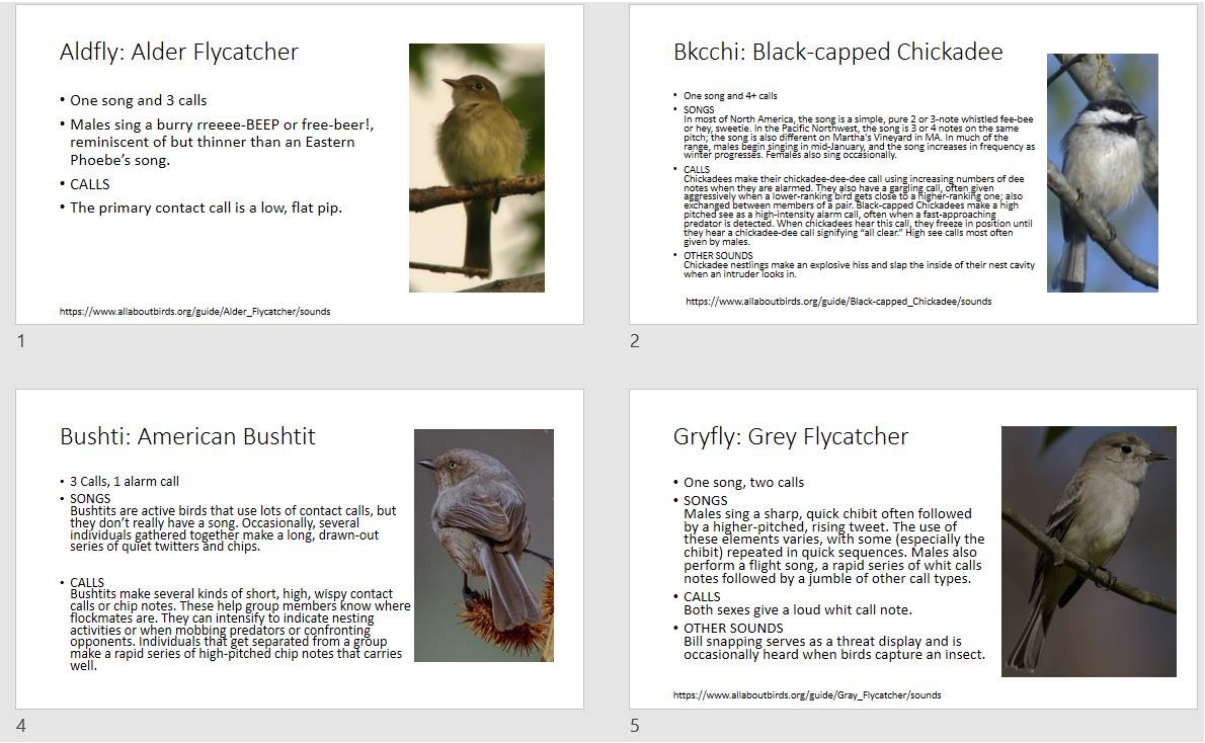
Elektor Labs

Project Ornithowave: Bird Species Identification

Avian diversity is a commonly used metric to assess environmental change and is growing in importance. In this project we use the MAX78000 to detect and classify birds based on their songs and calls. Thereby enabling remote sensing of avian diversity



Motivation & Overview

Avian diversity is a commonly used metric to assess environmental change and is growing in importance due to climate change and it's ecological effects. In the last few years, deep learning approaches have shown promise in reducing the expert labor intensive bird identification process. The 157 layer ResNet based BirdNet from Cornell University is currently the state of the art in this field and can identify almost a 1000 different bird species [1]. However implementations of this model are limited to areas where high bandwidth internet is available as the model is evaluated in the cloud.

Ecological Informatics 61 (2021) 101236

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

BirdNET: A deep learning solution for avian diversity monitoring

Stefan Kahl [a,*], Connor M. Wood [a], Maximilian Eibl [b], Holger Klinck [a]

[a] Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850, USA
[b] Technische Universität Chemnitz, D-09111 Chemnitz, Germany

We propose a simplified CNN based on the KWS-20 model to identify 20 different bird species based on their sounds and calls.
This project running on the ultra-low power MAX78000 can make bird monitoring possible in very remote locations without worrying about high bandwidth connectivity or big batteries. It can also be used by bird enthusiasts who want to keep track of bird species visiting their home or farm.

This Project Execution Comprised of 4 phases:
1. Literature Review & Open-Source Dataset Identification
2. Data Engineering
3. Model training & C-code synthesis using ai8x-training & ai8x-synthesis libraries
4. Testing on Hardware using audio samples from the test set.

Phase 1: Literature Review & Open-Source Dataset Identification

The Cornell Lab of Ornithology is an active research group in this space and regularly publish papers, and host competitions on Kaggle (BirdCLEF 2021: https://www.kaggle.com/c/birdclef-2021, CBL 2020: https://www.kaggle.com/c/birdsong-recognition/overview)

Their publications and datasets served as a solid foundation for this project. The networks that win these competition are very large and complex, for our project we used a subset of this dataset and retrained the KWS-20 model (ai85kws20) from maxim by replacing words with birds.

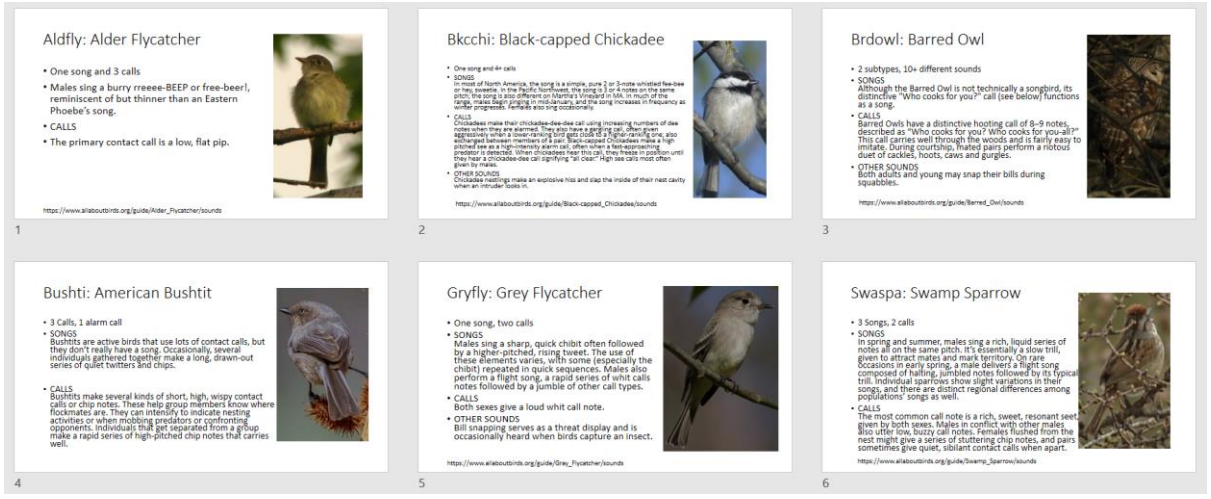
However, a major constraint is that all this data is weakly labelled and not standardized with varying lengths, quality and sampling frequency.

