Weekly Report

Wangwon Lee, 2019/03/16

This week

- **Audio Classification**
  - Fine tuning in previous work
  - Tune dropout rate in best case
  - Visualize filter

Next week

- **Audio Classification**
  - To visualize 1D-CNN
  - Understanding 1D-CNN model.
  - Mapping our brain data
  - To investigate 1D model more specific

Interesting and new finding

- Fine Tuning
- Filter in DSP (Digital Signal Processing)

The aim of this month / Discussion

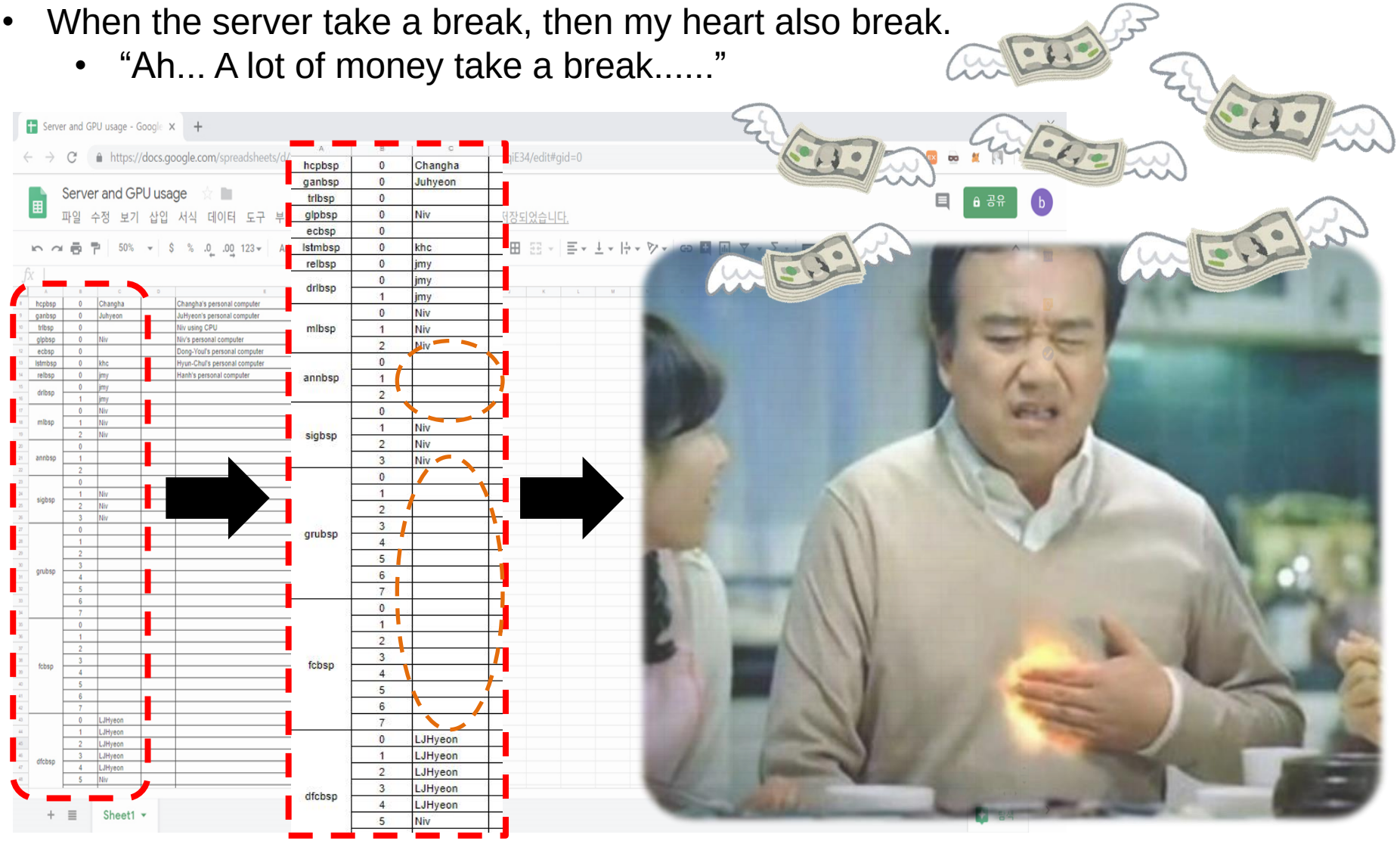
- **The aim of this month:** To study the brain and GLM, To investigate about CNN.

