

JavaScript와 Tensorflow.js로 배우는 머신러닝

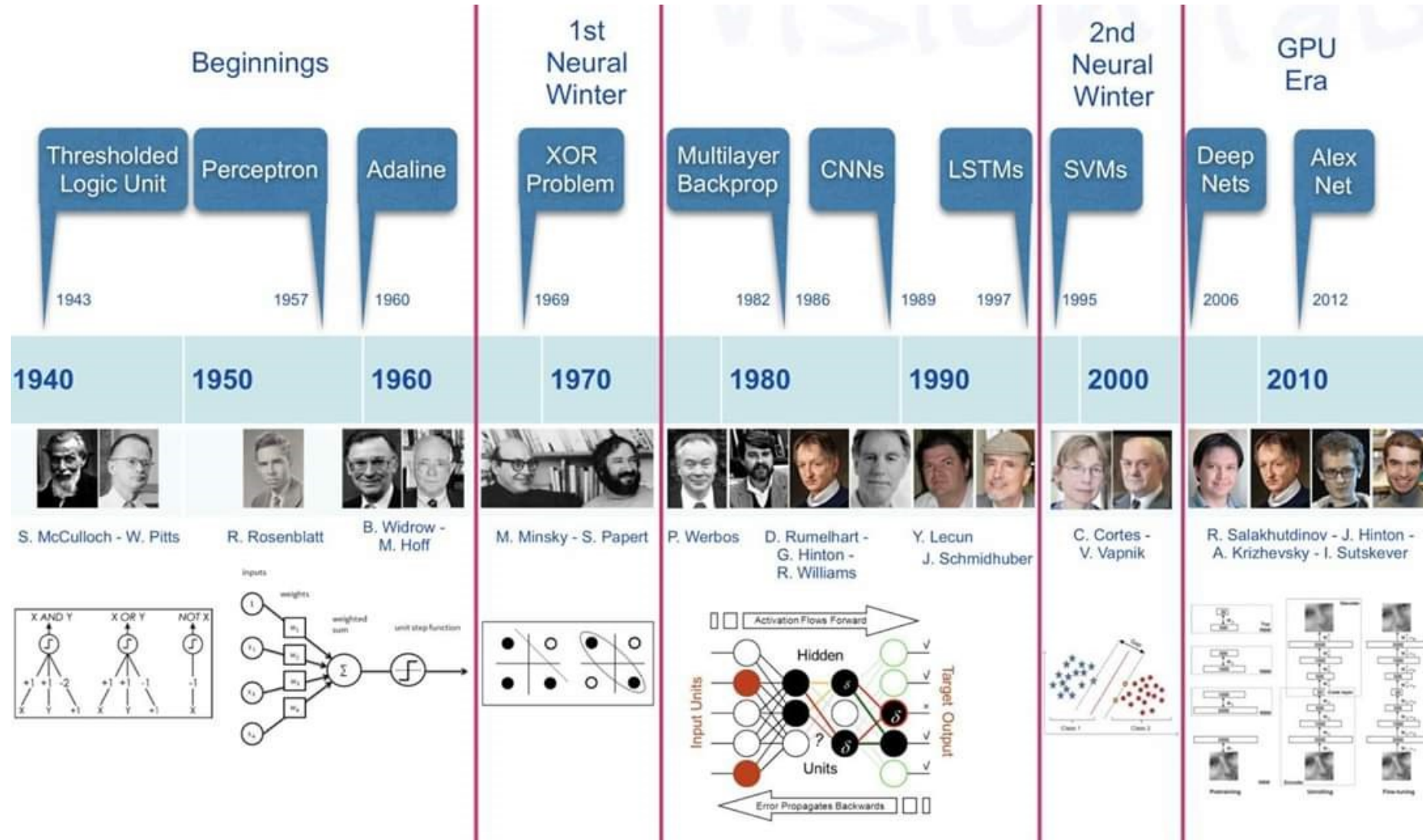
Day 3

2020. 8. 5

정준수 Ph.D.

<https://github.com/JSJeong-me/Sound>

◆ 인공지능(AI)의 발전과정

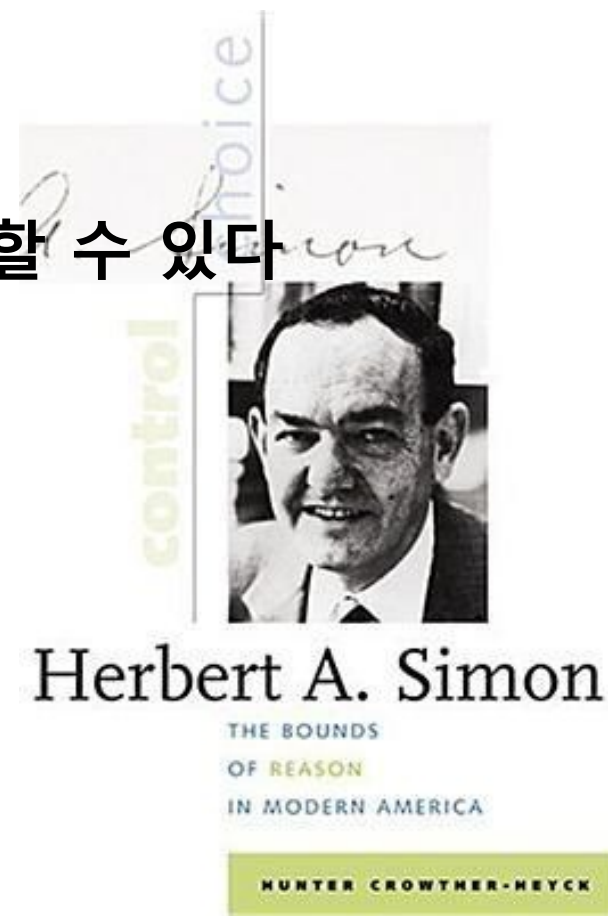


◆ 인공지능 SW란?

마음은 연산자!

