-bird area(16)
Box❷ Non-bird area(16)
Box❸ Crown(16)
Box❹ Back(10), Scapulars(3), Side(2), Breast(1)
Box❺ Breast(7), Scapulars(6), Throat(1), Belly(1), Wing coverts(1)
Box❻ Tail(16)
Box❼ Non-bird area(16)

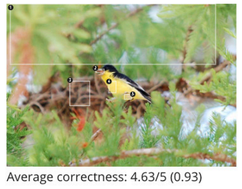
Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
dsc2A	5.00	11.00	9.31	1.45	2.09	16



Average correctness: 3.75/4 (0.94)

Box❶ Nape(16)
Box❷ Belly(13), Breast(3)
Box❸ Primaries(15), Secondaries(1)
Box❹ Thigh(16)

Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
dsc3B	4.00	8.00	6.63	0.93	0.86	16

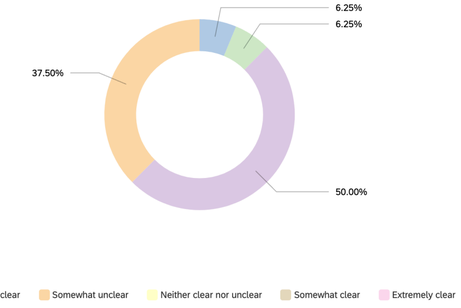


Average correctness: 4.63/5 (0.93)

Box❶ Non-bird area(16)
Box❷ Beak(15), Throat(1)
Box❸ Non-bird area(16)
Box❹ Breast(16), Side(1)
Box❺ Belly(12), Undertail coverts(1), Thigh(3)

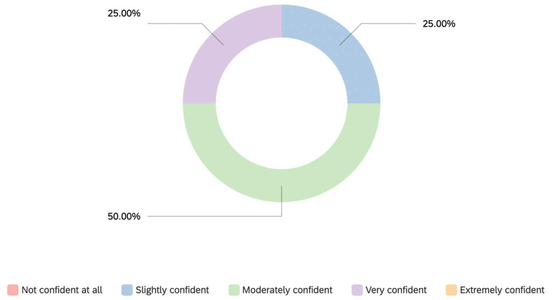
Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
dsc4B	3.00	5.00	4.63	0.60	0.36	16

Q29 - How clear were the description tasks for you?



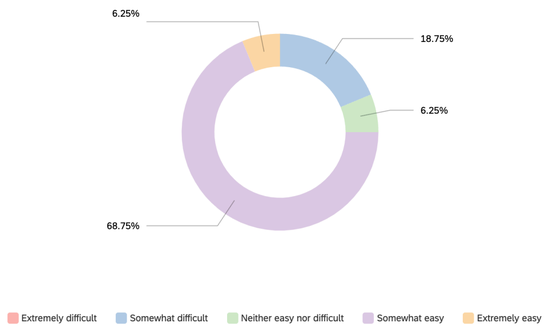
Answer	%	Count
Extremely unclear	0.00%	0
Somewhat unclear	6.25%	1
Neither clear nor unclear	6.25%	1
Somewhat clear	50.00%	8
Extremely clear	37.50%	6
Total	100%	16

Q30 - How confident were you in the descriptions you made?



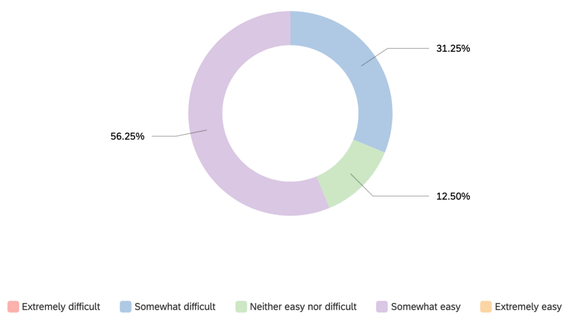
Answer	%	Count
Not confident at all	0.00%	0
Slightly confident	25.00%	4
Moderately confident	50.00%	8
Very confident	25.00%	4
Extremely confident	0.00%	0
Total	100%	16

Q27 - How easy is it for you to identify the body parts of the birds?



Answer	%	Count
Extremely easy	6.25%	1
Somewhat easy	68.75%	11
Neither easy nor difficult	6.25%	1
Somewhat difficult	18.75%	3
Extremely difficult	0.00%	0
Total	100%	16

Q28 - How easy is it for you to identify the colors of the highlighted areas?



Answer	%	Count
Extremely easy	0.00%	0
Somewhat easy	56.25%	9
Neither easy nor difficult	12.50%	2
Somewhat difficult	31.25%	5
Extremely difficult	0.00%	0
Total	100%	16

Q45 - What did you find difficult about the description task?

What did you find difficult about the description task?
finding the part in the list and picking a colour when there were to in equal proportions were the most difficult things, but still very easy
Hard to tell the body parts from one to another, especially for the tiny ones; and difficulties in color categorization as no lateral comparison showed as visual hints.
Choosing which two colors is dominant
bird in the diagram is not always the same shape/orientation as the bird in the photo
especially for the last lesser goldfinch, the bird on the image was far away and its parts were hard to distinguish. i also had of bit of a hard time deciding between olive/tan/grey a few times
ambiguity about which area was meant
some colors felt blended or unclear because the bird was small
some parts of the nody that we have to discribe have more than one color, but I can't choose both.
Identify colors
Some of the colors are very similar, like tan and buff or olive and tan and it's hard to differentiate.

Score of identification task II

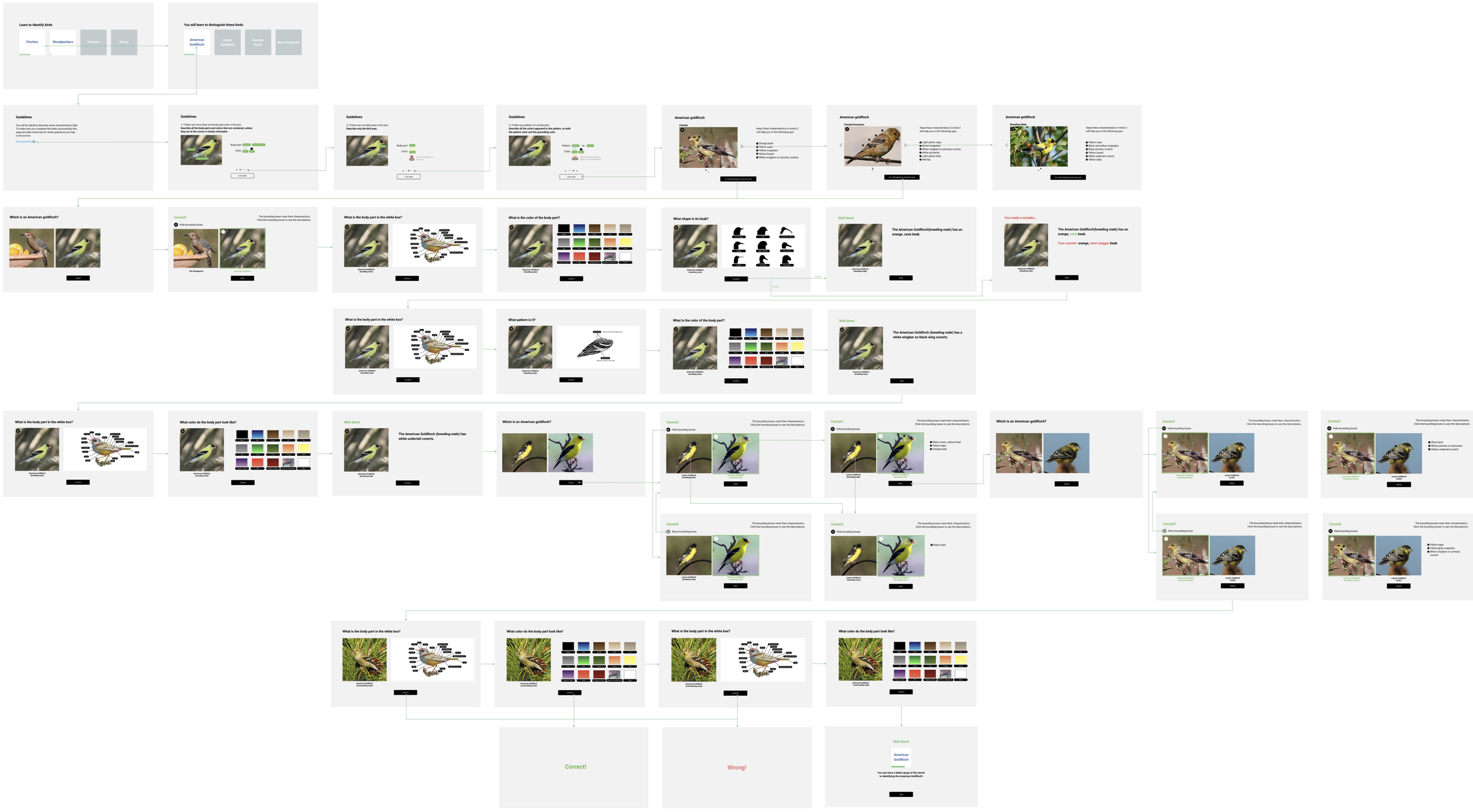
#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Identification2	4.00	10.00	7.56	2.03	4.12	16

