Over-sampling in a Deep Neural Network

Andrew J.R. Simpson, arXiv, 2015

강혁

Bigger network work better?

- A key factor in the success of the DNN is scalability.
- The reason for this scalability is not yet well understood.
- the DNN as a discrete system, of linear filters followed by nonlinear activations, that is subject to the laws of sampling theory.

Sampling theory

Sampling

신호 처리에서 표본화(標本化) 또는 샘플링(sampling)은 연속 신호(유동적인 신호)를 이산 신호(수치화된 신호)로 감소시키는 것을 말한다. 이를테면 파동 (연속 시간 신호)을 일련의 표본(이산 시간 신호)으로 바꾸는 것을 들 수 있다(Wikipedia)

Sampling frequency

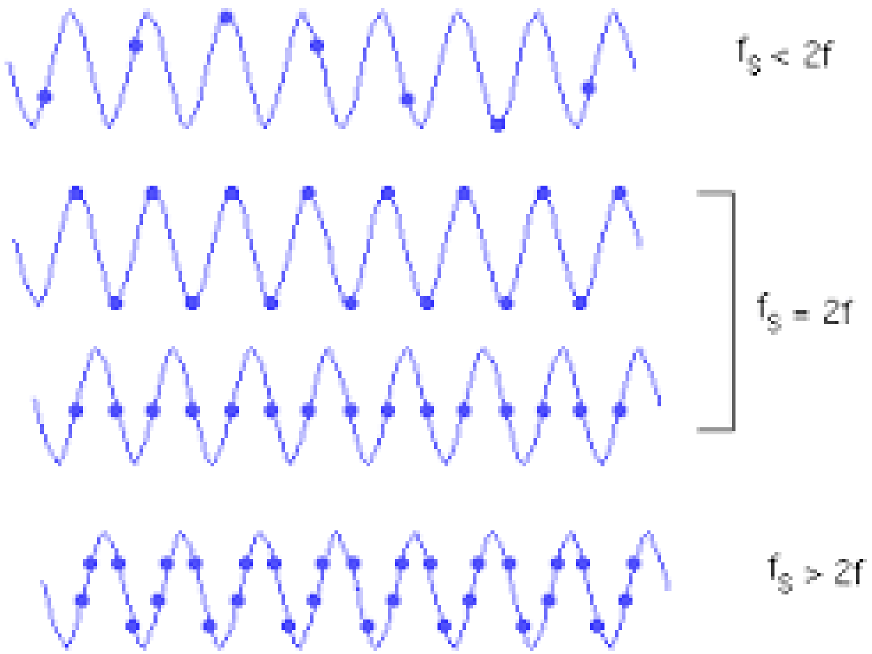
시간(초) 당 얻어지는 sample의 평균 개수(Wikipedia)