연산자란? 산술적 의미의 계산이 아니라,
그 과정의 세부 단계 절차들을 명확히 규정하여, 형식화할 수 있다

- 컴퓨터와 인간의 지능이 같은 원리가 적용되는 정보처리 시스템이라는 관점을 제시하여 컴퓨터와 인간의 마음 및 지능을 연결시키고,
- 오늘날 정보과학과 IT의 이론적 틀, 개념적 기초 제시
- 최초의 현대적 의미의 인공지능 프로그램 개념



◆ 인공지능(AI)의 출발점 - 인지과학은 IT(SW)의 모태적 학문

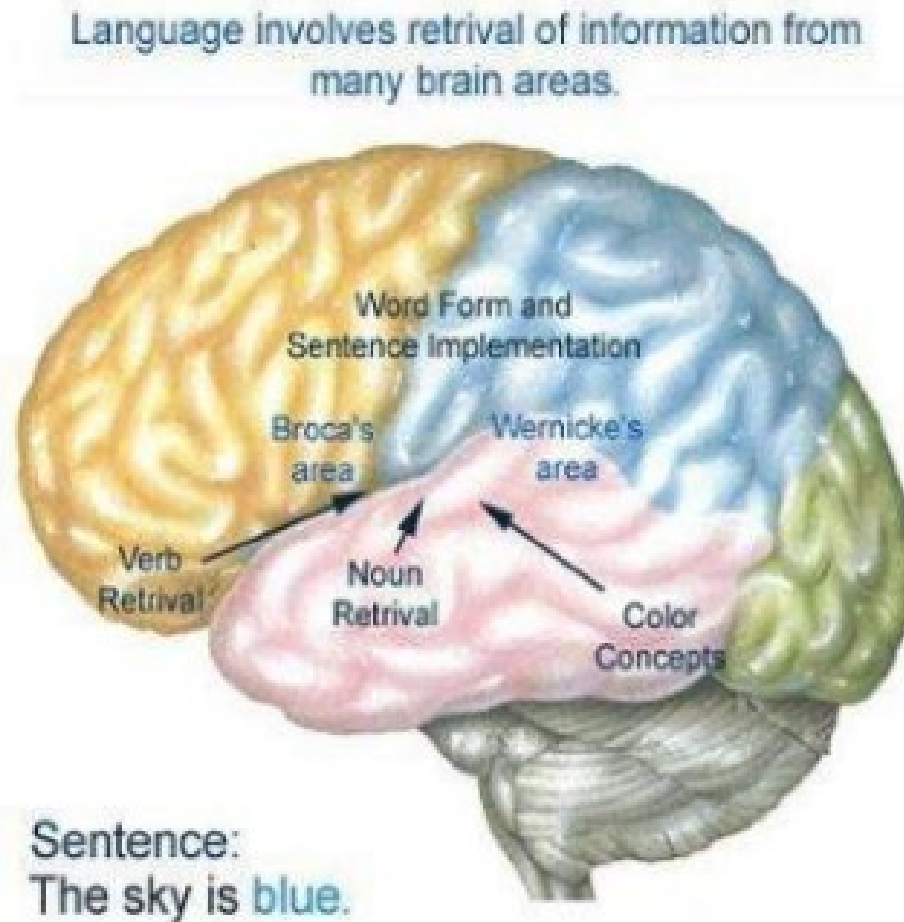
- **인지과학**은 인간의 마음에서 정보 처리 과정이 어떻게 이루어지는가에 대해 다양한 분야의 학제간 연결을 통해 통합적으로 연구하는 학문이다.
- 이전에는 단순한 숫자 처리 계산기에 지나지 않았던 계산기를 정보처리와 지능을 지닌 **컴퓨터**로 대변혁을 할 수 있게 한 이론적, 개념적 틀을 제공한 것이 **인지과학**이다.
- H. Simon, A. Newell 등의 인지과학자가 이러한 개념적 틀의 변혁을 초래한 장본인들이며, 구글 딥마인드 CEO 하사비스가 컴퓨터공학박사가 아니라, 인지(신경)과학 박사로서 인간의 기억 메카니즘 같은 인지(cognition) 연구를 수행하는 점은 시사하는 바가 크다.

◆ 인지과학의 형성은 여러 학문 분야의 융합적 사고 기반으로 형성

- 철학에서의 [형식주의\(formalism\)](#) 이론, 수학에서의 [계산 이론](#)의 발전과 [튜링기계](#) 이론, 디지털 컴퓨터의 발전과 컴퓨터과학에서의 [존 폰 노이만\(John von Neumann\)](#)의 '저장된 프로그램 이론', 1930년대 [커뮤니케이션 이론](#)의 발전과 [정보 이론](#)의 부상, 두뇌를 논리기계로 간주하는 생각의 발달, [사이버네틱스](#)라고 하는 두뇌-기계를 연결한 인공두뇌 이론과 일반체계 이론의 발전으로부터 형성 되었으며,
- 앨런 뉴웰(Allen Newell)과 [허버트 알렉산더 사이먼\(Herbert Alexander Simon\)](#)의 범용 목적적·물리적 기호체계 이론의 구체화, 에이브럼 노ם 촘스키(Avram Noam Chomsky)의 언어학 이론의 부상, 심리학 내에서의 정보 처리 틀 형성 등의 여러 사조와 학문적 사건들의 수렴에 의하여 인지과학이 출발하였다고 할 수 있다

THE “LAD” (Chomsky, 1965)

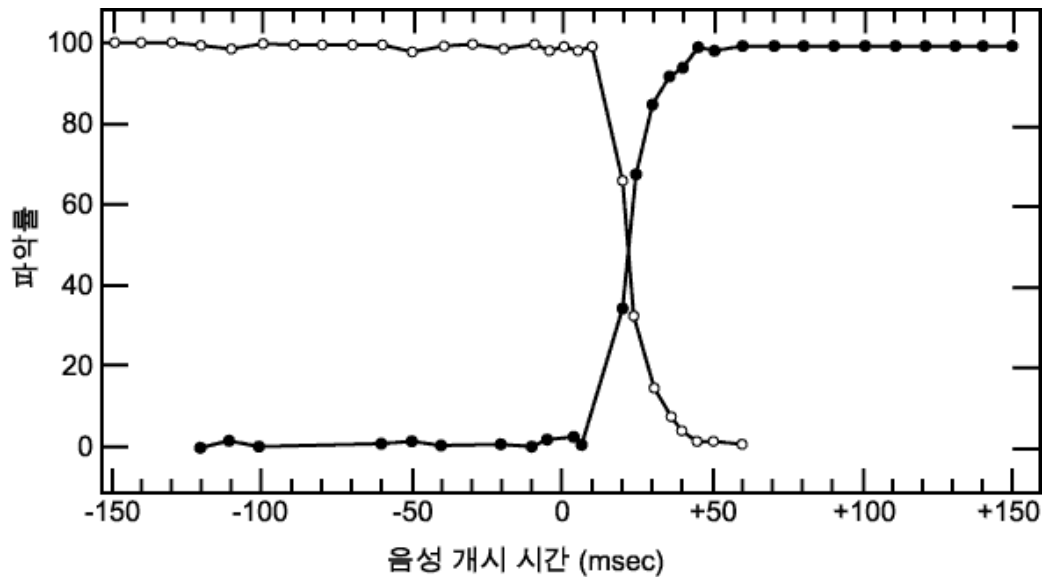
- The language acquisition Device (LAD) is a postulated organ of the brain that is supposed to function as a congenital device for learning symbolic language (i.e., language acquisition).