Scores broken down by people's knowledge

Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
None at all	4.00	9.00	6.00	1.83	3.33	6
A little	5.00	9.00	7.50	1.66	2.75	4
A moderate amount	8.00	10.00	9.00	1.00	1.00	4
A lot	9.00	10.00	9.50	0.50	0.25	2
I'm an expert in birds	0.00	0.00	0.00	0.00	0.00	0

Appendix L: Prototypes for Interface Tests

Preview of live prototypes (EN) (CH)



Appendix M: Evaluation Questionnaire for Interface Tests

Q1 How much do you agree with the following statement? Please rate.

	Strongly disagree (0)	Somewhat disagree (1)	Neither agree nor disagree (2)	Somewhat agree (3)	Strongly agree (4)
1. I think that I would like to use this system frequently. (1)	○	○	○	○	○
2. I found the system unnecessarily complex. (2)	○	○	○	○	○
3. I thought the system was easy to use. (3)	○	○	○	○	○
4. I think that I would need the support of a technical person to be able to use this system. (4)	○	○	○	○	○
5. I found the various functions in this system were well integrated. (5)	○	○	○	○	○
6. I thought there was too much inconsistency in this system. (6)	○	○	○	○	○
7. I would imagine that most people would learn to use this system very quickly. (7)	○	○	○	○	○
8. I found the system very cumbersome to use. (8)	○	○	○	○	○
9. I felt very confident using the system. (9)	○	○	○	○	○
10. I needed to learn a lot of things before I could get going with this system. (10)	○	○	○	○	○

End of Block: Default Question Block

Start of Block: Other questions

Q3 Do the terms and diagrams of the bird topography helpful for you?

- ○ Definitely not (1)
- ○ Probably not (2)
- ○ Might or might not (3)
- ○ Probably yes (4)
- ○ Definitely yes (5)

Q4 Do the chronological order of different information/tasks seem appropriate for you?

- ○ Definitely not (1)
- ○ Probably not (2)
- ○ Might or might not (3)
- ○ Probably yes (4)
- ○ Definitely yes (5)

Q5 How much do you think the set-up of the game is engaging to you?

- ○ Not at all (1)
- ○ Slightly (2)
- ○ Moderately (3)
- ○ Quite engaging (4)
- ○ Very engaging (5)

Q6 How well do you think the app helped you in learning to identify the birds?

- ○ Not well at all (1)
- ○ Slightly well (2)
- ○ Moderately well (3)
- ○ Very well (4)
- ○ Extremely well (5)

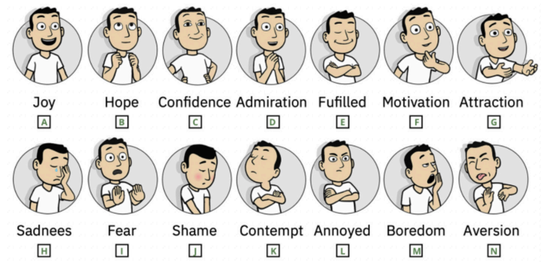
Q7 What aspects does it help you in learning about birds?

_____

End of Block: Other questions

Start of Block: Overall impression

Q8 How do you feel while using the product?



- ☐ A. Joy (1)
- ☐ B. Hope (2)
- ☐ C. Confidence (3)
- ☐ D. Admiration (4)
- ☐ E. Fulfilled (5)
- ☐ F. Motivation (6)
- ☐ G. Attraction (7)
- ☐ H. Sadness (8)
- ☐ I. Fear (9)
- ☐ J. Shame (10)
- ☐ K. Contempt (11)
- ☐ L. Annoyed (12)
- ☐ M. Boredom (13)
- ☐ N. Aversion (14)

End of Block: Overall impression

Appendix N: Interface Test Results

Results of questionnaires

SUS Score
Participant-1: 92.5
Participant-2: 90
Participant-3: 80

Average: 87.5

Information representation

1. Do the terms and diagrams helpful for you?[rating 1-5]
   Participant-1: 5
   Participant-2: 4
   Participant-3: 5

2. Do the chronological order of different information/tasks seem appropriate for you?[rating 1-5]
   Participant-1: 5
   Participant-2: 5
   Participant-3: 2

Engaging

3. How much do you think the set-up of the game is engaging to you?[rating 1-5]
   Participant-1: 4
   Participant-2: 4
   Participant-3: 1

   Participant-3: “I would prefer a lighter way of learning. Now there are too many details in it which make me fear.”

Learning

4. How well do you think the product helped you in learning to identify the birds? [rating 1-5]
   Participant-1: 5
   Participant-2: 4
   Participant-3: 4

5. What aspects did it help you learn about birds?
   Participant-1: The repeating. Seeing different photos. Learning about different parts of the bird.
   Participant-2: The aesthetics, the differences between the male/female, body parts of birds
   Participant-3: The detailed features of birds’ body parts have been marked out, which is very comprehensive.