Audio Classification

Few days ago…

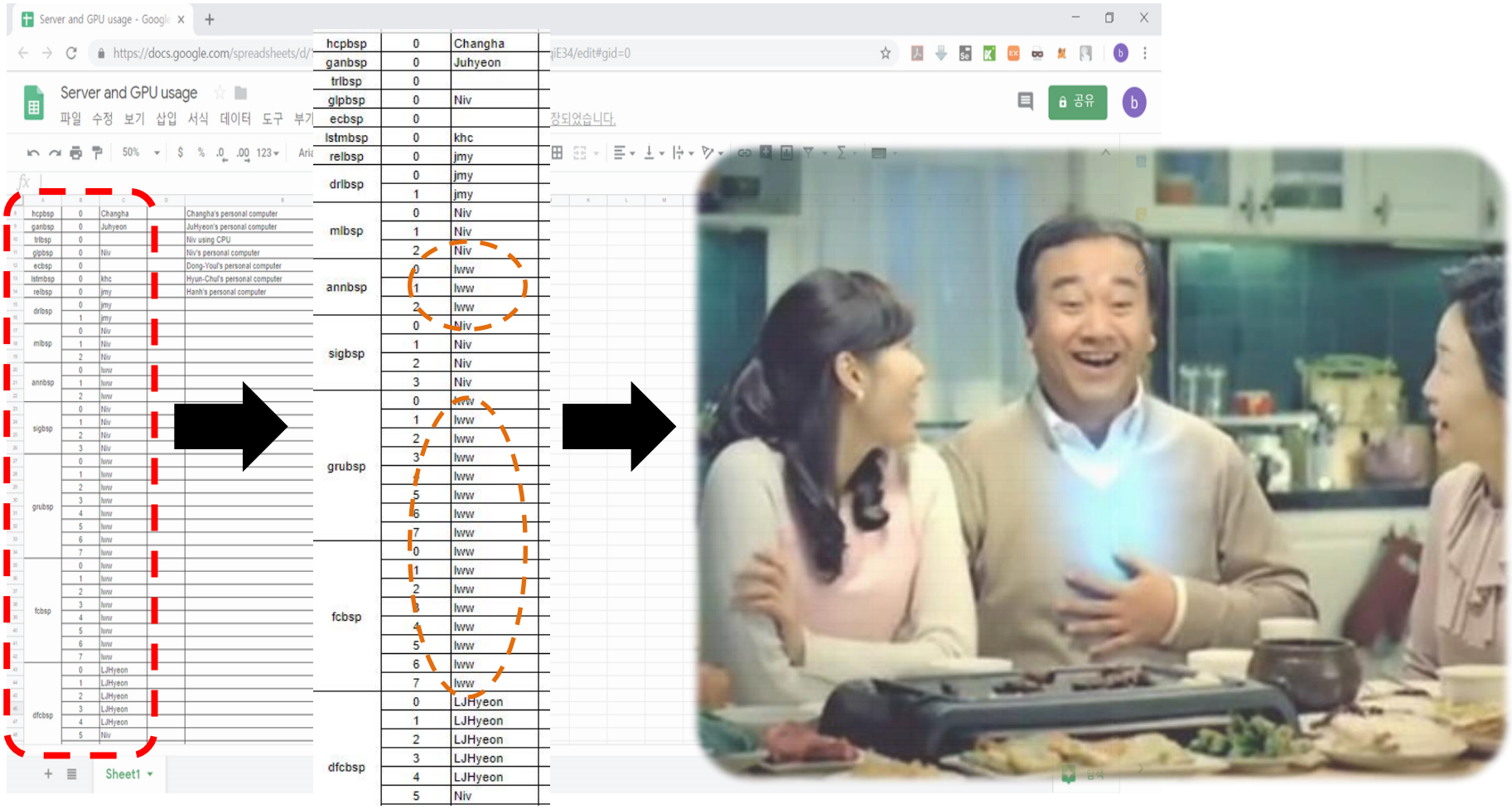
Audio Classification

- When the server take a break, then my heart also break.
  - “Ah... A lot of money take a break.....”



Audio Classification

- So I try the experiment as much as possible.
- A lot of experiment was tried.

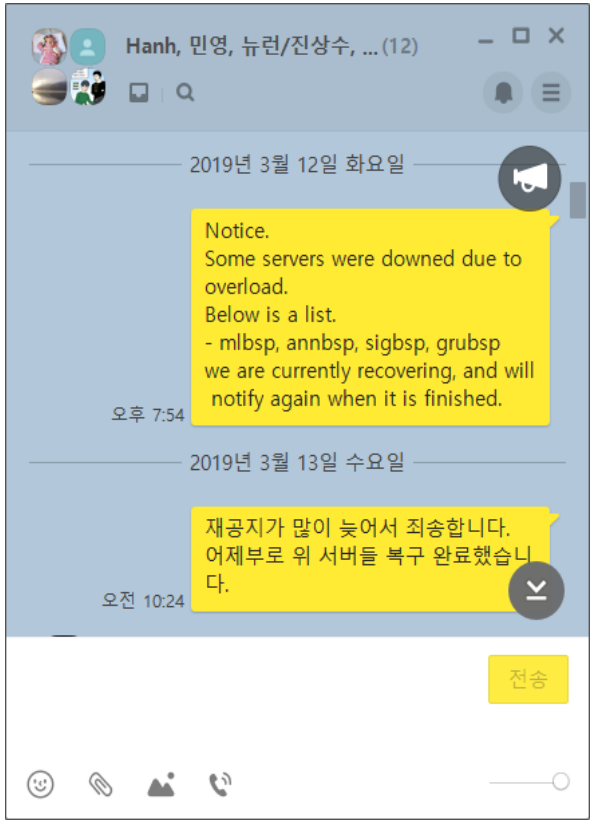


Audio Classification

And few hours later……

Audio Classification

- Some server were downed due to overload.
- I say “Don’t take a break”, But he understand only “~~Don’t take a~~ break”
- So, I felt the need to monitor the server status in real time.



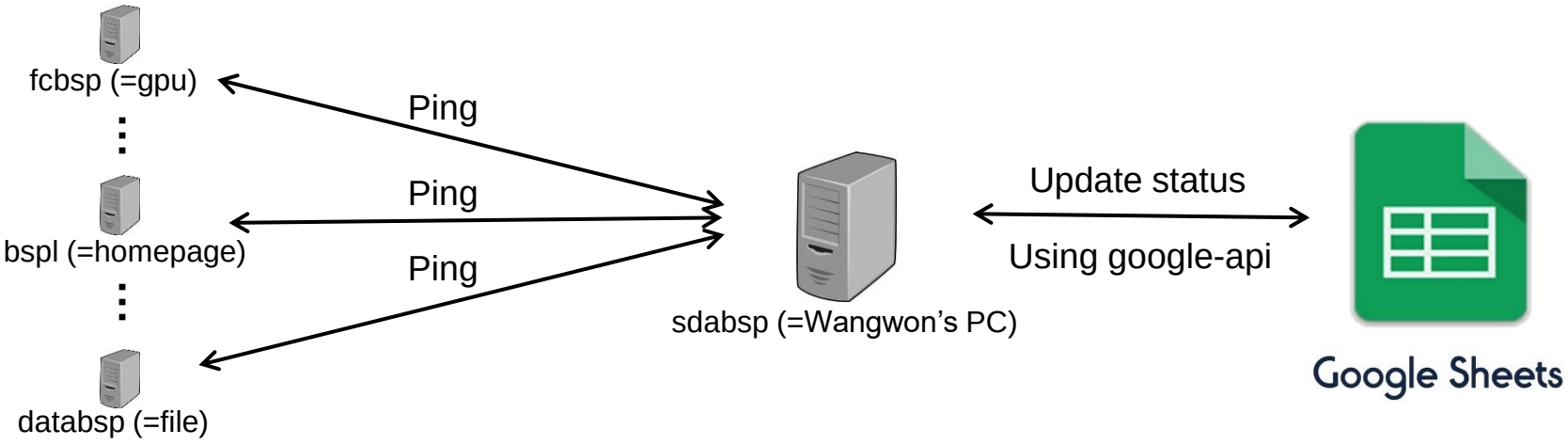
Audio Classification

- Frankly, many people are afraid of new things.
- So if we develop new things, Do not make new one, improve the old one.

