猫でも分かる
Variational AutoEncoder

2016/07/30

龍野　翔 (Sho Tatsuno)

Ishikawa Watanabe Lab
THE UNIVERSITY OF TOKYO
http://www.k2.t.u-tokyo.ac.jp/

今回の内容

- Variational Auto-Encoderの解説
  - 生成モデルそのものの概要
  - Variational Auto-Encoder(VAE)のなるべく噛み砕いた解説
  - その他生成モデル論文のざっくりした紹介

- 説明すること/しないこと
  - 説明すること
    - » 生成モデルの簡単な概要と事例
    - » Variational AutoEncoderの構造と数式的・直感的理解
  - 説明しないこと
    - » 生成モデルのその他のアルゴリズムの詳細(LDAとか)
    - » Deep Learningの基礎(Back Propagation・SGD等)
    - » 既存の最適化手法の詳細(MCMC・EMアルゴリズム等)

紹介元論文

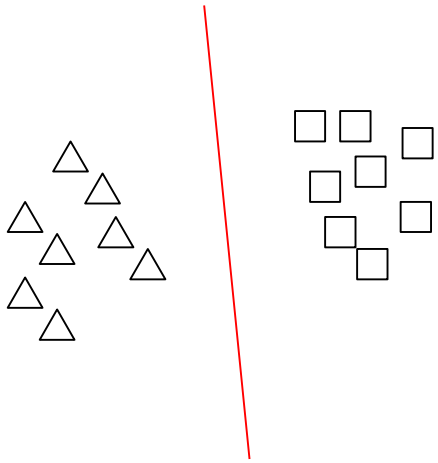
- Auto-Encoding Variational Bayes
  - Author: D. P. Kingma, 2013
  - URL: https://arxiv.org/pdf/1312.6114.pdf
  - Variational Auto-Encoderを最初に提唱した論文

- Tutorial on Variational Autoencoders
  - Author: Carl Doersch, 2016
  - URL: https://arxiv.org/abs/1606.05908
  - ニューラルネットによる生成モデルVariational Autoencoder(VAE)の紹介
    - » 変分ベイズの前提知識が不要
    - » 制約付きVAEであるConditional Variational Autoencoder(CVAE)についても紹介

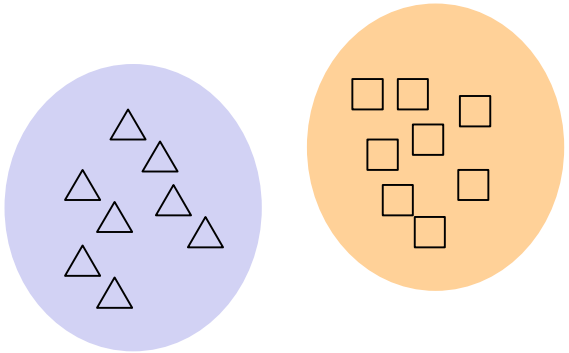
識別モデルと生成モデル

- 通常の機械学習は識別モデル
  - 各々を分けるための線を引く(識別する！)
- 生成モデルは識別(のみ)ではなく，範囲を考える

識別モデル

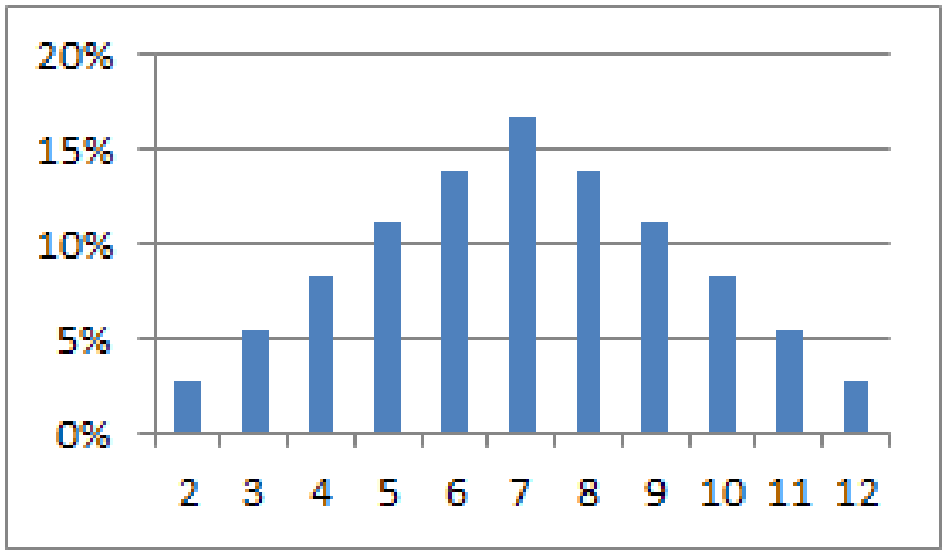


生成モデル



生成モデルって何？

- 生成モデル：観測データが得られる確率分布の推



サイコロを2回投げた時の
目の和の分布P(X)

↑

こいつを求めたい

↓ 実際にサイコロを2回振ってみた時

$P(X|\theta)$

分布を仮定->分布の裏側にあるパラメータθの最適化

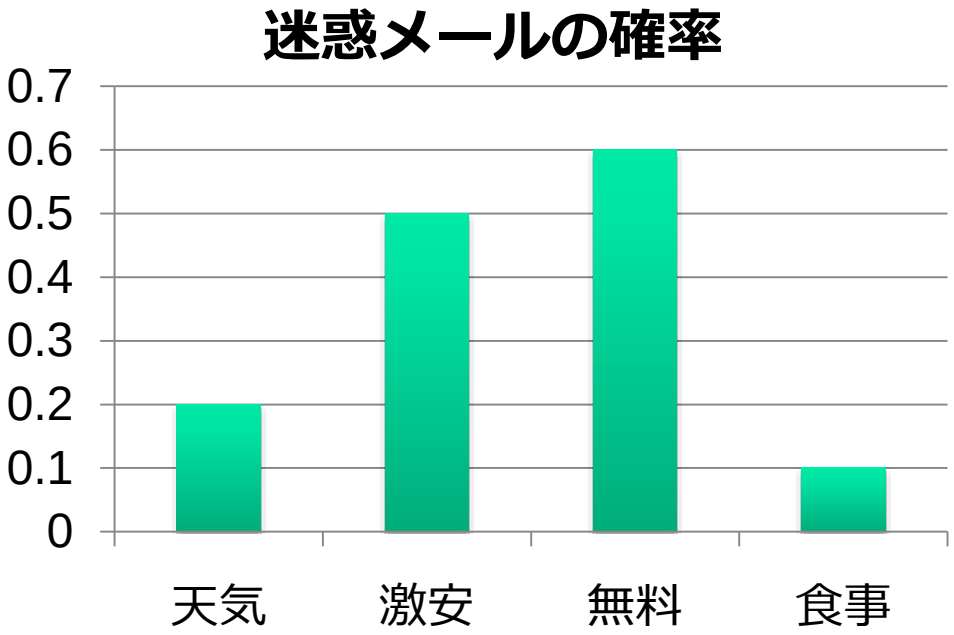
分布が分かると何が嬉しいのか

- 識別タスクなどに利用可能
  - Ex) Naive Bayes：迷惑メールの判別

各単語が迷惑メールに含まれる確率を元に迷惑メールの判別を行う

↓

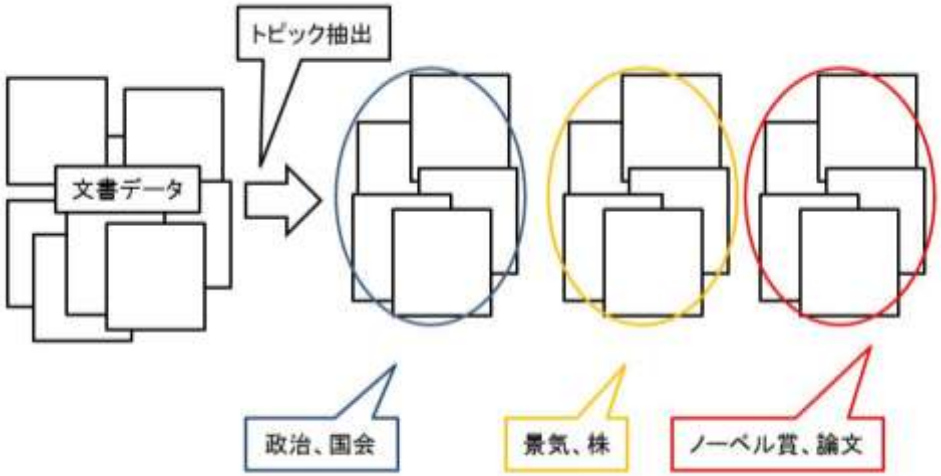
単語の「分布」をメールの「判別」に利用



既存の生成モデルの応用例

- 文章分類
  - NaiveBayes
    - » 迷惑メールかどうかの分類など
  - トピックモデル
    - » 文章のトピックを生成(政治の話・スポーツの話etc…)

- 異常値検出
  - ガウス混合分布
    - » 不正検出・侵入検出
  - 時系列モデル
    - » ウイルス・ワーム検出



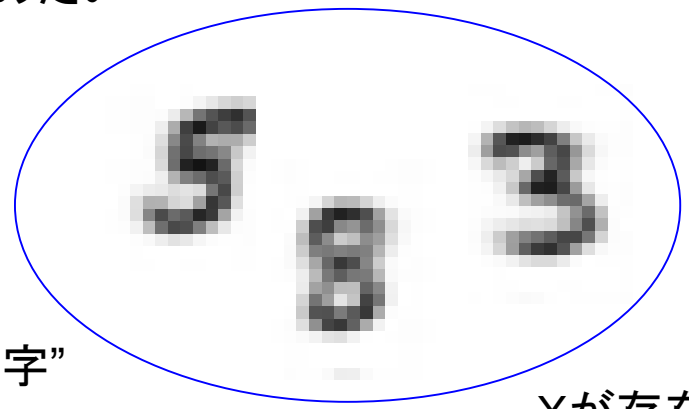
この辺の詳細説明は省略

画像の生成モデル

- 生成モデル
  - 画像の分布を求めたい
    - » データを元に未知のデータを作り出したい
    - » データが持つ抽象的な表現を捉えたい
      - 高次元なデータXが存在する確率分布P(X)を求めたい

画像らしい画像・文章らしい文章といえる部分を突き止めたい

例えば画像とか文章とか



画像の筆跡や数は違うが全て同じ”数字”