Overall impression

6. How do you feel while using the app?
   Participant-1: Joy, hope, confidence, admiration, fulfilled, motivation, attraction
   Participant-2: Fulfilled, motivation, attraction
   Participant-3: Joy, confidence, fear

Observation Result

- **Perfect success**
- **So-so success** (made nonfatal mistakes, or felt confused)
- **Fail** (didn't complete the task as expected)

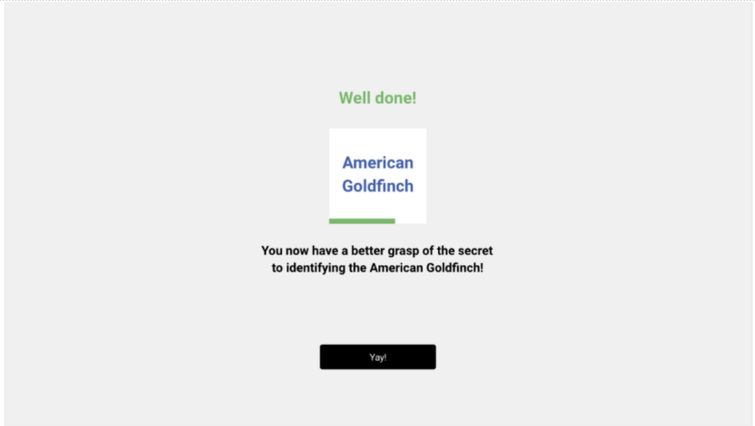
a. Ambiguous guidance
b. Bugs
c. Unnoticed indication

	Completion of tasks	Types of problems	Annotation
Select	P1 P2 P3		
Guidelines(3)	P1 P2 P3	a.	
Intro(3)	P1 P2 P3	a. c.	
Classification-1	P1 P2 P3	c.	
Annotation-1	P1 P2 P3	c.	P2: chose “short dagger” instead of “cone”
Annotation-2	P1 P2 P3	a.	P1 P2 P3: choose one color or both?
Annotation-3	P1 P2 P3		
Classification-2	P1 P2 P3		
Classification-3	P1 P2 P3	b.	
Annotation-4	P1 P2 P3		P3: not sure what to choose “breast”/ “throat”
Annotation-5	P1 P2 P3		P3: not sure what to choose “back”/ “nape”

Findings from interviews

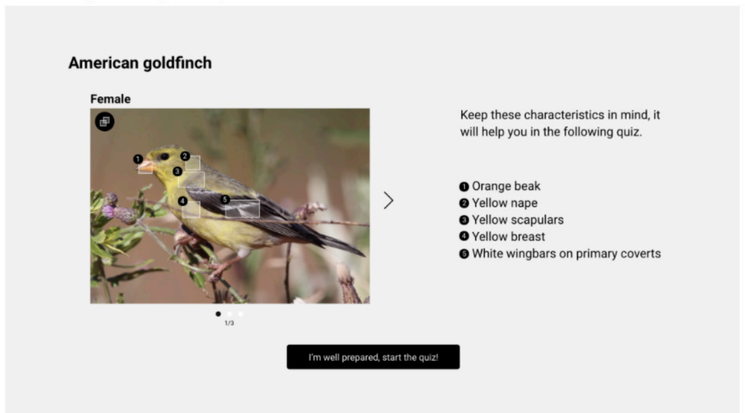
Positive feedback

- The repetition of the annotation act helps to learn
- Bringing attention to details
- Getting deeper
- The confirmation is cheering, making me feel I really achieved something



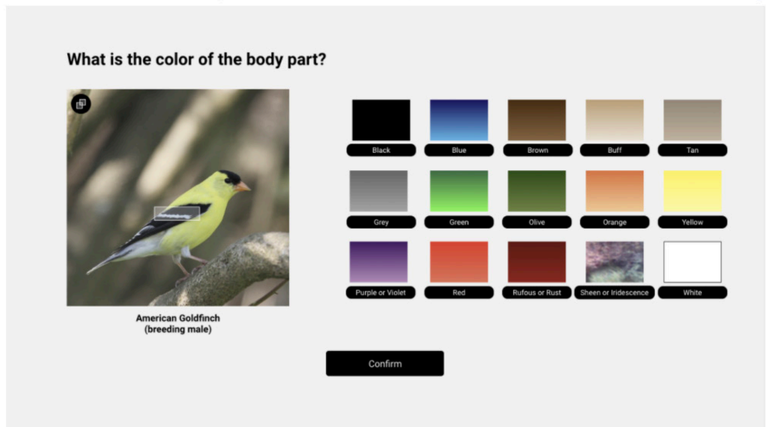
Recommendations

- On the guideline/intro page, provide a full example of the coming tasks. Participants were confused and a bit nervous at the guideline/intro page because they didn't know what to expect in the following tests.
- Enable guidelines on each page. All the participants didn't quite understand the guidelines and forgot about them in the tasks.
- Show a topography diagram along with/before the intro session. Participants felt confused when coming across the professional terms for the first time.
- Make the bounding box button and page indicator stand out more. They were ignored by most participants.



- Confusion on the color description.
  "Isn't it a yellow/grey/...?

- Confusion on how to annotate the pattern. For both participants B&C, it makes more sense to describe only the pattern but not the grounding colors. While all of them have forgotten what's said in the guidelines.
  "It should say 'select as many colors as possible' ".(participant A)
  "Because it says the color of the 'pattern', I think it should be white." (participant B)



- Bring more focus to the identification of different gender/ages.
  "It's still not super clear what are the differences between male/female, mature/immature birds." (participant-1)
  "It will be even funnier if there are tests of telling apart male/female birds."(participant-2)
- Transition sentences/page between different tasks.
  "It feels a bit unnatural when suddenly a classification task comes out after the annotation task. Maybe a transition sentence like 'now you are going to learn about how to identify female goldfinch' will be better." (participant-3)
- Participants still want to see feedback for the 4th and 5th annotation.

Other comments

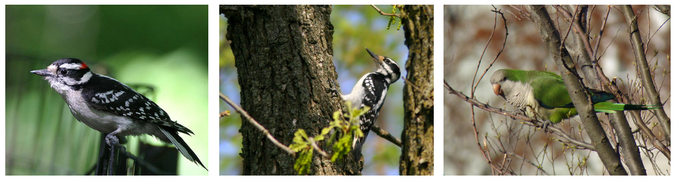
- Provide more in-depth information about birds.
  "There could also be sounds of the birds." (participant-1)
  "(I also want to see)Sound, flight, for example." (participant-2)

Appendix O: Information on the Classification Model and the Dataset

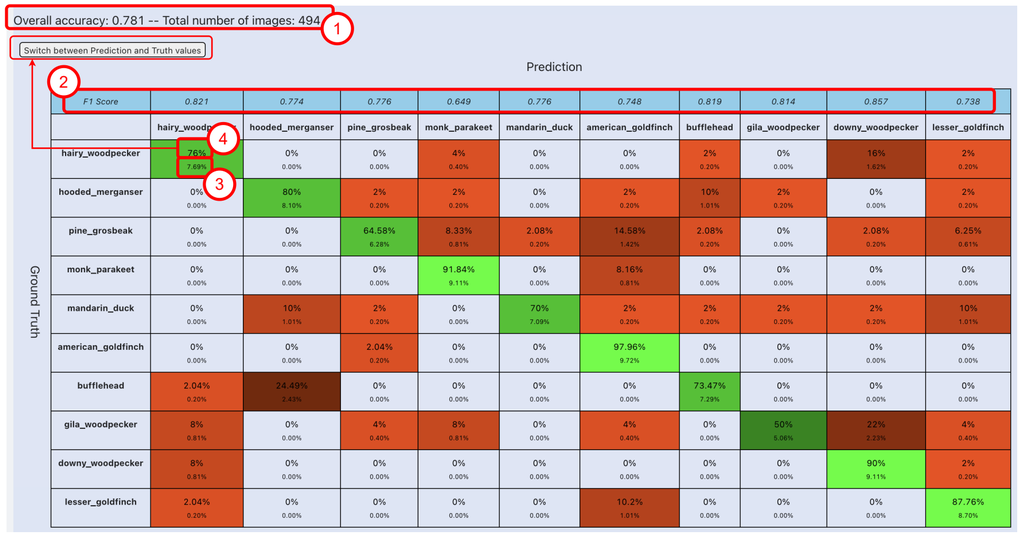
In this study, we used a 10-species bird image classification model developed by the SECA developers, as well as a dataset used to train and test this model. The dataset contains 1,470 testing photos and 1,291 training photos in total of the 10 species of birds.

