VQVAE

2019/11/15

Allen

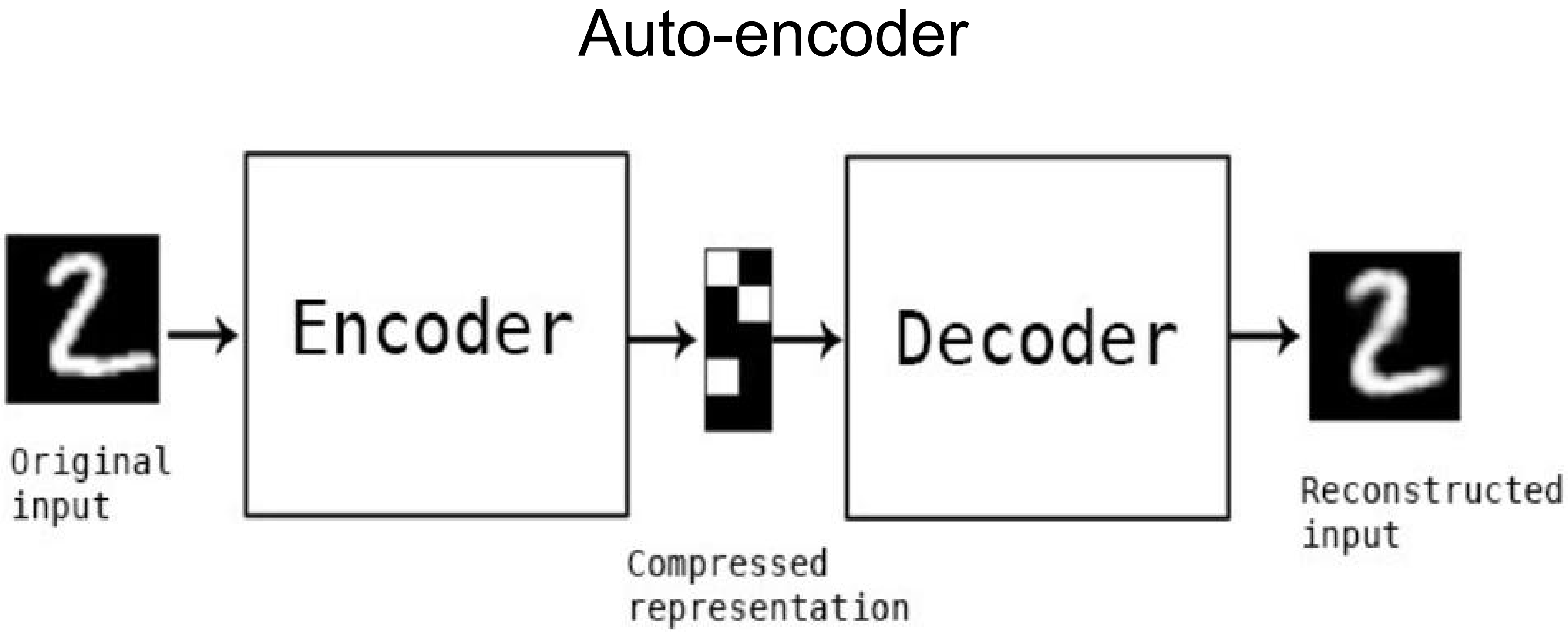
大綱

VQVAE? 設計原理?

VQVAE 架構

VQVAE 2 架構

AE ?



continous latent variable

VQVAE = VAE + discrete latent variable

Why discrete latent variable?

