MT-GCN For Multi-Label Audio Tagging With Noisy Labels

Shrivaslava, Harsh, et al.

*ICASSP 2020

음악오디오 연구실

이재준

*International Conference on Acoustics, Speech and Signal Processing

MT-GCN For Multi-Label Audio Tagging With Noisy Labels

MT-GCN For Multi-Label Audio Tagging With Noisy Labels

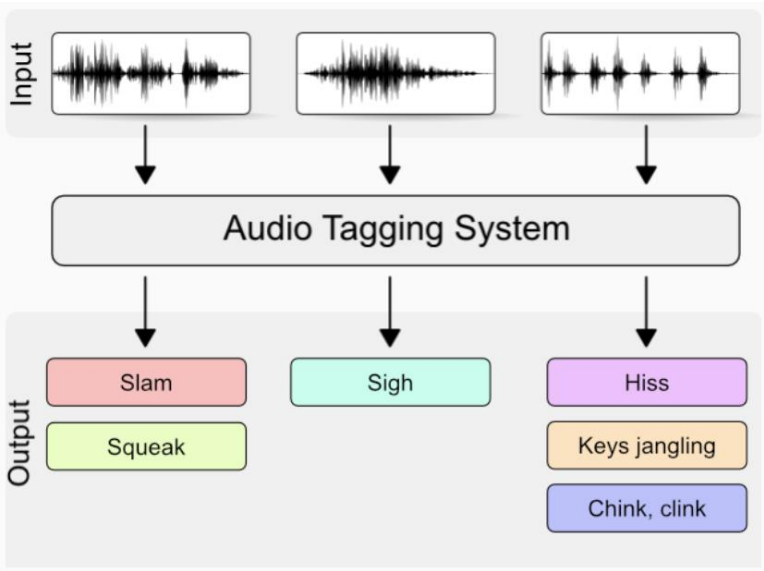


Figure - Overview of a multi-label tagging system. (http://dcase.community/challenge2019/task-audio-tagging)

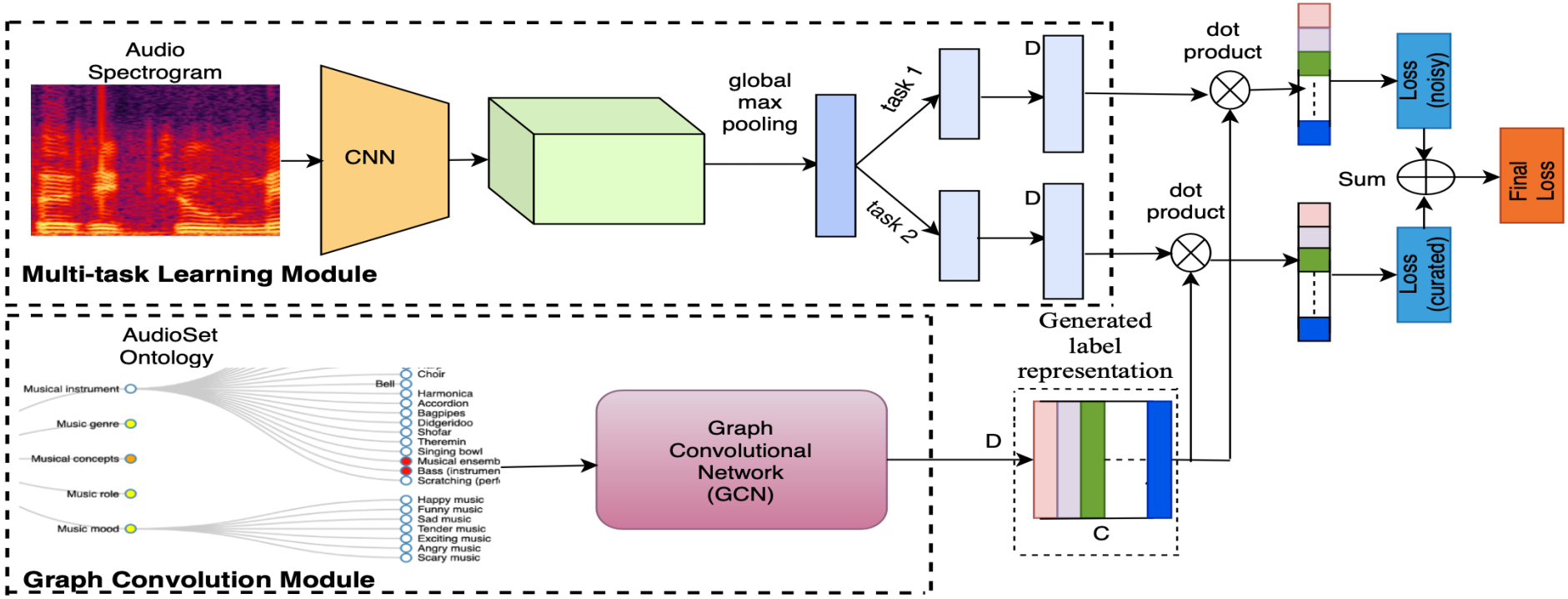
MT-GCN For Multi-Label Audio Tagging With Noisy Labels