List of bird species included in the classification model

- American Goldfinch
- Lesser goldfinch
- Gila woodpecker
- Pine Grosbeaks
- Mandarin duck
- Hooded merganser
- Bufflehead
- Downy Woodpeckers
- Hairy Woodpeckers
- Monk parakeet



Confusion Matrix of the model

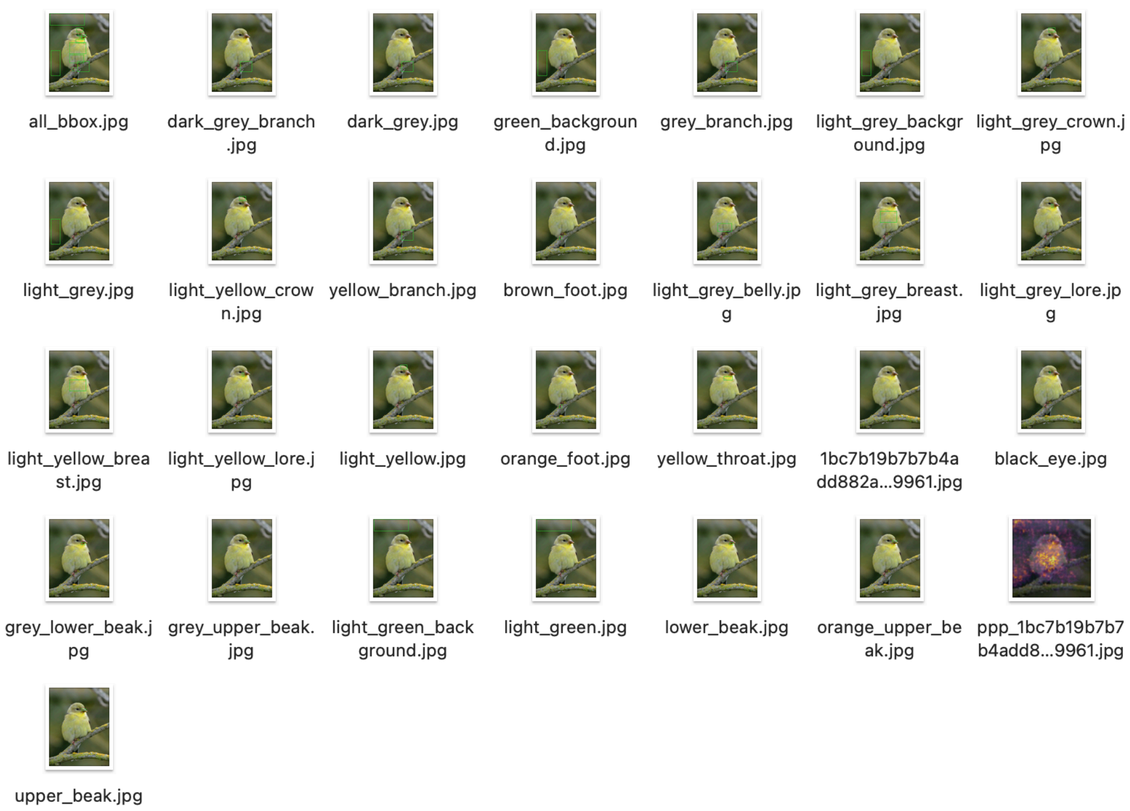


F1 Score	0.821	0.774	0.776	0.649	0.776	0.748	0.819	0.814	0.857	0.738
	hairy_woodp	hooded_merganser	pine_grosbeak	monk_parakeet	mandarin_duck	american_goldfinch	bufflehead	gila_woodpecker	downy_woodpecker	lesser_goldfinch
hairy_woodpecker	76% 7.69%	0% 0.00%	0% 0.00%	4% 0.40%	0% 0.00%	0% 0.00%	2% 0.20%	0% 0.00%	16% 1.62%	2% 0.20%
hooded_merganser	0% 0.00%	80% 8.10%	2% 0.20%	2% 0.20%	0% 0.00%	2% 0.20%	10% 1.01%	2% 0.20%	0% 0.00%	2% 0.20%
pine_grosbeak	0% 0.00%	0% 0.00%	64.58% 6.28%	8.33% 0.81%	2.08% 0.20%	14.58% 1.42%	2.08% 0.20%	0% 0.00%	2.08% 0.20%	6.25% 0.61%
monk_parakeet	0% 0.00%	0% 0.00%	0% 0.00%	91.84% 9.11%	0% 0.00%	8.16% 0.81%	0% 0.00%	0% 0.00%	0% 0.00%	0% 0.00%
mandarin_duck	0% 0.00%	10% 1.01%	2% 0.20%	0% 0.00%	70% 7.09%	2% 0.20%	2% 0.20%	2% 0.20%	2% 0.20%	10% 1.01%
american_goldfinch	0% 0.00%	0% 0.00%	2.04% 0.20%	0% 0.00%	0% 0.00%	97.96% 9.72%	0% 0.00%	0% 0.00%	0% 0.00%	0% 0.00%
bufflehead	2.04% 0.20%	24.49% 2.43%	0% 0.00%	0% 0.00%	0% 0.00%	0% 0.00%	73.47% 7.29%	0% 0.00%	0% 0.00%	0% 0.00%
gila_woodpecker	8% 0.81%	0% 0.00%	4% 0.40%	8% 0.81%	0% 0.00%	4% 0.40%	0% 0.00%	50% 5.06%	22% 2.23%	4% 0.40%
downy_woodpecker	8% 0.81%	0% 0.00%	0% 0.00%	0% 0.00%	0% 0.00%	0% 0.00%	0% 0.00%	0% 0.00%	90% 9.11%	2% 0.20%
lesser_goldfinch	2.04% 0.20%	0% 0.00%	0% 0.00%	0% 0.00%	0% 0.00%	10.2% 1.01%	0% 0.00%	0% 0.00%	0% 0.00%	87.76% 8.70%

Samples of saliency maps and bounding boxes materials provided by the developers

For every testing photo in the dataset of the target species, there was a folder containing the following materials as supplement:

- the original image
- the saliency map
- an image with all the bounding boxes
- k images with a single bounding box (k being the total number of possible bounding boxes for this image sample) so that you can also see each bounding box one by one
- for these k images, the name of the image correspondent to the label (annotation made by the developers) of the bounding box.



Confidence value table of predictions for test photos (the American and the Lesser Goldfinch)

