Bird identification in Quebec : prediction results on 2016-2020

LIS CNRS

Table des matières

Table 1 – Selected 43 species.

Selected species	
AlderFlycatcher	NashvilleWarbler
AmericanRobin	NorthernFlicker
AmericanTreeSparrow	NorthernParula
Black-and-whiteWarbler	NorthernWaterthrush
Blue-headedVireo	Olive-sidedFlycatcher
BlackburnianWarbler	Ovenbird
BlackpollWarbler	Rose-breastedGrosbeak
Black-throatedBlueWarbler	Ruby-crownedKinglet
Black-throatedGreenWarbler	Red-wingedBlackbird
CanadaGoose	SavannahSparrow
CanadaWarbler	SongSparrow
CommonYellowthroat	SwainsonsThrush
Chestnut-sidedWarbler	TennesseeWarbler
Dark-eyedJunco	TreeSwallow
EasternWood-Pewee	Veery
FoxSparrow	White-crownedSparrow
GreatCrestedFlycatcher	WinterWren
Golden-crownedKinglet	White-throatedSparrow
Gray-cheekedThrush	Yellow-belliedFlycatcher
LeastFlycatcher	Yellow-belliedSapsucker
LincolnsSparrow	Yellow-rumpedWarbler
MagnoliaWarbler	

1 Introduction

The dataset is made up of 15-minute recordings recorded every hour from April to August 2016, 2017, 2018, 2019. There are 75 recording sites across Quebec distributed in three main habitats : tundra, wetlands, and forests. 43 species were selected by bird experts to be submitted to the classification model. These species are not equally distributed across Quebec but represent species susceptible to climate change or specific to an habitat. The list of species is indicated in Table 1.

2 Methods

An energy detector analyzed all files to slice the recordings containing a signal in 10sec samples. Overlapping samples were discarded in order to reduce the number of samples with chorus (several species singing at the same time). However, this method can include chorus samples because the first detected signal can be a chorus. This step requires proper identification of chorus in future tests. Chorus detection and identification are major challenges in automatic bird classification because models are, up to now, not able to differentiate several acoustic sources.

10 sec samples are represented in a Mel spectrogram in the convolutional neural networks (CNN). This is one of the many representations of a sound signal in the time-frequency domain. Mel spectrograms are a good representation of human hearing. These 2D representations of the signal are the material on which CNN work (Figure 1).

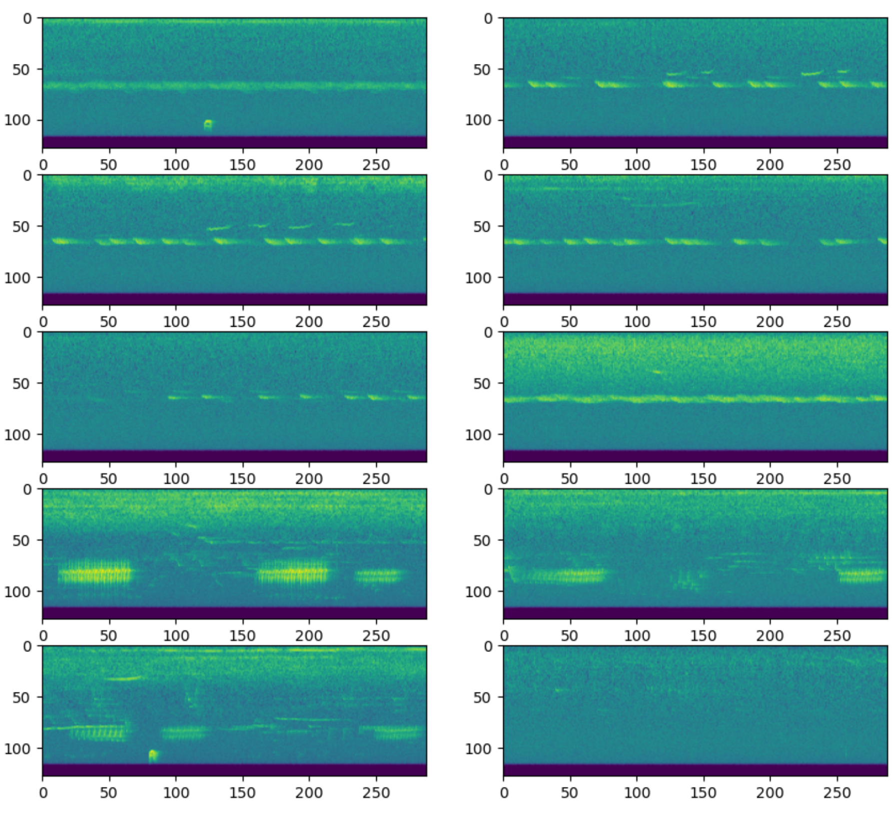

FIGURE 1 – Mel spectrograms used by the CNN to learn bird vocalizations

ResNet 18 model is the CNN used for bird identification due to its simplicity and proven accuracy.

The model is trained on xeno-canto dataset (recordings from all Canada). Xeno-canto data is used instead of a training directly on Quebec dataset to ensure a high quality of recordings (e.g. no chorus, no adverse weather conditions), with a clear and loud bird vocalization and a communitybased identification. The dataset is split into training and testing set. Once the testing accuracy is maximal on xeno-canto (82%), the CNN weights are saved.

The selected CNN with a good training accuracy is then used to identify samples from Quebec. A probability for 43 species is attributed by the model to each recording. The highest probability is chosen as the prediction from the model. Different methods were used in this step but to date, none has proven to be significantly more effective. However, when refining predictions given by the model, these methods could be used. For instance, data with low entropy or acoustic complexity index (ACI) could be selected as good quality data (with low surrounding noise and a clear single bird signal) which increases their good identification. Final predictions could be an weighted average of probabilities on the 43 species. Some potential weights are the time window (e. g. prediction on 3 sec windows in a 10 sec sample with a higher weight on prediction of the central 3 sec window) or the species ecology (e.g. tundra species cannot be detected in forest habitats).

3 Results

Predictions for Quebec recordings will be made available on `http://sabiod.univ-tln.fr/pub/`

4 Discussion

Presented results do not contain data augmentation (addition of environmental and abiotic noise to the recordings). Refining the results with models adapted to habitat or localization characteristics (e.g. tundra sites are more exposed to strong winds which decrease recordings quality) is one of the promising options that is being tested. Correct parameters for each model have yet to be identified.

Predictions should be taken with extreme caution as they are the result of an on-going process that is being corrected and refined. The background noise seems to be identified by the model rather the bird vocalization. The predicted species is not accurate in this case. This should be corrected by the use of several localization-based models. Future perspectives lie in the detection and suppression of chorus, the detection of good quality recordings that can accurately be predicted by the CNN, and in the creation of distinct testing sets (grouped by sites, month) to avoid background noise identification.