Xが存在しそうな領域

多様体論

- 下の二つの画像は色だけ見ると近しいが、明らかに違うものである
- 画像同士の間により低い次元での潜在的な意味が存在すると仮定する
  - ネコ・寿司

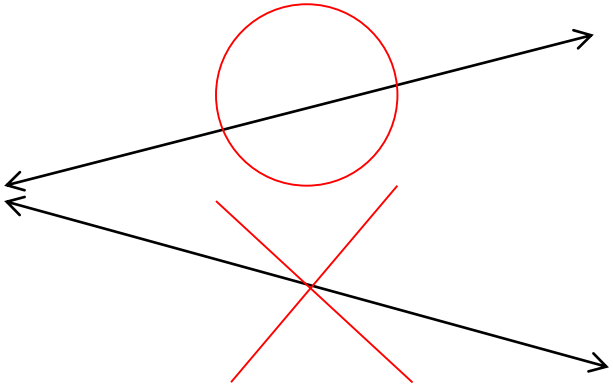
色だけ見ると識別困難

猫

寿司

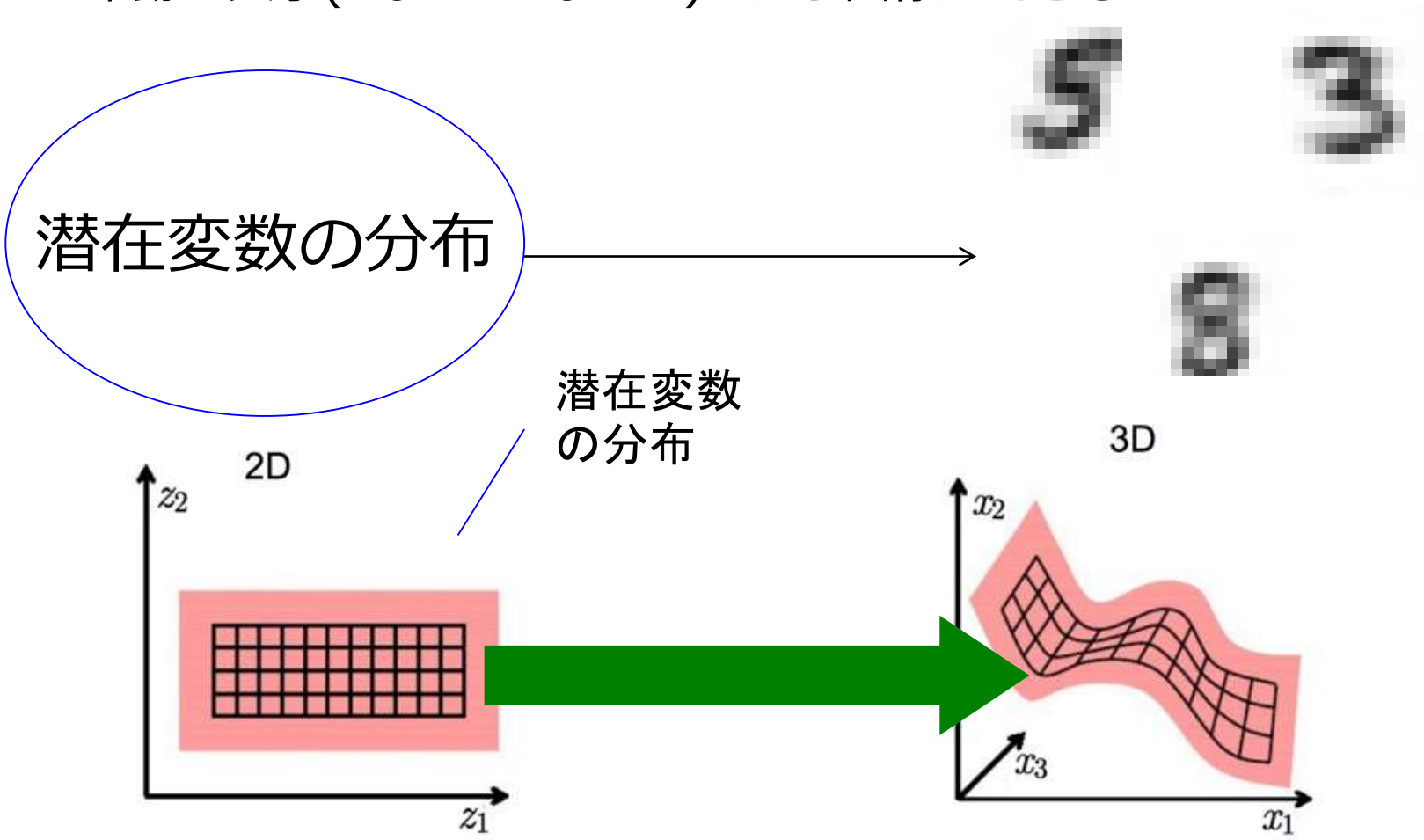
意味のある生成を

- 同じ意味の画像の認識
  - 画像間の潜在的な要素(潜在変数)を考える



潜在変数

- 例えば数字の潜在的な意味を考える
  - 筆跡？文字(3なのか5なのか)？から画像ができる

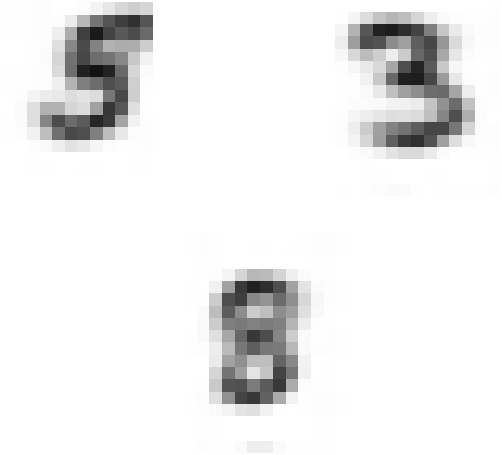


目的

- 画像の潜在空間の獲得と画像のバリエーション生成

顔の生成

表情・顔の造形が潜在変数？



数字の生成

文字・筆跡が潜在変数？

ここから先は数字を例にした解説を行う

既存の生成モデルの問題点

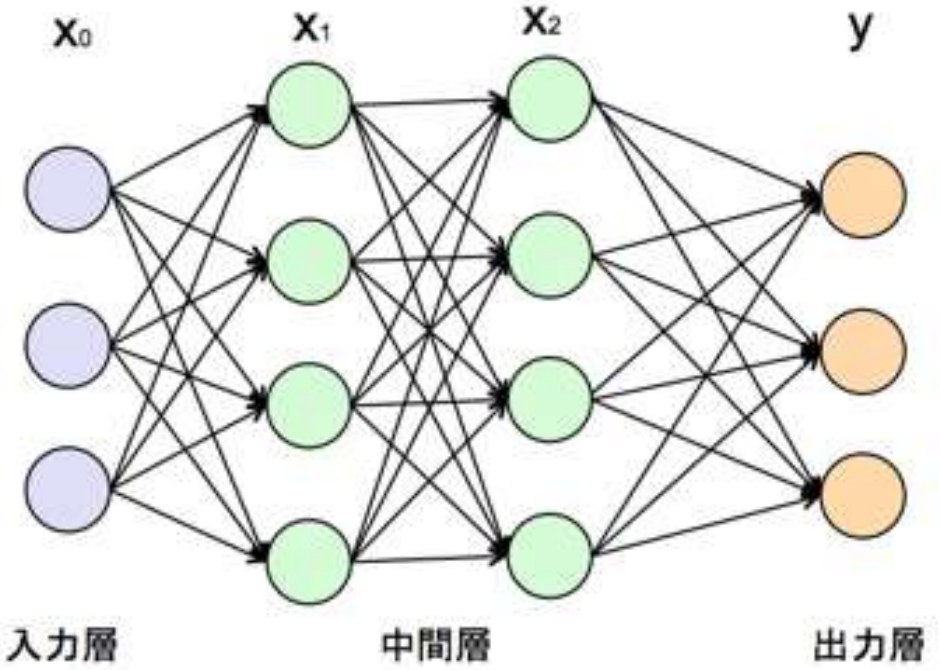
1. データ構造への強い仮定やモデルに強い近似が必要
   - こちらで何らかの分布を設定する必要性
   - 設定した分布にモデルが対応する必要性

2. 時間のかかる方法が必要
   - MCMC等、複数回サンプリングする必要性

詳細は省略

ニューラルネットの利用

- 単純なニューラルネットワークの例



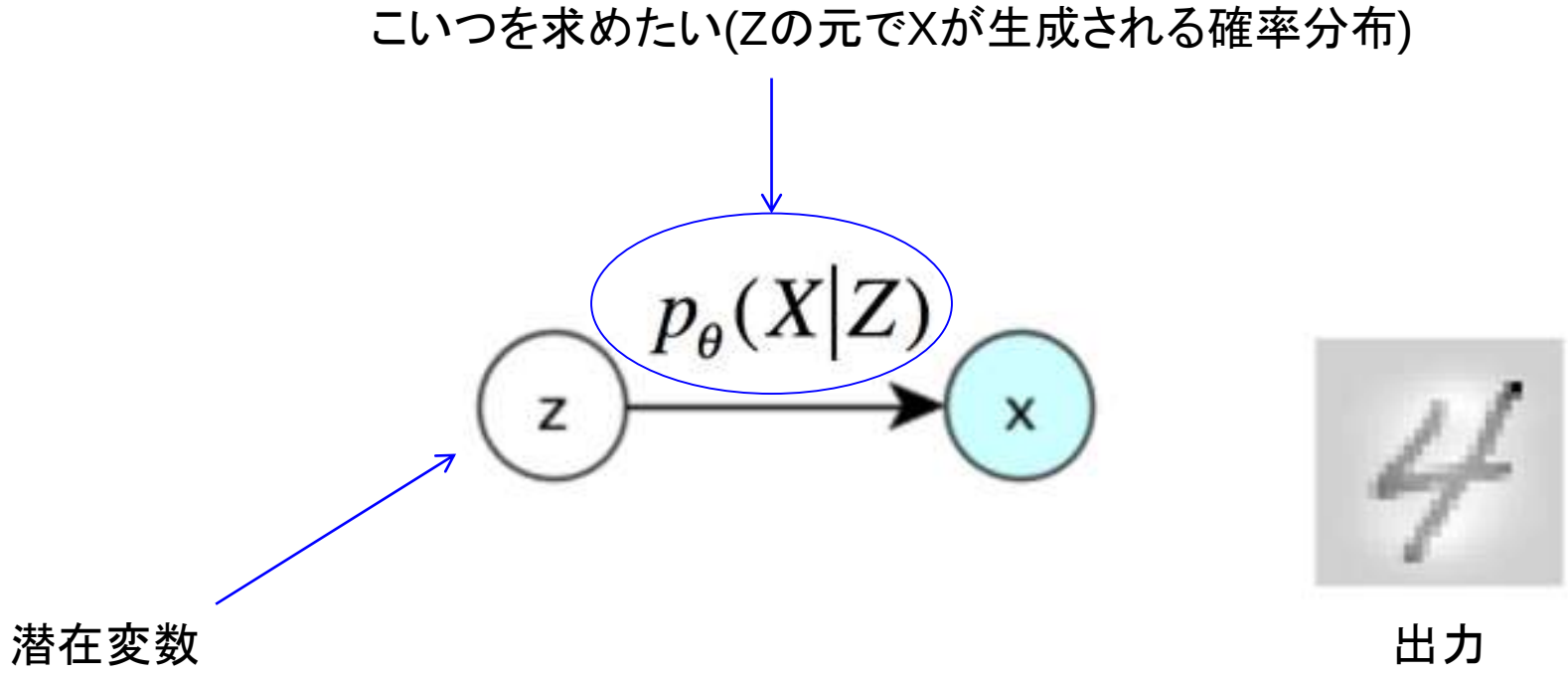
1. $y = f_2(w_2x_2+b_2)=f_2(w_2(f_1(w_1x_1+b_1))+b_2)=\cdots$