image_nam	category	predicted	confidence
fe63ca47af	american_g	american_g	{'american_goldfinch': 0.9998030066490173, 'bufflehead': 1.1346950259394362e-06, 'downy_woodpecker': 2.3198184706529
1ead7bb8b	american_g	american_g	{'american_goldfinch': 0.9982343912124634, 'bufflehead': 4.1994937305389
8982d57f5	american_g	american_g	{'american_goldfinch': 0.6683021187782288, 'bufflehead': 0.000260036264
3fb41eddd	american_g	american_g	{'american_goldfinch': 0.9999842643737793, 'bufflehead': 1.52204773229
1fe0cf9ea	american_g	american_g	{'american_goldfinch': 0.9971908926963806, 'bufflehead': 3.752562406589
323f091fc	american_g	american_g	{'american_goldfinch': 0.9995393753051758, 'bufflehead': 6.42126991579
c162fe6f6	american_g	american_g	{'american_goldfinch': 0.9928699493408203, 'bufflehead': 0.037078621923
3c9d0a84f	american_g	american_g	{'american_goldfinch': 0.9995797276496887, 'bufflehead': 3.772157924686
a8017b6f6	american_g	american_g	{'american_goldfinch': 0.7795418500900269, 'bufflehead': 5.109861376695
8eca02ad1	american_g	american_g	{'american_goldfinch': 0.9888085722923279, 'bufflehead': 1.143260851677
ffb950c506	american_g	american_g	{'american_goldfinch': 0.9999901056289673, 'bufflehead': 2.497803741619
3757769d5	american_g	american_g	{'american_goldfinch': 0.5113171339035034, 'bufflehead': 2.506500251797
b6e676dbe	american_g	american_g	{'american_goldfinch': 0.9967522621154785, 'bufflehead': 7.550686742519
6cfcf1541	american_g	american_g	{'american_goldfinch': 0.9998082518577576, 'bufflehead': 6.017539533331
b2b741572	american_g	american_g	{'american_goldfinch': 0.8581004142761223, 'bufflehead': 0.000757173052
8246b7b46	american_g	american_g	{'american_goldfinch': 0.9940420985221863, 'bufflehead': 3.002675839525
33872ac49	american_g	american_g	{'american_goldfinch': 0.7481778860092163, 'bufflehead': 0.002626025583
5dcb78bec	american_g	american_g	{'american_goldfinch': 0.9999990463256836, 'bufflehead': 7.087585345288
63636b797	american_g	american_g	{'american_goldfinch': 0.9999495747052, 'bufflehead': 1.2307907581998734
7a23fb9f8f	american_g	american_g	{'american_goldfinch': 0.9767122864723206, 'bufflehead': 0.002433673944
fd396c5830	american_g	american_g	{'american_goldfinch': 0.9999622106552124, 'bufflehead': 3.821698513206
579bce4be	american_g	american_g	{'american_goldfinch': 0.9937118291854858, 'bufflehead': 0.000108347441
9c840308a	american_g	american_g	{'american_goldfinch': 0.9645223617553711, 'bufflehead': 1.115486227831
04cba0c3d	american_g	american_g	{'american_goldfinch': 0.6098848581314087, 'bufflehead': 0.00891348253935
9f5d75d19	american_g	american_g	{'american_goldfinch': 0.9963164329520809, 'bufflehead': 3.021121756319
1bc7b19b7	american_g	american_g	{'american_goldfinch': 0.5390270352363586, 'bufflehead': 3.032451968465
ae80c48ba	american_g	american_g	{'american_goldfinch': 0.9928125739097595, 'bufflehead': 0.000718069088
0d300d61d	american_g	american_g	{'american_goldfinch': 0.9888095259666443, 'bufflehead': 1.718551914109
19f0b0482	american_g	american_g	{'american_goldfinch': 0.9992898702146, 'bufflehead': 1.6366784229976394
6a8861026	american_g	american_g	{'american_goldfinch': 0.9960506558418274, 'bufflehead': 2.218746857352
68c912898	american_g	american_g	{'american_goldfinch': 0.9986428618431091, 'bufflehead': 1.017026079352
ef5eef58a	american_g	american_g	{'american_goldfinch': 0.9789925217628479, 'bufflehead': 5.660577357957
a95615156	american_g	american_g	{'american_goldfinch': 0.9185003042221069, 'bufflehead': 6.595213108084
de9fb1901	american_g	american_g	{'american_goldfinch': 0.8380356431007385, 'bufflehead': 0.004793022293
f420b9889	american_g	american_g	{'american_goldfinch': 0.9999996423713313, 'bufflehead': 1.173737736409
49c5067471	american_g	american_g	{'american_goldfinch': 0.9983553290367126, 'bufflehead': 0.000168863727
1681ed5e9	american_g	american_g	{'american_goldfinch': 0.8492465019226074, 'bufflehead': 1.748085338704
dba312e91	american_g	american_g	{'american_goldfinch': 0.9967694282531738, 'bufflehead': 2.131563547891
01d284227	american_g	american_g	{'american_goldfinch': 0.9997951388359007, 'bufflehead': 3.010804221048
8806f7df87	american_g	american_g	{'american_goldfinch': 0.9813945293426514, 'bufflehead': 3.583299346132
d85815eaf	american_g	american_g	{'american_goldfinch': 0.9704319834709167, 'bufflehead': 0.000583722197
f8cd008f8f	american_g	pine_grosb	{'american_goldfinch': 0.029314370825886726, 'bufflehead': 1.705275991525923e-07, 'downy_woodpecker': 9.045362276083324e-06, 'gila_woodpecker': 9.725003913274577e-09, 'hairy_woodpecker': 8.770313684181019e-07, 'hooded_merganser': 9.421512123630047e-05, 'mandarin_duck': 0.04744098335504532, 'monk_parakeet': 5.219947851742290
0f1fda1cbd	american_g	american_g	{'american_goldfinch': 0.9848839044570923, 'bufflehead': 5.515908469533315e-06, 'downy_woodpecker': 0.0012155696749687195, 'gila_woodpecker': 0.0001618971436928974, 'hairy_woodpecker': 4.201729825581424e-05, 'hooded_merganser': 9.239711653208366e-05, 'mandarin_duck': 0.012030153535306454, 'monk_parakeet': 1.8735896446742117e-05, 'pine_grosbeak': 0.000224431772949174
67dccf8ec9	american_g	american_g	{'american_goldfinch': 0.9990302324295044, 'bufflehead': 2.1039944897438545e-08, 'downy_woodpecker': 9.917395793947956e-09, 'gila_woodpecker': 1.8665201423573308e-05, 'hairy_woodpecker': 3.4433386986165715e-07, 'hooded_merganser': 1.4917657608748414e-05, 'mandarin_duck': 0.000928355264477431
5f7290bcbd	american_g	american_g	{'american_goldfinch': 0.9769586920738223, 'bufflehead': 1.3035657502769027e-05, 'downy_woodpecker': 0.007624265737831592
7facdfb6fd	american_g	american_g	{'american_goldfinch': 0.9902502894440155, 'bufflehead': 2.9401837309706025e-05, 'downy_woodpecker': 0.0003964345378355653, 'gila_woodpecker': 0.0012222296791151166, 'hairy_woodpecker': 0.0004760645097121596
19bdcf5004	american_g	american_g	{'american_goldfinch': 0.9264947175979614, 'bufflehead': 0.0005290450062602758, 'downy_woodpecker': 0.0001788798690540716, 'gila_woodpecker': 3.798287025347256e-06, 'hairy_woodpecker': 0.0002734224944841
