**MODELS**

What to try:

* ATTNN paper + Regul
* ATT NN paper but with dsconv e gdsconv
* Riassunto cap 9.3.5
* 40 MFCC: DEEP RESIDUAL LEARNING FOR SMALL-FOOTPRINT KEYWORD SPOTTING
* TC-ResNet8 from Temporal Convolution for Real-time Keyword Spotting on Mobile Devices
* If particular erors: (no-on) train a different network based on those errors
* NO transformer for audio classification (<https://codeburst.io/how-to-use-transformer-for-audio-classification-5f4bc0d0c1f0>, <https://mc.ai/how-to-use-the-transformer-for-audio-classification%E2%80%8A-%E2%80%8Apart-2/>, <https://towardsdatascience.com/music-genre-classification-transformers-vs-recurrent-neural-networks-631751a71c58> )

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **#** | **Dataset** | **Acc** | **Bias** | **Var** | **Params** | **Network** | **LR** | **OPT** | **Reg** | **batchS** | **Comments** | **File** |
| 1 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.771 | 0 | 0.227 | 1,465,226 | LeNet5 con elu | / | Nadam | / | 32 | 0 Bias  High Variance  Overfitting |  |
| 2 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.687 | 0.013 | 0.333 | 123,542 | LeNet5 con elu  Senza i 2 FC layers | / | Nadam | / | 32 | 0 Bias  High Variance  Overfitting |  |
| 3 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.656 | 0 | 0.358 | 123,630 | LeNet5 con elu  Senza i 2 FC layers  Con Batch Norm | / | Nadam | / | 32 | Peggio del 2 | 2020-07-19\_15-13\_LeNet5-elu-1FC-BN |
| 4 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.691 | 0.002 | 0.312 | 123,542 | LeNet5 con elu  Senza i 2 FC layers  Regolariz | / | Nadam | L2  1e-4 | 32 | Leggermente meglio del 2  Overfitting | 2020-07-19\_15-36\_LeNet5-elu-1FC-Reg |
| 5 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.707 | 0.081 | 0.235 | 123,542 | LeNet5 con elu  Senza i 2 FC layers  Regolariz | / | Nadam | L2  1e-3 | 32 | Some Bias  High Variance  Leggermente meglio del 2 e 3  Overfitting | 2020-07-19\_15-44\_LeNet5-elu-1FC-Reg |
| 6 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.796 | 0 | 0.223 | 1,455,422 | LeNet5 con elu  Con FC 120  Regolariz | / | Nadam | L2  1e-3 | 32 | 0 Bias  High Variance  Meglio dei precedenti  Overfitting | 2020-07-19\_16-02\_LeNet5-elu-2FC-Reg |
| 7 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.796 | 0 | 0.223 | 1,455,422 | LeNet5 con elu  Con FC 120  Regolariz | / | Nadam | L2  1e-2 | 32 | Cambiato nulla dal 6 | 2020-07-19\_16-44\_LeNet5-elu-2FC-Reg |
| 8 | 10 cmd  10k-1k-1k  12mfcc +delta | 0.757 | 0 | 0.254 | 224,341 | ATTNETWORK paper | / | Nadam | / | 32 | 0 Bias  High Variance  Overfitting | 2020-07-19\_18-07\_AttRNNSpeechModel |