A. What is the old one?
→ We always check google-spreadsheets for using server

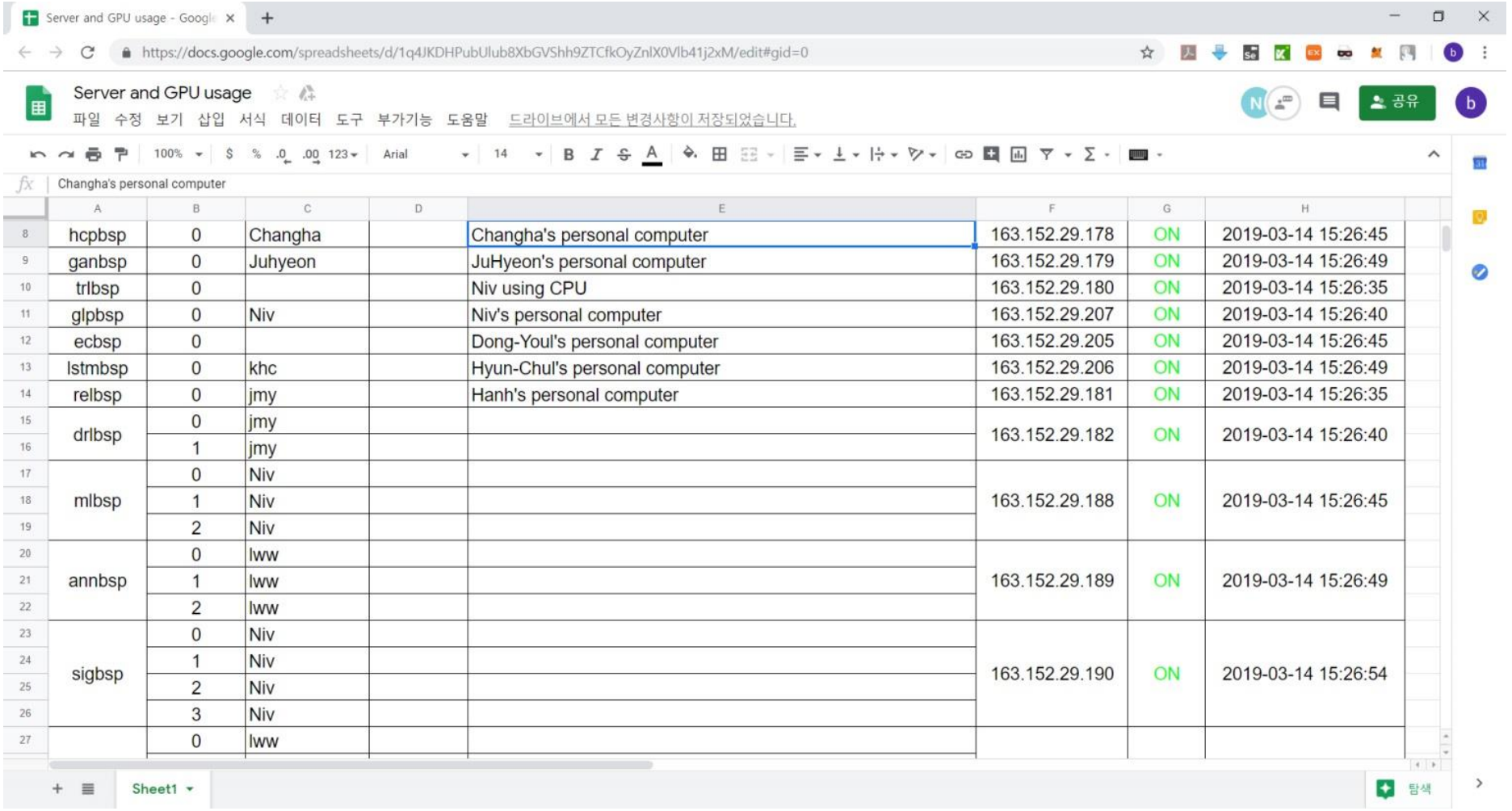
B. What do we need?
→ Check server status in realtime

- So I add new feature to check server status in realtime in google-sheet.



Audio Classification

- Now we can check server status in realtime ☺



Server and GPU usage - Google Sheets

https://docs.google.com/spreadsheets/d/1q4JKDHPubUlub8XbGVShh9ZTCfkOyZnlX0Vlb41j2xM/edit#gid=0

Server and GPU usage

파일 수정 보기 삽입 서식 데이터 도구 부가기능 도움말 드라이브에서 모든 변경사항이 저장되었습니다.

공유

Changha's personal computer

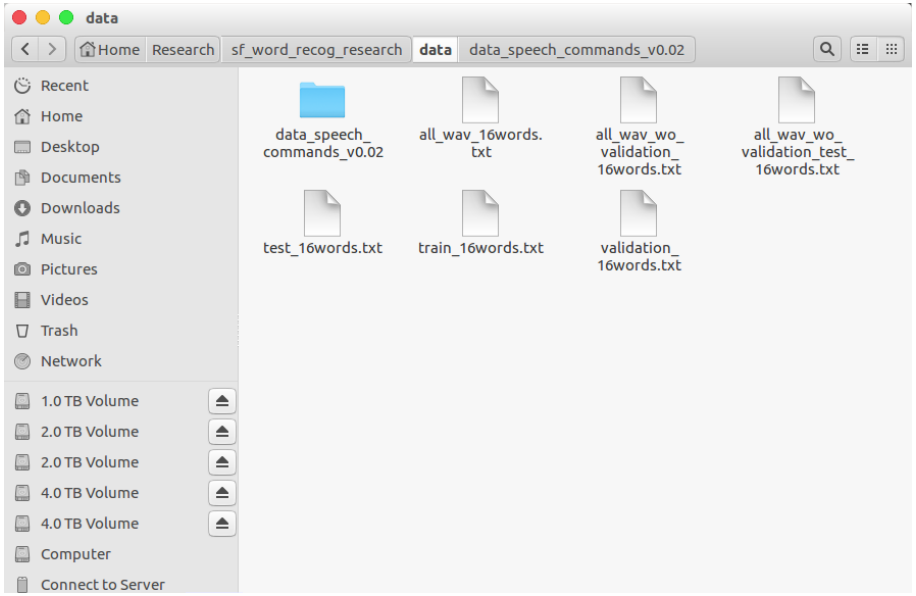
	A	B	C	D	E	F	G	H
8	hcpbsp	0	Changha		Changha's personal computer	163.152.29.178	ON	2019-03-14 15:26:45
9	ganbsp	0	Juhyeon		JuHyeon's personal computer	163.152.29.179	ON	2019-03-14 15:26:49
10	trlbsp	0			Niv using CPU	163.152.29.180	ON	2019-03-14 15:26:35
11	glpbsp	0	Niv		Niv's personal computer	163.152.29.207	ON	2019-03-14 15:26:40
12	ecbsp	0			Dong-Youl's personal computer	163.152.29.205	ON	2019-03-14 15:26:45
13	lstmbsp	0	khc		Hyun-Chul's personal computer	163.152.29.206	ON	2019-03-14 15:26:49
14	relbsp	0	jmy		Hanh's personal computer	163.152.29.181	ON	2019-03-14 15:26:35
15	drlbsp	0	jmy			163.152.29.182	ON	2019-03-14 15:26:40
16		1	jmy					
17	mlbsp	0	Niv			163.152.29.188	ON	2019-03-14 15:26:45
18		1	Niv					
19		2	Niv					
20	annbsp	0	lww			163.152.29.189	ON	2019-03-14 15:26:49
21		1	lww					
22		2	lww					
23	sigbsp	0	Niv			163.152.29.190	ON	2019-03-14 15:26:54
24		1	Niv					
25		2	Niv					
26		3	Niv					
27		0	lww					