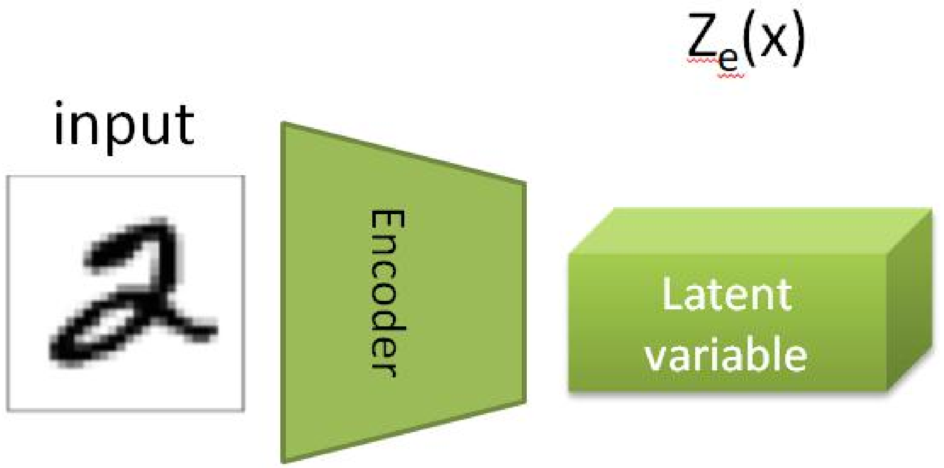
- 現實有很多資料都是屬於discrete的特性(words, phonemes(音階))
- 將資料embedding並且映射到discrete latent space

Neural Discrete Representation Learning
https://arxiv.org/abs/1711.00937

VQVAE 架構

Forward

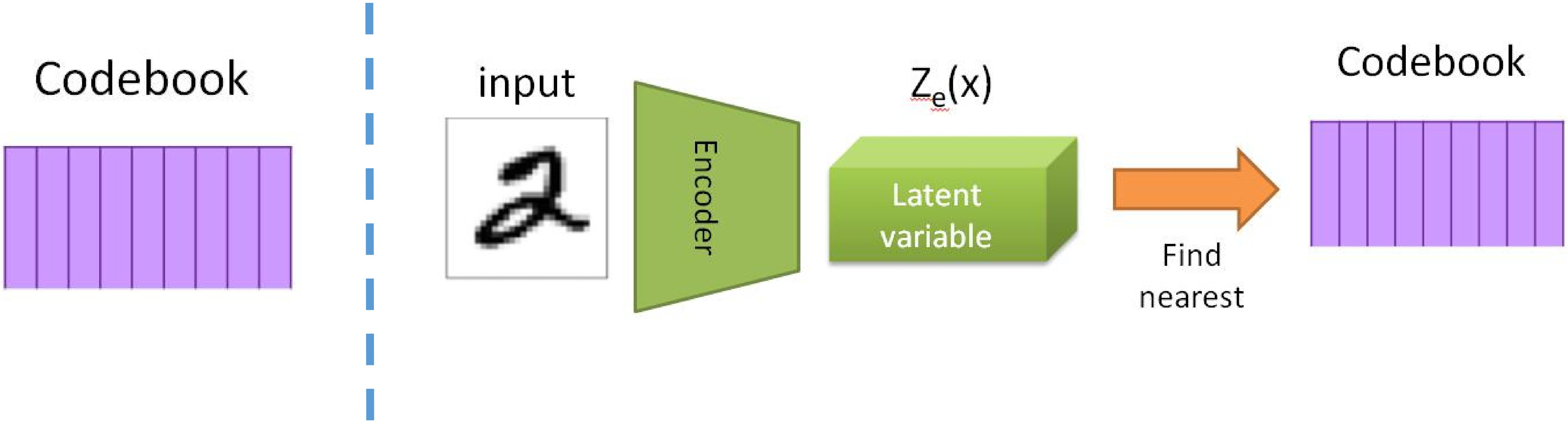
1. input => encoder => latent variable



VQVAE 架構

Forward

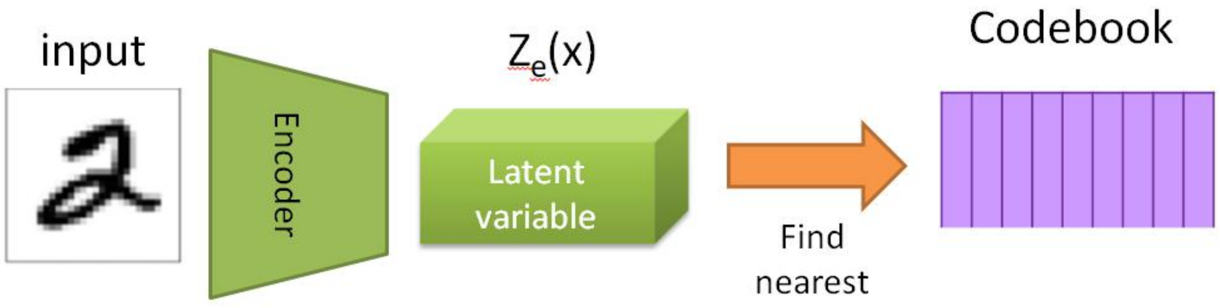
1. input => encoder => latent variable
2. 製作codebook embedding vector，並且計算與latent variable最近距離