| 9 | 10 cmd  10k-1k-1k  80 mel | 0.755 | 0 | 0.251 | 224,341 | ATTNETWORK paper | decay | Nadam | / | 32 | 0 Bias  High Variance  Overfitting  Confusion Matrix seems slightly better Than 8 | 2020-07-20\_14-58\_AttRNNSpeechModel |
| 10 | 10 cmd  10k-1k-1k  80 mel Normalized |  |  |  | 224,341 | ATTNETWORK paper | decay | Nadam | / | 32 | Very Bad |  |
| 11 | 10 cmd  10k-1k-1k  80 mel | 0.802 | 0 | 0.223 | 224,341 | ATTNETWORK paper | decay | adam | / | 32 | 0 Bias  High Variance  Overfitting  Better than 6 | 2020-07-20\_15-34\_AttRNNSpeechModel |
| 12 | 10 cmd  10k-1k-1k  40mfcc | 0.773 | 0 | 0.232 | 224,341 | ATTNETWORK paper | decay | adam | / | 32 | 0 Bias  High Variance  Overfitting | 2020-07-20\_16-40\_AttRNNSpeechModel |
| 13 | 10 cmd  20k-1k-1k  40mfcc +delta | 0.773 | 0.029 | 0.205 | 224,341 | ATTNETWORK paper | / | adam | / | 32 | Little Bias  High Variance  Overfitting  Little variance improvement | CLUSTER DEI  2020-07-20\_22-28\_AttRNNSpeechModel |
| 14 | 10 cmd  30k-3k-3k  80 mel | 0.874 | 0.008 | 0.128 | 224,341 | ATTNETWORK paper | / | Nadam | / | 32 | 0 Bias  Little variance | CLUSTER DEI  2020-07-21\_17-04\_AttRNNSpeechModel |
| 15 | 10 cmd  30k-3k-3k  80 mel | 0.909 | 0 | 0.099 | 224,341 | ATTNETWORK paper | decay | adam | / | 32 | 0 Bias  Little variance | CLUSTER DEI  2020-07-21\_19-28\_AttRNNSpeechModel |
| 16 | 10 cmd  30k-3k-3k  40mfcc +delta |  |  |  | 224,341 | ATTNETWORK paper | decay | adam | / | 32 |  | CLUSTER DEI |
| 17 | 10 cmd  1k-1k-1k  No preprocess | 0.643 | 0 | 0.363 | 256,058 | 1D CNN Paper  relu | / | adam | / | 32 | 0 Bias  High Variance  Overfitting  Solution: mor train data. 18 | 2020-07-22\_16-48\_DirectCNN |
| 18 | 10 cmd  10k-1k-1k  No preprocess | 0.858 | 0.006 | 0.149 | 256,058 | 1D CNN Paper  relu | / | adam | / | 32 | 0 Bias  Little variance  Solution: mor train data. 19 | 2020-07-22\_17-00\_DirectCNN |
| 19 | 10 cmd  20k-2k-2k  No preprocess | 0.888 | 0.018 | 0.1 | 256,058 | 1D CNN Paper  relu | / | adam | / | 32 | 0 Bias  Little variance  Solution: mor train data. 20 | 2020-07-22\_17-15\_DirectCNN |
| 20 | 10 cmd  30k-3k-3k  No preprocess | 0.912 | 0.015 | 0.086 | 256,058 | 1D CNN Paper  relu | / | adam | / | 32 | 0 Bias  Little variance  True – Predic: (off-up), (up-off), (go-down), (right-left), (yes-left), (no-go) | 2020-07-22\_17-26\_DirectCNN |
| 21 | 10 cmd  30k-3k-3k  No preprocess | 0.917 | 0.013 | 0.081 | 256,058 | 1D CNN Paper  relu | / | Nadam | / | 32 | 0 Bias  Little variance  True – Predic: (off-up), (up-off), | 2020-07-22\_17-51\_DirectCNN |
| 22 | 10 cmd  30k-3k-3k  No preprocess | 0.914 | 0.01 | 0.089 | 256,058 | 1D CNN Paper  relu | / | Nadam | / | 100 | 0 Bias  Little variance  True – Predic: (off-up), (up-off), (go-no), (no-go) | 2020-07-22\_18-14\_DirectCNN |