사람은 누구나 태어나면서부터 언어를 쉽게 터득할 수 있도록 언어습득장치(LAD)를 가지고 태어난다.

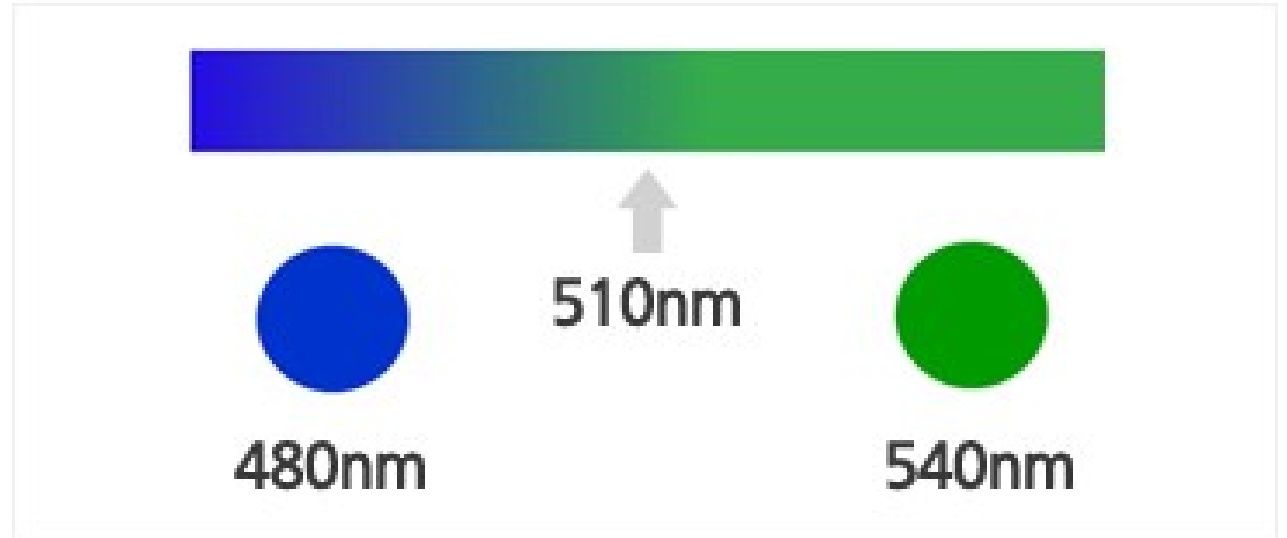
◆ 범주화 지각은 사람의 기본적 속성

소리 (Sound)



사람들은 음소들이 하나의 연속적 차원에서 다르더라도 이들이 별개의 범부에서 유래한다고 지각하는 경향이 있다.

색깔 (Color)



510nm에서 540nm로 변화할 경우 여전히 같은 색에 있다고 생각하지만 480nm로 변화하면 우리는 전혀 다른 범주의 색으로 느끼게 된다.

Innate Knowledge of Language

The Language Acquisition Device

'Universal Grammar'

Input

(Primary
Linguistic
Data)



LAD



Final State

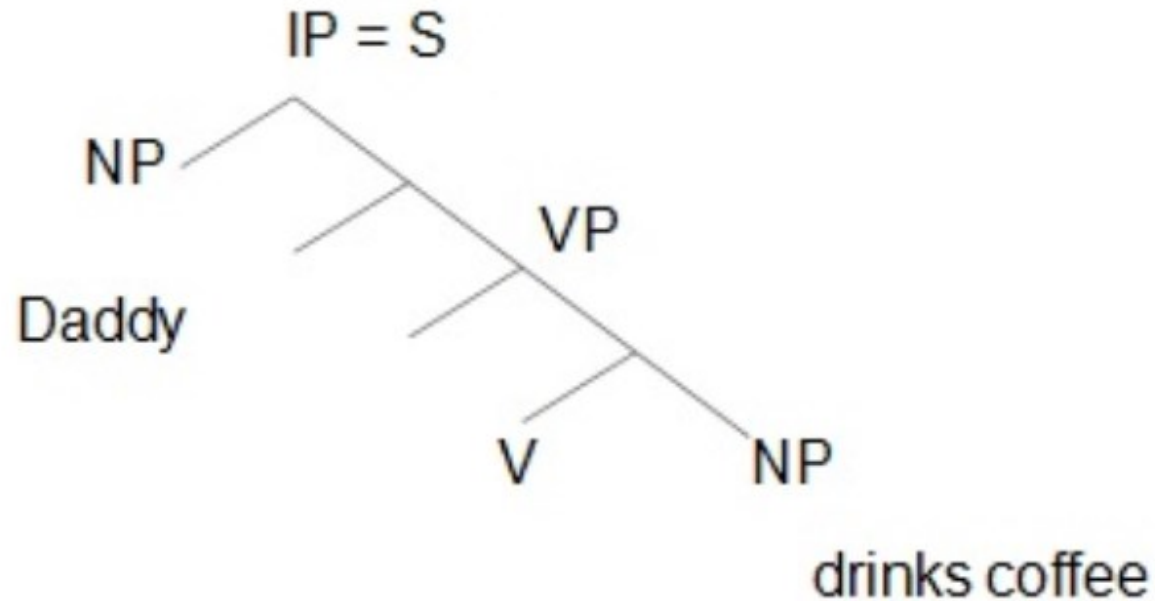
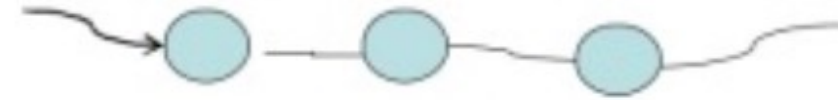


↑
The Mental
Grammar for
a particular
language

Computational System

UG gives us sentence representations that are hierarchical, not linear

'Daddy drinks coffee'



