

희연이와 아이들

김희연 신창민 이형준

독창성

사람은 오감 중 시각에 대한 의존도가 가장 높습니다. 저희는 이 점을 이용하여 주어진 소리 데이터를 보다 **직관적이고 뚜렷한 구분**을 할 수 있도록 이미지화 하였고, 추출한 이미지를 학습데이터로 사용하였습니다.

특징 추출

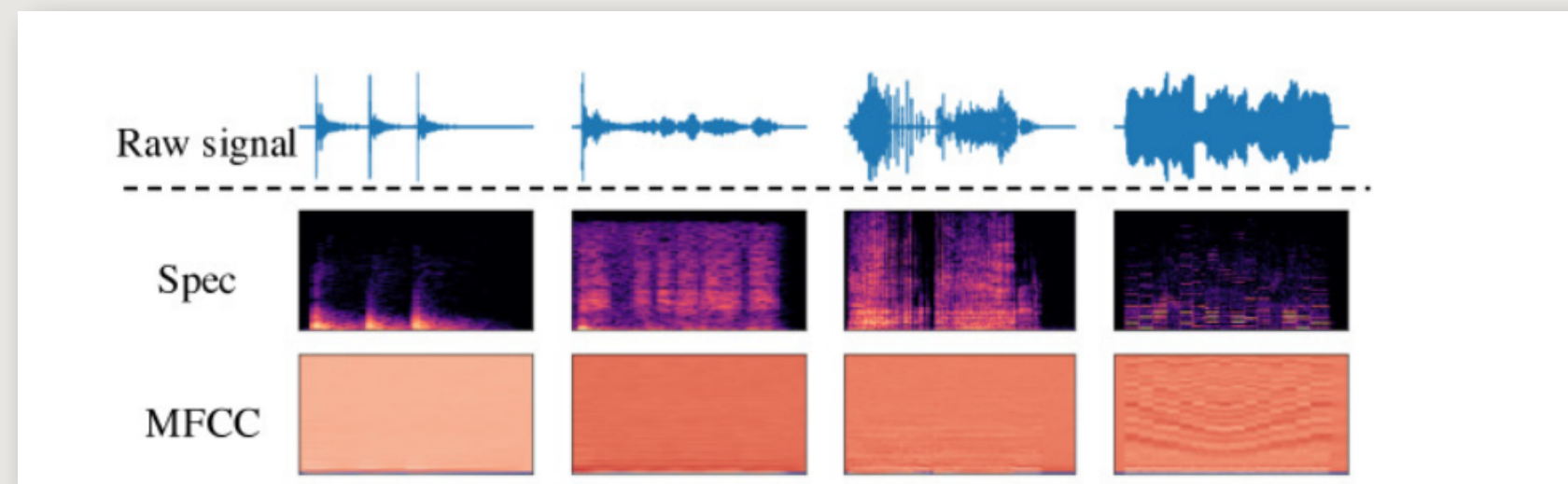


그림 1 다른 특징과 비교

이미지를 추출할 때 소리의 특징을 더 잘 담을수 있는 **Mel Spectrogram**을 선택하였습니다.

matplotlib로 mel spectrogram을 plot하여 그 이미지를 savefig(dpi = 400)으로 추출하였고, 해당 이미지는 (221,223,3) 차원을 가지며, 학습하기 위해 (64,64,3) 차원으로 resize를 해주었습니다.

모델 구조

Conv2D (32, 3, ST = 1, PD='same')
Conv2D (64, 3, ST = 1, PD='same'), MP (2), DO (0.25)
Conv2D (64, 3, ST = 1, PD='same')
Conv2D (64, 3, ST = 1, PD='same'), MP(2), DO (0.5)
Conv2D (128, 3, ST = 1, PD='same')
Conv2D (128, 3, ST = 1, PD='same'), MP(2), DO (0.5)
Dense (512), DO (0.5)
Dense (10), softmax

기타

loss categorical_crossentropy
optimizer tf.keras.optimizers.Adam(learning_rate=0.0001)
batch_size 128
epoch 300

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 64, 64, 3)]	0
conv2d (Conv2D)	(None, 64, 64, 32)	896
activation (Activation)	(None, 64, 64, 32)	0
conv2d_1 (Conv2D)	(None, 62, 62, 64)	18496
activation_1 (Activation)	(None, 62, 62, 64)	0
max_pooling2d (MaxPooling2D)	(None, 31, 31, 64)	0
dropout (Dropout)	(None, 31, 31, 64)	0
conv2d_2 (Conv2D)	(None, 31, 31, 64)	36928
activation_2 (Activation)	(None, 31, 31, 64)	0
conv2d_3 (Conv2D)	(None, 29, 29, 64)	36928
activation_3 (Activation)	(None, 29, 29, 64)	0
max_pooling2d_1 (MaxPooling2)	(None, 14, 14, 64)	0
dropout_1 (Dropout)	(None, 14, 14, 64)	0
conv2d_4 (Conv2D)	(None, 14, 14, 128)	73856
activation_4 (Activation)	(None, 14, 14, 128)	0
conv2d_5 (Conv2D)	(None, 12, 12, 128)	147584
activation_5 (Activation)	(None, 12, 12, 128)	0
max_pooling2d_2 (MaxPooling2	(None, 6, 6, 128)	0
dropout_2 (Dropout)	(None, 6, 6, 128)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 512)	2359808
activation_6 (Activation)	(None, 512)	0
dropout_3 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 10)	5130
Total params: 2,679,626		
Trainable params: 2,679,626		
Non-trainable params: 0		
Train on 6589 samples, validate on 18 samples		

그림 2 모델 구조 출력