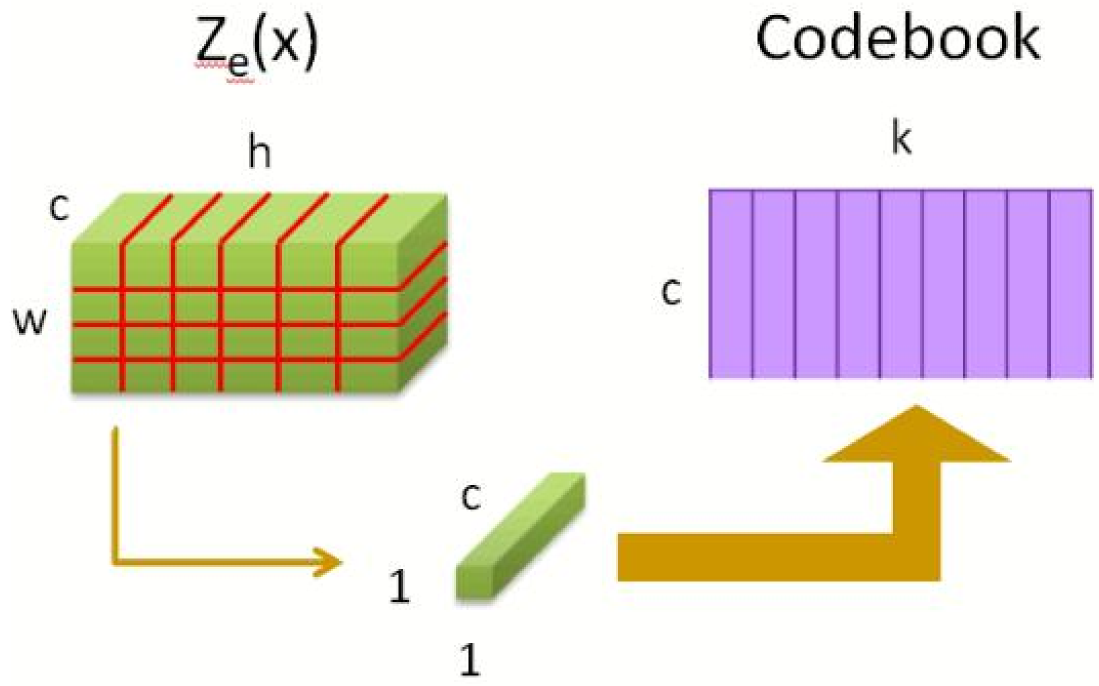


VQVAE 架構

Forward

1. input => encoder => latent variable
2. 製作codebook embedding vector，並且計算與latent variable最近距離