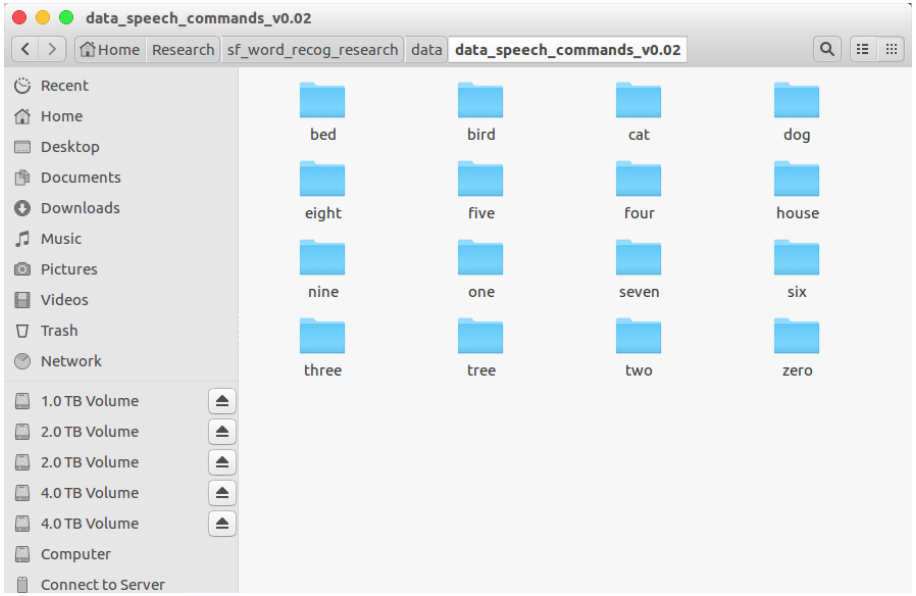
Sheet1

탐색

Audio Classification

- Data is low-waveform.
  - sec: 1, sampling rate: 16000, type: float32, channel: mono
- 16 class data.
  - 'zero', 'one', 'two', 'three', 'four', 'five', 'six', 'seven', 'eight', 'nine', 'bed', 'bird', 'tree', 'cat', 'house', 'dog'
- Train: 40851(≒80%), Validation: 4796(≒10%), Test: 5297(≒10%)

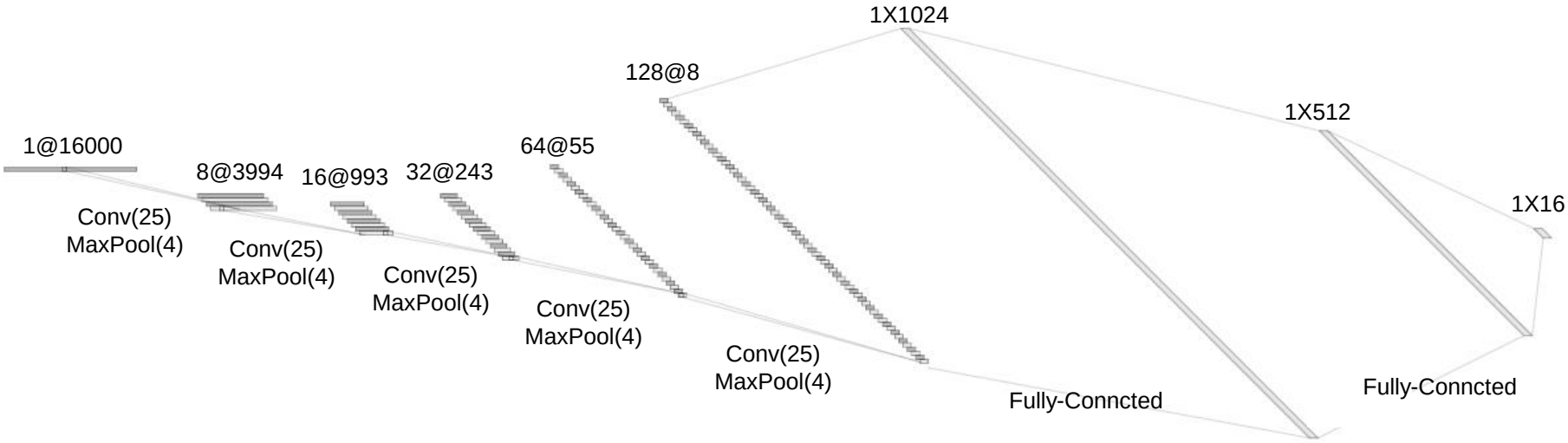




Audio Classification

- For example, ‘5Conv, 2FC’ baseline model’s detail.
- It just flatten 2D model. (5X5 filter->1X25 filter, 2X2 stride->1X4 stride)