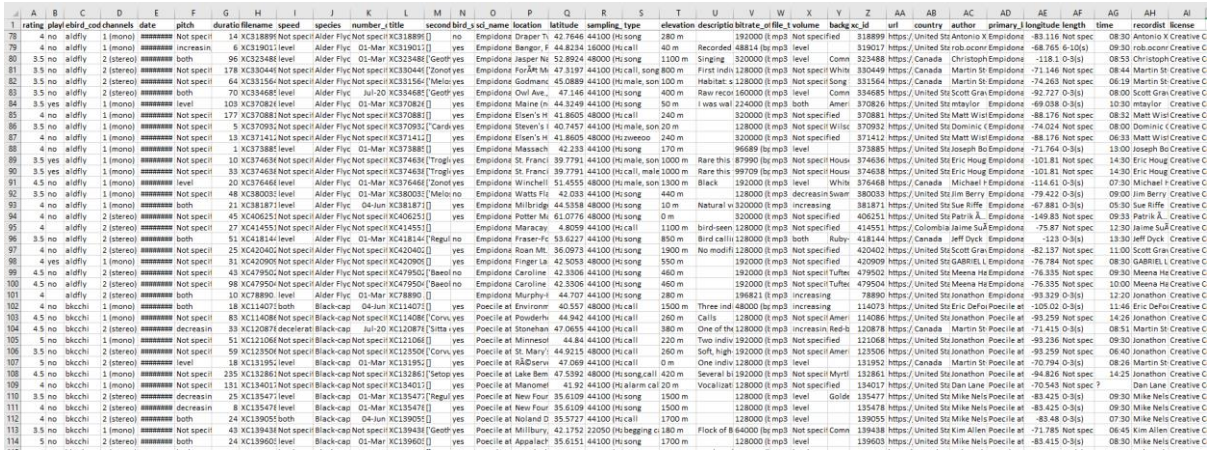
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1	rating	playl	ebird_cod	channels	date	pitch	duratio	filename	speed	species	number_r	title	second	bird_s	sci_name	location	latitude	sampling_	type	elevation	descriptio	bitrate_of	file_t	volume	backg	xc_id	url	country	author	primary_l	longitude	length	time	recordist	license
78	4	no	aldfly	1 (mono)	########	Not specif	14	XC318899	Not specif	Alder Flyc	Not specif	XC318899	[]	no	Empidona	Draper Tv	42.7646	44100 (Hz	song	280 m		192000 (b	mp3	Not specified		318899	https://.	United Sta	Antonio X	Empidona	-83.116	Not spec	08:30	Antonio X	Creative C
79	4	no	aldfly	1 (mono)	########	increasin	6	XC319011	level	Alder Flyc	01-Mar	XC319011	[]	yes	Empidona	Bangor, F	44.8234	16000 (Hz	call	40 m	Recorded	48814 (bp	mp3	level		319017	https://.	United Sta	rob.oconr	Empidona	-68.765	6-10(s)	09:30	rob.oconr	Creative C
80	3.5	no	aldfly	2 (stereo)	########	both	96	XC323488	level	Alder Flyc	01-Mar	XC323488	['Geoth	yes	Empidona	Jasper Ne	52.8924	48000 (Hz	song	1100 m	Singing	320000 (b	mp3	level	Comn	323488	https://.	Canada	Christoph	Empidona	-118.1	0-3(s)	08:53	Christoph	Creative C
81	3.5	no	aldfly	2 (stereo)	########	Not specif	178	XC330441	Not specif	Alder Flyc	Not specif	XC330441	['Zonot	yes	Empidona	ForÃªt Mo	47.3197	44100 (Hz	call, song	800 m	First indiv	128000 (b	mp3	Not specif	Whits	330449	https://.	Canada	Martin St	Empidona	-71.146	Not spec	08:44	Martin St	Creative C
82	3.5	no	aldfly	2 (stereo)	########	Not specif	64	XC331564	Not specif	Alder Flyc	Not specif	XC331564	['Melo	yes	Empidona	Godmanc	45.0889	44100 (Hz	male, son	100 m	Habitat: s	128000 (b	mp3	Not specif	Song	331564	https://.	Canada	Martin St	Empidona	-74.263	Not spec	06:19	Martin St	Creative C
83	3.5	no	aldfly	2 (stereo)	########	both	70	XC334685	level	Alder Flyc	Jul-20	XC334685	['Geoth	yes	Empidona	Owl Ave,	47.146	44100 (Hz	song	400 m	Raw recor	160000 (b	mp3	level	Comn	334685	https://.	United Sta	Scott Grai	Empidona	-92.727	0-3(s)	08:00	Scott Grai	Creative C
84	3.5	yes	aldfly	1 (mono)	########	level	103	XC370826	level	Alder Flyc	01-Mar	XC370826	[]	yes	Empidona	Maine (n	44.3249	44100 (Hz	song	50 m	I was wal	224000 (b	mp3	both	Amer	370826	https://.	United Sta	mtaylor	Empidona	-69.038	0-3(s)	10:30	mtaylor	Creative C
85	4	no	aldfly	1 (mono)	########	Not specif	177	XC370881	Not specif	Alder Flyc	Not specif	XC370881	[]	yes	Empidona	Elsen's H	41.8605	48000 (Hz	call	240 m		320000 (b	mp3	Not specified		370881	https://.	United Sta	Matt Wisl	Empidona	-88.176	Not spec	08:32	Matt Wisl	Creative C
86	3.5	no	aldfly	1 (mono)	########	Not specif	5	XC370932	Not specif	Alder Flyc	Not specif	XC370932	['Cardi	yes	Empidona	Steven's I	40.7457	44100 (Hz	male, son	20 m		128000 (b	mp3	Not specif	Wilse	370932	https://.	United Sta	Dominic (	Empidona	-74.024	Not spec	08:00	Dominic (	Creative C
87	4	no	aldfly	1 (mono)	########	Not specif	13	XC371412	Not specif	Alder Flyc	Not specif	XC371412	[]	yes	Empidona	Elsen's H	41.8605	48000 (Hz	zweeoo	240 m		320000 (b	mp3	Not specified		371412	https://.	United Sta	Matt Wisl	Empidona	-88.176	Not spec	06:33	Matt Wisl	Creative C
88	4	no	aldfly	1 (mono)	########	Not specif	1	XC373885	level	Alder Flyc	01-Mar	XC373885	[]	yes	Empidona	Massach	42.233	44100 (Hz	song	170 m		96689 (bp	mp3	level		373885	https://.	United Sta	Joseph Bo	Empidona	-71.764	0-3(s)	13:00	Joseph Bo	Creative C
89	3.5	yes	aldfly	1 (mono)	########	Not specif	10	XC374636	Not specif	Alder Flyc	Not specif	XC374636	['Trogl	yes	Empidona	St. Franci	39.7791	44100 (Hz	male, son	1000 m	Rare this	87990 (bp	mp3	Not specif	Hous	374636	https://.	United Sta	Eric Houg	Empidona	-101.81	Not spec	14:30	Eric Houg	Creative C
90	3.5	yes	aldfly	1 (mono)	########	Not specif	33	XC374638	Not specif	Alder Flyc	Not specif	XC374638	['Trogl	yes	Empidona	St. Franci	39.7791	44100 (Hz	call, male	1000 m	Rare this	99709 (bp	mp3	Not specif	Hous	374638	https://.	United Sta	Eric Houg	Empidona	-101.81	Not spec	14:30	Eric Houg	Creative C
91	4.5	no	aldfly	1 (mono)	########	level	20	XC376468	level	Alder Flyc	01-Mar	XC376468	['Zonot	yes	Empidona	Winchell	51.4555	48000 (Hz	male, son	1300 m	Black	192000 (b	mp3	level	Whits	376468	https://.	Canada	Michael F	Empidona	-114.61	0-3(s)	07:30	Michael F	Creative C
92	3.5	no	aldfly	1 (mono)	########	Not specif	48	XC380033	level	Alder Flyc	01-Mar	XC380033	['Melo	no	Empidona	Watts Fla	42.033	44100 (Hz	song	440 m		128000 (b	mp3	decreasin	Swar	380033	https://.	United Sta	Jim Berry	Empidona	-79.422	0-3(s)	09:00	Jim Berry	Creative C
93	4	no	aldfly	1 (mono)	########	both	21	XC381871	level	Alder Flyc	04-Jun	XC381871	[]	no	Empidona	Milbridge	44.5358	48000 (Hz	song	10 m	Natural vi	320000 (b	mp3	increasing		381871	https://.	United Sta	Sue Riffe	Empidona	-67.881	0-3(s)	05:30	Sue Riffe	Creative C
94	4	no	aldfly	2 (stereo)	########	Not specif	45	XC406251	Not specif	Alder Flyc	Not specif	XC406251	[]	yes	Empidona	Potter Ma	61.0776	48000 (Hz	song	0 m		320000 (b	mp3	Not specified		406251	https://.	United Sta	Patrik Ã…	Empidona	-149.83	Not spec	09:33	Patrik Ã…	Creative C
95	4		aldfly	2 (stereo)	########	Not specif	27	XC414551	Not specif	Alder Flyc	Not specif	XC414551	[]		Empidona	Maracay	4.8059	44100 (Hz	call	1100 m	bird-seen	128000 (b	mp3	Not specified		414551	https://.	Colombia	Jaime SuÃ	Empidona	-75.87	Not spec	12:30	Jaime SuÃ	Creative C
96	3.5	no	aldfly	2 (stereo)	########	both	51	XC418144	level	Alder Flyc	01-Mar	XC418144	['Regul	no	Empidona	Fraser-Fo	53.6227	44100 (Hz	song	850 m	Bird callin	128000 (b	mp3	both	Ruby-	418144	https://.	Canada	Jeff Dyck	Empidona	-123	0-3(s)	13:30	Jeff Dyck	Creative C
97	4	no	aldfly	2 (stereo)	########	Not specif	25	XC420402	Not specif	Alder Flyc	Not specif	XC420402	[]	yes	Empidona	Roan Mt,	36.0973	44100 (Hz	song	1900 m	No modifi	128000 (b	mp3	Not specified		420402	https://.	United Sta	Scott Grai	Empidona	-82.137	Not spec	11:00	Scott Grai	Creative C
98	4	yes	aldfly	1 (mono)	########	Not specif	31	XC420909	Not specif	Alder Flyc	Not specif	XC420909	[]	yes	Empidona	Finger La	42.5053	48000 (Hz	song	550 m		192000 (b	mp3	Not specified		420909	https://.	United Sta	GABRIEL L	Empidona	-76.784	Not spec	08:30	GABRIEL L	Creative C
99	4.5	no	aldfly	2 (stereo)	########	Not specif	43	XC479502	Not specif	Alder Flyc	Not specif	XC479502	['Baeol	no	Empidona	Caroline	42.3306	44100 (Hz	song	460 m		192000 (b	mp3	Not specif	Tufte	479502	https://.	United Sta	Meena He	Empidona	-76.335	Not spec	09:30	Meena He	Creative C
100	4.5	no	aldfly	2 (stereo)	########	Not specif	98	XC479504	Not specif	Alder Flyc	Not specif	XC479504	['Baeol	no	Empidona	Caroline	42.3306	44100 (Hz	song	460 m		192000 (b	mp3	Not specif	Tufte	479504	https://.	United Sta	Meena He	Empidona	-76.335	Not spec	10:00	Meena He	Creative C
101	4		aldfly	2 (stereo)	########	both	10	XC78890	level	Alder Flyc	01-Mar	XC78890	[]		Empidona	Murphy-H	44.707	44100 (Hz	song	280 m		196821 (b	mp3	increasing		78890	https://.	United Sta	Jonathon	Empidona	-93.329	0-3(s)	12:20	Jonathon	Creative C
102	4	no	bkcchi	1 (mono)	########	both	18	XC114071	both	Black-cap	04-Jun	XC114071	[]	yes	Poecile at	Environm	40.557	48000 (Hz	call	1500 m	Three indi	48000 (bp	mp3	increasing		114073	https://.	United Sta	Eric DeFoi	Poecile at	-105.02	0-3(s)	11:46	Eric DeFoi	Creative C
103	4.5	no	bkcchi	1 (mono)	########	Not specif	83	XC114086	Not specif	Black-cap	Not specif	XC114086	['Corvu	yes	Poecile at	Powderho	44.942	44100 (Hz	call	260 m	Calls	128000 (b	mp3	Not specif	Amer	114086	https://.	United Sta	Jonathon	Poecile at	-93.259	Not spec	14:26	Jonathon	Creative C
104	4.5	no	bkcchi	2 (stereo)	########	decreasin	33	XC120878	decelerat	Black-cap	Jul-20	XC120878	['Sitta	yes	Poecile at	Stonehan	47.0655	44100 (Hz	call	380 m	One of the	128000 (b	mp3	increasin	Red-b	120878	https://.	Canada	Martin St	Poecile at	-71.415	0-3(s)	08:51	Martin St	Creative C
105	5	no	bkcchi	1 (mono)	########	Not specif	51	XC121068	Not specif	Black-cap	Not specif	XC121068	[]	yes	Poecile at	Minnesot	44.84	44100 (Hz	call	220 m	Two indiv	192000 (b	mp3	Not specified		121068	https://.	United Sta	Jonathon	Poecile at	-93.236	Not spec	09:30	Jonathon	Creative C
106	3.5	no	bkcchi	2 (stereo)	########	Not specif	59	XC123506	Not specif	Black-cap	Not specif	XC123506	['Corvu	yes	Poecile at	St. Mary's	44.9215	48000 (Hz	call	260 m	Soft, high	192000 (b	mp3	Not specif	Amer	123506	https://.	United Sta	Jonathon	Poecile at	-93.259	Not spec	06:40	Jonathon	Creative C
107	5	no	bkcchi	2 (stereo)	########	level	18	XC131952	level	Black-cap	01-Mar	XC131952	[]	yes	Poecile at	RÃ©serv	47.069	44100 (Hz	call	0 m	One indiv	128000 (b	mp3	level		131952	https://.	Canada	Martin St	Poecile at	-70.794	0-3(s)	08:26	Martin St	Creative C
108	4.5	no	bkcchi	1 (mono)	########	Not specif	235	XC132861	Not specif	Black-cap	Not specif	XC132861	['Setop	yes	Poecile at	Lake Bem	47.5392	48000 (Hz	song,call	420 m	Several bi	192000 (b	mp3	Not specif	Myrtl	132861	https://.	United Sta	Jonathon	Poecile at	-94.826	Not spec	14:25	Jonathon	Creative C
109	4	no	bkcchi	1 (mono)	########	Not specif	131	XC134017	Not specif	Black-cap	Not specif	XC134017	[]	yes	Poecile at	Manomet	41.92	44100 (Hz	alarm call	20 m	Vocalizati	128000 (b	mp3	Not specified		134017	https://.	United Sta	Dan Lane	Poecile at	-70.543	Not spec	?		Dan Lane	Creative C
110	3.5	no	bkcchi	2 (stereo)	########	decreasin	25	XC135477	level	Black-cap	01-Mar	XC135477	['Regul	yes	Poecile at	New Four	35.6109	44100 (Hz	song	1500 m		128000 (b	mp3	level	Golde	135477	https://.	United Sta	Mike Nels	Poecile at	-83.425	0-3(s)	09:30	Mike Nels	Creative C
111	4	no	bkcchi	2 (stereo)	########	decreasin	8	XC135478	level	Black-cap	01-Mar	XC135478	[]	yes	Poecile at	New Four	35.6109	44100 (Hz	song	1500 m		128000 (b	mp3	level		135478	https://.	United Sta	Mike Nels	Poecile at	-83.425	0-3(s)	09:30	Mike Nels	Creative C
112	4	no	bkcchi	2 (stereo)	########	both	24	XC139055	both	Black-cap	04-Jun	XC139055	[]	yes	Poecile at	Noland D	35.5727	44100 (Hz	call	1700 m		128000 (b	mp3	level		139055	https://.	United Sta	Mike Nels	Poecile at	-83.48	0-3(s)	07:30	Mike Nels	Creative C
113	3.5	no	bkcchi	1 (mono)	########	Not specif	43	XC139438	Not specif	Black-cap	Not specif	XC139438	['Geoth	yes	Poecile at	Millbury,	42.1752	22050 (Hz	begging c	180 m	Flock of B	64000 (bp	mp3	Not specif	Comn	139438	https://.	United Sta	Kim Allen	Poecile at	-71.785	Not spec	06:45	Kim Allen	Creative C
114	5	no	bkcchi	2 (stereo)	########	both	24	XC139603	level	Black-cap	01-Mar	XC139603	[]	yes	Poecile at	Appalach	35.6151	44100 (Hz	song	1700 m		128000 (b	mp3	level		139603	https://.	United Sta	Mike Nels	Poecile at	-83.415	0-3(s)	08:30	Mike Nels	Creative C