◆ Generative modeling 적용 관점에서 사례 – VAE, GAN, etc

Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling

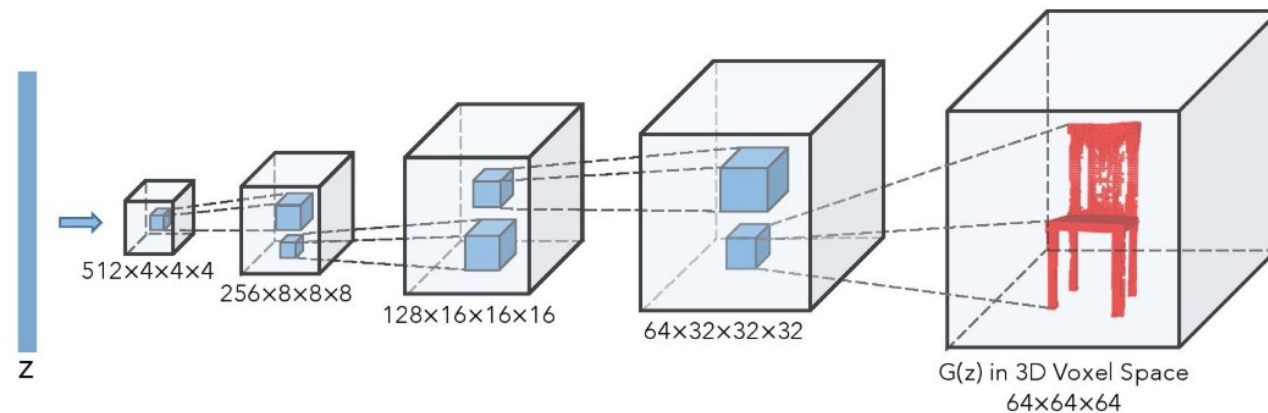


Figure 1: The generator of 3D Generative Adversarial Networks (3D-GAN)



Figure 2: Shapes synthesized by 3D-GAN

Generative modeling 적용 사례

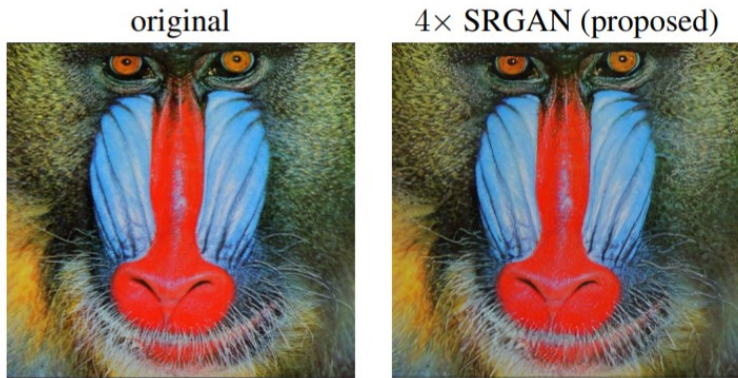


Figure 1: Super-resolved image (right) is almost indistinguishable from original (left) ($4\times$ downsampling).

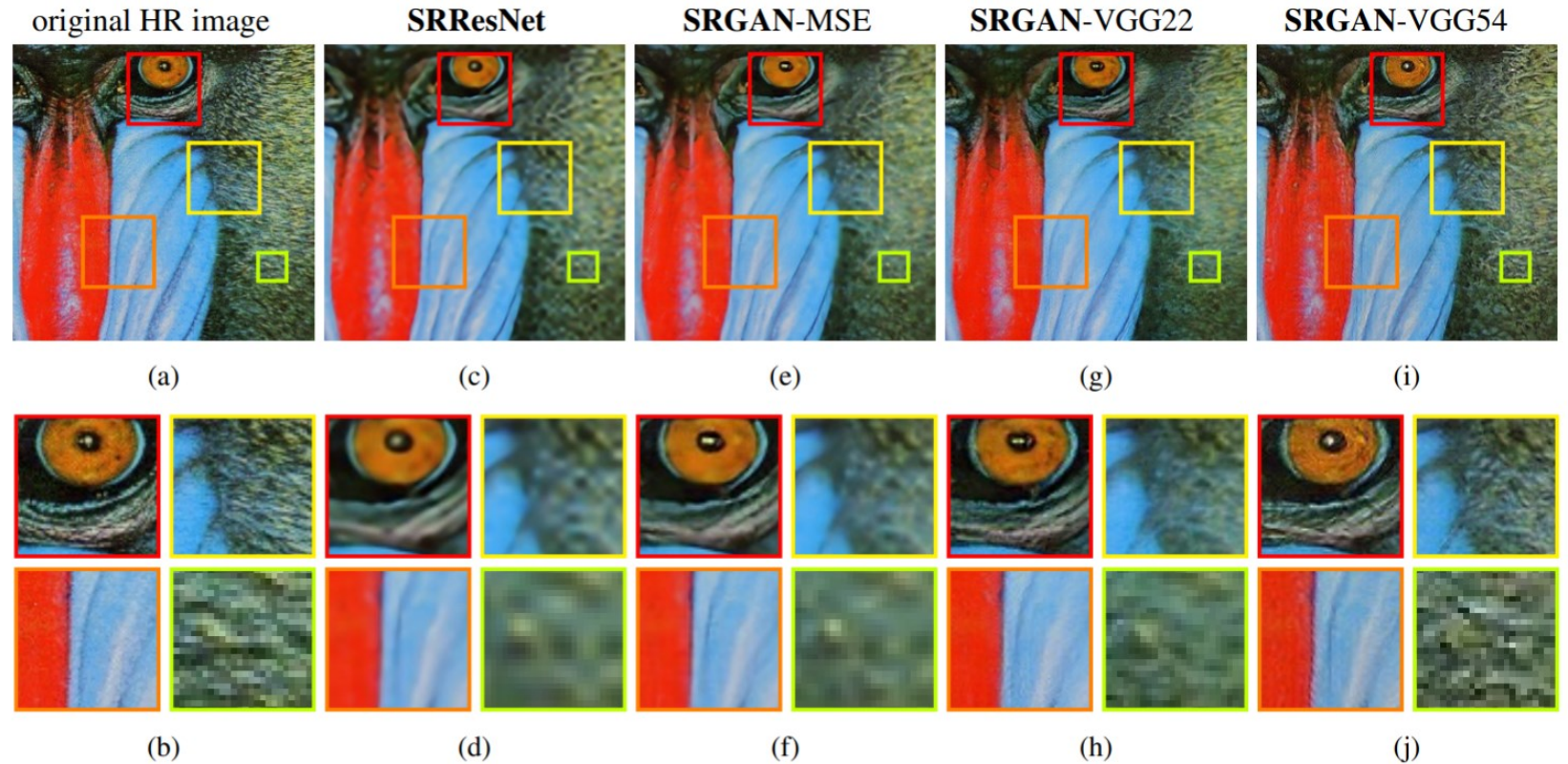


Figure 5: Reference HR image (left: a,b) with corresponding SRResNet (middle left: c,d), SRGAN-MSE (middle: e,f), SRGAN-VGG2.2 (middle right: g,h) and SRGAN-VGG54 (right: i,j) reconstruction results.