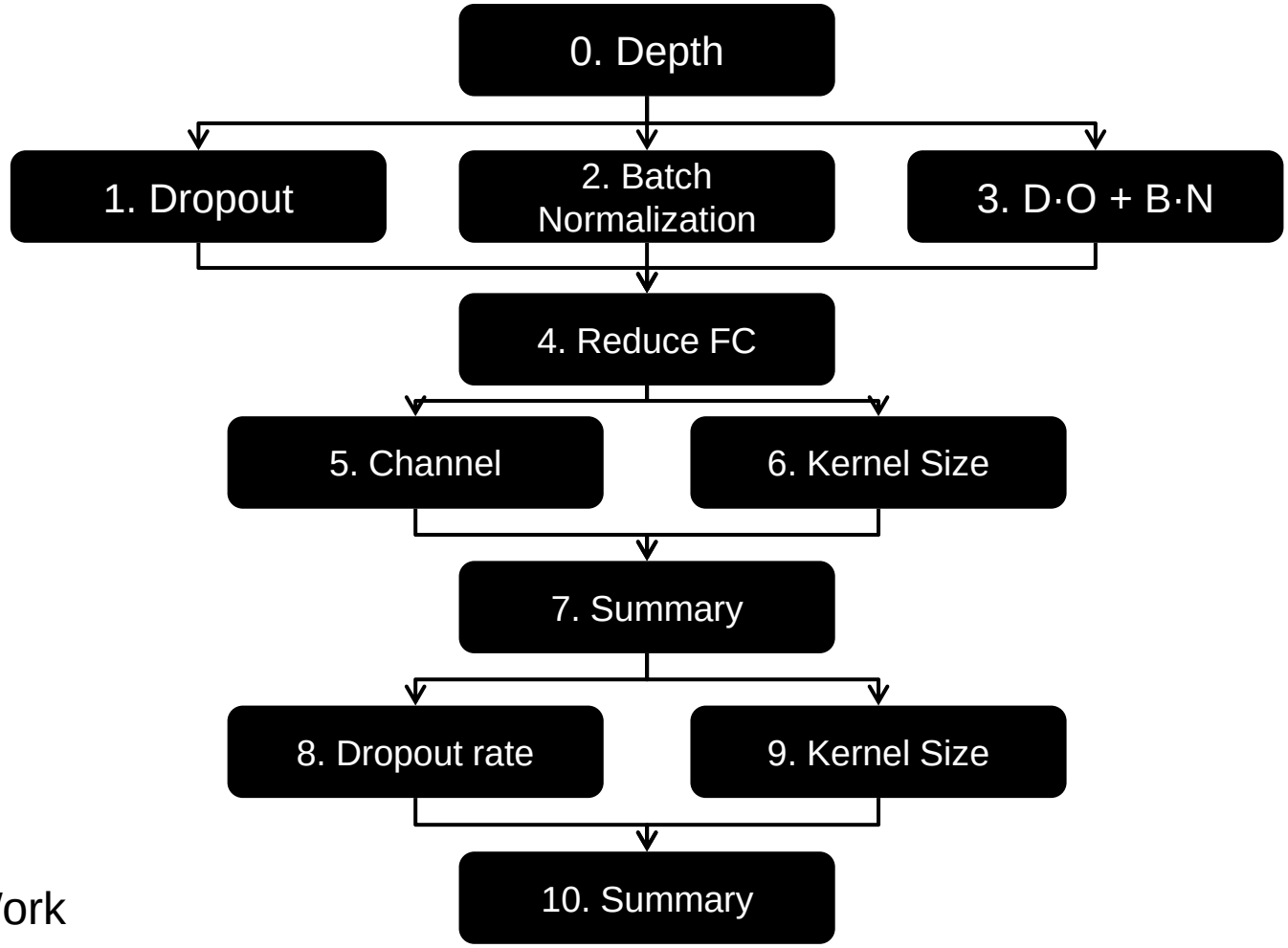
- Input: 16000X1 low waveform.
- Output:1x16 labeled one hot vector. ('zero', …, 'eight', …, 'house', 'dog')
- Loss: cross entropy loss
- Obtimizer: Adam



[Baseline Architecture]

Audio Classification

- Fine tuning task in 1D-CNN



Previous Work

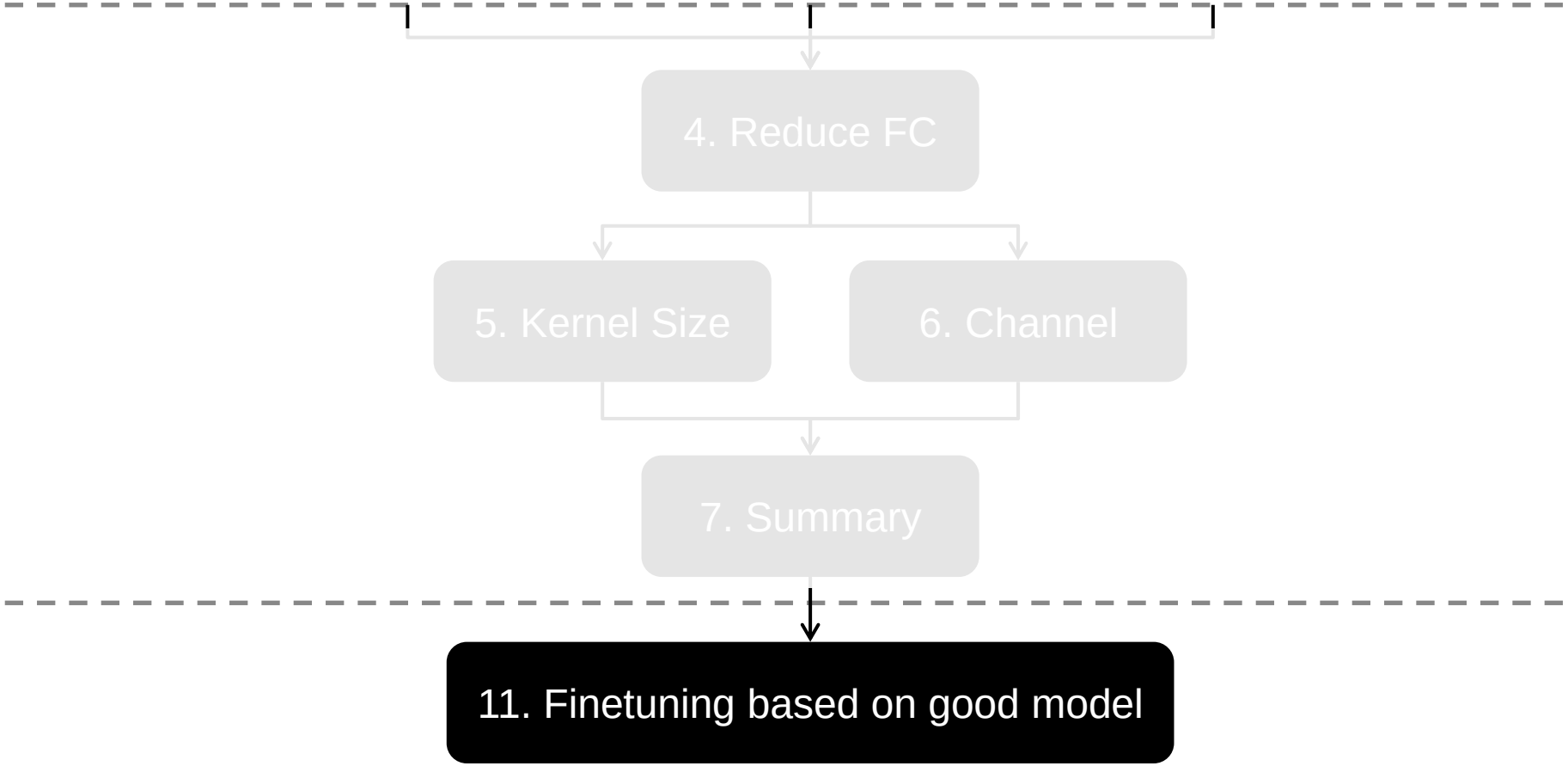
Audio Classification

- This is SOTA(State Of The Art) in previous research.