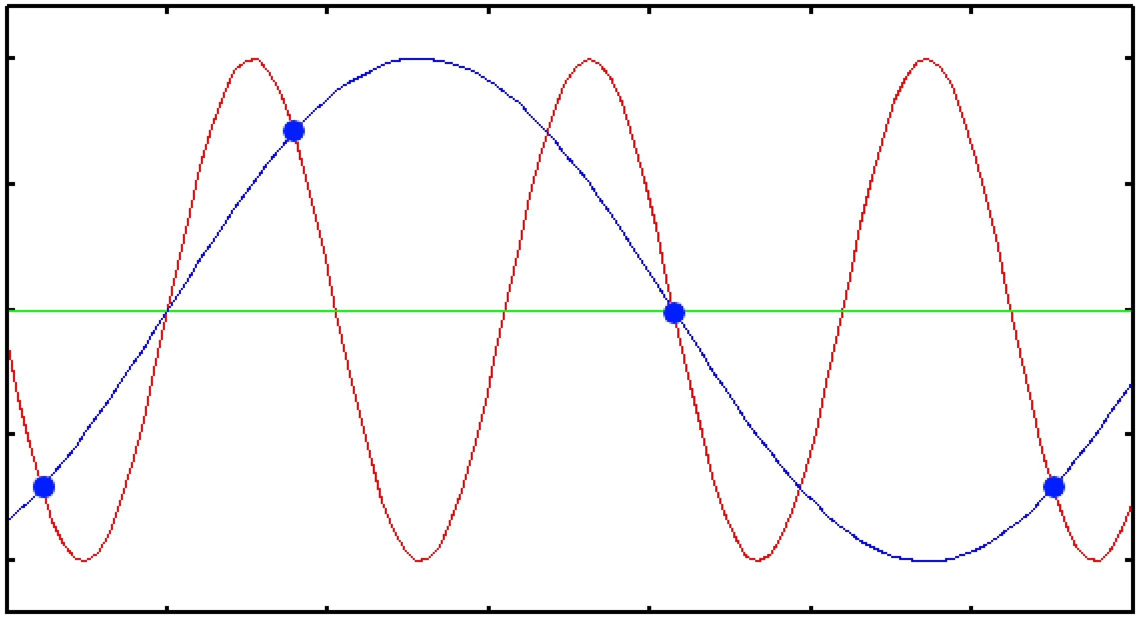
Over-sampling

신호 처리에서 오버샘플링(oversampling)은 두 배 이상의 대역폭, 또는 샘플링할 수 있는 최고의 샘플링 주파수로 신호를 샘플링하는 과정이다.
오버샘플링은 에일리어싱 방지, 해상도 향상, 노이즈감소에 효과적이다. (Wikipedia)

Nyquist frequency

샘플링 하려는 주파수보다 2배 높은 주파수로 샘플링해야만 정확하게 신호를 복원할 수 있다





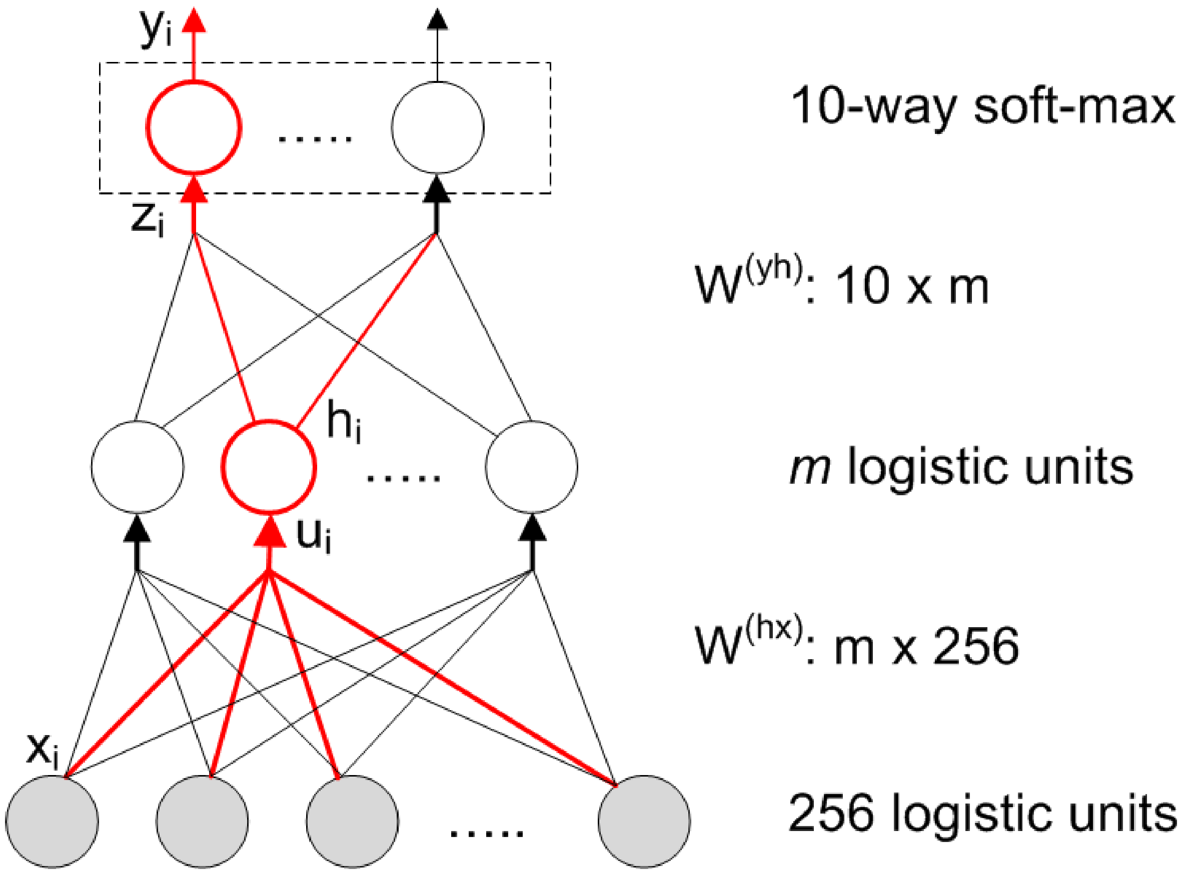
https://svi.nl/AliasingArtifacts

https://helpx.adobe.com/photoshop-elements/key-concepts/aliasing-anti-aliasing.html