Hair Designer (Beauty) 응용 사례

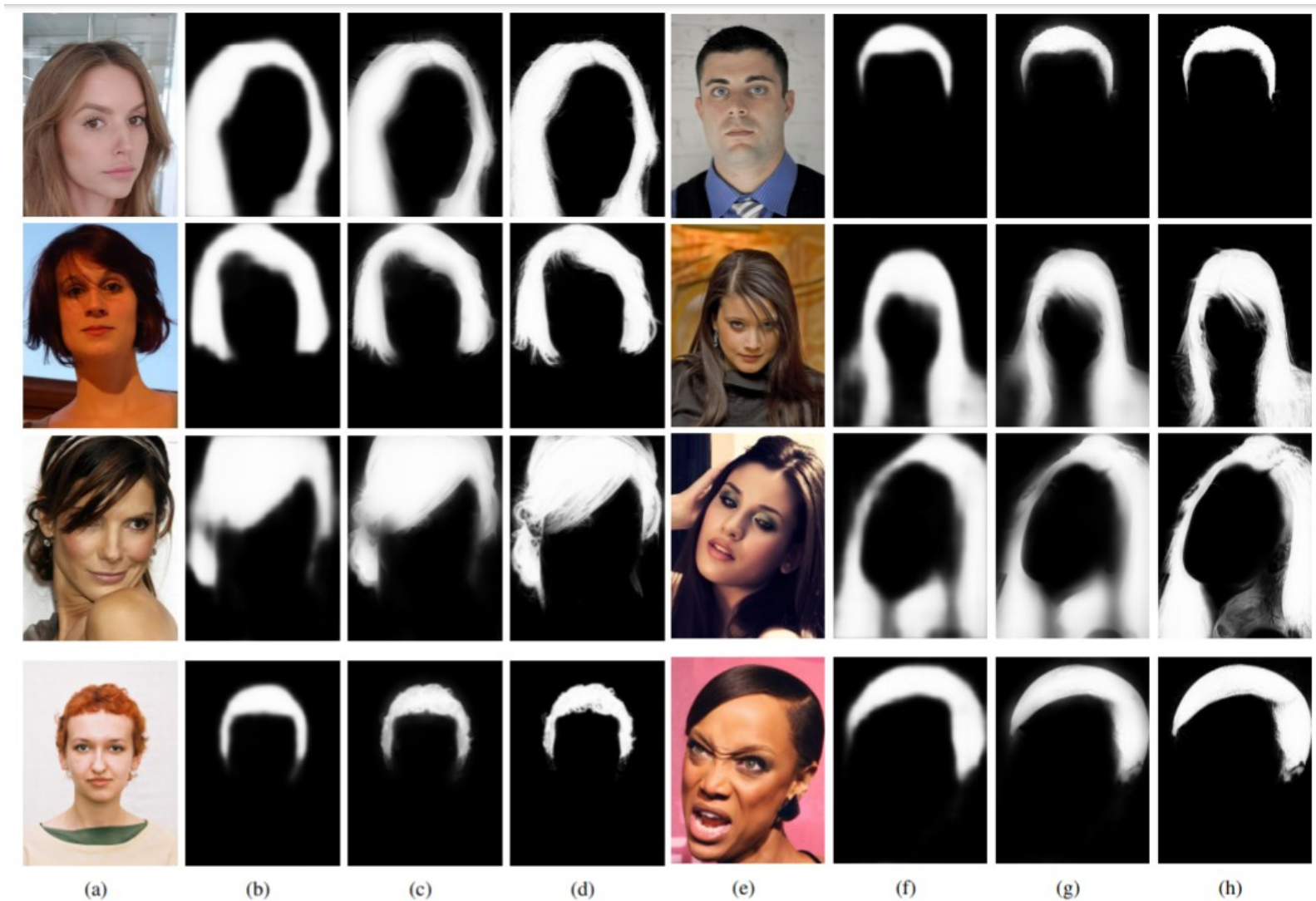


Figure 6. Qualitative evaluation. (a,e) Input image. (b,f) HairSegNet. (c,g) HairSegNet + Guided Filter. (d,h) HairMatteNet.

Hair Designer (Beauty) 전용 CNN (Transfer Learning)