->畳み込みの形でほぼ任意の関数表現が可能:モデルの制約を緩和可能

2. SGDを使えば1サンプルずつ最適化が可能

生成モデル最適化の前提

- そもそも論

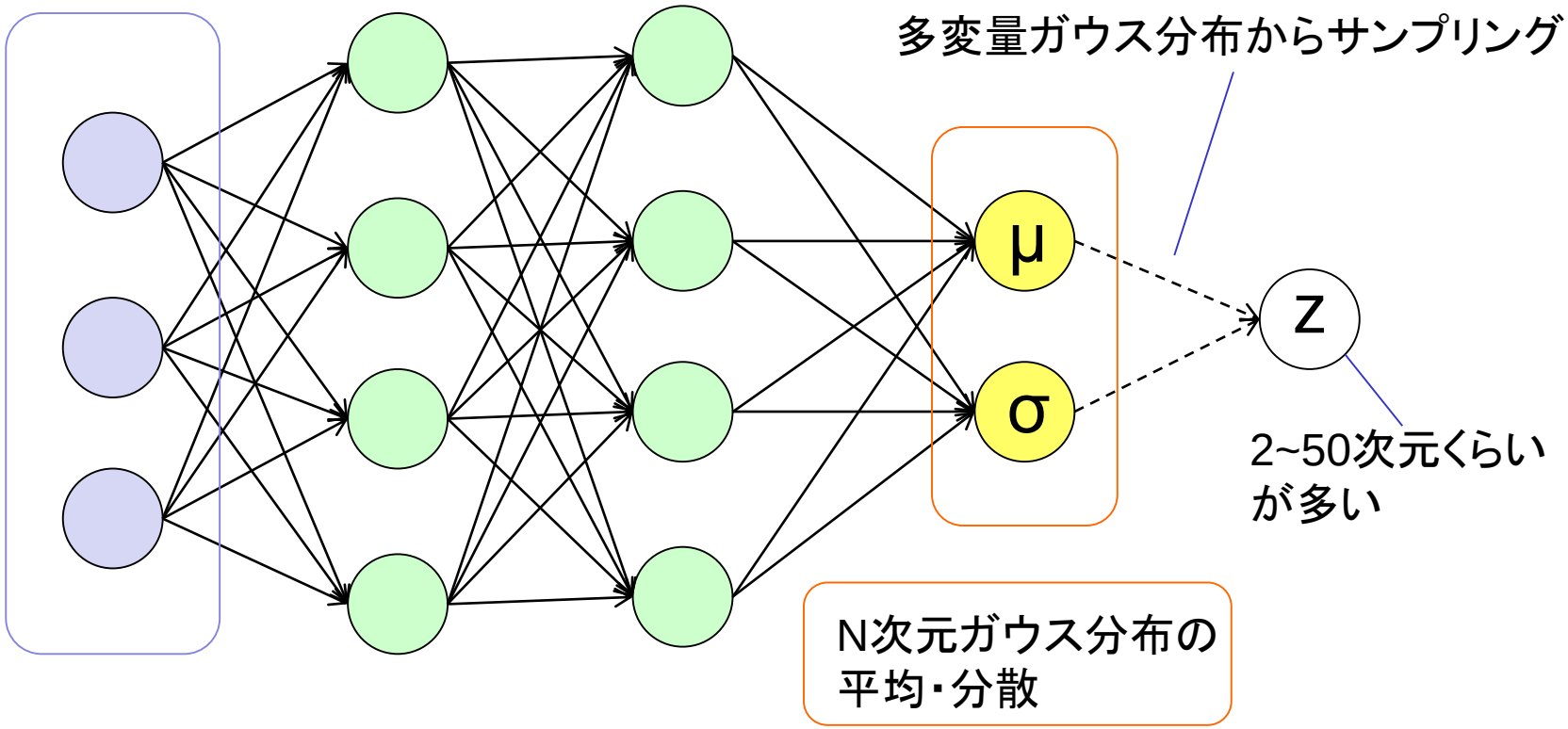


  - しかし,そのままp_θを求めるのは困難
    - » 入力(潜在変数z)に対応する答えが不明

入力から潜在変数zへの分布の仮定

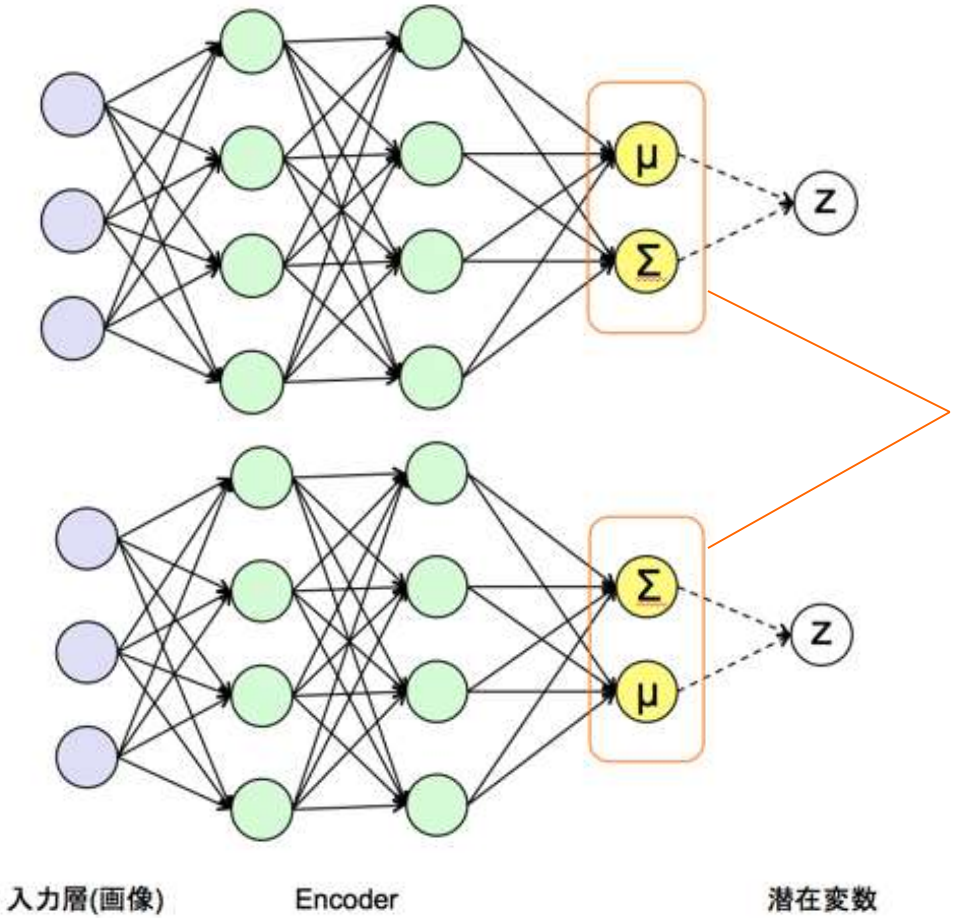
- Encoder
  - 潜在変数のガウス分布性を仮定する

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^m \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})\right)$$



分布の母数の妥当性

- μとΣの決め方に妥当性はあるのか
  - どちらがμ・Σでも良い：μとΣが最適となるように各層を最適化する

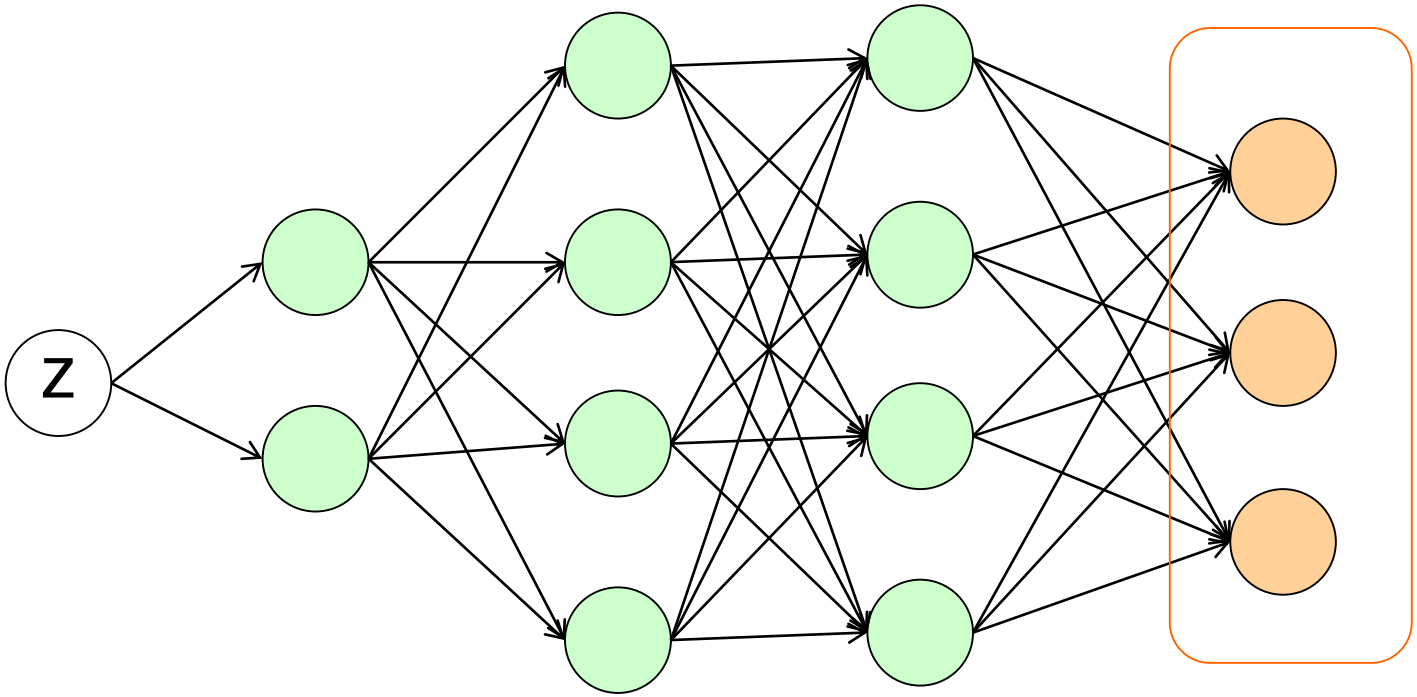


どちらでも大丈夫(予めμ・Σが定義づけられているわけではない)