Wider network

1) Greater bandwidth of input data may be represented
2) Higher-order filter may be learned
3) Less aliasing of high order distortion products

10-unit softmax output layer



Decimated image

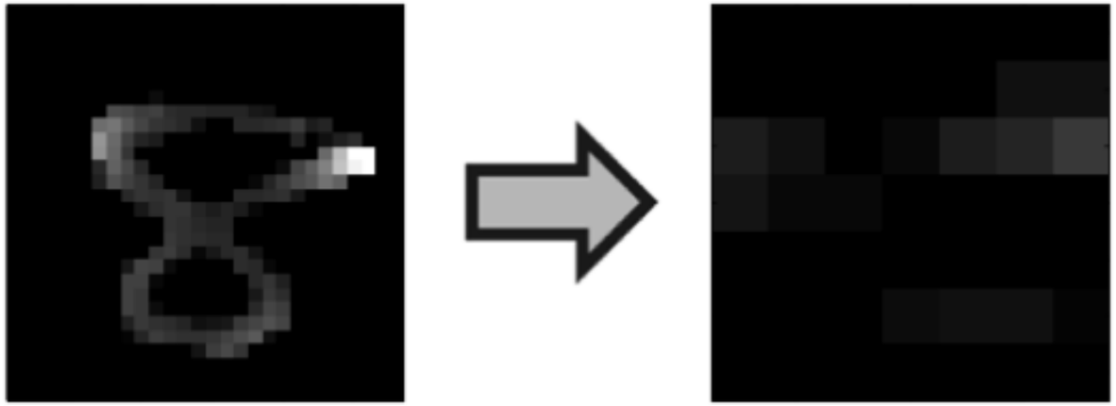
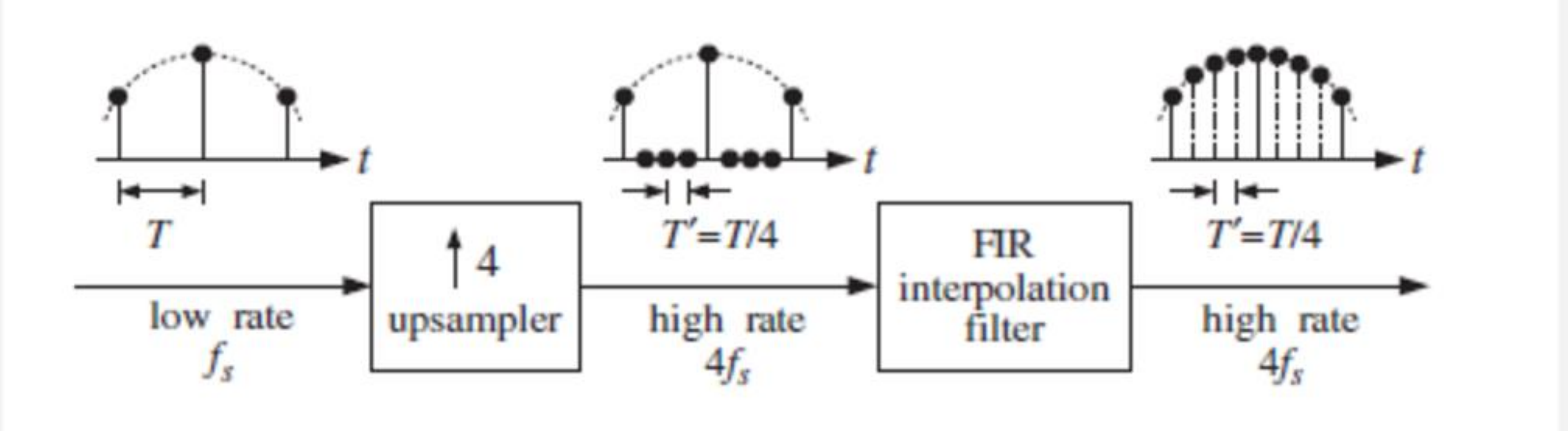


Fig. 1. Example MNIST image decimated by factor of 16. We took the 28x28 pixel images, unpacked them into a vector and decimated the vectors by a factor of 16, yielding an effective 7x7 pixel representation (represented here, for illustration, as a matrix re-wrapped from the vector for illustration).