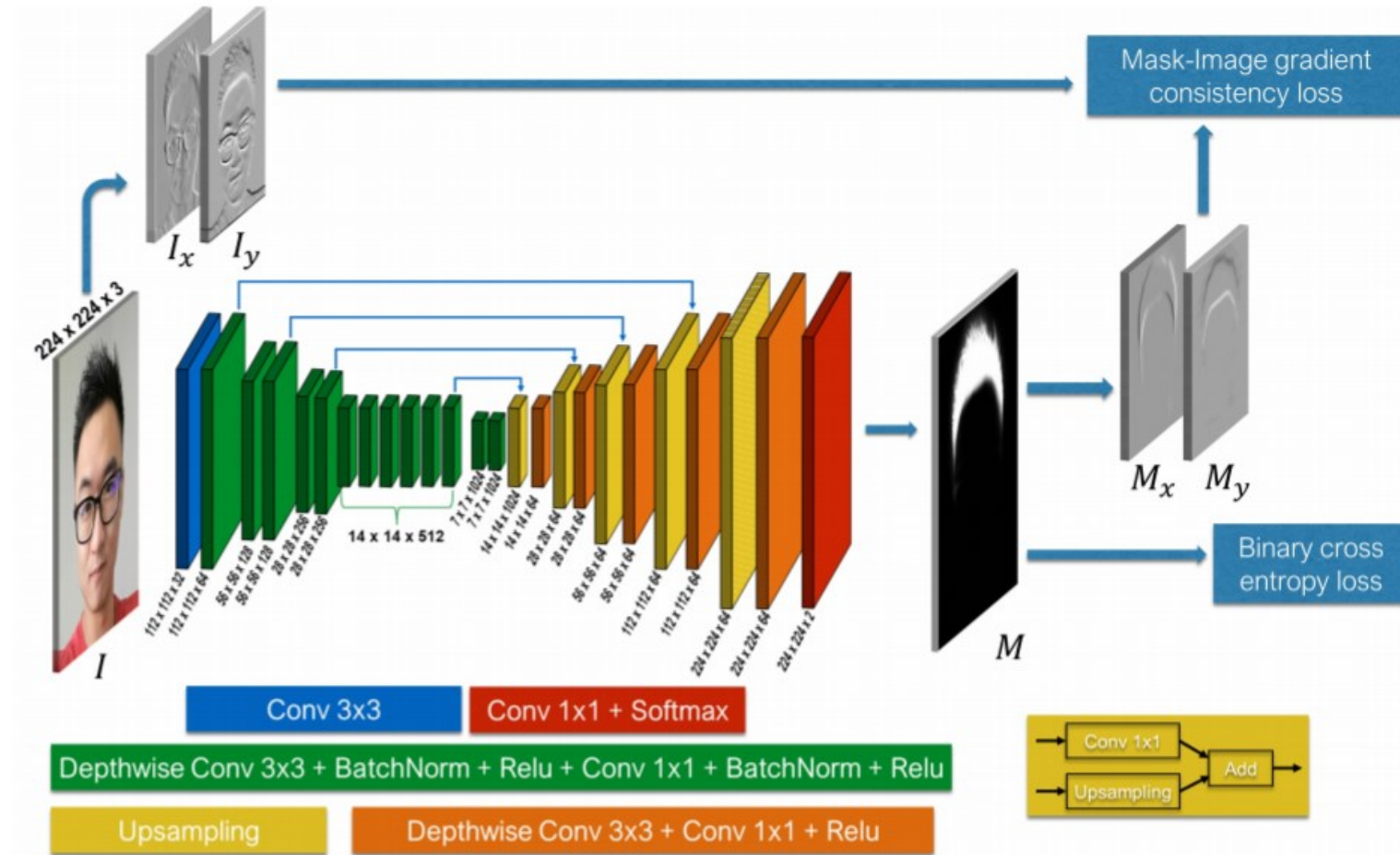


Figure 4. Fully Convolutional MobileNet Architecture for Hair Matting. Skip connections are added to increase the network capacity for capturing high resolution detail. Mask-image gradient consistency loss is added alongside with standard binary cross entropy loss to promote detailed matting results.

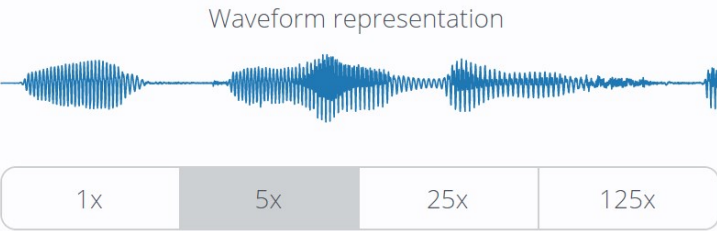
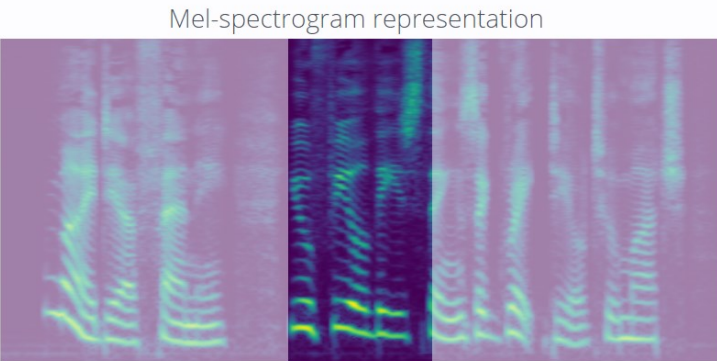
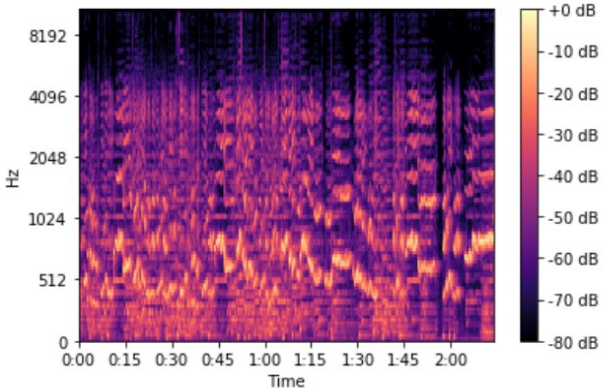


<https://www.youtube.com/watch?v=s9PQ7qPklUM>

The Mel Spectrogram

```
[9]: S = librosa.feature.melspectrogram(whale_song, sr=sr, n_fft=n_fft,
                                     hop_length=hop_length,
                                     n_mels=n_mels)

S_DB = librosa.power_to_db(S, ref=np.max)
librosa.display.specshow(S_DB, sr=sr, hop_length=hop_length,
                        x_axis='time', y_axis='mel');
plt.colorbar(format='%+2.0f dB');
```