後でμ・Σに対応するように学習させる

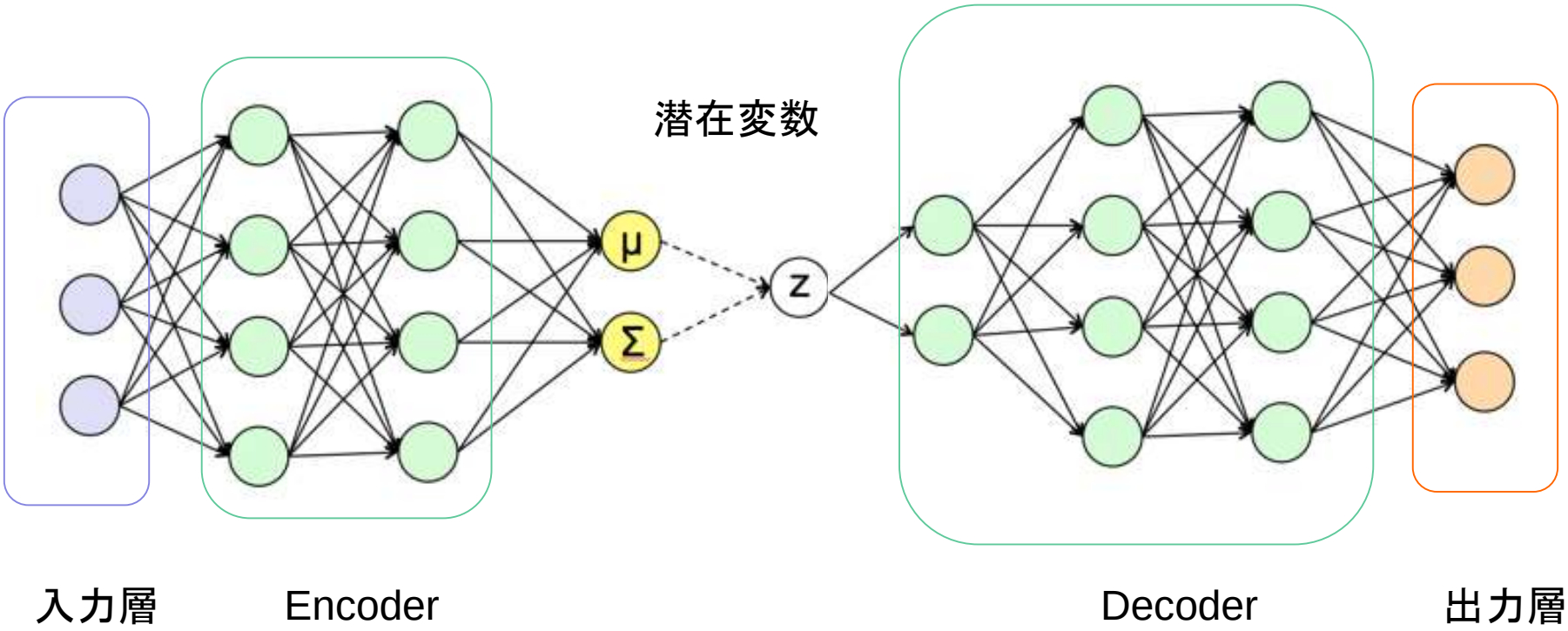
Variational AutoEncoder

- Decoder
  - こちらはzから出力層までにニューラルネットを組めばOK



Variational AutoEncoder

- Total Structure

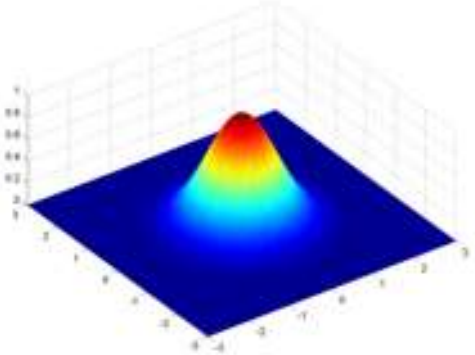


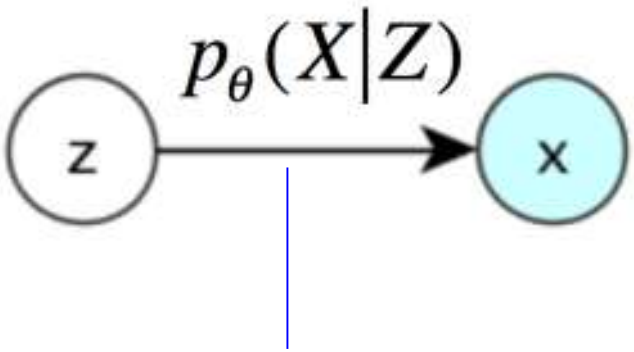
潜在変数の仮定

- 潜在変数は多次元のガウス分布を仮定
  - 扱いやすいから
  - 今回の場合、潜在変数に文字の筆跡や形を想定->ガウス分布？

・z ~ p(z)：pの事前分布として簡単な形(今回は多項標準ガウス分布)を考える

潜在変数の分布



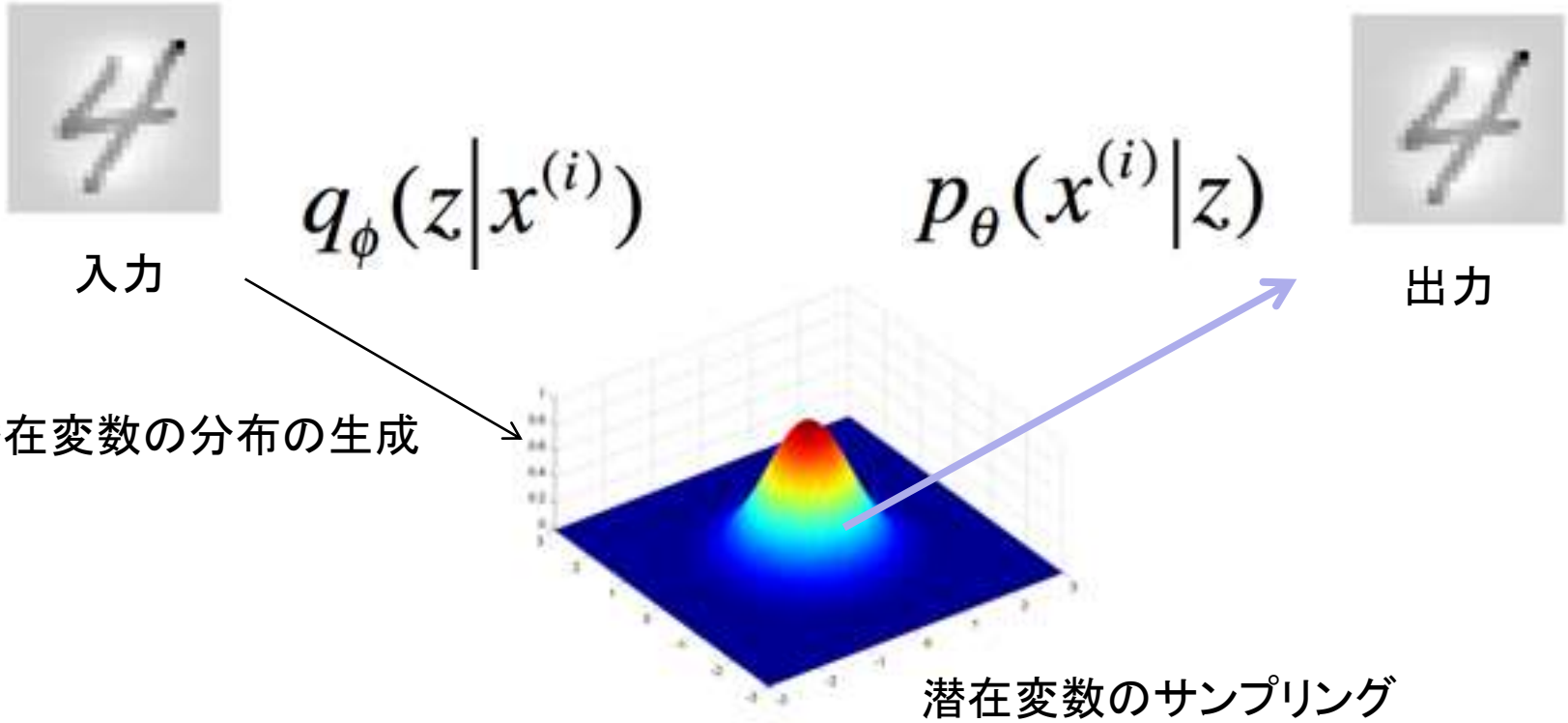


出力画像

潜在変数zから画像Xを生成(θは母数)

VAEのグラフィカルな理解

- 入力->潜在変数の分布を生成 $q_\phi(z|x^{(i)})$
- 潜在変数からのサンプリング->入力に近しい出力の生成 $p_\theta(x^{(i)}|z)$



最適化の必要性

- で、これってどうやって最適化するの？
  - 最尤推定：周辺尤度log(pθ(x))の最大化から考える

    母数θを定めた時に取りうるxの周辺確率が最も高くなるように設定する
  - 周辺尤度log(pθ(x))は以下のように分解できる

一般的な変分下界における数式的な展開

$$\log p_\theta(x) = D_{KL}(q_\phi(z|x)||p_\theta(z|x)) + \mathcal{L}(\theta, \phi, x)$$

$$\geq \mathcal{L}(\theta, \phi, x)$$

変分下界: θ, φの汎関数

$$D_{KL}(q_\phi(z|x)||p_\theta(z)) \geq 0$$

(p=qの時、等号成立)
pとqを近づけることが目的なので変分下界を最大化する必要がある

ref) PRML下巻9.4

変分下界の展開

- 変分下界の展開
  - 以下の式変形により、最適にすべき項が導出できる

$$
\begin{aligned}
\mathcal{L}(\theta, \phi, x) &= \mathbb{E}_{q_\phi(z|x)} \left[\log p_\phi(x, z) - \log q_\phi(z|x) \right] \\
&= \mathbb{E}_{q_\phi(z|x)} \left[\log p_\phi(x|z) - p_\phi(z) - \log q_\phi(z|x) \right] \\
&= - D_{KL}(q_\phi(z|x) || p_\phi(z)) + \mathbb{E}_{q_\phi(z|x)} \left[\log p_\theta(x|z) \right]
\end{aligned}
$$

正則化項：KL Divergence
(Regularization Parameter)

復元誤差
(Reconstruction Error)

この二つの和を最大化すれば良い

正則化項：KL Divergence

- KL Divergenceの計算

$$D_{KL}(q_\phi(z|x)||p_\theta(z))$$

~N(μ, Σ)　　~N(0, I)

$$=D_{KL}(N(\mu,\Sigma)||N(0,I))$$

$$=-\frac{1}{2}\sum_{j=1}^{J}(1+\log(\sigma_j^2)-\mu_j^2+\sigma_j^2)$$

ref) 細かい式の導出は原論文のAPPENDIX C) 参照

復元誤差：Reconstruction Error

- Reconstruction Errorは以下のように近似できる

$$\mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right] \simeq \frac{1}{L}\sum_{l=1}^{L} \log p_\theta(x|z)$$

- 画像のピクセルを0~1に調整した時にベルヌーイ分布を仮定すると
logp(x|z)は以下のように表すことができる
(yはzの潜在変数を全結合層に通した最終層の変数)

$$\log p(x|z) = \sum_{i=1}^{D} x_i \log y_i + (1 - x_i) \cdot \log(1 - y_i)$$

VAEのグラフィカルな理解

- 最適化関数
  - KL Divergence: p(z)とq(z|x)の情報的な距離・正則化項：⟷
  - Reconstruction error: 入出力の差：⟷

