

Musical Genre Classification via CNN



AutoML 연구실
석사 과정 조성현
21.10.05

The standard statistics of the datasets

Dataset	W	Avg. Length	Suggested Domain	Used Domain	C	Avg. Size of Class	Mel Spectrogram
Ballroom	698	30	Genre	Genre	8	87.25 ± 17.56	O
Ballroom-Extended	4180	30	Genre	Genre	13	321.54 ± 195.70	X
emoMusic-G	1000	45	Emotion	Genre	8	125.00 ± 0.0	X
Emotify-G	400	43	Emotion	Genre	4	100.00 ± 0.0	O
FMA-MEDIUM	24984	30	Genre	Genre	16	1561.50 ± 2069.36	X
FMA-SMALL	7996	30	Genre	Genre	8	999.50 ± 0.50	X
GiantStepsKey-G	604	80	Key	Genre	24	24.9130 ± 28.0274	O
GMD-G	1042	300	Groove	Genre	18	57.89 ± 70.07	O
GTZAN	1000	30	Genre	Genre	10	100.000 ± 0.000	O
HOMBURG	1886	10	Genre	Genre	9	209.56 ± 134.87	O
ISMIR04	2187	30	Genre	Genre	6	121.17 ± 95.60	X
MICM	1137	360	Genre	Genre	7	162.43 ± 119.72	$\triangle$
Tropical Genres	1500	30	Genre	Genre	5	300.000 ± 0.000	$\triangle$

|W| : the number of patterns

Avg. Length : the average length of music excerpts in the collection

Suggested Domain : the domain suggested originally by the creator of the collection

Used Domain : the domain used for creating the corresponding dataset

|C| : the number of classes or tags

Avg. Size of Class : the average number of patterns per class and its standard deviation

Mel Spectrogram : whether converted to mel spectrogram (O : Done, X : Scheduled, $\triangle$: Ongoing)

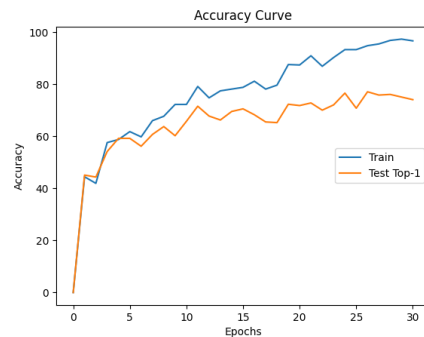
Experiment Settings

Dataset	Avg. Length
Ballroom	30
Ballroom-Extended	30
emoMusic-G	45
Emotify-G	43
FMA-MEDIUM	30
FMA-SMALL	30
GiantStepsKey-G	80
GMD-G	300
GTZAN	30
HOMBURG	10
ISMIR04	30
MICM	360
Tropical Genres	30

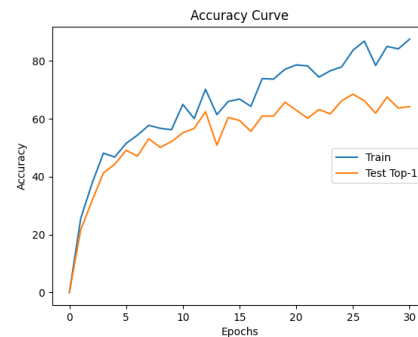
- Overall preprocessings and settings
 1. Convert to .wav file
 2. Split to 30s length
 3. Convert to mel spectrogram
 4. Train : Test = 6 : 4

Comparison with other CNN models

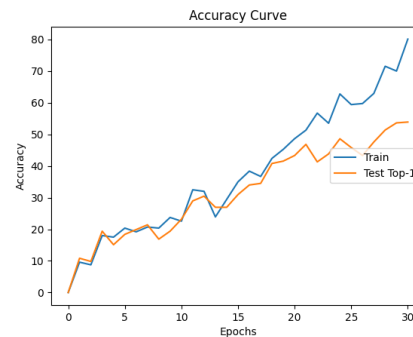
Dataset	Proposed	ResNet18	ResNet50	AlexNet	DenseNet	ShuffleNet
Ballroom	-	-	-	-	-	-
Ballroom-Extended	-	-	-	-	-	-
emoMusic-G	-	-	-	-	-	-
Emotify-G	-	-	-	-	-	-
FMA-MEDIUM	-	-	-	-	-	-
FMA-SMALL	-	-	-	-	-	-
GiantStepsKey-G	-	-	-	-	-	-
GMD-G	-	-	-	-	-	-
GTZAN	-	74.0554	64.2317	53.9043	66.7506	66.4987
HOMBURG	-	-	-	-	-	-
ISMIR04	-	-	-	-	-	-
MICM	-	-	-	-	-	-
Tropical Genres	-	-	-	-	-	-



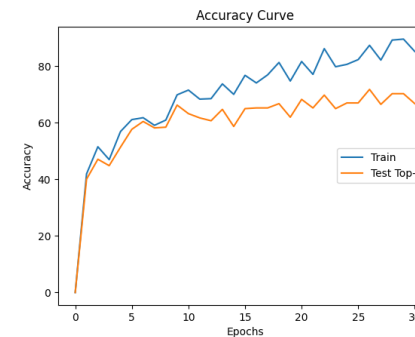
ResNet18



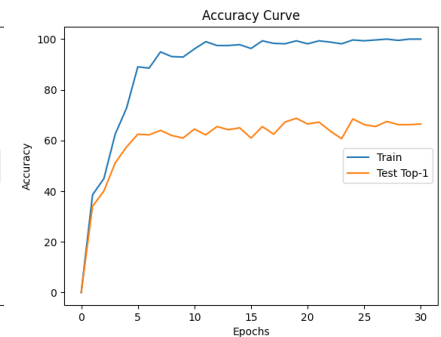
ResNet50



AlexNet



DenseNet



ShuffleNet

Further Discussion Points

Some datasets known as 30s length has some clips with less than 30s.

(e.g. Ballroom, Ballroom-Extended, Emotify-G, FMA-MEDIUM, FMA-SMALL, GMD-G, GTZAN, ISMIR04, MICM)

1. Keep it excluded
 - May differ from the known number of datasets.
 - Datasets consist of less than 30s length are excluded (e.g. HOMBURG).
2. Change numpy shape to a 10-second long data format.
 - May affect classification accuracy.
 - Previously excluded data need to be newly checked and reshaped.