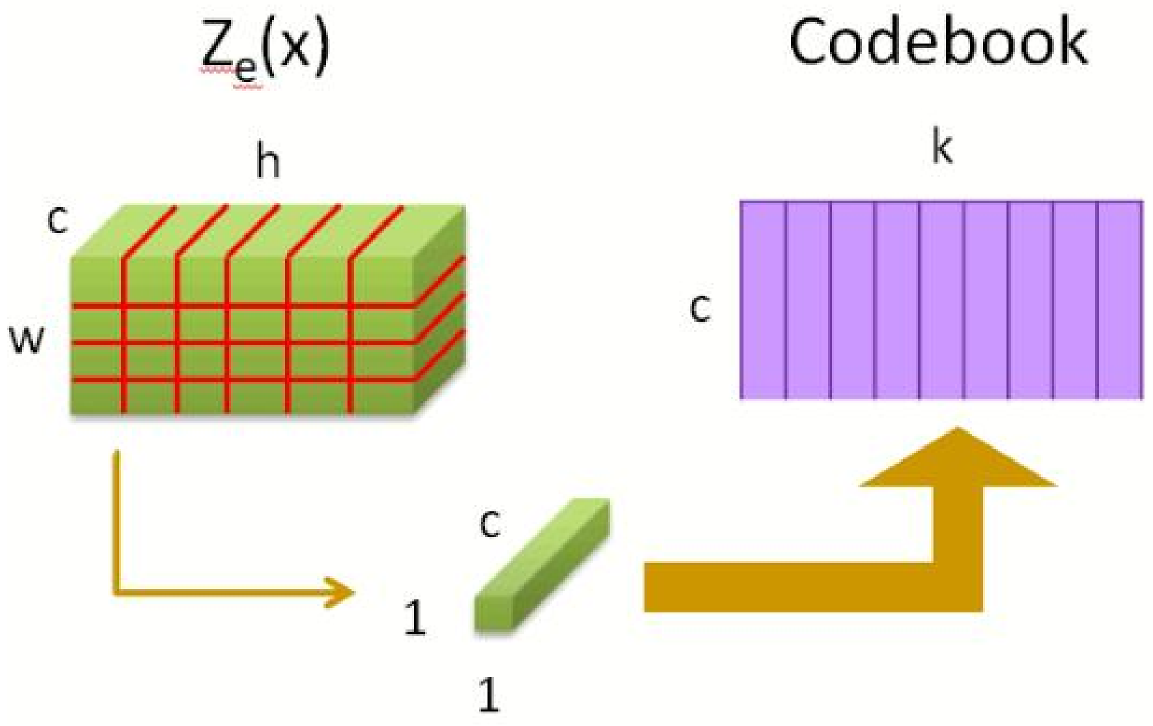




VQVAE 架構

Forward

1. input => encoder => latent variable
2. 製作codebook embedding vector，並且計算與latent variable最近距離