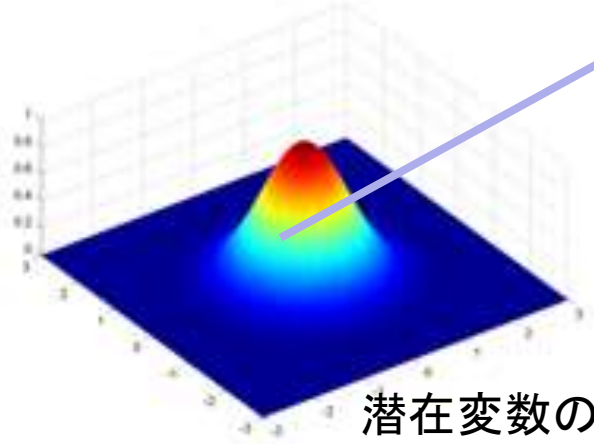
入力

$q_\phi(z|x^{(i)})$

$p_\theta(x^{(i)}|z)$

出力

潜在変数の分布の生成

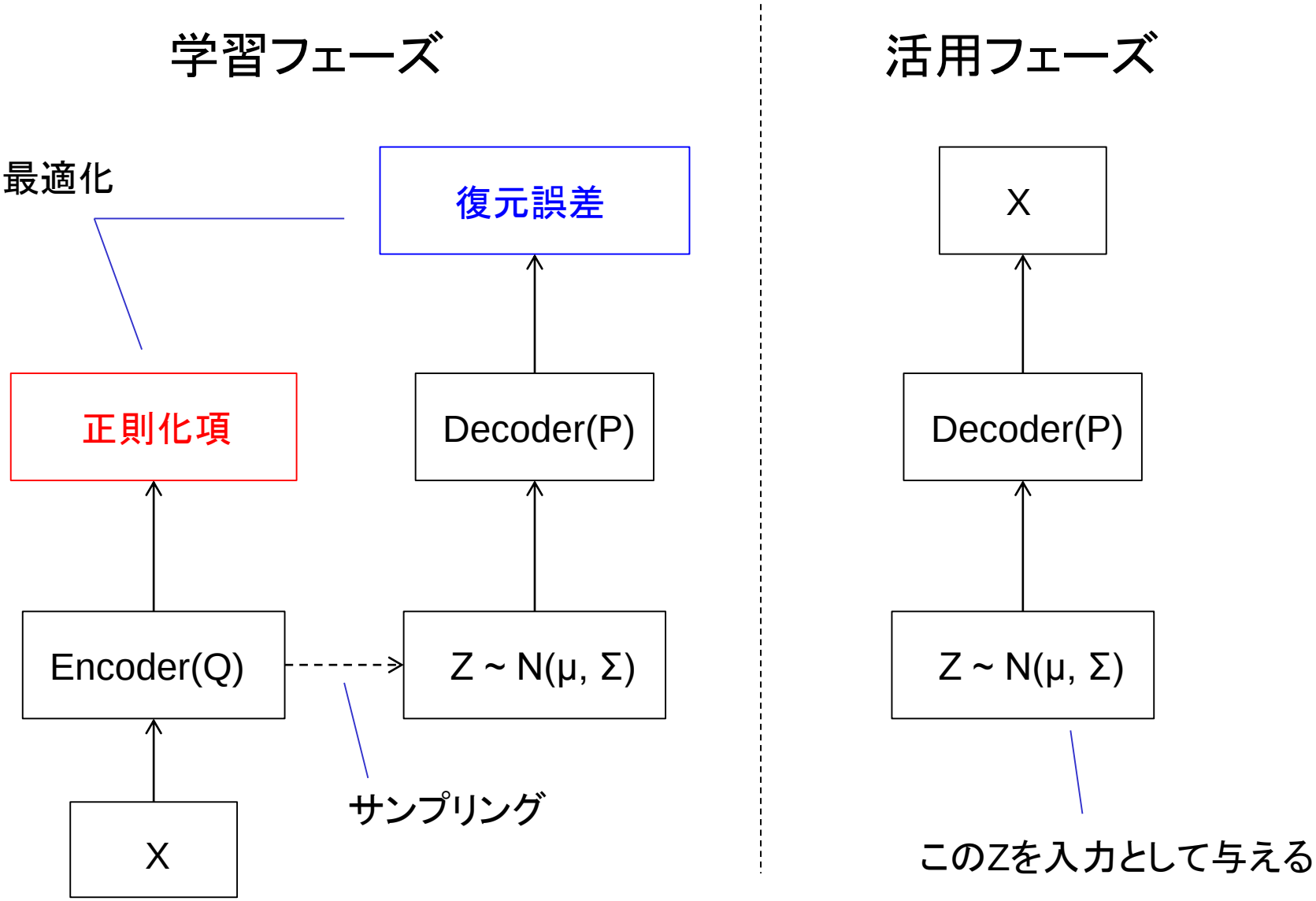


$p_\phi(z)$

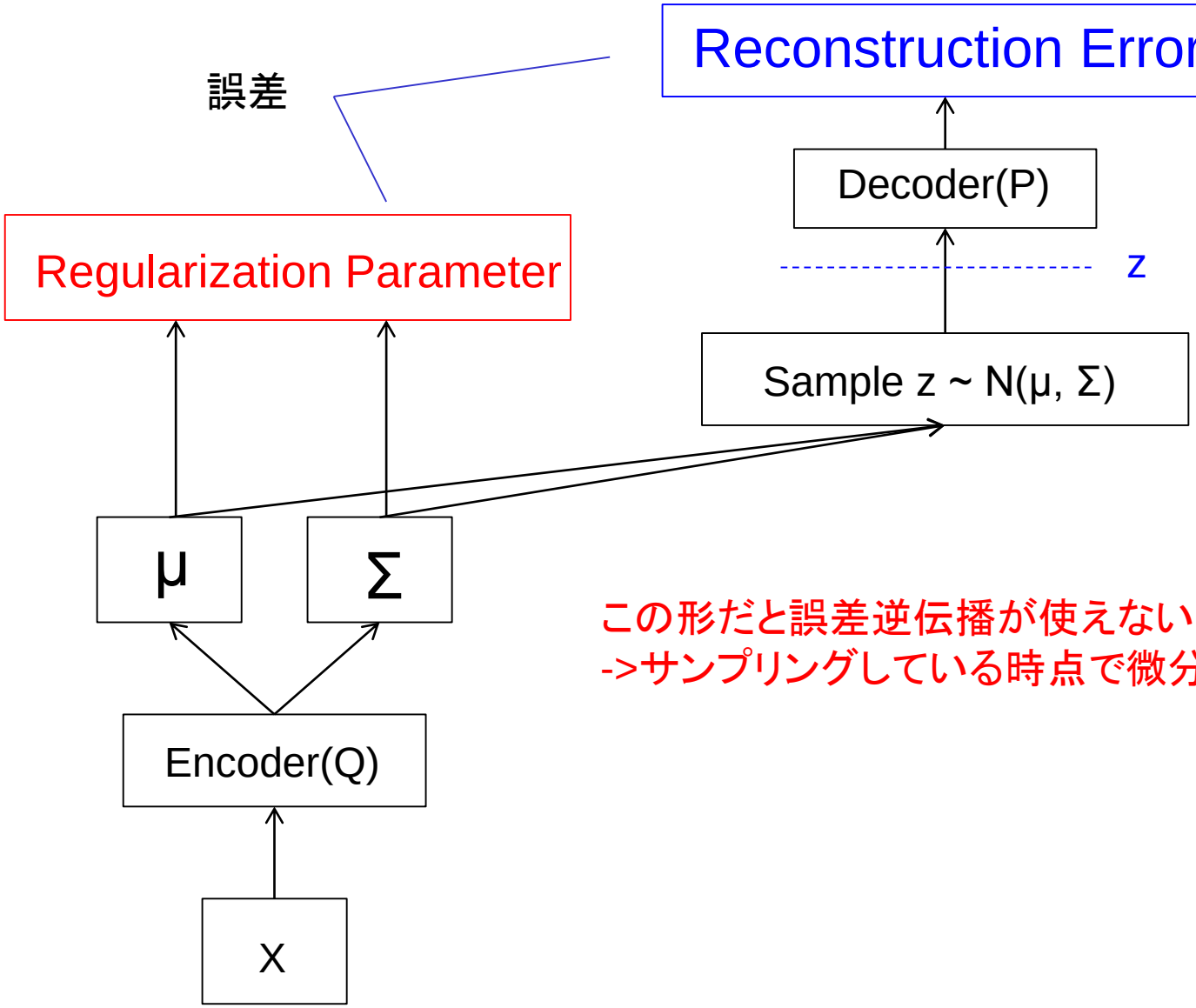
潜在変数の事前分布

潜在変数のサンプリング

VAE全体のブロック図

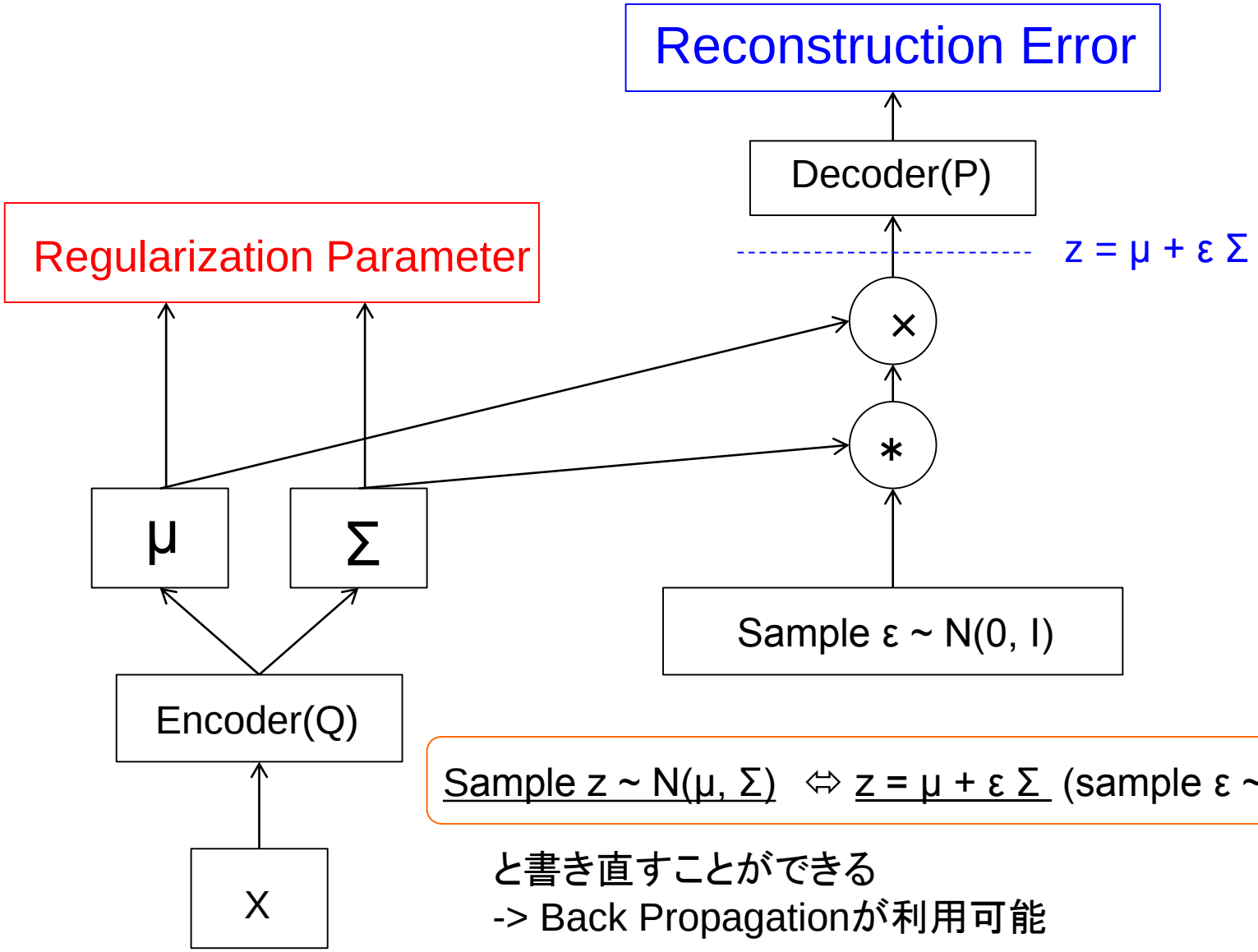


学習フェーズのより詳細な構造



この形だと誤差逆伝播が使えない！
->サンプリングしている時点で微分の計算が不可

Reparametrization Trick



Sample z ~ N(μ, Σ) ⇔ z = μ + ε Σ (sample ε ~ N(0, I))

と書き直すことができる
-> Back Propagationが利用可能

zの変換について

- 一次元の場合の簡単な証明

Sample z ~ N(μ, σ) ⇔ z = μ + ε σ (sample ε ~ N(0, 1))

εは標準正規分布なので確率密度関数は $f(\epsilon) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{\epsilon^2}{2})$

$z = \mu + \epsilon \cdot \sigma \Leftrightarrow \epsilon = \frac{z-\mu}{\sigma}$ と変換できるので代入・周辺化して

$$f(z) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(z-\mu)^2}{2\sigma^2})$$

これは正規分布z~(N(μ, σ))からのサンプリングに他ならない

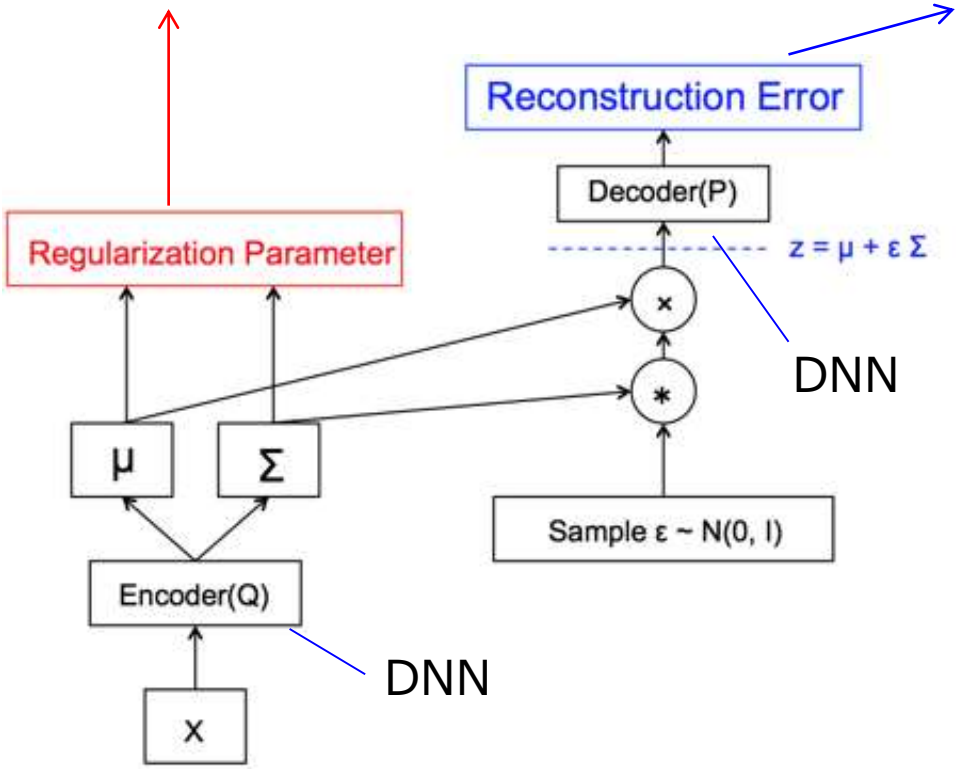
次数が2次以上の場合も同様

VAEの最適化のフローのまとめ

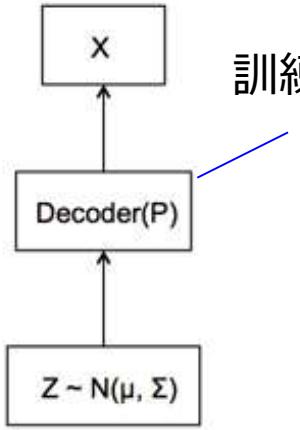
・訓練時

$$\mathcal{L}(\theta, \phi, x)$$

$$= -\frac{1}{2}\sum_{j=1}^{J}(1 + \log(\sigma_j^2) - \mu_j^2 + \sigma_j^2) + \mathbb{E}\left[\sum_{i=1}^{D}(x_i \log y_i + (1 - x_i) \cdot \log(1 - y_i))\right]$$