ae8eef3b6	american_g	american_g	{'american_goldfinch': 0.9714066386222839, 'bufflehead': 0.0003656627377449516, 'downy_woodpecker': 0.0003914680797606707, 'gila_woodpecker': 0.0061026411113102436, 'hairy_woodpecker': 2.976724999825819
9f39dce805	american_g	american_g	{'american_goldfinch': 0.9917761087417603, 'bufflehead': 0.0004224589793011546, 'downy_woodpecker': 0.0003025020274549425, 'gila_woodpecker': 9.070962551049888e-05, 'hairy_woodpecker': 3.884354373441464e-05, 'hooded_merganser': 0.0001746174015913
3fc49d96ef	lesser_gold	lesser_gold	{'american_goldfinch': 0.0003749109455384314, 'bufflehead': 1.3960236300292195e-09, 'downy_woodpecker': 2.703476980059349
6349d1f3c3	lesser_gold	lesser_gold	{'american_goldfinch': 0.03407139703631401, 'bufflehead': 0.0001711665536276996
edfe68b6d4	lesser_gold	lesser_gold	{'american_goldfinch': 0.0011714977445080876, 'bufflehead': 1.1307984095765278e-06, 'downy_woodpecker': 2.787886721478116
0450365d7	lesser_gold	lesser_gold	{'american_goldfinch': 0.3911843597888465, 'bufflehead': 0.000130373082356527
d7f4e039a3	lesser_gold	lesser_gold	{'american_goldfinch': 0.009469252079725266, 'bufflehead': 0.00141403218731284
2025a3cbb	lesser_gold	american_g	{'american_goldfinch': 0.4756784141063695, 'bufflehead': 0.012424483895301819
b2c377cff1	lesser_gold	lesser_gold	{'american_goldfinch': 0.0001167380905826576, 'bufflehead': 0.000668150256387
62d72ad31	lesser_gold	lesser_gold	{'american_goldfinch': 0.2566740214824676
4615291f72	lesser_gold	lesser_gold	{'american_goldfinch': 0.08982377499341965, 'bufflehead': 2.369837784499356
0285fb1e44	lesser_gold	lesser_gold	{'american_goldfinch': 0.14071141183376312, 'bufflehead': 1.858867837434
3f73d0506a	lesser_gold	lesser_gold	{'american_goldfinch': 0.9995797276496887, 'bufflehead': 1.935637374117
32741aeb6c	lesser_gold	lesser_gold	{'american_goldfinch': 0.3079405128955841, 'bufflehead': 0.022347129881
acc682a47	lesser_gold	lesser_gold	{'american_goldfinch': 0.03166729956865311, 'bufflehead': 1.013978180708
b2f5c00451	lesser_gold	lesser_gold	{'american_goldfinch': 0.009711230173707008, 'bufflehead': 3.48765581748
61efb2f6a3	lesser_gold	lesser_gold	{'american_goldfinch': 0.0005491172196343541, 'bufflehead': 3.870118892024
a7ee677d0	lesser_gold	lesser_gold	{'american_goldfinch': 0.0084196068465004, 'bufflehead': 2.401046003797
2de76fe72c	lesser_gold	lesser_gold	{'american_goldfinch': 0.026053229346871376, 'bufflehead': 0.004988768137
fba3c697e2	lesser_gold	lesser_gold	{'american_goldfinch': 0.0665468722518634, 'bufflehead': 2.059935923170
3670ddf552	lesser_gold	american_g	{'american_goldfinch': 0.8241626602092664, 'bufflehead': 0.004327343776
827e16941	lesser_gold	lesser_gold	{'american_goldfinch': 0.17412705719470978, 'bufflehead': 0.001753037213
0628a95a5	lesser_gold	lesser_gold	{'american_goldfinch': 0.3950268626213074, 'bufflehead': 0.000251312681
deb25ce6a	lesser_gold	lesser_gold	{'american_goldfinch': 0.07808101922273636, 'bufflehead': 2.75162383331
656dd0e9a	lesser_gold	lesser_gold	{'american_goldfinch': 0.011595059186220169, 'bufflehead': 3.596601683142
cbfcfeb093	lesser_gold	lesser_gold	{'american_goldfinch': 0.07771868258714676, 'bufflehead': 0.014647281728
8ffe254a70	lesser_gold	lesser_gold	{'american_goldfinch': 0.002403575228571883, 'bufflehead': 7.982484362401
a980f50f88	lesser_gold	lesser_gold	{'american_goldfinch': 0.021895634010434315, 'bufflehead': 0.000715434085
bf2ad79e7f	lesser_gold	lesser_gold	{'american_goldfinch': 0.0001776225544745105, 'bufflehead': 3.74184054408
c5a824bac	lesser_gold	lesser_gold	{'american_goldfinch': 0.0010716228280216455, 'bufflehead': 1.360998425070
d8e98a299	lesser_gold	lesser_gold	{'american_goldfinch': 0.010628196410837
0d300d61d	lesser_gold	lesser_gold	{'american_goldfinch': 0.05463556945323944, 'bufflehead': 1.330119836495
f061dcd7e5	lesser_gold	american_g	{'american_goldfinch': 0.8412225246429443, 'bufflehead': 0.002186139114201
d68ac1373	lesser_gold	american_g	{'american_goldfinch': 0.7037954330444336, 'bufflehead': 0.001949480967596
a1b9d78cc	lesser_gold	lesser_gold	{'american_goldfinch': 0.01264851912856102, 'bufflehead': 0.000208326629945
df2a9454f3	lesser_gold	lesser_gold	{'american_goldfinch': 0.0014669166412204504, 'bufflehead': 4.92215583471553
7f69149b17	lesser_gold	lesser_gold	{'american_goldfinch': 0.0002728712570388787, 'bufflehead': 0.000428263883
d16303d6fe	lesser_gold	lesser_gold	{'american_goldfinch': 0.002455943264067173, 'bufflehead': 1.241694125297
079728241	lesser_gold	lesser_gold	{'american_goldfinch': 0.05142182856798172, 'bufflehead': 0.002107056536
cb1517f56d	lesser_gold	lesser_gold	{'american_goldfinch': 0.08716000616550446, 'bufflehead': 4.143365686104
f8a45c8d14	lesser_gold	lesser_gold	{'american_goldfinch': 0.005599236115813255, 'bufflehead': 7.26575208866
9cbfb349f6	lesser_gold	lesser_gold	{'american_goldfinch': 0.002982260659337048, 'bufflehead': 2.461495023453
31bbee17d	lesser_gold	lesser_gold	{'american_goldfinch': 0.03307780623435974, 'bufflehead': 1.334555221887
72e79e4b2	lesser_gold	lesser_gold	{'american_goldfinch': 0.15419804559005572e-05, 'bufflehead': 1.448717199536
f1d2da4cf8	lesser_gold	lesser_gold	{'american_goldfinch': 0.41579349020033035e-05, 'bufflehead': 7.926106263766997
60ba30673f	lesser_gold	american_g	{'american_goldfinch': 0.4985435307025909
efeb04c313	lesser_gold	hairy_wood	{'american_goldfinch': 0.14442776143550873, 'bufflehead': 3.081922113779001e-05, 'downy_woodpecker': 0.007624265737831592
9ab5032f8d	lesser_gold	lesser_gold	{'american_goldfinch': 0.46271994709968567, 'bufflehead': 0.000775836000686
a8cd28109f	lesser_gold	lesser_gold	{'american_goldfinch': 0.0003107868542426906, 'bufflehead': 5.285016300149437e-07, 'downy_woodpecker': 1.24112693459336e-08, 'gila_woodpecker': 4.710735268531607e-08, 'hairy_woodpecker': 0.0384320059972936e-10, 'hooded_merganser': 0.0005458001978695393, 'mandarin_duck': 0.6065223361656535e-06, 'monk_parakeet': 0.000611622329
fbbecc803c	lesser_gold	lesser_gold	{'american_goldfinch': 0.2022351324558258, 'bufflehead': 0.019066713750362396, 'downy_woodpecker': 0.0013874048599973
5364254fea	lesser_gold	lesser_gold	{'american_goldfinch': 0.0016134515171787059, 'bufflehead': 5.347737442207296e-06, 'downy_woodpecker': 9.515511192148551e-05, 'hairy_woodpecker': 5.228235750109888