Phase 2: Data Engineering

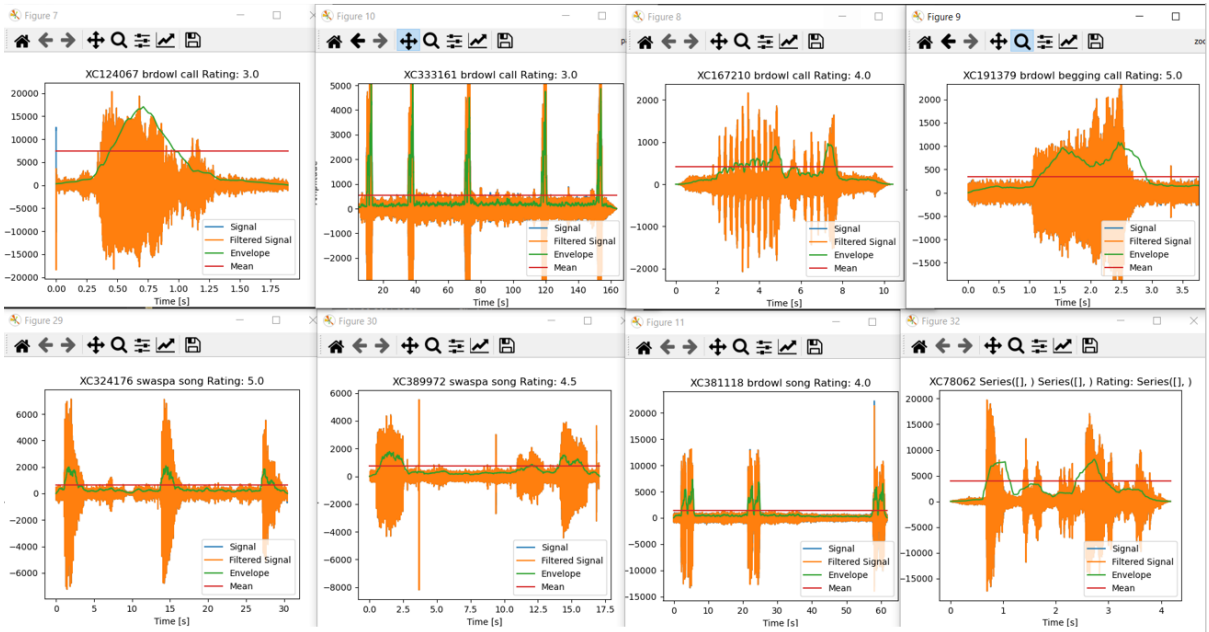
This the most challenging phase.
Upon closer inspection it was observed that all of the data in the open-source datasets come from this one website: https://www.xeno-canto.org/

Data on this website is collected by birders all over the world and hence the quality can vary from sample to sample. There is no standardization of sample rates and sometimes the bird sound is not very clear. Each bird can also make multiple different calls, sounds and alarms, adding to the complexity.
A summary of a subset of the dataset (6 birds) can be found in image 6-birds.png, train-6.xls and dataset-meta.png

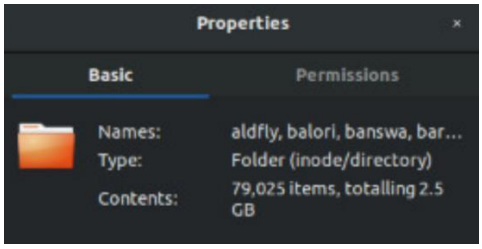


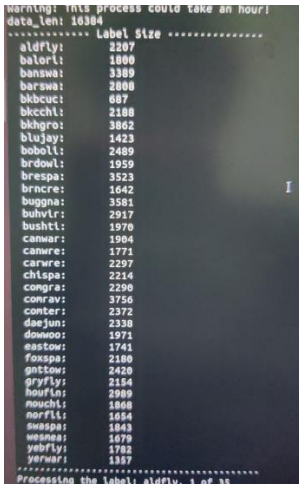


To standardize the dataset and make it similar to the KWS dataset, two steps were taken:
1. Convert mp3 to wav and resample all data to 16kHz (same as KWS dataset, and bird calls are typically between 1-8kHz in frequency so 16k is plenty): Refer code convert.py and convert16k.py
2. Segment the audio into 1 second clips: The recordings in the dataset vary from 1.2 seconds to 900 seconds in length, with anywhere from 1 to 100 relevant bird sounds within a recording. A processing script was written to first filter and then identify and extract useful 1 second long segments. Refer to code bird_slicer.py and image slice_ops.png to get an intuition on how this algorithm works.
Note: The code convert16k.py and bird_slicer.py can be found in folder python-code.zip



The outcome of this process is a dataset of 79025 unique samples (refer data-properties.png, with the following distribution. (refer data-distribution.jpeg). This data has not been manually cleaned after segmentation as the process is too time consuming and requires expert annotation

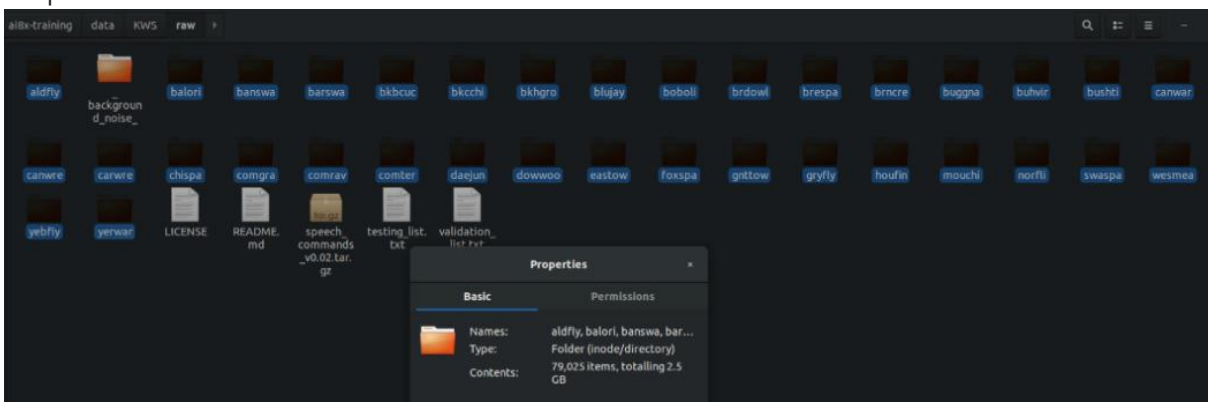




Phase 3: Model training & C-code synthesis using ai8x-training & ai8x-synthesis libraries

Once the data was ready, only a few changes needed to be made to the training scripts to train on the new data, the steps were as follows:

- Update data folder with new data



- Remove data unzip operations from kws20.py (__extract_archive function)

```
234
235     def __extract_archive(self, from_path,  # pylint: disable=no-self-use
236                           to_path=None, remove_finished=False):
237 #         if to_path is None:
238 #             to_path = os.path.dirname(from_path)
239 
240 #         if from_path.endswith('.tar.gz'):
241 #             with tarfile.open(from_path, 'r:gz') as tar:
242 #                 tar.extractall(path=to_path)
243 
244 #         else:
245 #             raise ValueError("Extraction of {} not supported".format(from_path))
246 
247 #         if remove_finished:
248 #             os.remove(from_path)
249         print("Not extracting")
250 
```

- Update class arrays & dictionaries (multiple lines)