・生成時

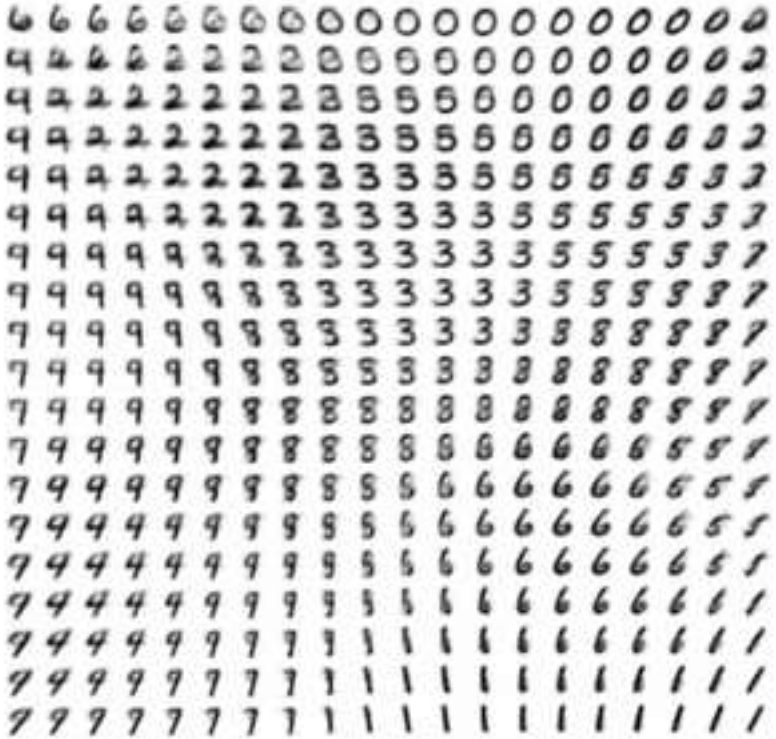


訓練時のものを利用

Result

- 潜在空間に対応する画像の生成

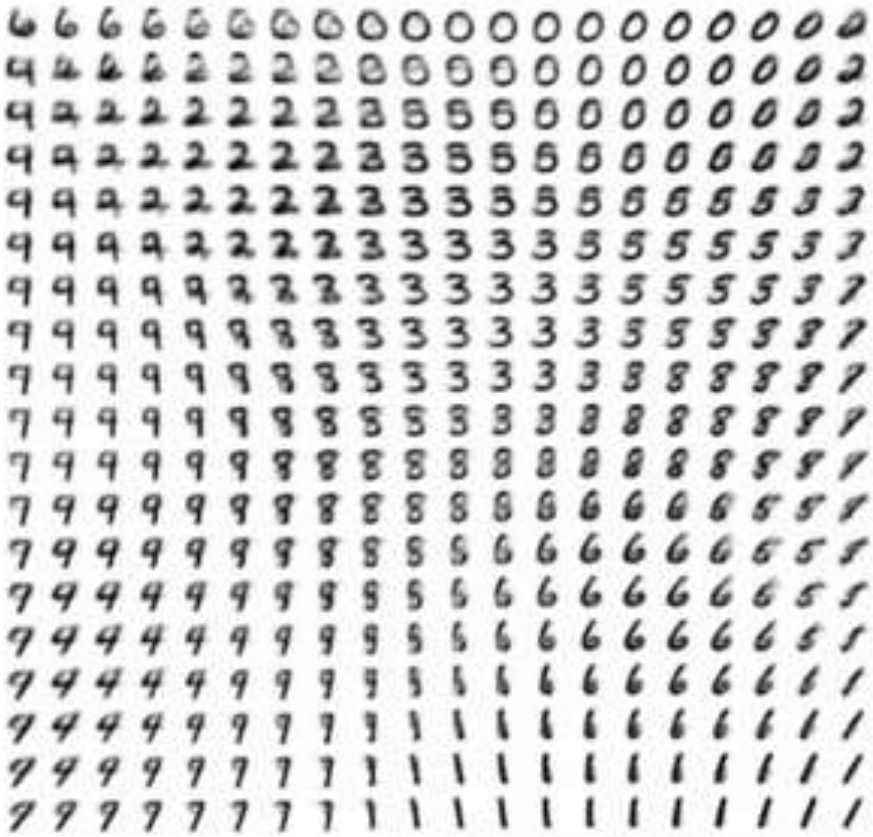
表情の生成



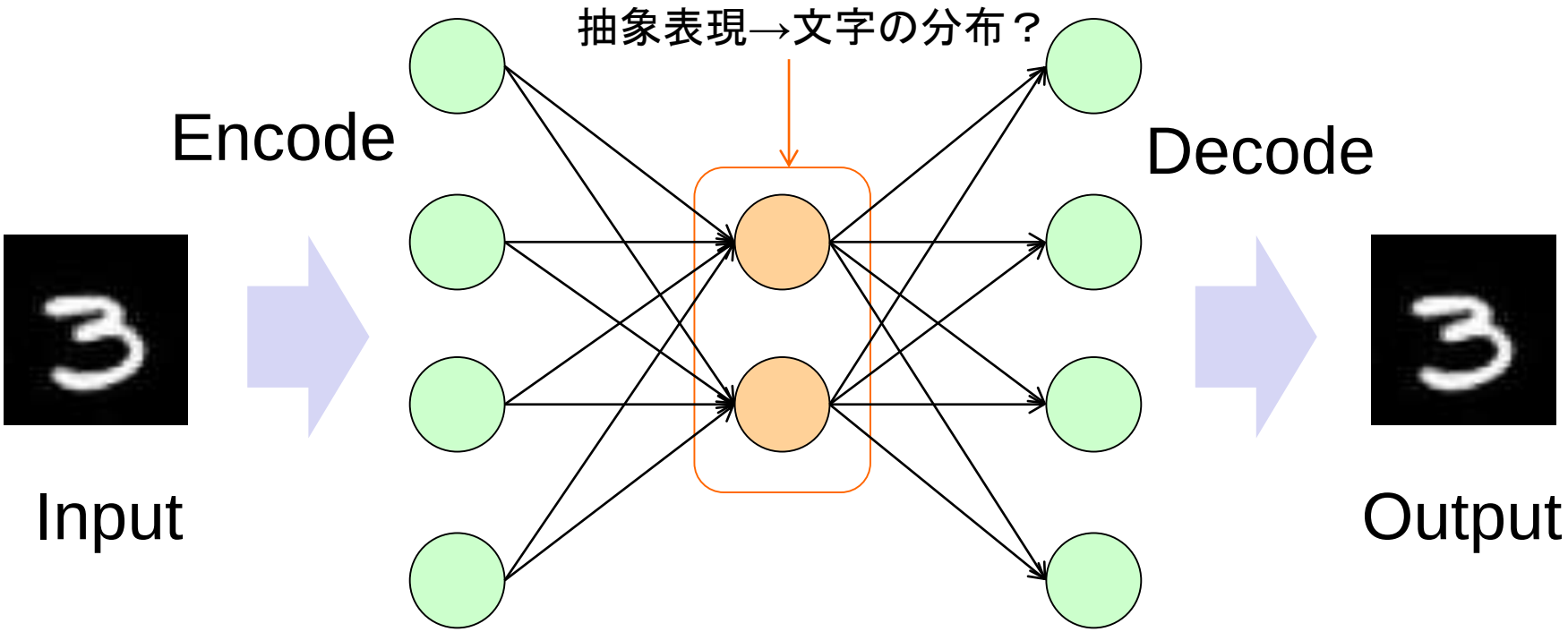
文字の生成

VAEの結論

- Deep Learningを生成モデルに適用
  - 筆跡や表情といった潜在変数の分布を組み込む形を考案し、データセットに存在しない自然な画像の生成を可能にした

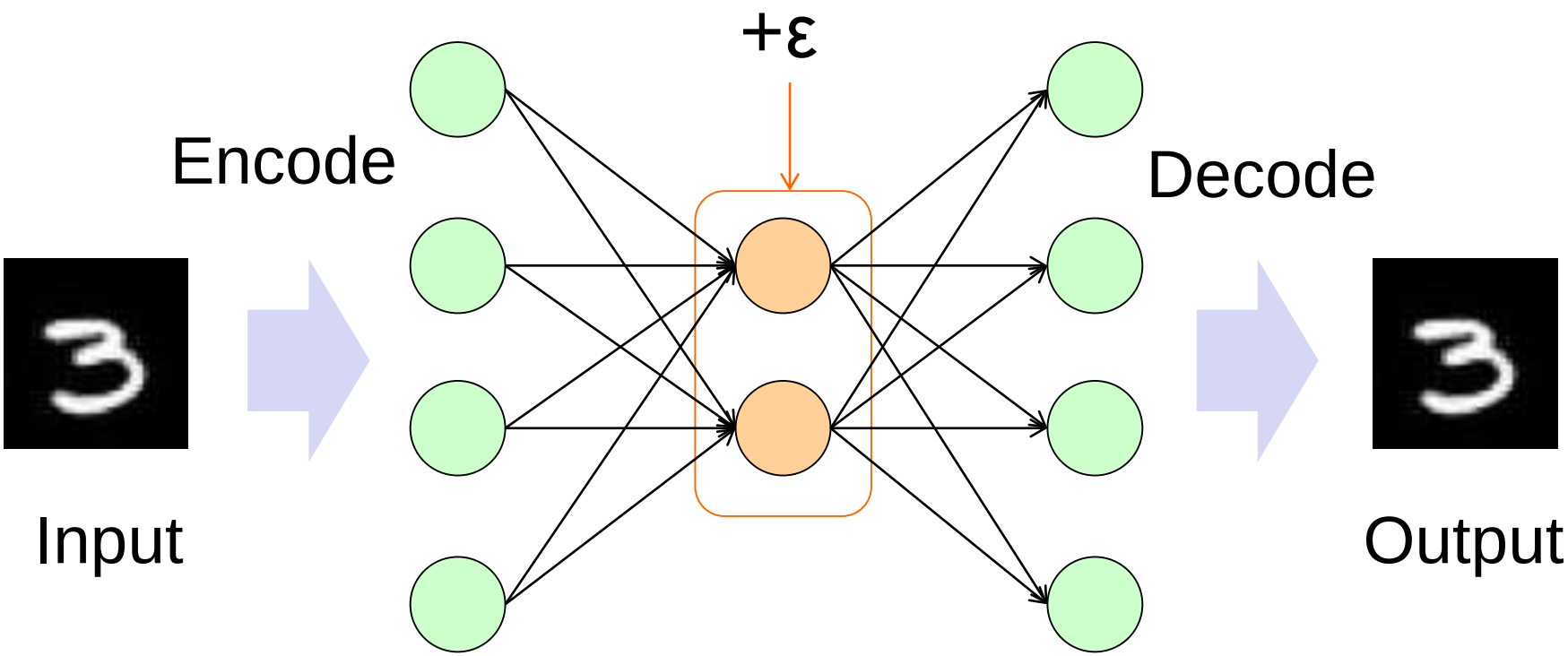


おまけ：AutoEncoder



・AutoEncoderによる画像の圧縮・再構成
・中間層での画像の抽象表現の獲得

Valuational AutoEncoderの実態



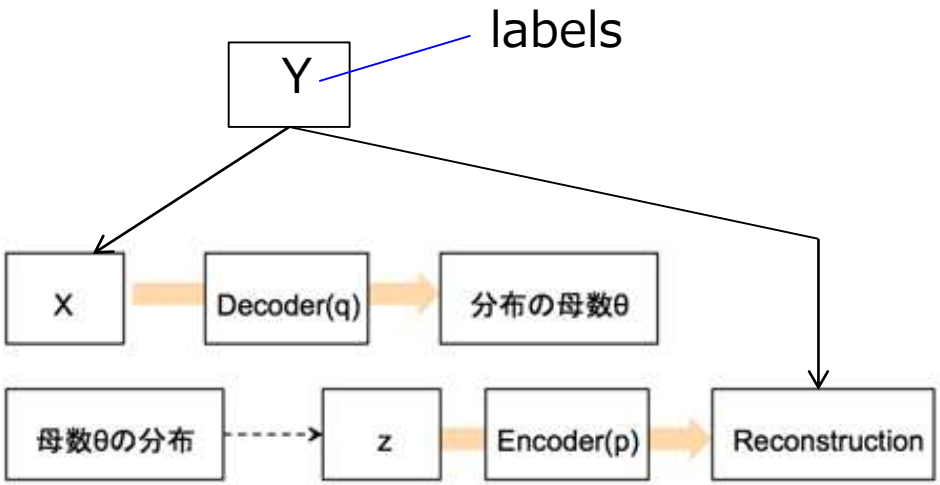
・構造はAutoEncoderの中間層にノイズを入れただけ
・loss関数に正則化項を加えた
・構造，名前は非常に似ているが由来は異なる