The highlighting colors indicate which task the photos were used in the annotation task (Chapter 6).
Green & yellow: introduction;
Blue: identification tasks;
Orange: description tasks.

Appendix P: The Project Brief

DESIGN FOR our future

TU Delft

IDE Master Graduation

Project team, Procedural checks and personal Project brief

This document contains the agreements made between student and supervisory team about the student's IDE Master Graduation Project. This document can also include the involvement of an external organisation, however, it does not cover any legal employment relationship that the student and the client (might) agree upon. Next to that, this document facilitates the required procedural checks. In this document:

- The student defines the team, what he/she is going to do/deliver and how that will come about.
- SSC E&SA (Shared Service Center, Education & Student Affairs) reports on the student's registration and study progress.
- IDE's Board of Examiners confirms if the student is allowed to start the Graduation Project.

USE ADOBE ACROBAT READER TO OPEN, EDIT AND SAVE THIS DOCUMENT
Download again and reopen in case you tried other software, such as Preview (Mac) or a webbrowser.

STUDENT DATA & MASTER PROGRAMME
Save this form according the format "IDE Master Graduation Project Brief_familyname_firstname_studentnumber_dd-mm-yyyy".
Complete all blue parts of the form and include the approved Project Brief in your Graduation Report as Appendix 1 !

family name: Zeng
initials: W. given name: Wei
student number: 5103347
street & no.:
zipcode & city:
country pone:
email:

Your master programme (only select the options that apply to you):
IDE master(s): ( ) IPD ( ) Dfl ( ) SPD
2nd non-IDE master:
individual programme: - - (give date of approval)
honours programme: ( ) Honours Programme Master
specialisation / annotation: ( ) Medisign ( ) Tech. in Sustainable Design ( ) Entrepeneurship

SUPERVISORY TEAM **
Fill in the required data for the supervisory team members. Please check the instructions on the right !

** chair: Alessandro Bozzon dept. / section: SDE-IOT
** mentor: Dave Murray-Rust dept. / section: HCD-HICD
2nd mentor:
organisation:
city: country:
comments (optional)

Chair should request the IDE Board of Examiners for approval of a non-IDE mentor, including a motivation letter and c.v..

Second mentor only applies in case the assignment is hosted by an external organisation.

Ensure a heterogeneous team. In case you wish to include two team members from the same section, please explain why.

TU Delft

Procedural Checks - IDE Master Graduation

APPROVAL PROJECT BRIEF
To be filled in by the chair of the supervisory team.

chair Prof. Alessandro Bozzon date 13 - 05 - 2021 signature Alessandro Bozzon Digitally signed by Alessandro Bozzon Date: 2021.05.13 11:13:27 +02'00'

CHECK STUDY PROGRESS
To be filled in by the SSC E&SA (Shared Service Center, Education & Student Affairs), after approval of the project brief by the Chair.
The study progress will be checked for a 2nd time just before the green light meeting.

Master electives no. of EC accumulated in total: _____ EC
Of which, taking the conditional requirements into account, can be part of the exam programme _____ EC
List of electives obtained before the third semester without approval of the BoE

( ) YES all 1st year master courses passed
( ) NO missing 1st year master courses are:

name date - - signature

FORMAL APPROVAL GRADUATION PROJECT
To be filled in by the Board of Examiners of IDE TU Delft. Please check the supervisory team and study the parts of the brief marked **.
Next, please assess, (dis)approve and sign this Project Brief, by using the criteria below.

- Does the project fit within the (MSc)-programme of the student (taking into account, if described, the activities done next to the obligatory MSc specific courses)?
- Is the level of the project challenging enough for a MSc IDE graduating student?
- Is the project expected to be doable within 100 working days/20 weeks ?
- Does the composition of the supervisory team comply with the regulations and fit the assignment ?

Content: ( ) APPROVED ( ) NOT APPROVED
Procedure: ( ) APPROVED ( ) NOT APPROVED
comments

name date - - signature

Enabling Experts-in-the-loop Machine Learning Interpretability Methods — project title

Please state the title of your graduation project (above) and the start date and end date (below). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

start date 10 - 05 - 2021 — 27 - 09 - 2021 end date

INTRODUCTION **

Please describe, the context of your project, and address the main stakeholders (interests) within this context in a concise yet complete manner. Who are involved, what do they value and how do they currently operate within the given context? What are the main opportunities and limitations you are currently aware of (cultural- and social norms, resources (time, money,...), technology, ...).

Computer vision is a field of artificial intelligence that trains computers to interpret and understand the visual world. Image classification is a basic task in the field of computer vision. Among all the different cases that image classification could be useful for, the topic of bird species identification is chosen in this project. Birds are a crucial part of the ecosystem. Applying image classification in bird species recognition makes it more efficient in helping bird lovers identify birds, and could benefit the research and protection of birds, therefore is one of the research topics that naturalist communities and data scientists are interested in.

While with the AI bird species identification algorithms, the end users can only know the final outcome of the model's prediction, without having enough clue to know whether to trust the prediction or not. This is because they are short of knowledge about the mechanism of the classification models. By presenting textual or visual artefacts that explain the relationship between features of the input and the prediction result, the classification process will no longer remain a "black box" to the users. This is referred to as an interpretability method.

SECA (Semantic Concept Extraction and Analysis) is a human-in-the-loop framework that provides semantically diverse, informative and relevant explanations for people to understand machine learning models' behaviors. [Balayn et al. 2021] According to the SECA framework, the following steps will take place in order to get a model with interpretability: (1) generate saliency map based on image dataset and the learning model to explain, (2) have participants annotate on the saliency map, (3) extract classification rules based on the annotation. When interpretability is developed, explanations on machine learning models will be provided to the end users.

Currently, the SECA method hasn't been evaluated with real stakeholders, and the output of SECA is not designed to allow usage by people without a computer science background. The developers of the SECA framework are interested in knowing how interpretability can help the end users or other stakeholders to understand the model's capabilities and internal biases, to then study how trust in such models develop and evolve.

In this project, prototypes of a digital bird species identification platform with interpretability will be developed for exploration and evaluation. Target users of this platform include both amateur bird lovers and bird experts who have abundant knowledge in bird taxonomy. By doing so, this project will explore what are their separate needs for interpretability to trust the classification models, and evaluate among the target users what forms of information representation will be most appropriate for their needs.

space available for images / figures on next page

introduction (continued): space for images

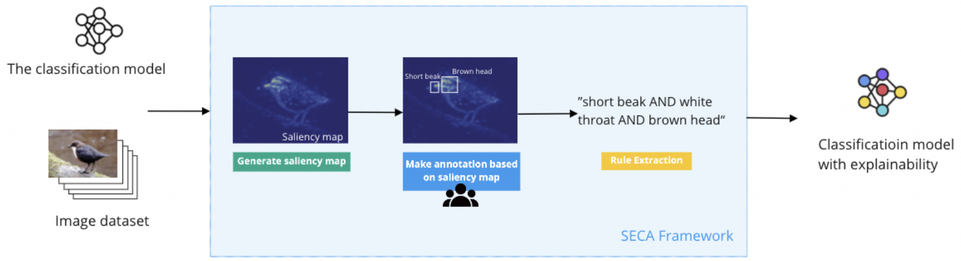


image / figure 1: Overview of the SECA framework

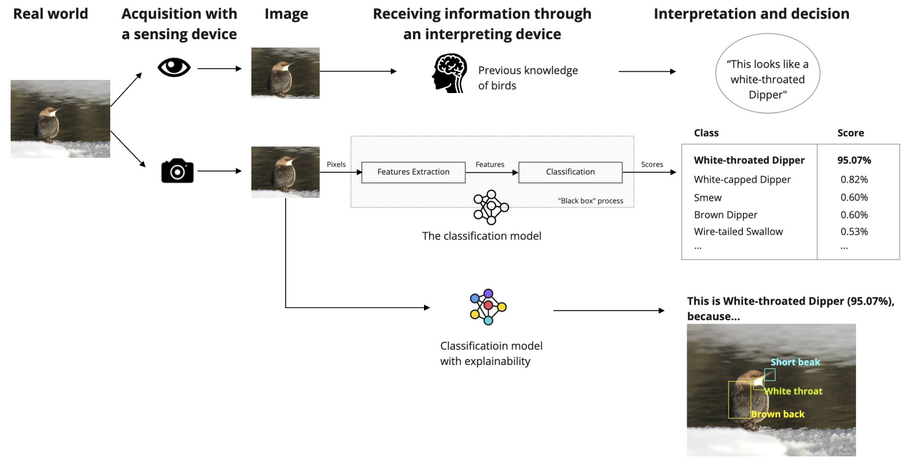


image / figure 2: Typical human and computer vision pipeline for species identification

PROBLEM DEFINITION **

Limit and define the scope and solution space of your project to one that is manageable within one Master Graduation Project of 30 EC (= 20 full time weeks or 100 working days) and clearly indicate what issue(s) should be addressed in this project.