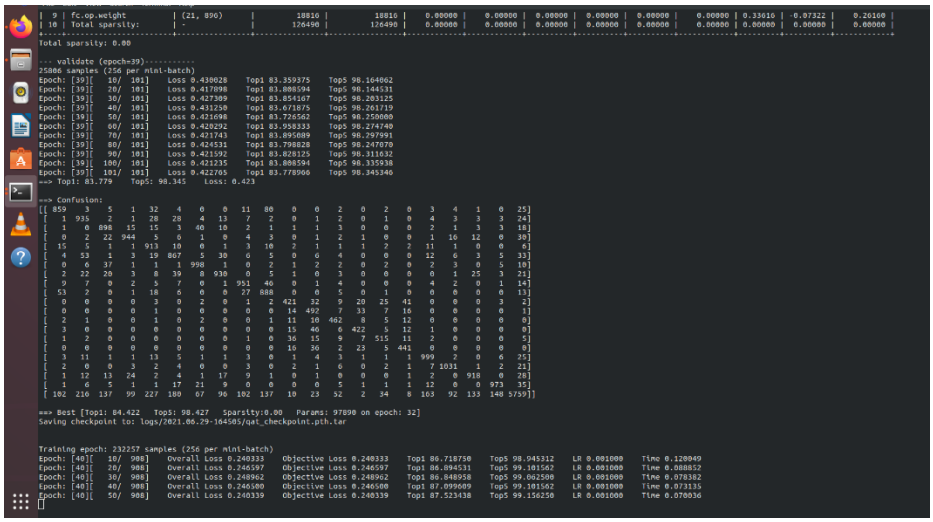
```
70      class_dict = {'aldfly': 0, 'balori': 1, 'banswa': 2, 'barswa': 3, 'bkbcuc': 4, 'bkcchi': 5,
71                    'bkhgro': 6, 'blujay': 7, 'boboli': 8, 'brdowl': 9, 'brespa': 10, 'brncre': 11,
72                    'buggna': 12, 'buhvir': 13, 'bushti': 14, 'canwar': 15, 'canwre': 16, 'carwre': 17,
73                    'chispa': 18, 'comgra': 19, 'comrav': 20, 'comter': 21, 'daejun': 22, 'dowwoo': 23,
74                    'eastow': 24, 'foxspa': 25, 'gnttow': 26, 'gryfly': 27, 'houfin': 28, 'mouchi': 29,
75                    'norfli': 30, 'swaspa': 31, 'wesmea': 32, 'yebfly': 33, 'yerwar': 34}
76 
```

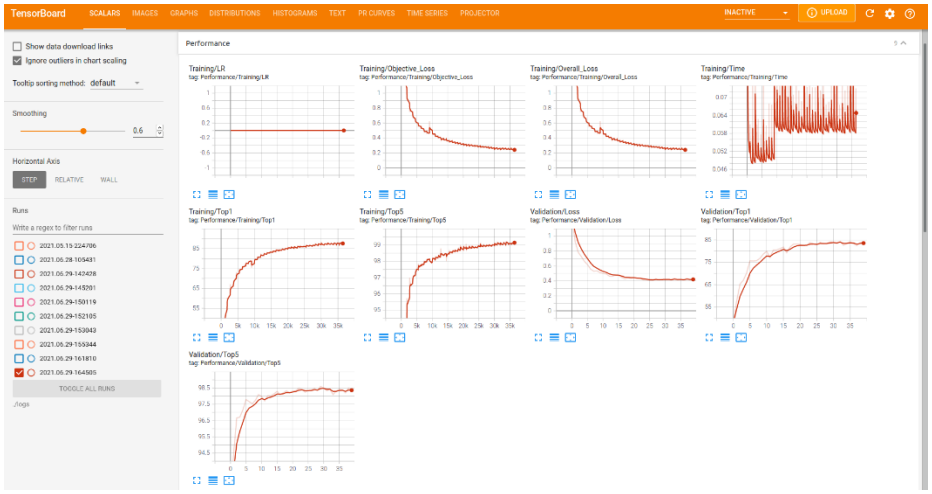
```
599     elif num_classes == 20:
600         classes = ['aldfly', 'banswa', 'barswa', 'bkcchi', 'bkhgro', 'boboli', 'brdowl', 'brespa', 'buggna', 'buhvir',
601          'bushti', 'chispa', 'comgra', 'comrav', 'comter', 'daejun', 'gnttow', 'gryfly', 'houfin', 'swaspa']
602     elif num_classes == 35:
```

```
652     {
653         'name': 'KWS_20',  # 20 keywords
654         'input': (128, 128, 1),
655         'output': (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20),
656         'weight': (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0.14),
657         'loader': KWS_20_get_datasets,
658     },
```

Run the script train.sh but reduce the number of epochs to 50.
The network: ai85kws20net that is found in ai85net-kws20.py was used.

We first tried to train a mix of birds and words to see how the network performs on the hybrid dataset.
Screenshots of the training process can be found in:
- Words&Birds-ep10-confusion
- Words&Birds-ep40-confusion
- Words&Birds-ep40-tensorboard
In the confusion matrix you can clearly identify the cluster of birds as they perform different from the words.