生成モデルの論文紹介

- 画像の自動生成
- 3次元モデルの自動生成
- 行動の予測
- (生成モデルではないけどおまけ)DNNを騙す画像の生成

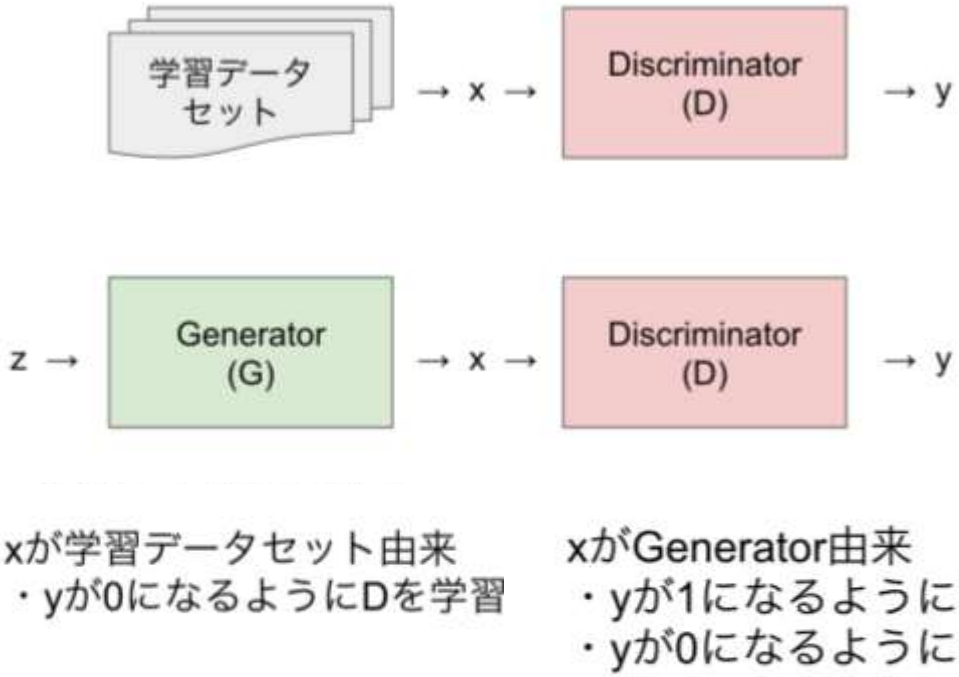
教師ありVAE

- Semi-supervised Learning with Deep Generative Models (14' M. Welling)
  - 教師ありVAE, セミ教師ありVAEの提案
  - 同じ筆跡の別の文字などの生成も可能に



GAN

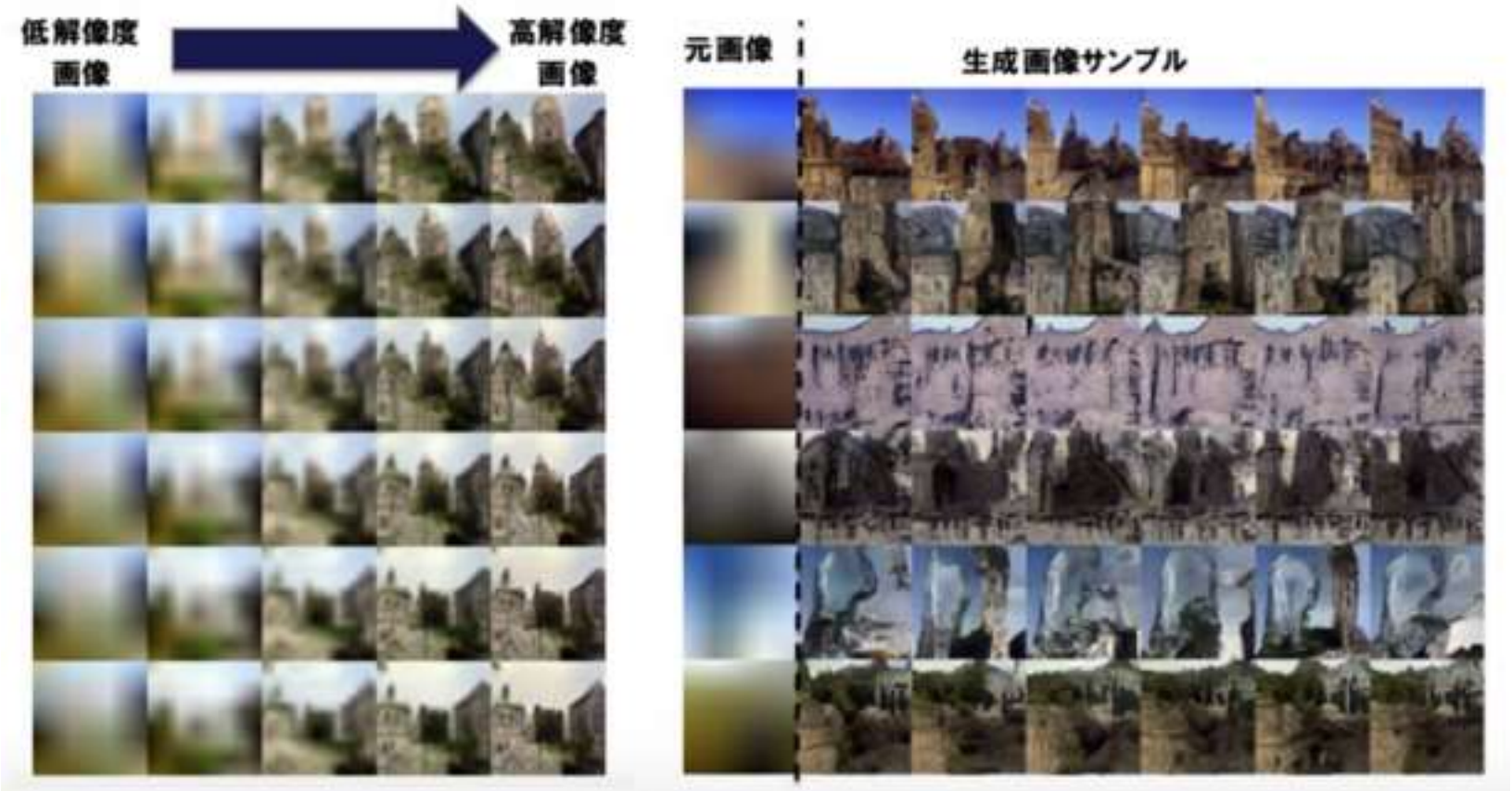
- Generative Advisarial Net(14’ I. J. Goodfellow)
  - 学習データに似たイメージを作るGenerator
  - 学習データかGeneratorが作成したデータか見分けるDiscriminator
    - » GenratorとDiscriminatorでイタチごっこをする



一番右が近いイメージ

LAPGAN

- Deep Generative Image Models using a Laplacian Pylamid of Adversarial Networks(15’ E. Denton)
  - 周波数ごとのGANを作り高解像の画像を生成する手法



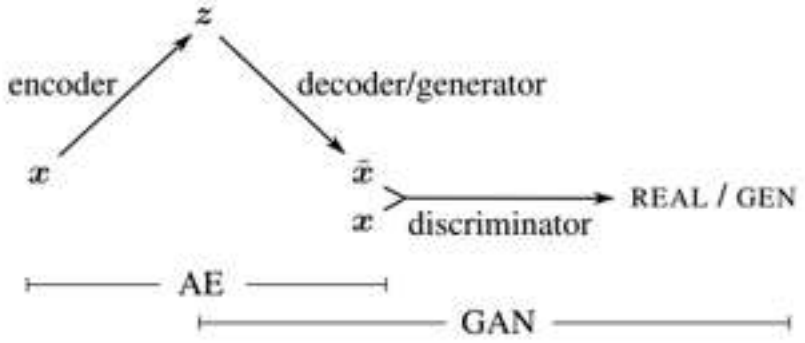
DCGAN

- Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks(16’ A.Radford)
- GANにCNN, Leaky ReLU, Batch Normalizationを加えた手法
  - Leaky Relu: 0以下にも勾配をつけたReLU
  - Batch Normalization: バッチごとの平均・分散を用いて正規化

精細な画像，潜在変数のベクトル演算

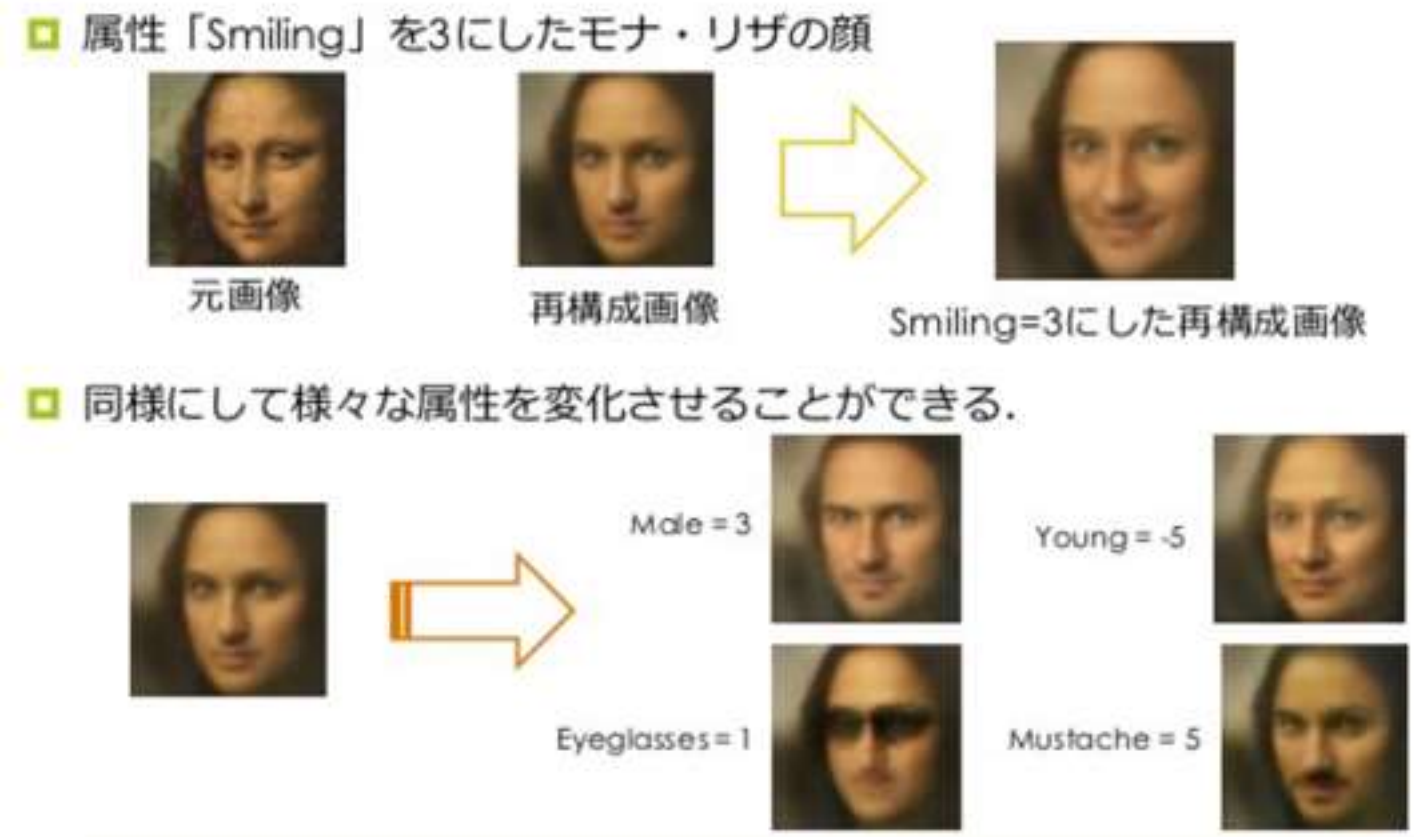
VAEとGANの統合

- Autoencoding beyond pixels using a learned similarity metric (15’ A. B. L. Larsen)
  - VAEの後ろ部分にGANをくっつけたもの
  - VAEのreconstructionとGANの精細さを両立