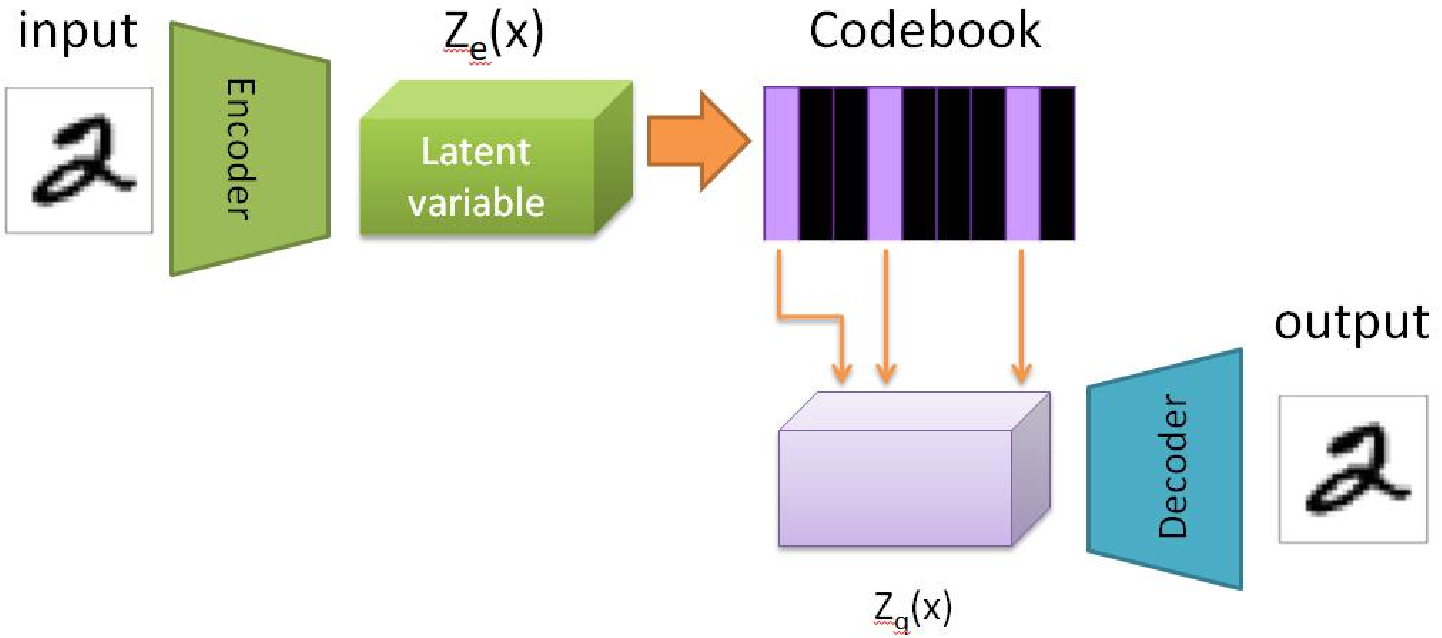


$$z_q(x) = e_k, \quad \text{where} \quad k = \text{argmin}_j \|z_e(x) - e_j\|_2$$

VQVAE 架構

Forward

1. input => encoder => latent variable
2. 製作codebook embedding vector，並且計算與latent variable最近距離
3. 做好的$Z_q(x)$就由decoder重建



VQVAE Loss function (for gradient)

(1) (2) (3)

$$L = \log p(x|z_q(x)) + \|\text{sg}[z_e(x)] - e\|_2^2 + \beta\|z_e(x) - \text{sg}[e]\|_2^2$$

x = input
e = codebook
sg = stop gradient
Beta = hyper parameter, 0.25~2

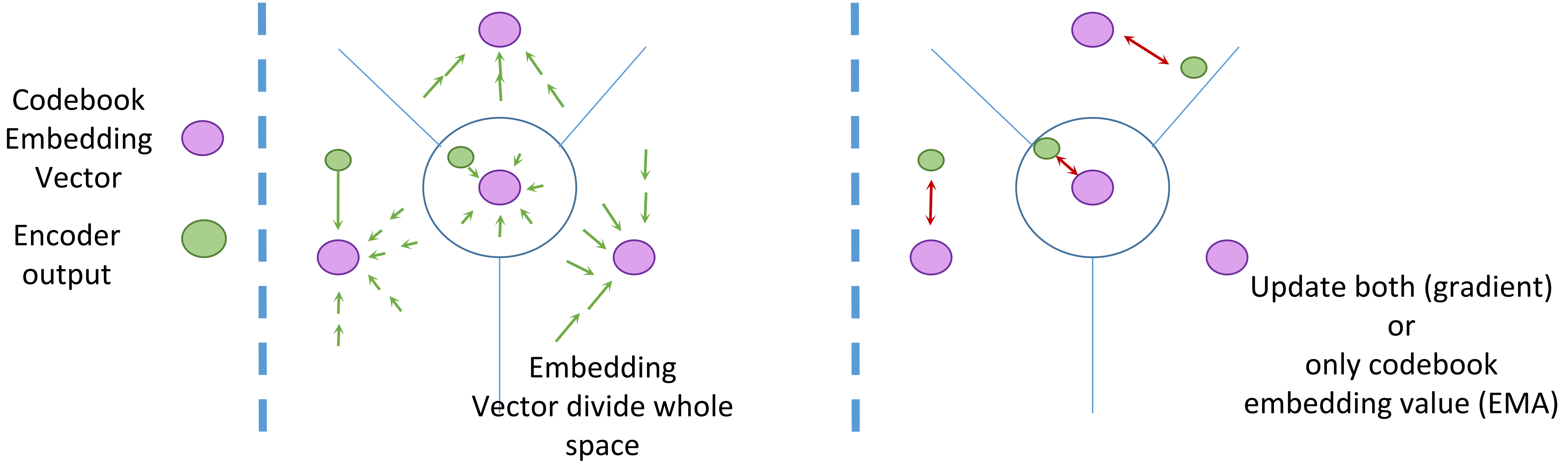
要怎麼做到梯度斷掉的情形下更新？

1. 拆計算圖再去更新
2. 利用stop gradient的特性，將前後接在一起

Zq = Ze + tf.stop_gradient(Zq - Ze)
forward: Zq = Ze + Zq - Ze →→ Zq = Zq
backward: Zq = Ze →→ 能將gradient傳遞到Ze

Vector Quantization變化

I. 透過比較 latent variable 和 codebook embedding vector 去尋找 representative embedding vector.
II. Move codebook to input value(optional: also input close to codebook value)



VQVAE Result from paper

原圖　　　　　　　　　　　　　　重建

Figure 2: Left: ImageNet 128x128x3 images, right: reconstructions from a VQ-VAE with a 32x32x1 latent space, with K=512.