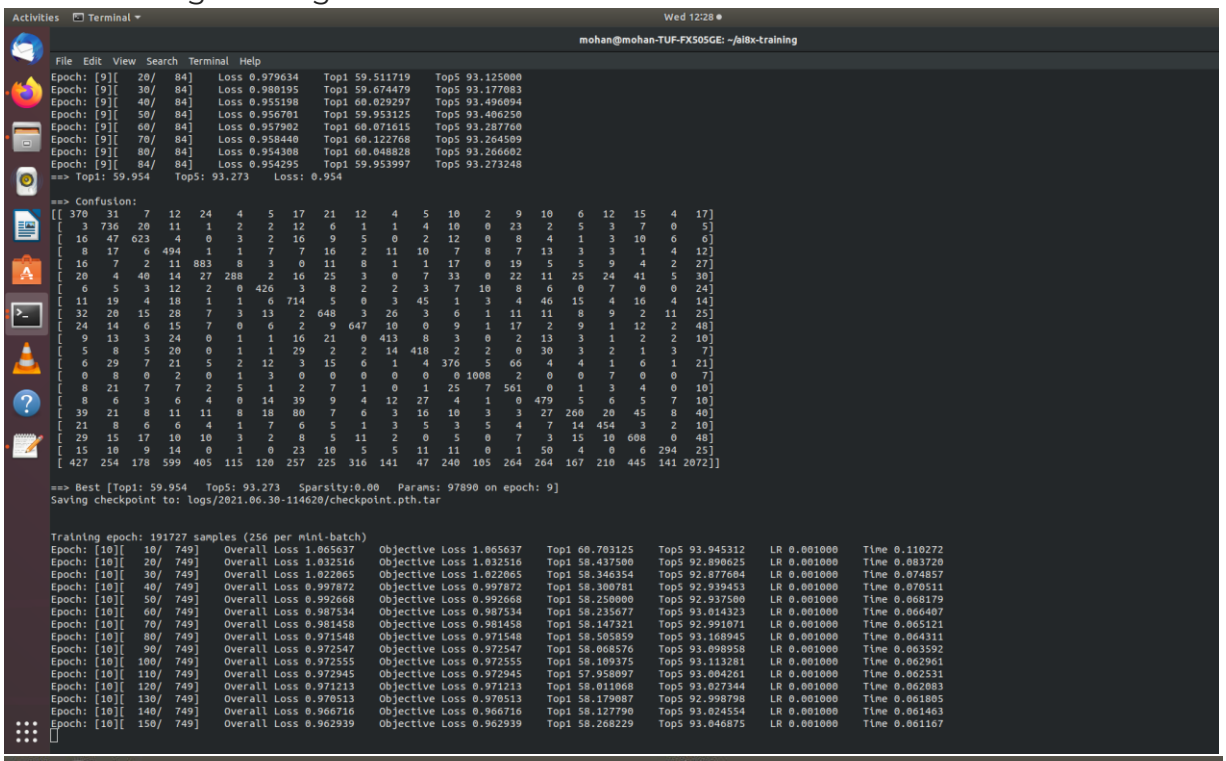


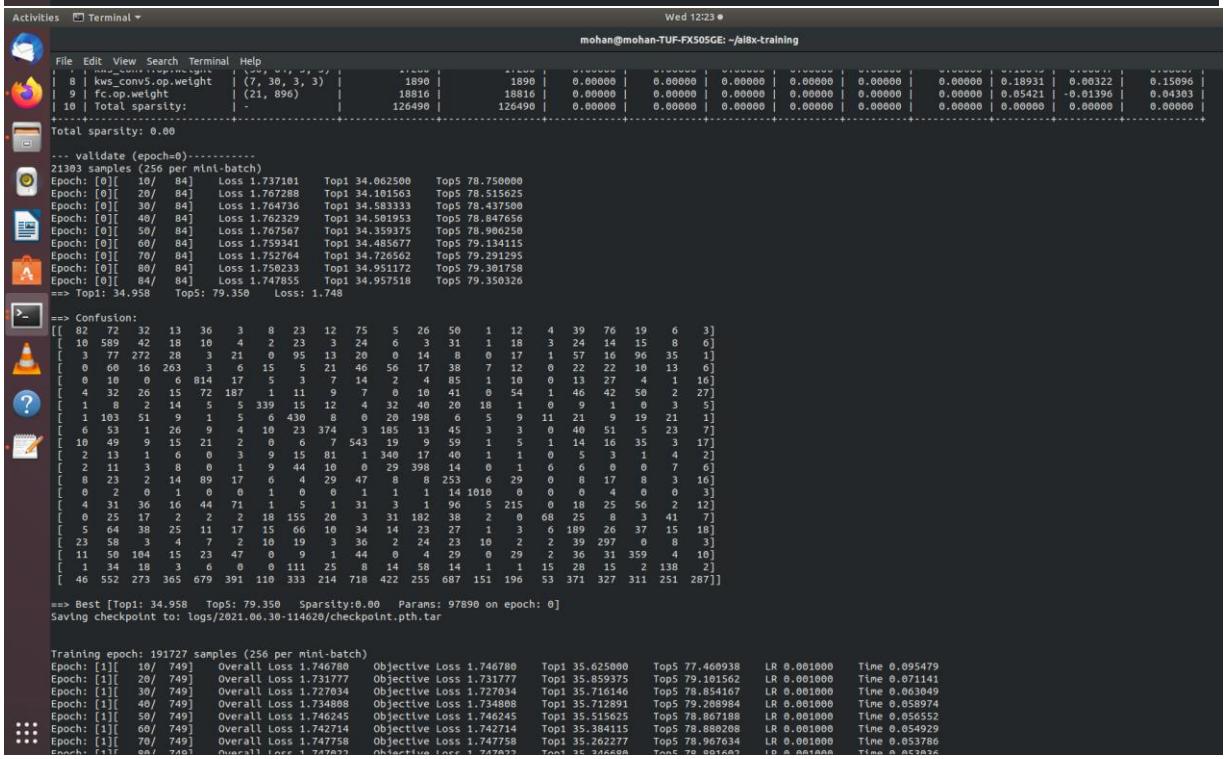


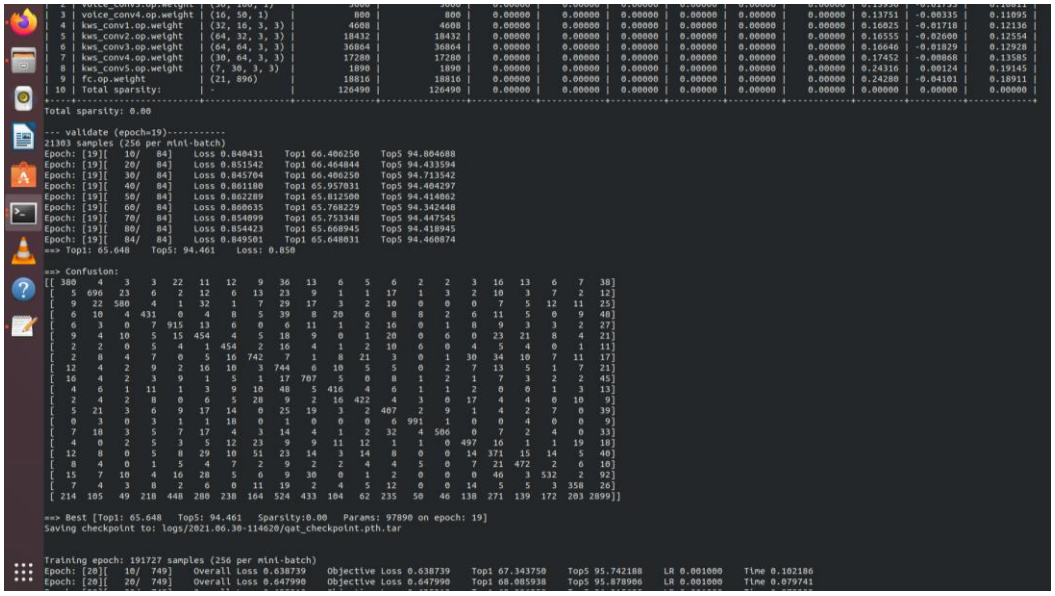
Satisfied with this PoC we trained the network on only birds.
We found that the network starts to overtrain after 50 epochs, so we stopped the training process then.

The following images document the training output:

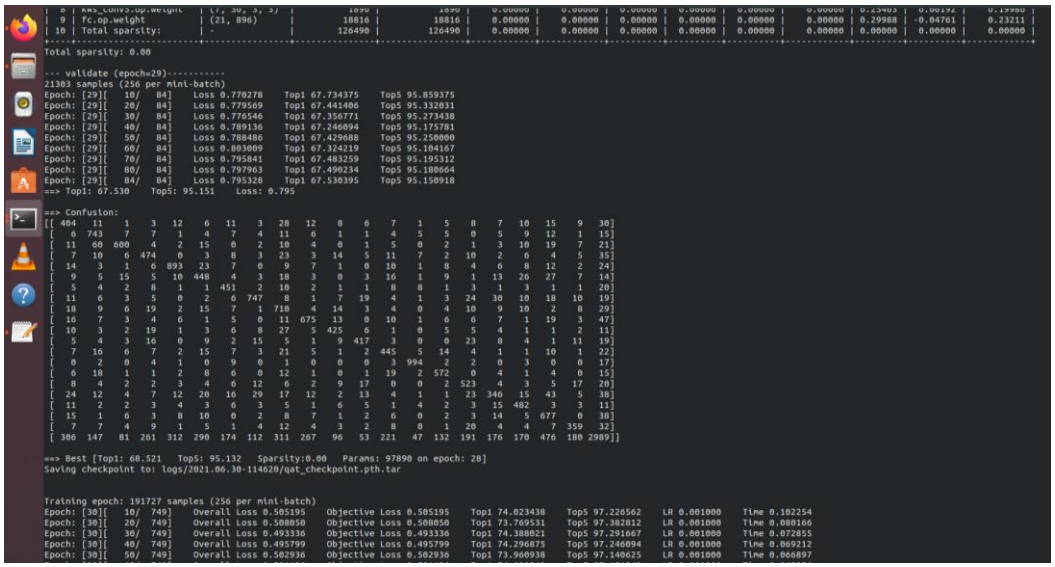
- PNG files e1, e10, e20, e30, e40 and e50 show the progression of the confusion matrix during training



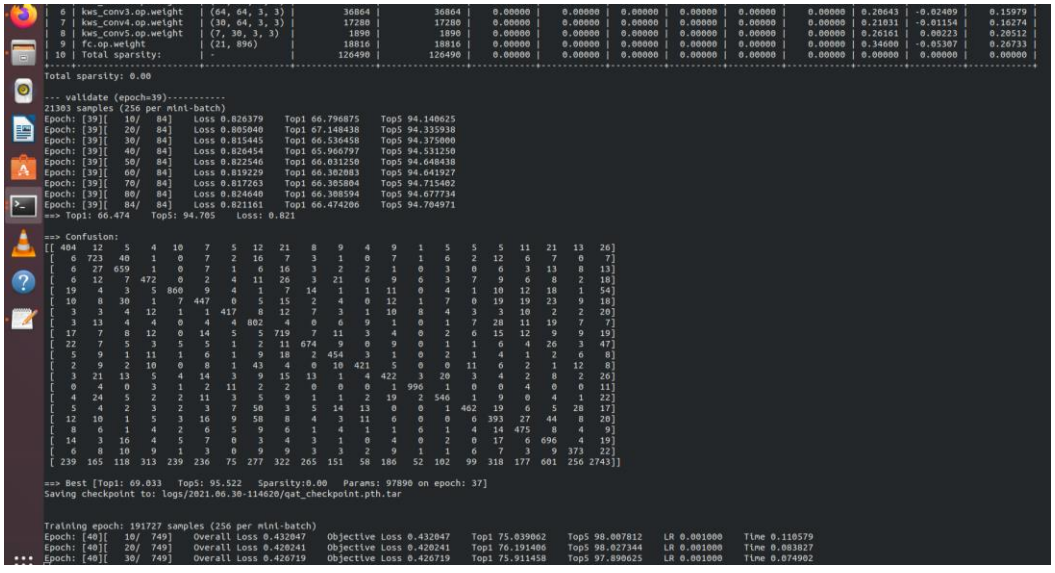




```
| 3 | voice_conv4.op.weight | (16, 50, 1)       |          800 |          800 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.13751 | -0.00335 |  0.11095 |
| 4 | kws_conv1.op.weight   | (32, 16, 3, 3)    |         4608 |         4608 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.16825 | -0.01718 |  0.12136 |
| 5 | kws_conv2.op.weight   | (64, 32, 3, 3)    |        18432 |        18432 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.16555 | -0.02600 |  0.12554 |
| 6 | kws_conv3.op.weight   | (64, 64, 3, 3)    |        36864 |        36864 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.16646 | -0.01829 |  0.12928 |
| 7 | kws_conv4.op.weight   | (30, 64, 3, 3)    |        17280 |        17280 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.17452 | -0.00868 |  0.13585 |
| 8 | kws_conv5.op.weight   | (7, 30, 3, 3)     |         1890 |         1890 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.24316 |  0.00124 |  0.19145 |
| 9 | fc.op.weight          | (21, 896)         |        18816 |        18816 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.24280 | -0.04101 |  0.18911 |
| 10 | Total sparsity:      | -                 |       126490 |       126490 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |
+----+----------------------+-------------------+--------------+--------------+----------+----------+----------+----------+----------+----------+----------+----------+
Total sparsity: 0.00

--- validate (epoch=19)----------
21303 samples (256 per mini-batch)
Epoch: [19][  10/  84]    Loss 0.840431    Top1 66.406250    Top5 94.804688
Epoch: [19][  20/  84]    Loss 0.851542    Top1 66.464844    Top5 94.433594
Epoch: [19][  30/  84]    Loss 0.845704    Top1 66.406250    Top5 94.713542
Epoch: [19][  40/  84]    Loss 0.861180    Top1 65.957031    Top5 94.404297
Epoch: [19][  50/  84]    Loss 0.862289    Top1 65.812500    Top5 94.414062
Epoch: [19][  60/  84]    Loss 0.860635    Top1 65.768229    Top5 94.342448
Epoch: [19][  70/  84]    Loss 0.854099    Top1 65.753348    Top5 94.447545
Epoch: [19][  80/  84]    Loss 0.854423    Top1 65.668945    Top5 94.418945
Epoch: [19][  84/  84]    Loss 0.849501    Top1 65.648031    Top5 94.460874
==> Top1: 65.648    Top5: 94.461    Loss: 0.850

==> Confusion:
[[ 380    4    3    3   22   11   12    9   36   13    6    5    6    2    2    3   16   13    6    7   38]
 [   5  696   23    6    2   12    6   13   23    9    1    1   17    1    3    2   10    3    7    2   12]
 [   9   22  580    4    1   32    1    7   29   17    3    2   10    0    0    0    7    5   12   11   25]
 [   6   10    4  431    0    4    8    5   39    8   20    6    8    8    2    6   11    5    0    9   40]
 [   6    3    0    7  915   13    6    0    6   11    1    2   16    0    1    8    9    3    3    2   27]
 [   9    4   10    5   15  454    4    5   18    9    0    1   20    0    6    0   23   21    8    4   21]
 [   2    2    0    5    4    1  454    2   16    4    1    2   10    6    0    4    5    4    0    1   11]
 [   2    8    4    7    0    5   16  742    7    1    8   21    3    0    1   30   34   10    7   11   17]
 [  12    4    2    9    2   16   10    3  744    6   10    5    5    0    2    7   13    5    1    7   21]
 [  16    4    2    3    9    1    5    1   17  707    5    0    8    1    2    1    7    3    2    2   45]
 [   4    6    1   11    1    3    9   10   48    5  416    4    6    1    1    2    0    0    1    3   13]
 [   2    4    2    8    0    6    5   20    9    2   16  422    4    4    3    0   17    4    4    0   10    9]
 [   5   21    3    6    9   17   14    0   25   19    3    2  407    2    9    1    4    2    7    0   39]
 [   0    3    0    3    1    1   18    0    1    0    0    0    6  991    1    0    0    4    0    0    9]
 [   7   18    3    5    7   17    4    3   14    4    1    2   32    4  506    0    7    2    4    0   33]
 [   4    0    2    5    3    5   12   23    9    9   11   12    1    1    0  497   16    1    1   19   18]
 [  12    8    0    5    8   29   10   51   23   14    3   14    8    0    0   14  371   15   14    5   40]
 [   8    4    0    1    5    4    7    2    9    2    2    4    4    5    0    7   21  472    2    6   10]
 [  15    7   10    4   16   28    5    6    9   30    0    1    2    0    0    0   46    3  532    2   92]
 [   7    4    3    8    2    6    0   11   19    2    4    5   12    0    0   14    5    5    3  358   26]
 [ 214  105   49  218  448  280  238  164  524  433  104   62  235   50   46  138  271  139  172  203 2899]]

==> Best [Top1: 65.648   Top5: 94.461   Sparsity:0.00   Params: 97890 on epoch: 19]
Saving checkpoint to: logs/2021.06.30-114620/qat_checkpoint.pth.tar


Training epoch: 191727 samples (256 per mini-batch)
Epoch: [20][  10/  749]    Overall Loss 0.638739    Objective Loss 0.638739    Top1 67.343750    Top5 95.742188    LR 0.001000    Time 0.102186
Epoch: [20][  20/  749]    Overall Loss 0.647990    Objective Loss 0.647990    Top1 68.085938    Top5 95.878906    LR 0.001000    Time 0.079741
```