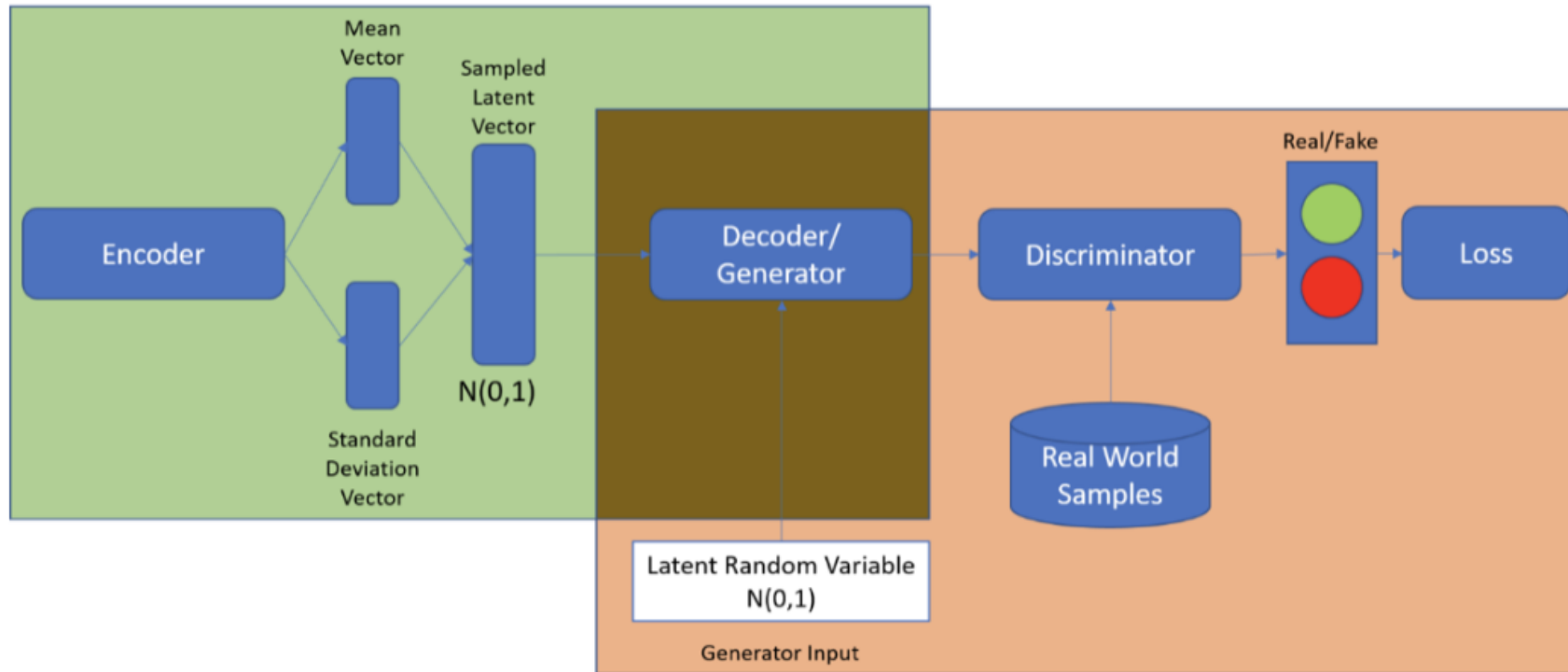


First rows

	0	1	10	100	101	102	103	104	105
0	0.071488	0.112931	0.093710	0.003623	0.002465	0.000812	0.001944	0.004218	0.004891
1	0.327453	0.696546	1.089640	0.010338	0.006793	0.005620	0.006476	0.013521	0.017855
2	0.310112	0.253482	0.210603	0.009782	0.007759	0.017301	0.011055	0.008583	0.010366
3	0.010208	0.109415	0.129118	0.020580	0.003772	0.006510	0.011296	0.012814	0.016841
4	0.053919	0.066492	0.029974	0.007042	0.011837	0.012552	0.007069	0.012861	0.014812
5	0.254680	0.086527	0.016534	0.011019	0.005026	0.008153	0.011811	0.008895	0.008429
6	0.208996	0.060032	0.019137	0.016377	0.009610	0.007126	0.014477	0.010154	0.013721
7	0.392639	0.159828	0.190023	0.123125	0.096298	0.053775	0.071976	0.059147	0.069463
8	0.320323	0.092376	0.031002	0.028073	0.022313	0.012524	0.011378	0.007347	0.061672
9	0.108713	0.042663	0.011409	0.142361	0.072162	0.047233	0.037520	0.051558	0.181629

MelNet combines various representational and modelling improvements to yield a highly expressive, broadly applicable, and fully end-to-end generative model of audio.

Variational Autoencoder



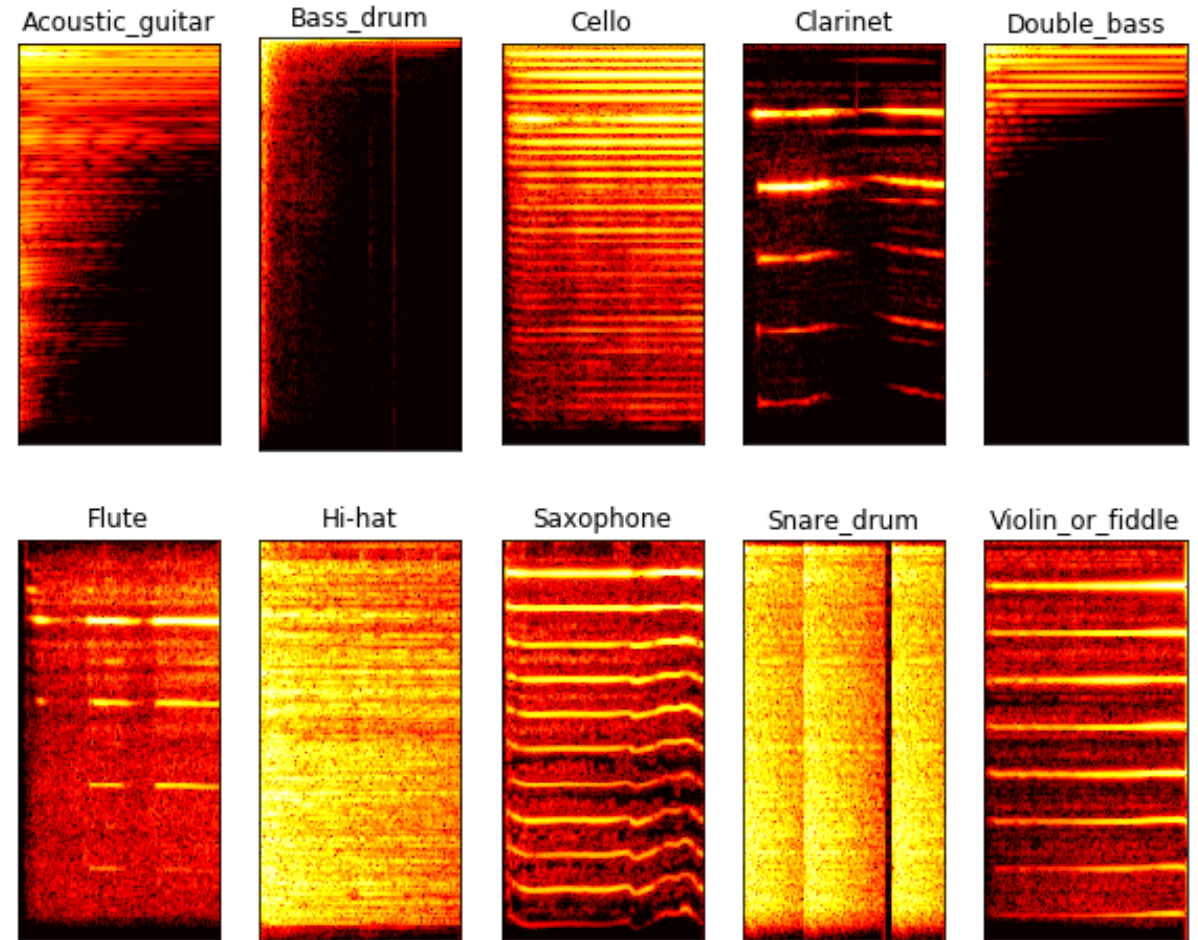
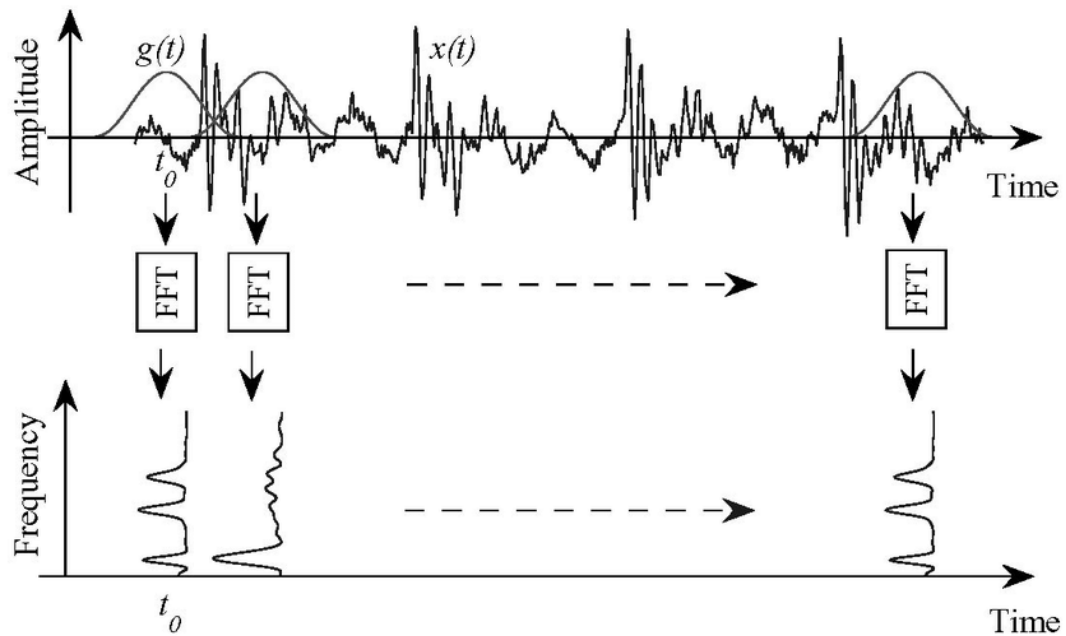
Generative Adversarial Network