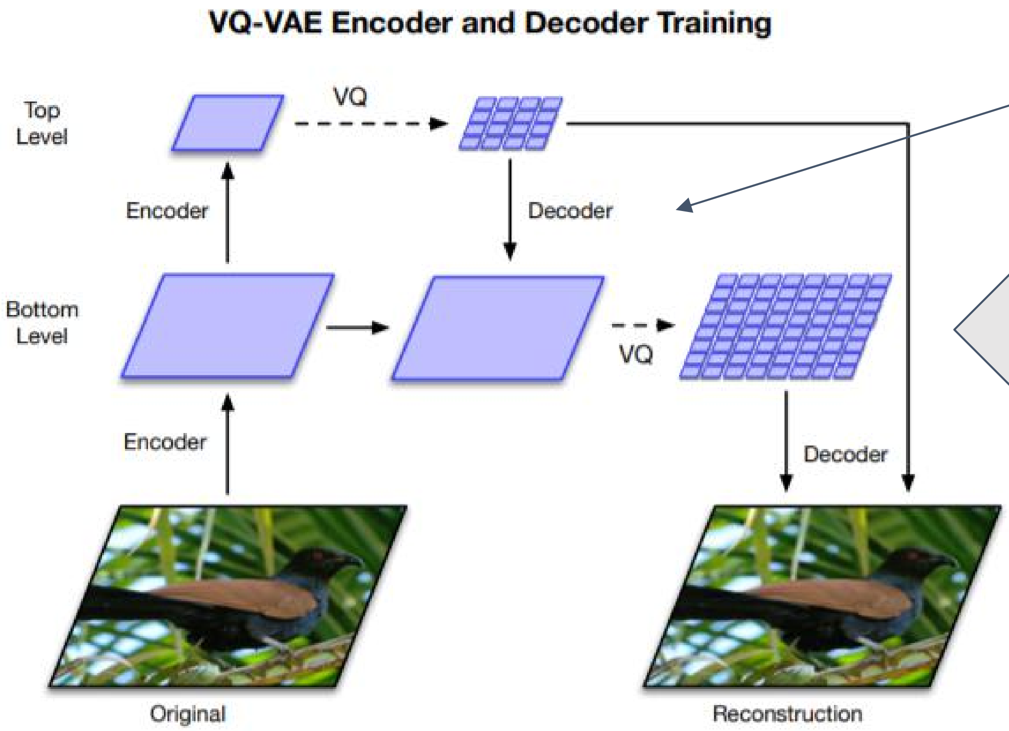
VQVAE 2

VQVAE2

- 概念:
  - 組合多個VQVAE
  - 不同Codebook對應不同尺度的特徵

top要作為
bottom的輸入

global
feature



local feature

Generating Diverse High-Fidelity Images with VQ-VAE-2
https://arxiv.org/abs/1906.00446

VQVAE2表現如何?

h_{top}　　$h_{\mathrm{top}}, h_{\mathrm{middle}}$　　$h_{\mathrm{top}}, h_{\mathrm{middle}}, h_{\mathrm{bottom}}$　　Original

- 和只有top(VQVAE1)比起來，使用多層會增加圖片的精細度

總結

Conclusion

VQVAE

- 1種discrete latent variable
- 組合多個VQVAE圖片可以做得更精細

reference

- Neural Discrete Representation Learning:

https://papers.nips.cc/paper/7210-neural-discrete-representation-learning.pdf
- Generating Diverse High-Fidelity Images with VQ-VAE-2:
https://arxiv.org/pdf/1906.00446.pdf
- Towards a better understanding of Vector Quantized Autoencoders
https://openreview.net/pdf?id=HkGGfhC5Y7
- Fast Decoding in Sequence Models Using Discrete Latent Variables:
https://arxiv.org/pdf/1803.03382.pdf