kenders2000

As I provided the test dataset labels there is a conflict of interest, but I was hoping as long as that is declared in the write up that'd be ok?

Here is a short description of the method:

BAD Methodology: (results in 84.85% on http://lsis-argo.lsis.org/scores **)**

Features Gammatone-Spectrogram

1024 sample long non-overlapping windows, gamatone filter bank, 64 filters from 50Hs to Fs/2 using this implementation. Prior to computation each each sample was normalised to rms level.

http://www.ee.columbia.edu/ln/rosa/matlab/gammatonegram/

Data-augmentation (this is **not** new data augmentation based on transformations of warblrb10k and ff1010bird data) each spectrogram was randomly shifted by 20% either way in time, and 10% either way in frequency, this random time and frequency shift was re-applied over each mini-batch.

Learning model – Convolutional Neural Network (CNN)

Used Keras, with Theano Backend, trained on a Nvidia GTX 960 with 2 Gb of RAM, 80 epochs took 6 hours to train. Used both warblrb10k and ff1010bird datasets to train, with a 1% validation set to check for overtraining (157 examples), minibatch size was 32

CNN structure:

Input feature shape: 434 x 64

convolution Layer(3x3, 32 filters) – relu activation – with zero padding– Ouput size (64, 434, 32)

convolution Layer(3x3, 32 filters) – relu activation - Ouput size (62, 432, 32)

Max pooling Layer(2x2) – Ouput size (31, 216, 32)

Dropout Layer (25%)

convolution Layer(3x3, 64 filters) – relu activation – with zero padding - Ouput size (31, 216, 64)

convolution Layer(3x3, 64 filters) – relu activation – Ouput size (29, 214 ,64)

Max pooling Layer(2x2) – Ouput size (14, 107, 64)

convolution Layer(3x3, 64 filters) – relu activation – with zero padding - Ouput size (14, 107, 64)

convolution Layer(3x3, 64 filters) – relu activation – Ouput size (12, 105, 64)

Max pooling Layer(2x2) – Ouput size (3, 26, 64)

Dropout Layer (25%)

Flatten layer (4992x1)

Fully connected layer– relu activation (512 units) 512 outputs

Dropout Layer (50%)

Fully connected layer– SoftMax activation (2 units) 2 outputs

Dr Paul Kendrick

Lecturer in Audio, Acoustics and Broadcast Engineering

Acoustics Research Centre

Room 105, Newton Building, University of Salford, Salford M5 4WT

t: +44 (0) 161 295 4618 **| m:** +44 (0)7917303197

p.kendrick@salford.ac.uk **|** www.salford.ac.uk

www.acoustics.salford.ac.uk