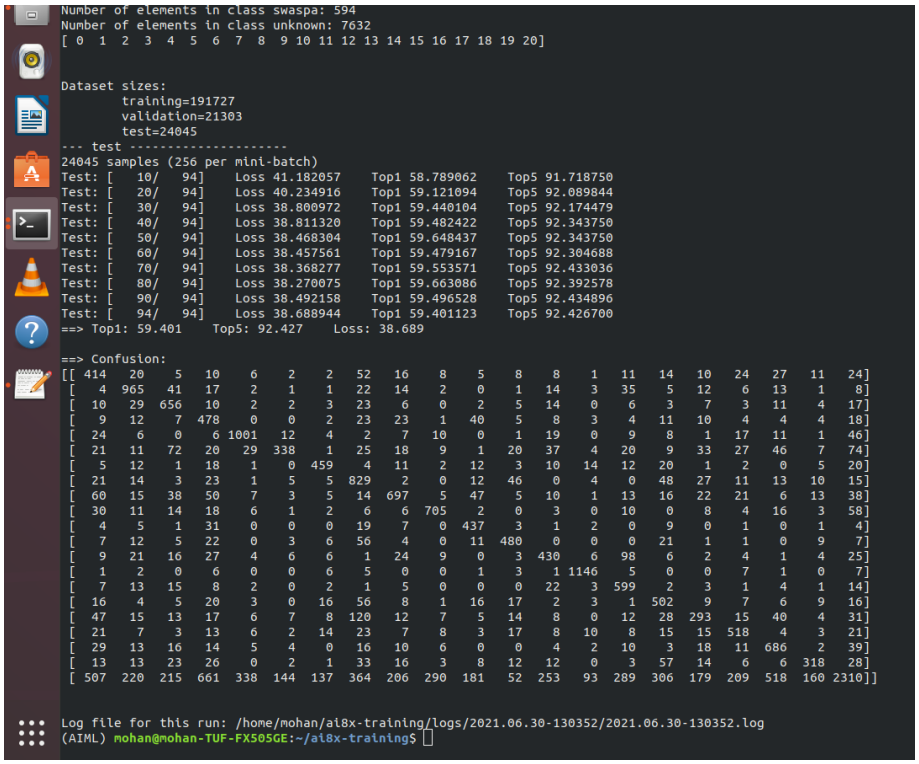


```
| 8 | kws_conv5.op.weight   | (7, 30, 3, 3)     |         1890 |         1890 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.23403 |  0.00192 |  0.19980 |
| 9 | fc.op.weight          | (21, 896)         |        18816 |        18816 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.29988 | -0.04761 |  0.23211 |
| 10 | Total sparsity:      | -                 |       126490 |       126490 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |
+----+----------------------+-------------------+--------------+--------------+----------+----------+----------+----------+----------+----------+----------+----------+
Total sparsity: 0.00

--- validate (epoch=29)----------
21303 samples (256 per mini-batch)
Epoch: [29][  10/  84]    Loss 0.770278    Top1 67.734375    Top5 95.859375
Epoch: [29][  20/  84]    Loss 0.779569    Top1 67.441406    Top5 95.332031
Epoch: [29][  30/  84]    Loss 0.776546    Top1 67.356771    Top5 95.273438
Epoch: [29][  40/  84]    Loss 0.789136    Top1 67.246094    Top5 95.175781
Epoch: [29][  50/  84]    Loss 0.788486    Top1 67.429688    Top5 95.250000
Epoch: [29][  60/  84]    Loss 0.803009    Top1 67.324219    Top5 95.104167
Epoch: [29][  70/  84]    Loss 0.795841    Top1 67.483259    Top5 95.195312
Epoch: [29][  80/  84]    Loss 0.797963    Top1 67.490234    Top5 95.180664
Epoch: [29][  84/  84]    Loss 0.795328    Top1 67.530395    Top5 95.150918
==> Top1: 67.530    Top5: 95.151    Loss: 0.795

==> Confusion:
[[ 404   11    1    3   12    6   11    3   28   12    8    6    7    1    5    8    7   10   15    9   30]
 [   6  743    7    7    1    4    7    4   11    6    1    1    1    4    5    5    0    5    9   12    1   15]
 [  11   60  600    4    2   15    0    2   10    4    0    1    5    0    2    1    3   10   19    7   21]
 [   7   10    6  474    0    3    8    3   23    3   14    5   11    7    2   10    2    6    4    5   35]
 [  14    3    1    6  893   23    7    0    9    7    1    0   10    1    8    4    6    8   12    2   24]
 [   9    5   15    5   10  448    4    3   18    3    0    3   16    1    9    1   13   26   27    7   14]
 [   5    4    2    8    1    1  451    2   10    2    1    1    8    8    1    3    1    3    1    1   20]
 [  11    6    3    5    0    2    6  747    8    1    7   19    4    1    3   24   30   10   18   10   19]
 [  18    9    6   19    2   15    7    1  710    4   14    3    4    0    4   10    9   10    2    8   29]
 [  16    7    3    4    6    1    5    0   11  675   13    0   10    1    6    6    7    1   19    3   47]
 [  10    3    2   19    1    3    6    8   27    5  425    6    1    0    5    5    4    1    1    2   11]
 [   5    4    3   16    0    9    2   15    5    1    9  417    3    0    0   23    8    4    1   11   19]
 [   7   16    6    7    2   15    7    3   21    5    1    2  445    5   14    4    1    1   10    1   22]
 [   0    2    0    4    1    0    9    0    1    0    0    0    3  994    2    2    0    3    0    0   17]
 [   6   18    1    1    2    8    6    0   12    1    0    1   19    2  572    0    4    1    4    0   15]
 [   8    4    2    2    3    4    6   12    6    2    9   17    0    0    2  523    4    3    5   17   20]
 [  24   12    4    7   12   20   16   29   17   12    2   13    4    1    1   23  346   15   43    5   38]
 [  11    2    2    3    4    3    6    3    5    1    6    5    1    4    2    3   15  482    3    3   11]
 [  15    1    6    3    8   10    0    2    8    7    1    2    6    0    2    3   14    5  677    0   38]
 [   7    7    4    9    1    5    1    4   12    4    3    2    8    0    1   20    4    4    7  359   32]
 [ 306  147   81  261  312  290  174  112  311  267   96   53  221   47  132  191  176  170  476  180 2989]]

==> Best [Top1: 68.521   Top5: 95.132   Sparsity:0.00   Params: 97890 on epoch: 28]
Saving checkpoint to: logs/2021.06.30-114620/qat_checkpoint.pth.tar


Training epoch: 191727 samples (256 per mini-batch)
Epoch: [30][  10/  749]    Overall Loss 0.505195    Objective Loss 0.505195    Top1 74.023438    Top5 97.226562    LR 0.001000    Time 0.102254
Epoch: [30][  20/  749]    Overall Loss 0.508050    Objective Loss 0.508050    Top1 73.769531    Top5 97.382812    LR 0.001000    Time 0.080166
Epoch: [30][  30/  749]    Overall Loss 0.493336    Objective Loss 0.493336    Top1 74.388021    Top5 97.291667    LR 0.001000    Time 0.072855
Epoch: [30][  40/  749]    Overall Loss 0.495799    Objective Loss 0.495799    Top1 74.296875    Top5 97.246094    LR 0.001000    Time 0.069212
Epoch: [30][  50/  749]    Overall Loss 0.502936    Objective Loss 0.502936    Top1 73.960938    Top5 97.140625    LR 0.001000    Time 0.066897
```