	Architecture (i = 0,1,2…)	1D DO(0.5)	1D BN	1D DO+BN	Params
	baseline				
base model	5 Conv(25, 8*2^i), 5 Pool(4), 2 FC	0.9090	0.9072	0.9240	1,855,056
	Accuracy and Number of parameters				
Custom channel 32	8 CONV(5, 64)	0.9477	0.9285	0.9391	94,768
Custom channel 64 DO(0.75)	8 CONV(5, 128)	0.9589	X	0.9497	363,600
Custom VGG style DO(0.75)	16 CONV(3, 128) , 8 Pool	0.9620	0.9423	0.9136	470,736
	Only Accuracy				
Custom channel 128	9 CONV(5, 512)	0.9396	0.9632	0.9674	2,071,184

Audio Classification

- Fine tuning task in 1D-CNN



Added Work

Audio Classification

- I tune the dropout's rate in 'custom ch32' model
- The accuracy was exceeded 0.95

Architecture	DO(0.25)	DO(0.25)+BN	DO(0.75)	DO(0.75)+BN
1 CONV(5, 32)	X	X	X	X
2 CONV(5, 32)	X	X	X	X
3 CONV(5, 32)	0.5034	0.5277	0.6260	0.6816
4 CONV(5, 32)	0.6395	0.6712	0.7983	0.8019
5 CONV(5, 64)	0.7248	0.7458	0.8432	0.8461
6 CONV(5, 64)	0.8650	0.8536	0.9101	0.8951
7 CONV(5, 64)	0.9267	0.9215	0.9425	0.9252
8 CONV(5, 64)	0.9452	0.9369	0.9533	0.9458
9 CONV(5, 128)	0.9362	0.9385	0.9506	0.9491

Audio Classification

- This is SOTA(State Of The Art) in current research.
- Despite less than 100,000 parameters, it was exceeded 0.95.

	Architecture (i = 0,1,2…)	1D DO(0.5)	1D BN	1D DO+BN	Params
	baseline				
base model	5 Conv(25, $8*2^i$), 5 Pool(4), 2 FC	0.9090	0.9072	0.9240	1,855,056
	Accuracy and Number of parameters				
Custom channel 32	8 CONV(5, 64)	0.9477	0.9285	0.9391	94,768
Custom channel 64 DO(0.75)	8 CONV(5, 128)	0.9589	X	0.9497	363,600
Custom VGG style DO(0.75)	16 CONV(3, 128) , 8 Pool	0.9620	0.9423	0.9136	470,736
	Only Accuracy				
Custom channel 128	9 CONV(5, 512)	0.9396	0.9632	0.9674	2,071,184
	And, here is new challenger				
Custom channel 32 DO(0.75)	8 CONV(5, 64)	0.9533	X	0.9458	94,768

Audio Classification

- Confusion matrix
- Compare ‘baseline’ and this model
- In ‘Three’ and ‘Tree’ case, it is around two times better.

Actual class

Predict Class

	Zero	One	Two	Three	Fore	Five	Six	Seven	Eight	Nine	Bed	Bird	Cat	Dog	House	Tree
Zero	367	0	8	0	5	0	0	0	1	1	1	1	0	0	0	1
One	1	344	0	0	3	7	0	1	0	6	1	0	0	0	0	1
Two	5	0	359	0	7	0	0	2	1	1	1	1	1	4	1	1
Three	1	0	5	333	0	4	2	3	5	1	1	2	0	0	0	20
Fore	3	2	3	2	344	3	0	2	0	0	2	1	0	6	0	0
Five	1	4	0	4	0	388	1	0	1	4	0	1	1	2	1	0
Six	0	0	0	1	0	1	367	1	2	1	0	0	1	0	0	0
Seven	4	1	0	1	1	0	5	355	0	4	0	1	4	0	0	0
Eight	0	1	0	5	2	1	3	1	349	4	5	2	3	0	0	0
Nine	0	2	0	1	0	3	0	0	0	364	2	1	0	4	0	0
Bed	2	3	1	1	0	1	4	3	4	0	153	3	8	0	0	0
Bird	1	2	0	2	1	3	0	0	0	5	4	133	1	0	0	1
Cat	0	0	3	0	1	1	0	1	0	1	1	0	157	1	2	0
Dog	1	1	2	0	0	1	0	0	1	4	1	5	0	176	0	0
House	0	0	0	0	0	2	1	0	0	1	0	0	3	0	159	1
Tree	1	0	4	32	0	0	0	0	3	1	0	0	0	0	0	122

base model (Acc: 0.9090)

Actual class

	Zero	One	Two	Three	Fore	Five	Six	Seven	Eight	Nine	Bed	Bird	Cat	Dog	House	Tree
Zero	369	0	5	1	3	0	1	4	0	1	0	0	1	0	0	0
One	1	347	0	0	4	1	1	0	0	7	2	0	0	1	0	0
Two	8	0	369	2	0	0	0	0	1	0	0	0	2	2	0	0
Three	1	0	2	356	0	1	1	2	6	0	0	0	0	0	0	8
Fore	2	2	1	1	359	2	0	0	0	0	0	0	1	0	0	0
Five	0	6	0	3	4	386	0	2	1	3	0	1	2	0	0	0
Six	0	0	0	3	1	0	366	1	2	0	1	0	0	0	0	0
Seven	3	0	0	0	1	0	2	367	0	0	3	0	0	0	0	0
Eight	1	0	1	2	1	0	0	0	367	0	2	0	1	0	0	1
Nine	0	6	0	0	0	3	0	0	0	364	3	1	0	0	0	0
Bed	1	0	2	0	0	0	0	0	6	0	166	5	2	1	0	0
Bird	0	1	1	1	0	0	2	0	0	2	7	137	0	2	0	0
Cat	0	0	0	0	0	0	0	1	0	0	1	0	161	4	1	0
Dog	0	1	5	0	0	0	0	0	0	1	1	0	2	182	0	0
House	0	0	2	0	0	1	1	0	0	1	0	0	5	1	156	0
Tree	2	0	2	15	0	0	1	0	4	1	0	0	0	0	0	138

Custom channel 32 DO(0.75) **(Acc: 0.9533)**

Audio Classification

- I tune the dropout's rate in 'custom ch128' model
- 0.97……!?