| 23 | 10 cmd  30k-3k-3k  No preprocess | 0.93 | 0.017 | 0.058 | 257,018 | 1D CNN Paper  Relu  BatchNorm after each CNN  Dropout 0.25 after 128 e 64 | / | Nadam | / | 32 | 0 Bias  Little variance  Solution: regularization. 24  True – Predic: (off-up), (up-off), (go-no), (no-go) | 2020-07-23\_13-11\_DirectCNNBatchDrop |
| 24 | 10 cmd  30k-3k-3k  No preprocess | 0.875 | 0.147 | 0.015 | 257,018 | 1D CNN Paper  Relu  BatchNorm after each CNN  Dropout 0.25 after 128 e 64 | / | Nadam | L2  1e-2 | 32 | High bias  0 Variance  Very bad plot for validation  Solution: reduce regularization. 25 | 2020-07-23\_14-08\_DirectCNNBatchDrop |
| 25 | 10 cmd  30k-3k-3k  No preprocess | 0.926 | 0.079 | 0.088 | 257,018 | 1D CNN Paper  Relu  BatchNorm after each CNN  Dropout 0.25 after 128 e 64 | / | Nadam | L2  1e-3 | 32 | Little Bias  Little Variance  Very bad plot for validation  Solution: reduce regularization. 26 | 2020-07-23\_14-41\_DirectCNNBatchDropRegu |
| 26 | 10 cmd  30k-3k-3k  No preprocess | 0.925 | 0.046 | 0.053 | 257,018 | 1D CNN Paper  Relu  BatchNorm after each CNN  Dropout 0.25 after 128 e 64 | / | Nadam | L2  1e-4 | 32 | Little Bias  Little Variance  Very bad plot for validation  Solution: reduce regularization. 27 | 2020-07-23\_15-30\_DirectCNNBatchDropRegu |
| 27 | 10 cmd  30k-3k-3k  No preprocess | 0.919 | 0.03 | 0.061 | 257,018 | 1D CNN Paper  Relu  BatchNorm after each CNN  Dropout 0.25 after 128 e 64 | / | Nadam | L2  1e-5 | 32 | Little Bias  Little Variance  Meglio il 23 senza regolarizzazione | 2020-07-23\_15-56\_DirectCNNBatchDropRegu |
| 28 | 10 cmd  30k-3k-3k  No preprocess | 0.929 | 0.013 | 0.112 | 257,018 | 1D CNN Paper  elu  BatchNorm after each CNN.  Dropout 0.25 after 128 e 64 | / | Nadam | / | 32 |  | 2020-07-23\_16-33\_DirectCNNBatchDropELU |
| 29 | 10 cmd  3k-3k-3k  40mfcc | 0.784 | 0 | 0.213 | 333,970 | DS CNN paper  Relu  3 DSCnn  1° Cnn “valid” | / | Nadam | / | 32 | 0 bias  High Variance  At the end was improving, maybe it needs more epochs | 2020-07-24\_15-30\_DSConvModel |
| 30 | 10 cmd  3k-3k-3k  40mfcc | 0.895 | 0 | 0.108 | 571,330 | DS CNN paper  Relu  5 DSCnn  1° Cnn “same” | / | Nadam | / | 32 | 0 bias  High Variance | 2020-07-24\_16-11\_DSConvModel |
| 31 | 10 cmd  30k-3k-3k  40mfcc | 0.953 | 0.004 | 0.052 | 571,330 | DS CNN paper  Relu  5 DSCnn  1° Cnn “same” | / | Nadam | / | 32 | 0 Bias  Little variance  True – Predic: ??? | 2020-07-24\_16-57\_DSConvModel |
| 32 | 10 cmd  30k-3k-3k  120mfcc | 0.939 | 0.007 | 0.06 | 1,178,530 | DS CNN paper  Relu  5 DSCnn  1° Cnn “same” | / | Nadam | / | 32 |  | 2020-08-02\_13-30\_DSConvModel10 |
| 33 | 10 cmd  30k-3k-3k  40mfcc | 0.929 | 0.004 | 0.072 | 127,818 | DSConvSmall | / | Nadam | / | 32 | 0 Bias  Little variance  True – Predic: (down-no), (down-go), (off-up), (go,no), (up-off) | 2020-07-25\_10-30\_DSConvModelSmall |