Large but noisy labeled data / Small but precisely curated data
=> Could be overfitted to noisy labeled data

MT-GCN For Multi-Label Audio Tagging With Noisy Labels

Multi-task Learning based Graph Convolutional Network
that learns domain knowledge from ontology
=> Regularization effect

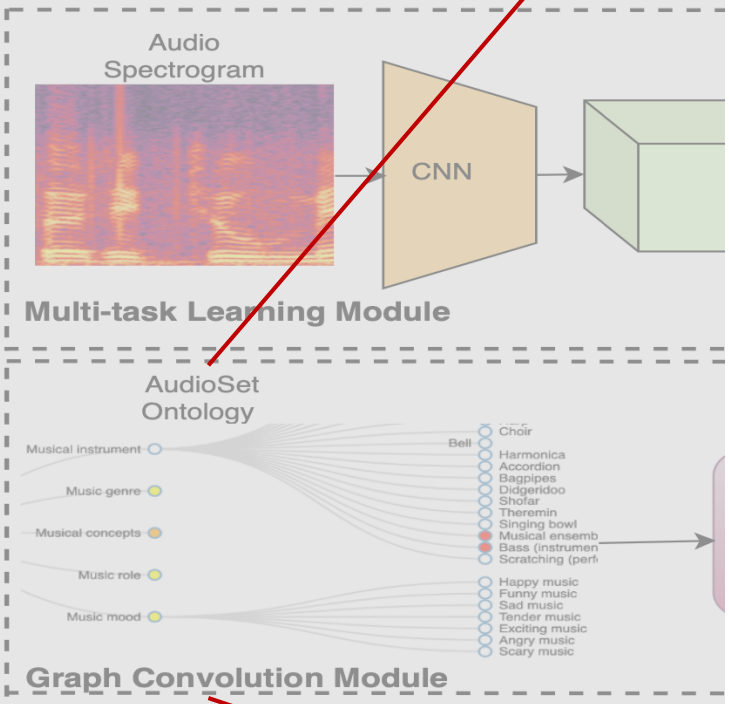
Block diagram of MT-GCN



Benchmark dataset : FSDKaggle2019 in DCASE2019 Challenge (Not Google AudioSet)

Multi-task learning network : ResNet-101

Block diagram of MT-GCN



Dataset : FSDKaggle2019 in DCASE2019 Challe
Network : ResNet-101

Human sounds
- Human voice
- Whistling
- Respiratory sounds
- Human locomotion
- Digestive
- Hands
- Heart sounds, heartbeat
- Otoacoustic emission
- Human group actions

Source-ambiguous sounds
- Generic impact sounds
- Surface contact
- Deformable shell
- Onomatopoeia
- Silence
- Other sourceless

Animal
- Domestic animals, pets
- Livestock, farm animals, working animals
- Wild animals

Sounds of things
- Vehicle
- Engine
- Domestic sounds, home sounds
- Bell
- Alarm
- Mechanisms
- Tools
- Explosion
- Wood
- Glass
- Liquid
- Miscellaneous sources
- Specific impact sounds

Music
- Musical instrument
- Music genre
- Musical concepts
- Music role
- Music mood

Natural sounds
- Wind
- Thunderstorm
- Water
- Fire

Channel, environment and background
- Acoustic environment
- Noise
- Sound reproduction

Figure – Google AudioSet Ontology. (https://research.google.com/audioset/ontology/index.html)

Experiment and Results

Type of correlation (=Adjacency=Affinity) Matrix for MT-GCN

1) **MT-GCN_1** : Co-occurrence based method one - (method of *Chen, Zhao-Min, et al.)
   => Use only curated dataset
2) **MT-GCN_2** : Co-occurrence based method one - (method of *Chen, Zhao-Min, et al.)
   => Use curated and noisy dataset
3) **MT-GCN_3** : Ontology-based method one (Using Google AudioSet)
   => Train GCN using only *n* nodes in AudioSet
4) **MT-GCN_4** : Ontology-based method two (Using Google AudioSet)
   => Train GCN using all *N* nodes in AudioSet and slices out only *n* nodes
   (N - Google AudioSet class number, n - benchmark dataset class number (*N* >>> *n*)

Methods	Overall Lwlrap
MTN	0.6794
MT-GCN_1	0.6941
MT-GCN_2	0.7178
MT-GCN_3	0.7244
MT-GCN_4	**0.7405**

Lwlrap : label-weighted label-ranking average precision (0 bad ~ 1 good)

*Chen, Zhao-Min, et al. "Multi-label image recognition with graph convolutional networks", CVPR 2019

Thank you