Architecture	DO(0.25)	DO(0.25)+BN	DO(0.75)	DO(0.75)+BN
1 CONV(5, 128)	X	X	X	X
2 CONV(5, 128)	X	X	X	X
3 CONV(5, 128)	0.5522	0.5124	0.6476	0.6457
4 CONV(5, 128)	0.6768	0.6866	0.7803	0.7867
5 CONV(5, 256)	0.7601	0.7445	0.8386	0.8449
6 CONV(5, 256)	0.8906	0.8789	0.9130	0.8970
7 CONV(5, 256)	0.9369	0.9171	0.9576	0.9302
8 CONV(5, 256)	0.9570	0.9547	0.9618	0.9580
9 CONV(5, 512)	0.9535	0.9701	0.9645	0.9595

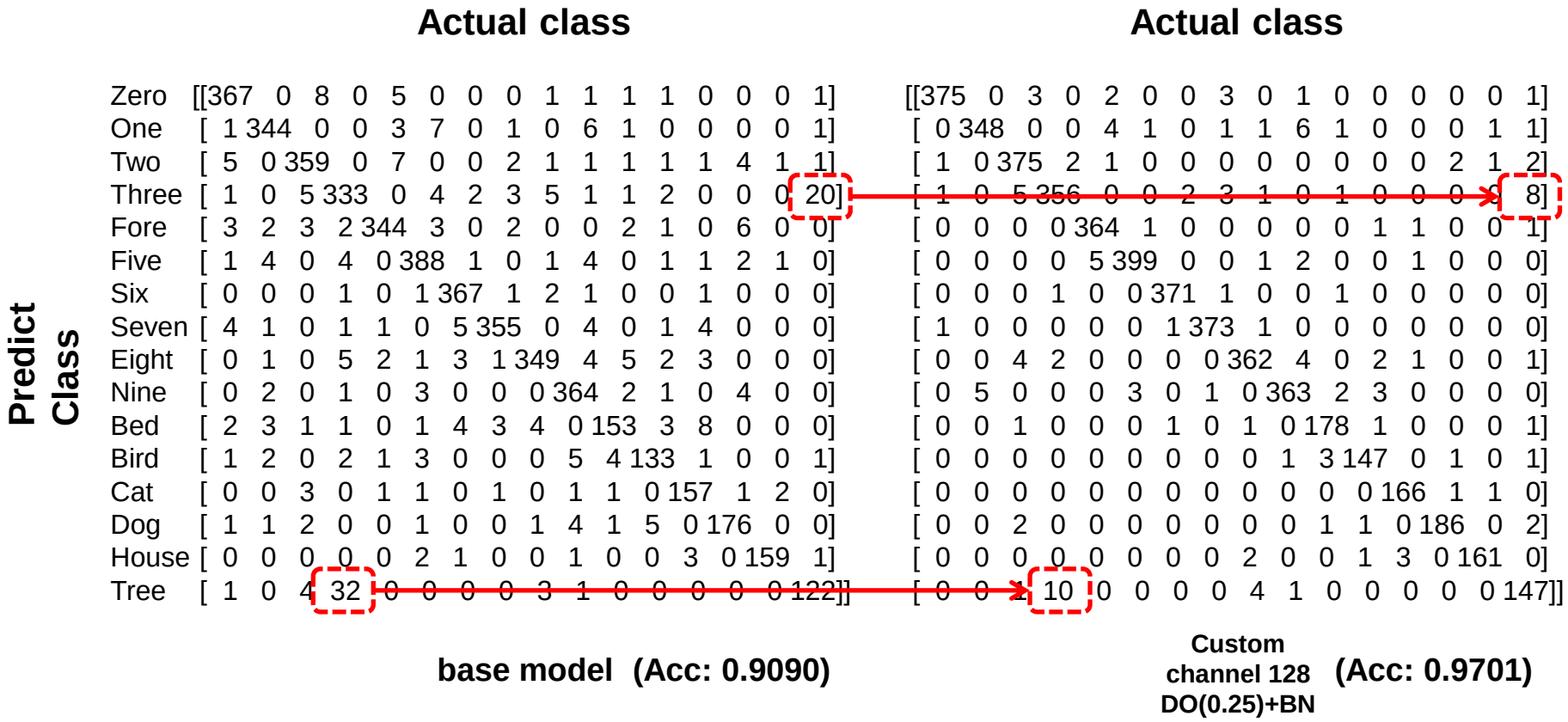
Audio Classification

- This is SOTA(State Of The Art) in current research.
- The accuracy achieve 0.97

	Architecture (i = 0,1,2…)	1D DO(0.5)	1D BN	1D DO+BN	Params
	baseline				
base model	5 Conv(25, $8*2^i$), 5 Pool(4), 2 FC	0.9090	0.9072	0.9240	1,855,056
	Accuracy and Number of parameters				
Custom channel 32 DO(0.75)	8 CONV(5, 64)	0.9533	0.9285	0.9391	94,768
Custom channel 64 DO(0.75)	8 CONV(5, 128)	0.9589	X	0.9497	363,600
Custom VGG style DO(0.75)	16 CONV(3, 128) , 8 Pool	0.9620	0.9423	0.9136	470,736
	Only Accuracy				
Custom channel 128	9 CONV(5, 512)	0.9396	0.9632	0.9674	2,071,184
	And, here is new challenger				
Custom channel 128 DO(0.25)+BN	9 CONV(5, 512)	0.9535	X	0.9701	2,071,184