| 34 | 10 cmd  30k-3k-3k  40mfcc | 0.95 | 0.003 | 0.054 | 262,998 | DSConvMedium | / | Nadam | / | 32 | 0 Bias  Little variance  True – Predic: (off-up), (go,no), (no-down) | 2020-07-25\_11-22\_DSConvModelMedium |
| 35 | 10 cmd  30k-3k-3k  80 mels | 0.96 | 0.001 | 0.052 | 874,930 | DS CNN paper  Relu  5 DSCnn  1° Cnn “same” | / | Nadam | / | 32 | 0 Bias  Little variance | 2020-07-25\_13-50\_DSConvModel |
| 36 | 10 cmd  30k-3k-3k  80 mels | 0.925 | 0.004 | 0.138 | 300,618 | DSConvSmall | / | Nadam | / | 32 | 0 Bias  Little variance | 2020-07-29\_15-42\_DSConvModelSmall |
| 37 | 10 cmd  30k-3k-3k  80 mels | 0.943 | 0.003 | 0.059 | 469,398 | DSConvMedium | / | Nadam | / | 32 | 0 Bias  Little variance | 2020-07-29\_16-51\_DSConvModelMedium |
|  | 10 cmd  30k-3k-3k  120mfcc | 0.925 | 0.004 | 0.076 | 473,418 | DSConvSmall | / | Nadam | / | 32 |  | 2020-08-02\_16-59\_DSConvModelSmall10 |
|  | 10 cmd  30k-3k-3k  120mfcc | 0.948 | 0.006 | 0.067 | 675,798 | DSConvMedium | / | Nadam | / | 32 |  | 2020-08-02\_15-35\_DSConvModelMedium10 |
|  | 10 cmd  30k-3k-3k  120mfcc | 0.955 | 0.005 | 0.044 |  | DSConv | / | Nadam | / | 32 |  | 2020-08-08\_14-24\_DSConvModel10 |
| 38 | 21 cmd  3k-3k-3k  40mfcc | 0.441 | 0 | 0.564 | 241,877 | DSConvSmall | / | Nadam | / | 32 | 0 Bias  High Variance  There are much more examples in Unknown x8 | 2020-07-25\_16-15\_DSConvModelSmall |
| 39 | 21 cmd  30k-9.9k-11k  40mfcc | 0.774 | 0 |  | 241,877 | DSConvSmall | / | Nadam | / | 32 | 0 Bias  High Variance  There are much more examples in Unknown x8 | 2020-07-25\_16-42\_DSConvModelSmall |
| 40 | 21 cmd  70k-9.9k-11k  No preproces | 0.89 | 0.052 | 0.044 | 257,733 | 1D CNN Paper  Relu  BatchNorm after each CNN  Dropout 0.25 after 128 e 64 | / | Nadam | / | 32 | Little Bias  Little Variance  If training more the bias would be 0.  There are much more examples in Unknown x8 | 2020-07-26\_10-54\_directCNN |
| 41 | 21 cmd  70k-9.9k-11k  80mels | 0.871 | 0.003 | 0.114 | 224,704 | ATTNETWORK paper | decay | Nadam | / | 32 | 0 Bias  High variance | 2020-07-26\_14-53\_AttRNNSpeechModel21 |
| 42 | 21 cmd  84k-9.9k-11k  40 mfcc | 0.809 | 0.001 | 0.171 | 224,704 | ATTNETWORK paper | decay | adam | / | 32 | 0 Bias  High variance | 2020-07-26\_16-21\_AttRNNSpeechModel21 |
| 43 | 21 cmd  84k-9.9k-11k  40 mfcc Delta | 0.797 | 0.033 | 0.143 | 224,704 | ATTNETWORK paper | / | Nadam | / | 32 | CLUSTER 18091 | CLUSTER DEI  ?????? |
| 44 | 21 cmd  84k-9.9k-11k  40 mfcc | 0.892 | 0.006 | 0.101 | 241,877 | DSConvSmall | / | Nadam | / | 32 | 0 Bias  High variance | 2020-07-27\_15-42\_DSConvModelSmall21 |