Interpolation filter



introduction to signal processing sophocles j. orfanidis

Transfer function

$$X(s) = \mathcal{L}\left\{x(t)\right\} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x(t)e^{-st}\,dt,$$

$$Y(s) = \mathcal{L}\left\{y(t)\right\} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} y(t)e^{-st}\,dt.$$

Then the output is related to the input by the transfer function $H(s)$ as

$$Y(s) = H(s)X(s)$$

and the transfer function itself is therefore

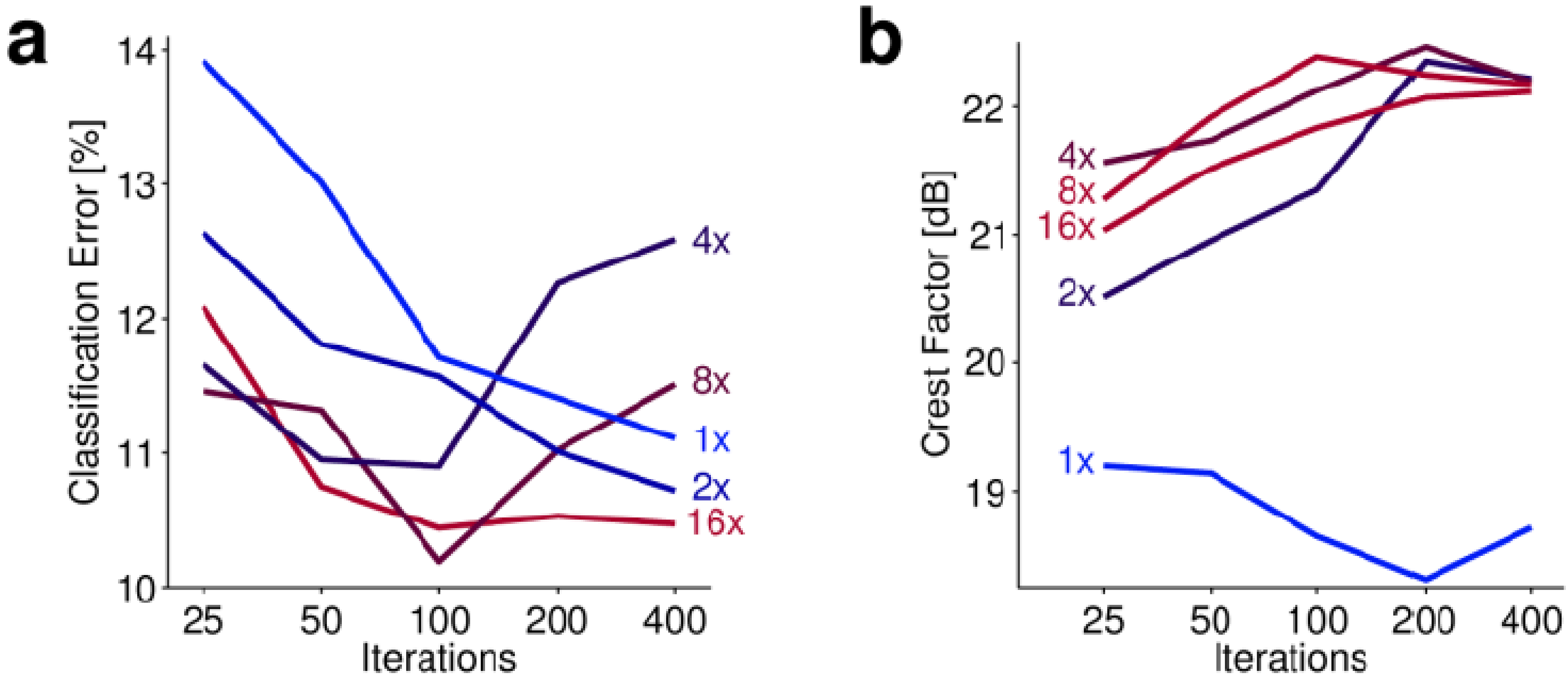
$$H(s) = \frac{Y(s)}{X(s)}.$$

Gain of transfer function

$$g = 10log10\left(\frac{max(H)}{\min(H)}\right)$$

$$\mathrm{SNR_{dB}} = 10\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right).$$

Result



a
14
13
12
11
10
Classification Error [%]
4x
8x
1x
2x
16x
25
50
100
200
400
Iterations
b
22
21
20
19
Crest Factor [dB]
4x
8x
16x
2x
1x
25
50
100
200
400
Iterations

Discussion

- Learning rate scales with over-sampling rate
- Onset of over-fitting is delayed longer at higher degrees of over-sampling

Discussion

- Over-sampling may be a useful alternative to regularization by dropout
- Results suggest that the over-fitting demonstrated here is at least partly the result of aliasing, rather than rare feature coincidences in the data

Discussion

- aliasing provides a very abrupt and arbitrary projection of high order distortion products back into the Nyquist band.
- For example, consider energy near 10x the Nyquist limit. If this energy shifts in frequency a small amount, its aliased location in the Nyquist range shifts a very large amount.