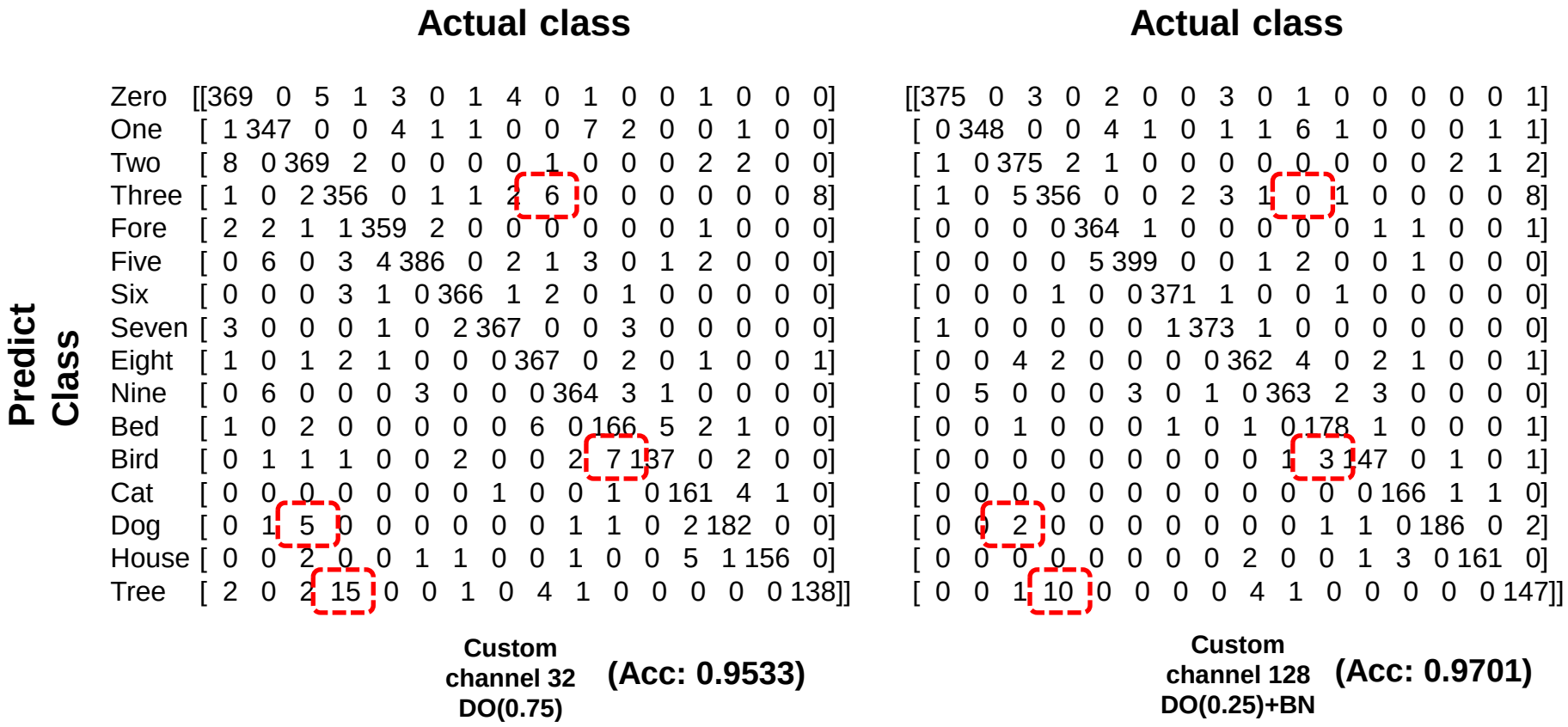
Audio Classification

- Confusion matrix
- Compare ‘baseline’ and this model
- In ‘Three’ and ‘Tree’ case, it is around two or three times better.



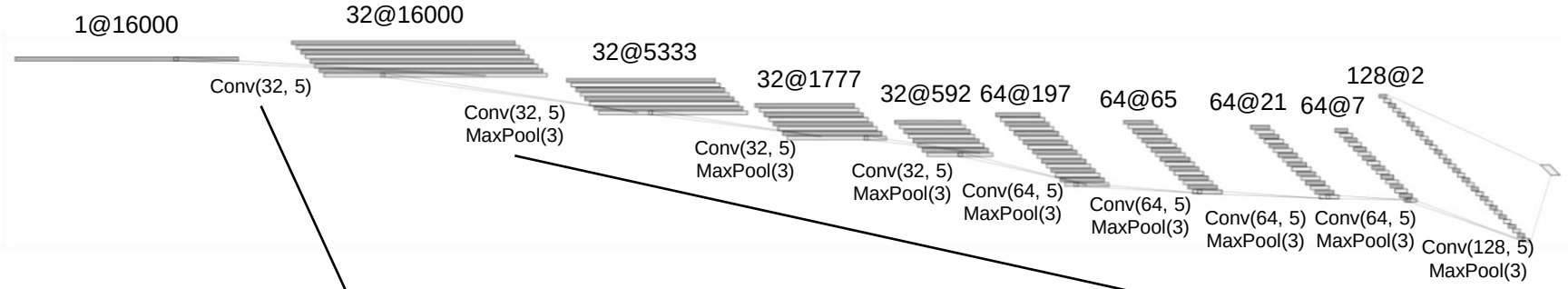
Audio Classification

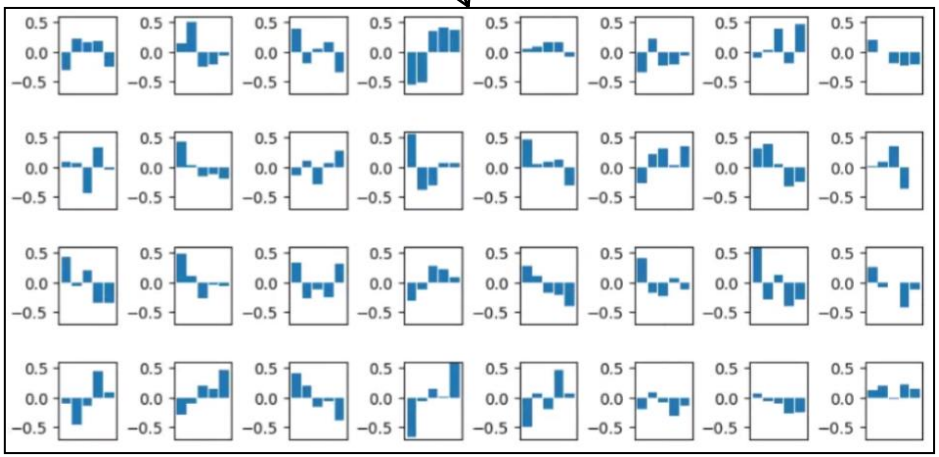
- Confusion matrix
- Compare two best model.
- Not much different, but the right model is a little better.



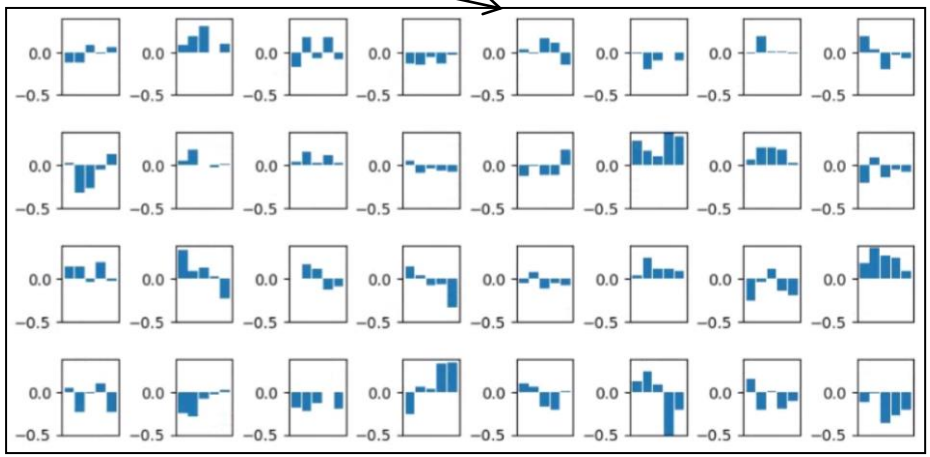
Audio Classification

- Visualize the filter map. (Custom channel 32 DO(0.75) Model)
- Of course, Small number of parameters is easy to analyze
- There was a shape to know, but most of the shape was hard to understand.
- Next time, I will prepare the feature map and analyze it more detail.





First layer's filter



Second layer's filter

Any Question?

Thank you