*Design goal
-Design, propose and evaluate a bird species identification platform that features interpretability methods. The platform is intended to help stakeholders know what to trust about the bird species classification models.

Target users are mainly involved in the annotation step as annotators, and the explanation step as the audience. Throughout the design of such a platform, the following research questions concerning explanation and annotation will be answered.

*Overall research questions:
-How can domain experts contribute their domain knowledge to the interpretability methods, in order to provide explanations that indicate when the prediction of the bird classification model can be trusted?
-What forms of information representation will be most appropriate for them to know what to trust about the predictions?

*Sub questions:
-What are amateur bird lovers' and bird experts' separate needs for interpretability so they know when to trust the classification models?
-What forms of annotation are needed in order to develop the expected interpretability in relation to the various needs?
-How should the interaction be designed for the users to transfer their existing knowledge into information supported by the model?

ASSIGNMENT **

State in 2 or 3 sentences what you are going to research, design, create and / or generate, that will solve (part of) the issue(s) pointed out in " problem definition ". Then illustrate this assignment by indicating what kind of solution you expect and / or aim to deliver, for instance: a product, a product-service combination, a strategy illustrated through product or product-service combination ideas, ... In case of a Specialisation and/or Annotation, make sure the assignment reflects this/these.

Design a bird species identification platform with interpretability. The goal is to create a proper representation of explanation that can facilitate the target users' understanding and judgment of the identification algorithm.

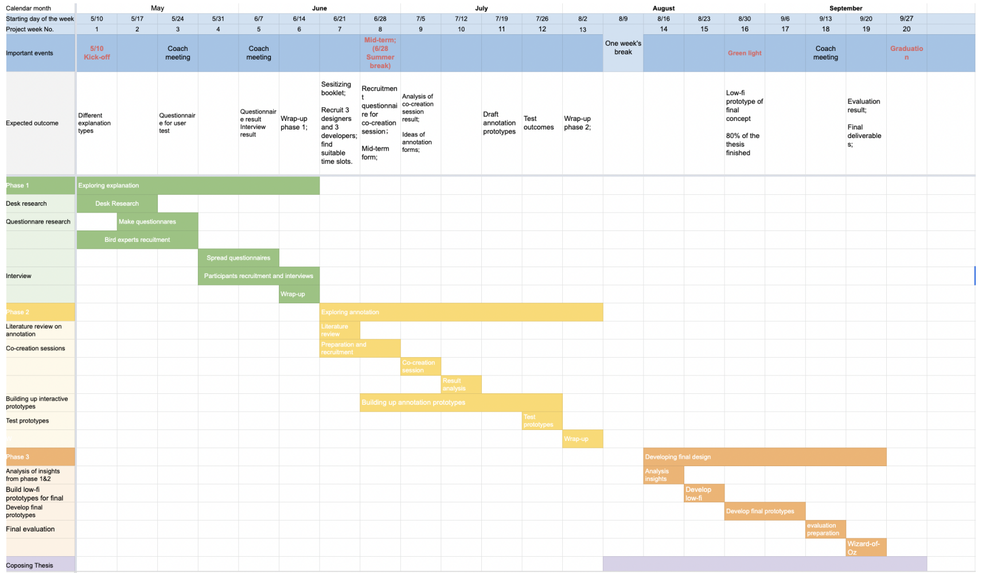
*Research
-Literature review on interpretability types, bird taxonomy and state-of-art bird identification models to understand: (1) what are the different ways to provide explanation for species classification; (2) what are the factors that can indicate when a classification model could be trusted.
-Conduct user research on what kind of explanation is wanted by different end users (amateur bird lovers and bird experts).

*Prototype and evaluation
-Develop interactive prototypes with P5.js for exploring annotation and explanation around the pre-trained people bird species identification models.
-Develop low and high fidelity prototypes with Figma for the final design of explainable bird species identification platform.
-Evaluate final interaction and interface design outcomes among target users.

PLANNING AND APPROACH **

Include a Gantt Chart (replace the example below - more examples can be found in Manual 2) that shows the different phases of your project, deliverables you have in mind, meetings, and how you plan to spend your time. Please note that all activities should fit within the given net time of 30 EC = 20 full time weeks or 100 working days, and your planning should include a kick-off meeting, mid-term meeting, green light meeting and graduation ceremony. Illustrate your Gantt Chart by, for instance, explaining your approach, and please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any, for instance because of holidays or parallel activities.

start date 10 - 5 - 2021 27 - 9 - 2021 end date



This project consists of 3 phases: Exploration of explanation, exploration of annotation and development of final design.

*Phase 1: Exploring explanation (6 weeks)
-Desk research: literature review of the state-of-art species classification models, interpretability types, and bird taxonomy
-Develop online questionnaire, and spread among 10-20 target users
-Interviews on 2 target users
-Ideation of explanation forms

*Phase 2: Exploring annotation (7 weeks)
-Co-creation session with developers and designers (4-6 participants)
-Building up of interactive prototypes with p5.js
-Test prototypes among 4 target users and get feedback

*Phase 3: Developing final design (7 weeks)
-Analysis of insights from phase 1&2
-Ideation of final concepts
-Building up final prototype
-Final evaluation

Personal Project Brief - IDE Master Graduation

TU Delft

MOTIVATION AND PERSONAL AMBITIONS

Explain why you set up this project, what competences you want to prove and learn. For example: acquired competences from your MSc programme, the elective semester, extra-curricular activities (etc.) and point out the competences you have yet developed. Optionally, describe which personal learning ambitions you explicitly want to address in this project, on top of the learning objectives of the Graduation Project, such as: in depth knowledge a on specific subject, broadening your competences or experimenting with a specific tool and/or methodology, ... Stick to no more than five ambitions.

*Competences
Have been coding for 5 years back when I was in middle school, I have always been fascinated by the beauty of algorithms design. When I first heard of machine learning years ago, I grew super interested in the ideas behind it.

With a background in Internet and New Media, I have sufficient experience in user interaction and interface design for digital platform. Courses from DFI master have equipped me with various design methods and approaches in conducting user research, developing iterative prototypes, and conducting usability tests.

During my third semester at IDE, I took two electives that are relevant to the chosen topic. In the elective Experimental Form-giving of Visual Information, I was exposed to multiple machine learning models in visual processing, and investigated in how to carry out visual experiments with open AI platforms. In another elective Machine Learning for Intelligent Products, I had a closer look at different categories of machine learning algorithms and learnt about how they can be applied in industry.

As a designer interested in both technology and human psychology, it is an attractive topic for me to explore how people get along with the artificial intelligent, as well as the strengths and constraints of applying AI in human everyday lives.

Thus, this graduation project on machine learning interpretability greatly meets my personal interests and aligns with the skills I have acquired in my study.

In this project, I will also investigate in how to merge the technology perspective into a design process, to explore my role as a designer in the technology field.

*Ambitions
-Learn to collaborate with developers to build experiential prototypes involving machine learning features.
-Develop my skills in building interactive low and high fidelity digital prototypes with P5.js and Figma.
-Learn to carry out user research and user tests under Covid-19 restrictions.
-Learn to share and communicate insights effectively with stakeholders from different backgrounds, such as computer science, ornithology, etc.

FINAL COMMENTS

In case your project brief needs final comments, please add any information you think is relevant.