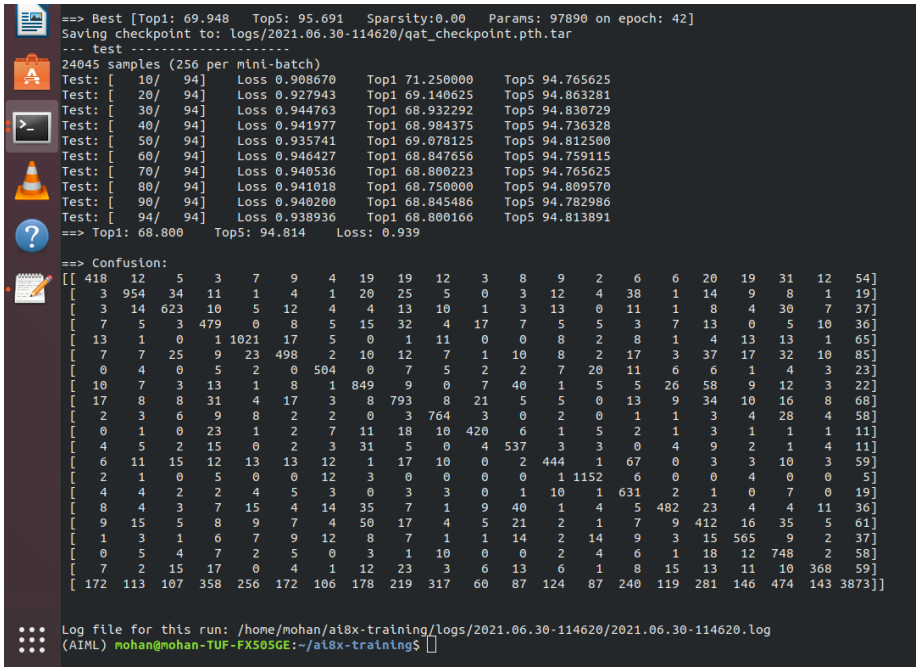


```
| 6 | kws_conv3.op.weight   | (64, 64, 3, 3)    |        36864 |        36864 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.20643 | -0.02409 |  0.15979 |
| 7 | kws_conv4.op.weight   | (30, 64, 3, 3)    |        17280 |        17280 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.21031 | -0.01154 |  0.16274 |
| 8 | kws_conv5.op.weight   | (7, 30, 3, 3)     |         1890 |         1890 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.26161 |  0.00223 |  0.20512 |
| 9 | fc.op.weight          | (21, 896)         |        18816 |        18816 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.34600 | -0.05307 |  0.26733 |
| 10 | Total sparsity:      | -                 |       126490 |       126490 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |  0.00000 |
+----+----------------------+-------------------+--------------+--------------+----------+----------+----------+----------+----------+----------+----------+----------+
Total sparsity: 0.00

--- validate (epoch=39)----------
21303 samples (256 per mini-batch)
Epoch: [39][  10/  84]    Loss 0.826379    Top1 66.796875    Top5 94.140625
Epoch: [39][  20/  84]    Loss 0.805040    Top1 67.148438    Top5 94.335938
Epoch: [39][  30/  84]    Loss 0.815445    Top1 66.536458    Top5 94.375000
Epoch: [39][  40/  84]    Loss 0.826454    Top1 65.966797    Top5 94.531250
Epoch: [39][  50/  84]    Loss 0.822546    Top1 66.031250    Top5 94.648438
Epoch: [39][  60/  84]    Loss 0.819239    Top1 66.302083    Top5 94.641927
Epoch: [39][  70/  84]    Loss 0.817263    Top1 66.305804    Top5 94.715402
Epoch: [39][  80/  84]    Loss 0.824640    Top1 66.308594    Top5 94.677734
Epoch: [39][  84/  84]    Loss 0.821161    Top1 66.474206    Top5 94.704971
==> Top1: 66.474    Top5: 94.705    Loss: 0.821

==> Confusion:
[[ 404   12    5    4   10    7    5   12   21    8    9    4    9    1    5    5    5   11   21   13   26]
 [   6  723   40    1    0    7    2   16    7    3    1    8    7    1    6    2   12    6    7    0    7]
 [   6   27  659    1    0    7    1    6   16    3    2    2    1    0    3    0    6    3   13    8   13]
 [   6   12    7  472    0    2    4   11   26    3   21    6    9    6    3    7    9    6    8    2   18]
 [  19    4    3    5  860    9    4    1    7   14    1    1   11    0    4    1   10   12   18    1   54]
 [  10    8   30    1    7  447    0    5   15    2    4    0   12    1    7    0   19   19   23    9   18]
 [   3    3    4   12    1    1  417    8   12    7    3    1   10    8    4    3    3   10    2    2   28]
 [   3   13    4    4    0    4    4  802    4    0    6    9    1    0    1    7   28   11   19    7    7]
 [  17    7    8   12    0   14    5    5  719    7   11    3    4    0    2    6   15   12    9    9   19]
 [  22    7    5    3    5    5    1    2   11  674    9    0    9    0    1    1    6    4   26    3   47]
 [   5    9    1   11    1    6    1    9   18    2  454    3    1    0    2    1    4    1    2    6    8]
 [   2    9    2   10    0    8    1   43    4    0   10  421    5    0    0   11    6    2    1   12    8]
 [   3   21   13    5    4   14    3    9   15   13    1    4  422    3   20    3    4    2    8    2   26]
 [   0    4    0    3    1    2   11    2    2    0    0    1  996    1    0    0    4    0    0   11]
 [   4   24    5    2    2   11    3    5    9    1    1    2   19    2  546    1    9    0    4    1   22]
 [   5    4    2    3    2    3    7   50    3    5   14   13    0    0    1  462   19    6    5   28   17]
 [  12   10    1    5    3   16    9   58    8    4    3   11    6    0    0    6  393   27   44    8   20]
 [   8    6    1    4    2    6    5    9    6    1    4    1    1    6    1    4   14  475    8    4    9]
 [  14    3   16    4    5    7    0    3    4    3    1    0    4    0    2    0   17    6  696    4   19]
 [   6    8   10    9    1    3    0    9    9    3    3    2    9    1    1    6    7    3    9  373   22]
 [ 239  165  118  313  239  236   75  277  322  265  151   58  186   52  102   99  318  177  601  256 2743]]

==> Best [Top1: 69.033   Top5: 95.522   Sparsity:0.00   Params: 97890 on epoch: 37]
Saving checkpoint to: logs/2021.06.30-114620/qat_checkpoint.pth.tar


Training epoch: 191727 samples (256 per mini-batch)
Epoch: [40][  10/  749]    Overall Loss 0.432047    Objective Loss 0.432047    Top1 75.039062    Top5 98.007812    LR 0.001000    Time 0.110579
Epoch: [40][  20/  749]    Overall Loss 0.428241    Objective Loss 0.428241    Top1 76.191406    Top5 98.027344    LR 0.001000    Time 0.083827
Epoch: [40][  30/  749]    Overall Loss 0.426719    Objective Loss 0.426719    Top1 75.911458    Top5 97.890625    LR 0.001000    Time 0.074902
```

- PNG files eval and test show the results of the evaluation script and test run after training