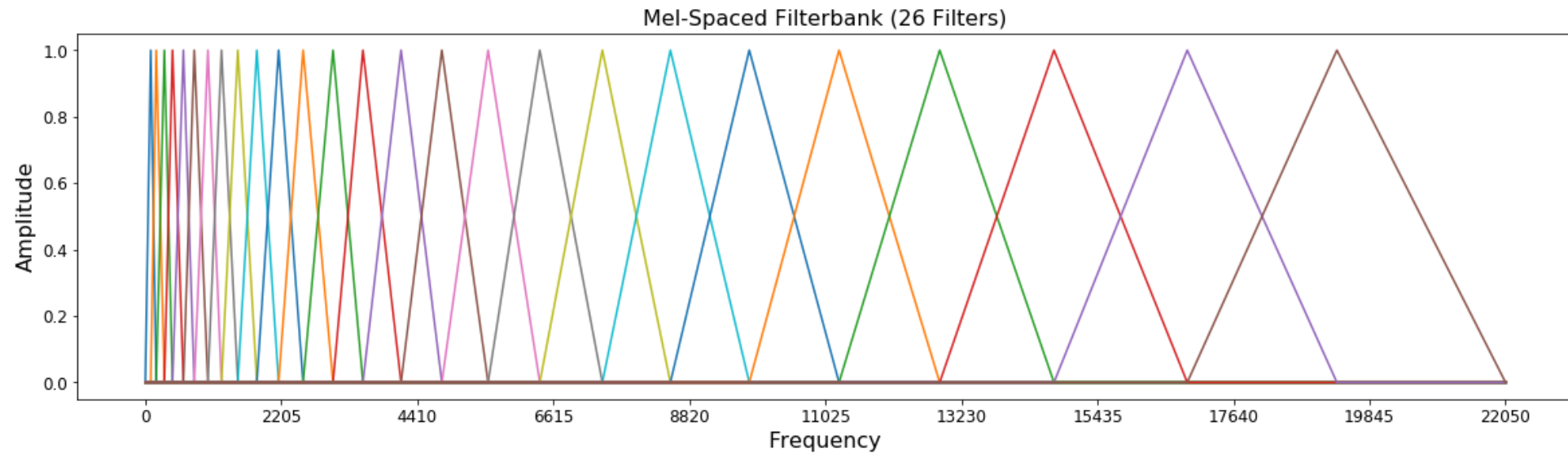
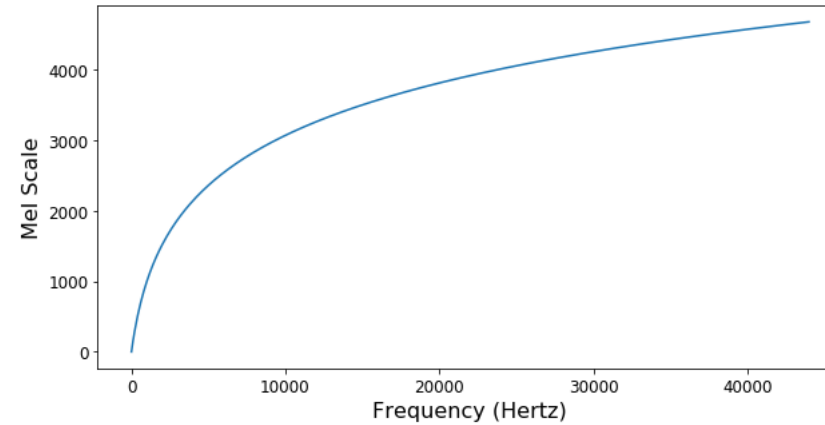
Jackie-Evancho(88sec).midi

Short-Time Fourier Transform

Sampling Rate = 16 kHz
Window Length = 25 ms = 400 samples
Step Size = 10 ms = 160 samples
N FFT = 512 samples



Mel Filterbank





에어버스 [A-380](#) 은 상당한 양의 소프트웨어를 사용하여 "종이 없는" 조종석을 창조하였다. 소프트웨어 공학으로 항공기 소프트웨어를 이루는 수백만행의 소스코드를 변환하고 계획하였다. 20

정 준 수 Ph.D.

heinem@naver.com

010-5359-3644



소프트웨어를 아는 자가 미래를 연다!