| 45 | 21 cmd  84k-9.9k-11k  40 mfcc | 0.922 | 0.002 | 0.063 | 399,233 | DSConvMedium | / | Nadam | / | 32 | 0 Bias  Little variance | 2020-07-27\_15-50\_DSConvModelMedium21 |
| 46 | 21 cmd  84k-9.9k-11k  40 mfcc | 0.937 | 0.01 | 0.052 | 738,321 | DSConv | / | Nadam | / | 32 | DA DARE A LUCA | 2020-08-07\_12-47\_DSConvModel21 |
| 47 | 21 cmd  84k-9.9k-11k  80mels | 0.9 | 0.003 | 0.091 | 604,757 | DSConvSmall | / | Nadam | / | 32 | 0 Bias  Little variance | 2020-07-28\_14-18\_DSConvModelSmall21 |
| 48 | 21 cmd  84k-9.9k-11k  80mels | 0.927 | 0.006 | 0.067 | 832,673 | DSConvMedium | / | Nadam | / | 32 | 0 Bias  Little variance | 2020-07-30\_12-03\_DSConvModelMedium21 |
| 49 | 21 cmd  84k-9.9k-11k  80mels | 0.934 | 0.004 | 0.059 | 1,375,881 | DSConv | / | Nadam | / | 32 | 0 Bias  Little variance | 2020-07-30\_14-37\_DSConvModel21 |
| 50 | 21 cmd  84k-9.9k-11k  40 mfcc Delta | 0.865 | 0.006 | 0.155 | 967,637 | DSConvSmall | / | Nadam | / | 32 |  | 2020-08-07\_14-41\_DSConvModelSmall21 |
| 51 | 21 cmd  84k-9.9k-11k  40 mfcc Delta | 0.917 | 0.007 | 0.066 | 1,266,113 | DSConvMedium | / | Nadam | / | 32 |  | 2020-08-08\_11-43\_DSConvModelMedium21 |
| 52 | 21 cmd  84k-9.9k-11k  40 mfcc Delta | 0.927 | 0.004 | 0.064 | 2,013,441 | DSConv | / | Nadam | / | 32 | CLUSTER 18140 | 2020-07-28\_10-36\_DSConvModel21 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Ensemble 1 |  |  |  |  |  |  |  |  |  |  |  |
|  | Ensemble 2 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

**Datasets**:

* 10 cmd
* 20 cmd + unknown

**Preprocessing**:

* No. Directly train on the 16000-element vector
* Mel spectrogram with 80mels
* 40 MFCC
* 40MFCC +delta +deltadelta
* Solo 12 MFCC????? Non si puo fare DSConv

**Architectures:**

* Raw 1D CNN
* DSConv
* ATTNN

**Metrics**:

* Accuracy (in a problem where there is a large class imbalance, a model can predict the value of the majority class for all predictions and achieve a high classification accuracy. So, further performance measures are needed such as F1 score and Brier score, but since in our dataset the classes are balanced we can still use Accuracy)
* Prediction speed (ms)

Comparison:

* DEEP RESIDUAL LEARNING FOR SMALL-FOOTPRINT KEYWORD SPOTTING

Project (60 points):

* originality (10)
* data preprocessing techniques (10)
* learning architectures (20)
* comparison against other/existing approaches (10)
* live demo of the code (10)

Written report (40 points):

* clarity of exposition (10)
* completeness (10)
* analysis of results (number and type of metrics used) (20)

Oral exposition (20 points):

* duration (your talk must be shorter than 25 minutes, using slides) (10)
* clarity of exposition (10)

The final grade will be computed as grade = (points\*30)/100