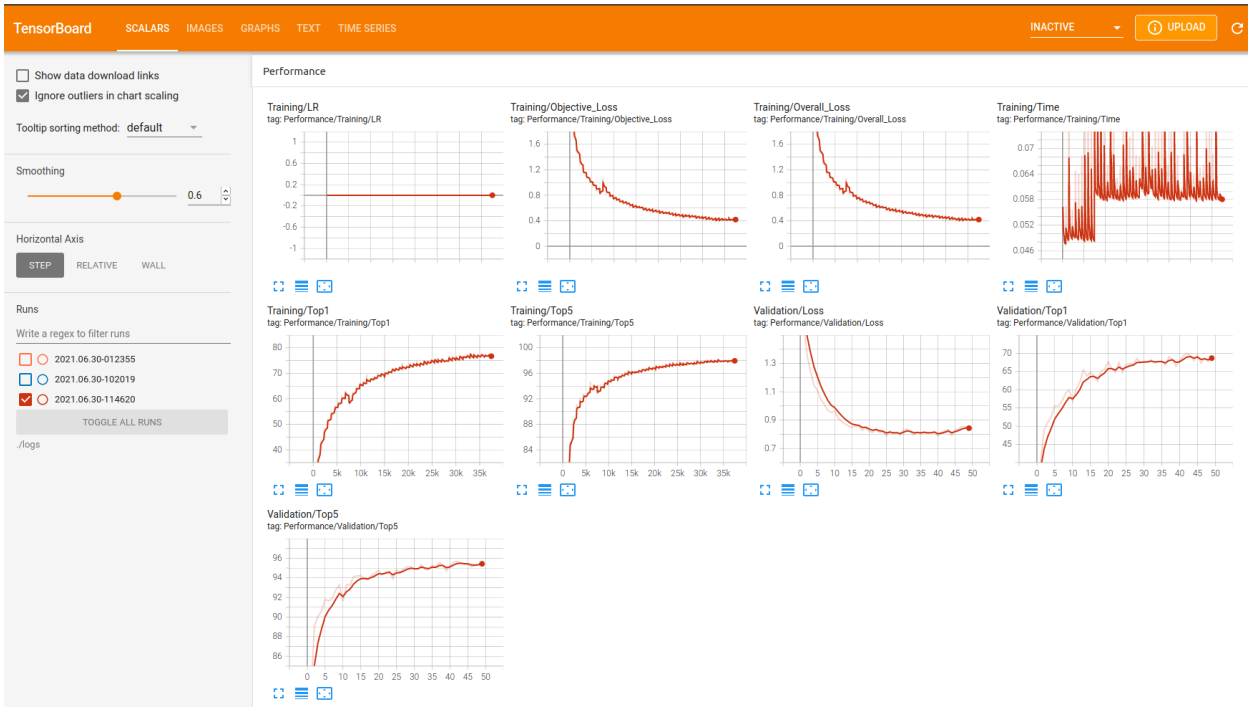


```
Number of elements in class swaspa: 594
Number of elements in class unknown: 7632
[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20]


Dataset sizes:
       training=191727
       validation=21303
       test=24045
--- test --------------------
24045 samples (256 per mini-batch)
Test: [   10/   94]    Loss 41.182057    Top1 58.789062    Top5 91.718750
Test: [   20/   94]    Loss 40.234916    Top1 59.121094    Top5 92.089844
Test: [   30/   94]    Loss 38.800972    Top1 59.440104    Top5 92.174479
Test: [   40/   94]    Loss 38.811320    Top1 59.482422    Top5 92.343750
Test: [   50/   94]    Loss 38.468304    Top1 59.648437    Top5 92.343750
Test: [   60/   94]    Loss 38.457561    Top1 59.479167    Top5 92.304688
Test: [   70/   94]    Loss 38.368277    Top1 59.553571    Top5 92.433036
Test: [   80/   94]    Loss 38.270075    Top1 59.663086    Top5 92.392578
Test: [   90/   94]    Loss 38.492158    Top1 59.496528    Top5 92.434896
Test: [   94/   94]    Loss 38.688944    Top1 59.401123    Top5 92.426700
==> Top1: 59.401    Top5: 92.427    Loss: 38.689

==> Confusion:
[[ 414   20    5   10    6    2    2   52   16    8    5    8    8    1   11   14   10   24   27   11   24]
 [   4  965   41   17    2    1    1   22   14    2    0    1   14    3   35    5   12    6   13    1    8]
 [  10   29  656   10    2    2    3   23    6    0    2    5   14    0    6    3    7    3   11    4   17]
 [   9   12    7  478    0    0    2   23   23    1   40    5    8    3    4   11   10    4    4    4   18]
 [  24    6    0    6 1001   12    4    2    7   10    0    1   19    0    9    8    1   17   11    1   46]
 [  21   11   72   20   29  338    1   25   18    9    1   20   37    4   20    9   33   27   46    7   74]
 [   5   12    1   18    1    0  459    4   11    2   12    3   10   14   12   20    1    2    0    5   20]
 [  21   14    3   23    1    5    5  829    2    0   12   46    0    4    0   48   27   11   13   10   15]
 [  60   15   38   50    7    3    5   14  697    5   47    5   10    1   13   16   22   21    6   13   38]
 [  30   11   14   18    6    1    2    6    6  705    2    0    3    0   10    0    8    4   16    3   58]
 [   4    5    1   31    0    0    0   19    7    0  437    3    1    2    0    9    0    1    0    1    4]
 [   7   12    5   22    0    3    6   56    4    0   11  480    0    0    0   21    1    1    0    9    7]
 [   9   21   16   27    4    6    6    1   24    9    0    3  430    6   98    6    2    4    1    4   25]
 [   1    2    0    6    0    0    6    5    0    0    1    3    1 1146    5    0    0    7    1    0    7]
 [   7   13   15    8    2    0    2    1    5    0    0    0   22    3  599    2    3    1    4    1   14]
 [  16    4    5   20    3    0   16   56    8    1   16   17    2    3    1  502    9    7    6    9   16]
 [  47   15   13   17    6    7    8  120   12    7    5   14    8    0   12   28  293   15   40    4   31]
 [  21    7    3   13    6    2   14   23    7    8    3   17    8   10    8   15   15  518    4    3   21]
 [  29   13   16   14    5    4    0   16   10    6    0    0    4    2   10    3   18   11  686    2   39]
 [  13   13   23   26    0    2    1   33   16    3    8   12   12    0    3   57   14    6    6  318   28]
 [ 507  220  215  661  338  144  137  364  206  290  181   52  253   93  289  306  179  209  518  160 2310]]


Log file for this run: /home/mohan/ai8x-training/logs/2021.06.30-130352/2021.06.30-130352.log
(AIML) mohan@mohan-TUF-FX505GE:~/ai8x-training$
```



```
==> Best [Top1: 69.948   Top5: 95.691   Sparsity:0.00   Params: 97890 on epoch: 42]
Saving checkpoint to: logs/2021.06.30-114620/qat_checkpoint.pth.tar
--- test --------------------
24045 samples (256 per mini-batch)
Test: [   10/   94]    Loss 0.908670    Top1 71.250000    Top5 94.765625
Test: [   20/   94]    Loss 0.927943    Top1 69.140625    Top5 94.863281
Test: [   30/   94]    Loss 0.944763    Top1 68.932292    Top5 94.830729
Test: [   40/   94]    Loss 0.941977    Top1 68.984375    Top5 94.736328
Test: [   50/   94]    Loss 0.935741    Top1 69.078125    Top5 94.812500
Test: [   60/   94]    Loss 0.946427    Top1 68.847656    Top5 94.759115
Test: [   70/   94]    Loss 0.940536    Top1 68.800223    Top5 94.765625
Test: [   80/   94]    Loss 0.941018    Top1 68.750000    Top5 94.809570
Test: [   90/   94]    Loss 0.940200    Top1 68.845486    Top5 94.782986
Test: [   94/   94]    Loss 0.938936    Top1 68.800166    Top5 94.813891
==> Top1: 68.800    Top5: 94.814    Loss: 0.939

==> Confusion:
[[ 418   12    5    3    7    9    4   19   19   12    3    8    9    2    6    6   20   19   31   12   54]
 [   3  954   34   11    1    4    1   20   25    5    0    3   12    4   38    1   14    9    8    1   19]
 [   3   14  623   10    5   12    4    4   13   10    1    3   13    0   11    1    8    4   30    7   37]
 [   7    5    3  479    0    8    5   15   32    4   17    7    5    5    3    7   13    0    5   10   36]
 [  13    1    0    1 1021   17    5    0    1   11    0    0    8    2    8    1    4   13   13    1   65]
 [   7    7   25    9   23  498    2   10   12    7    1   10    8    2   17    3   37   17   32   10   85]
 [   0    4    0    5    2    0  504    0    7    5    2    2    7   20   11    6    6    1    4    3   23]
 [  10    7    3   13    1    8    1  849    9    0    7   40    1    5    5   26   58    9   12    3   22]
 [  17    8    8   31    4   17    3    8  793    8   21    5    5    0   13    9   34   10   16    8   68]
 [   2    3    6    9    8    2    2    0    3  764    3    0    2    0    1    1    3    4   28    4   58]
 [   0    1    0   23    1    2    7   11   18   10  420    6    1    5    2    1    3    1    1    1   11]
 [   4    5    2   15    0    2    3   31    5    0    4  537    3    3    0    4    9    2    1    4   11]
 [   6   11   15   12   13   13   12    1   17   10    0    2  444    1   67    0    3    3   10    3   59]
 [   2    1    0    5    0    0   12    3    0    0    0    0    1 1152    6    0    0    4    0    0    5]
 [   4    4    2    2    4    5    3    0    3    3    0    1   10    1  631    2    1    0    7    0   19]
 [   8    4    3    7   15    4   14   35    7    1    9   40    1    4    5  482   23    4    4   11   36]
 [   9   15    5    8    9    7    4   50   17    4    5   21    2    1    7    9  412   16   35    5   61]
 [   1    3    1    6    7    9   12    8    7    1    1   14    2   14    9    3   15  565    9    2   37]
 [   0    5    4    7    2    5    0    3    1   10    0    0    2    4    6    1   18   12  748    2   58]
 [   7    2   15   17    0    4    1   12   23    3    6   13    6    1    8   15   13   11   10  368   59]
 [ 172  113  107  358  256  172  106  178  219  317   60   87  124   87  240  119  281  146  474  143 3873]]


Log file for this run: /home/mohan/ai8x-training/logs/2021.06.30-114620/2021.06.30-114620.log
(AIML) mohan@mohan-TUF-FX505GE:~/ai8x-training$
```

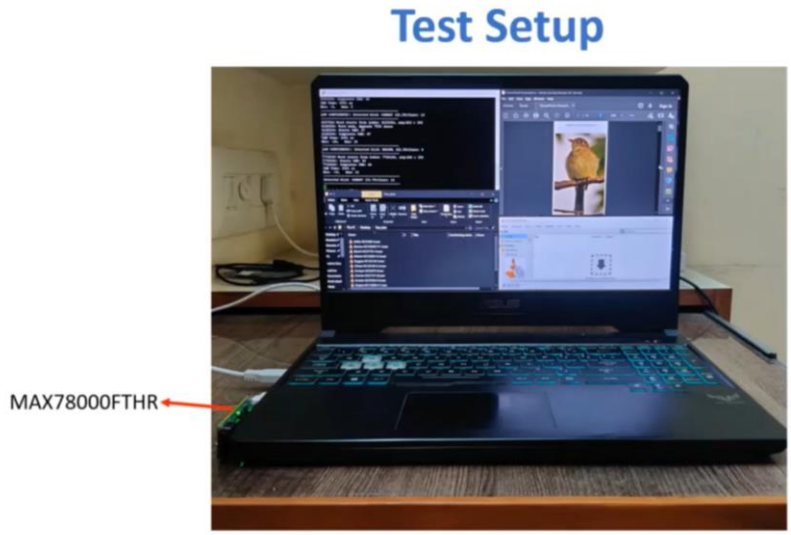
- tensorboard-plots.png shows the output of the tensorboard console



The model was then synthesized using the ai8x-synthesis library, following standard synthesis procedure. Synthesized code can be found in Synthesized-C-Code.rar

Phase 4: Testing on Hardware using audio samples from the test set.

The code was then flashed onto the MAX78000 feather board using the KWS20 codebase as a reference. A demo of the same can be found in the video.



Please download the zip file brd20_demo.zip (based on kws_20) to be able to run the project on your own board. We have also included the test samples that were used in the video demo. No additional hardware is required to test this code as we have used the on board microphone.

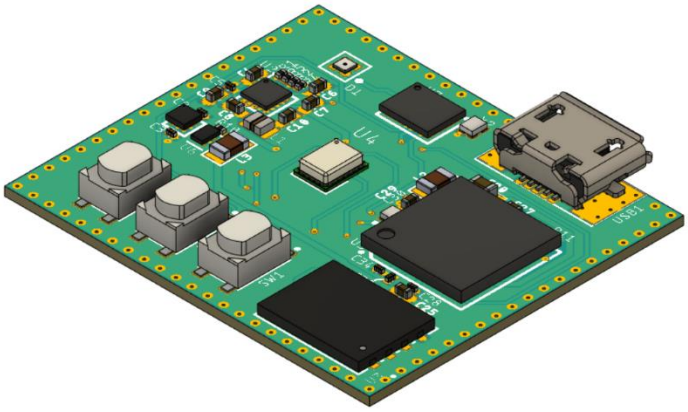
Hardware Design

We have built a custom hardware based off of the MAX78000FTHR schematic. The major changes made to the design are

1. Removal of all video related hardware.

2. Addition of a 32MB flash memory to store collected audio data and improve detection accuracy over time.

3. Cost-effective BoM and PCB design to facilitate easy manufacturing and scale-up.



Future Scope:
- Clean dataset: The generated dataset is purely code based with no manual cleaning. As the xeno-canto data can sometimes be of questionable quality with a lot of background noise, manual cleaning and labelling can greatly improve accuracy.
- Enhance dataset: More birds, more samples, improve 'unknown' class by adding human and non-bird animal sounds as well.
- Larger input size: Increasing the input length from 1s to 2.5s is shown to improve accuracy.
- Neural Architecture Search: Did not play around with this much due to resource constraints. Hope to do this next
- Improve hardware: No additional hardware was required for this project as the on-board mic was sufficient. However it is reasonable to believe that a better mic can lead to better results. Since this a remote sensing solution, a battery pack and GSM module can also add value.

Other Uploads:
I have uploaded all the source code that I wrote and modified for this project and will be quite excited to work with someone who wants to build on this work. Unfortunately the dataset was too large to upload but I'm happy to share the same.
The demo video also walks through some of the execution steps and code modifications made.

Acknowledgements:
A big thank you to Elektor for hosting and Maxim for sponsoring this competition, it has been a great learning experience.
Many thanks to my friend & colleague Mr. Mohan Prabhakar for helping me by out training all these models on his gaming laptop (16GB Ram, i7-8th gen, GTX1050-Ti). He also helped me test the model on the MAX78000FTHR board and make the demo video.

References:
[1] Stefan Kahl, Connor M. Wood, Maximilian Eibl, Holger Klinck,
BirdNET: A deep learning solution for avian diversity monitoring,
Ecological Informatics,
Volume 61,2021,101236, ISSN 1574-9541,
https://doi.org/10.1016/j.ecoinf.2021.101236.
(https://www.sciencedirect.com/science/article/pii/S1574954121000273)
[2] Audio Based Bird Species Identification using Deep Learning Techniques, ETH Zurich, http://ceur-ws.org/Vol-1609/16090547.pdf
[3] KEYWORDS SPOTTING USING THE MAX78000 (AN7359)
https://www.maximintegrated.com/en/design/technical-documents/app-notes/7/7359.html