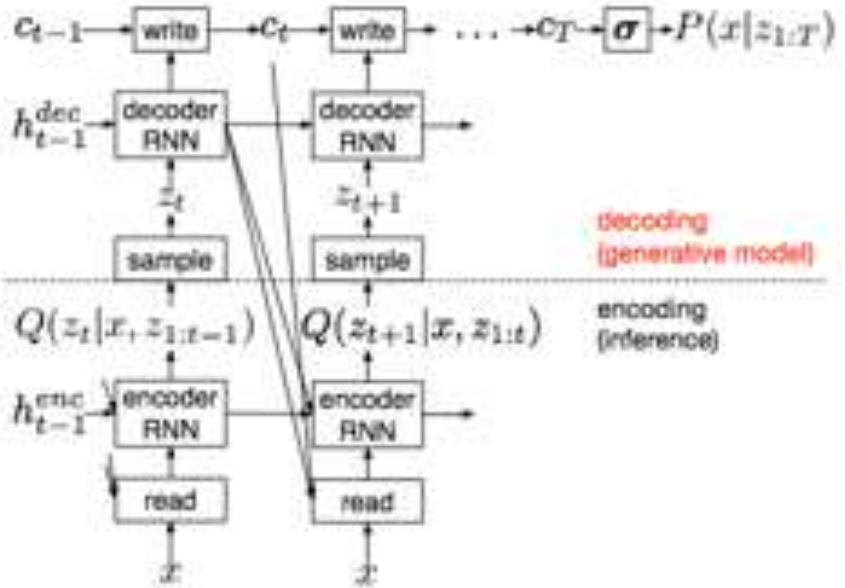
文章と画像の統合

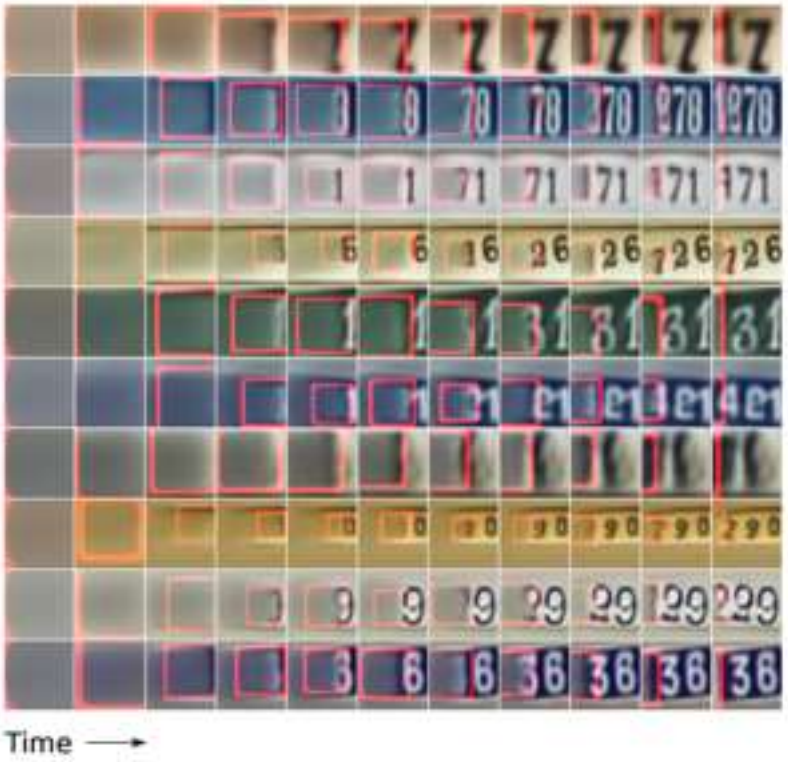
- 深層生成モデルを用いたマルチモーダル学習(16’ 鈴木)
  - 複数のモダリティ情報を統合
  - 例えば，モナリザにsmile要素を加える



RNN+VAE

- DRAW: A Recurrent Neural Network For Image Generation (15’ K.Gregor)
  - RNNとVAEを組み合わせて徐々に画像を生成していく手法
  - 上に塗っていくイメージ





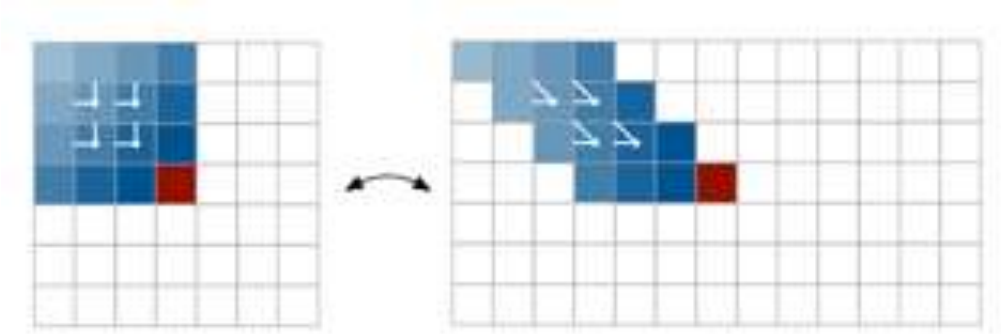
画像圧縮

- Towards Conceptual Compression(16’ K.Gregor)
  - Convolutional DRAWを提案し,画像圧縮に応用
  - 上からJPEG,JPEG2000,下二つが分散あり・なしのCDRAW
  - 圧縮率を上げても自然な圧縮を実現

圧縮率20%

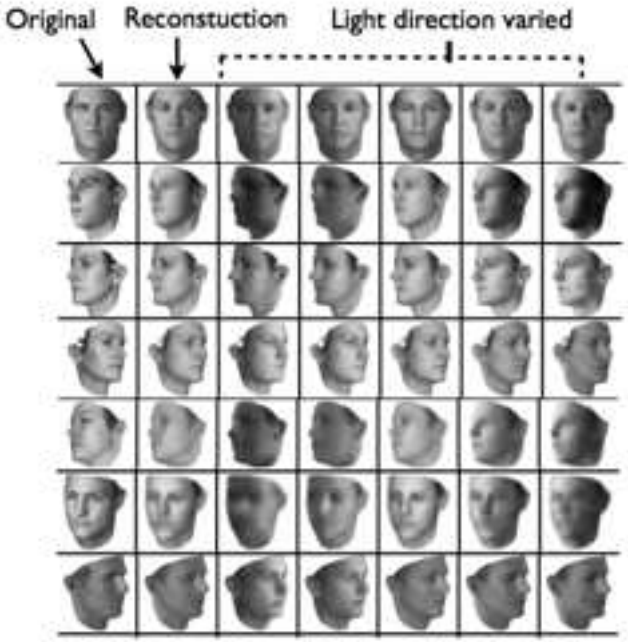
Pixelごとの復元

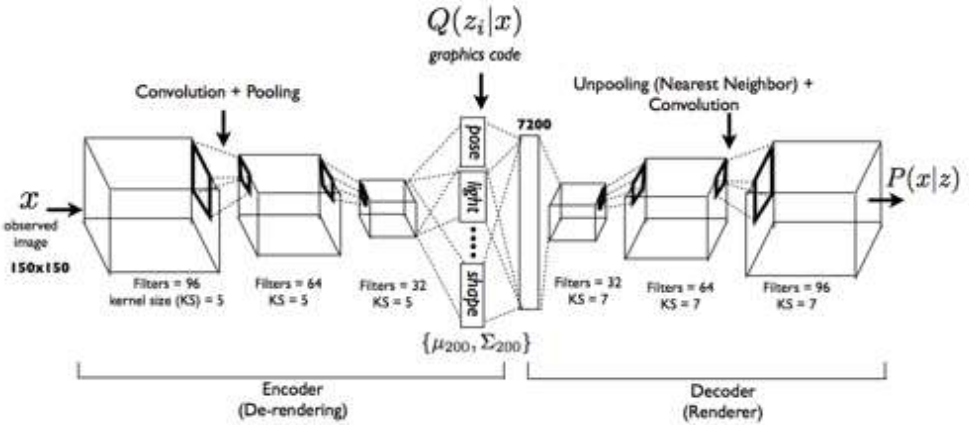
- Pixel Recurrent Neural Networks(16’ A. Oord)
  - RNNを用いて近傍のPixel群から周りのpixel群を予測して復元



3次元版のVAE

- Deep Convolutional Inverse Graphics Network(15’ TD. Kuulkarni)
  - VAEを3次元に拡張
  - 人の顔のモデルのバリエーションを生成





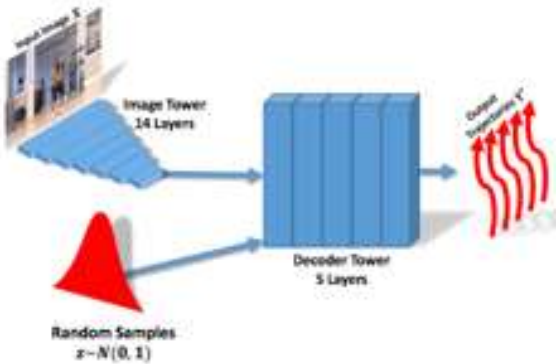
2次元画像->3次元モデル

- Unsupervised Learning of 3D Structure from Images (16' D.J.Rezende)
  - 二次元画像から三次元モデルを復元する
  - 教師データとしての三次元モデルは一切与えない

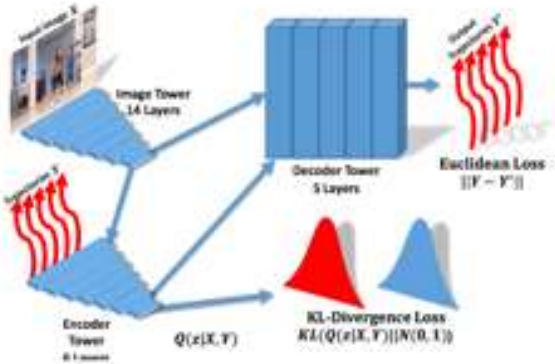
2次元画像から復元した3次元モデル

運動ベクトルの推定

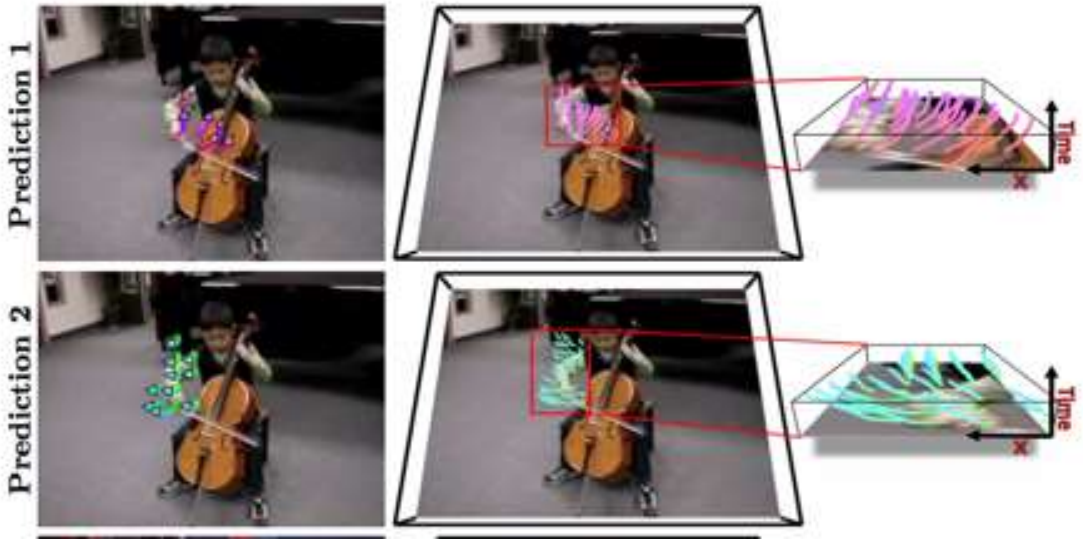
- An Uncertain Future: Forecasting from Static Images using Variational Autoencoders (16’ J. Walker)
  - 画像と動きのベクトルを学習させて画像のどの部分が動くか予測



(a) Testing Architecture

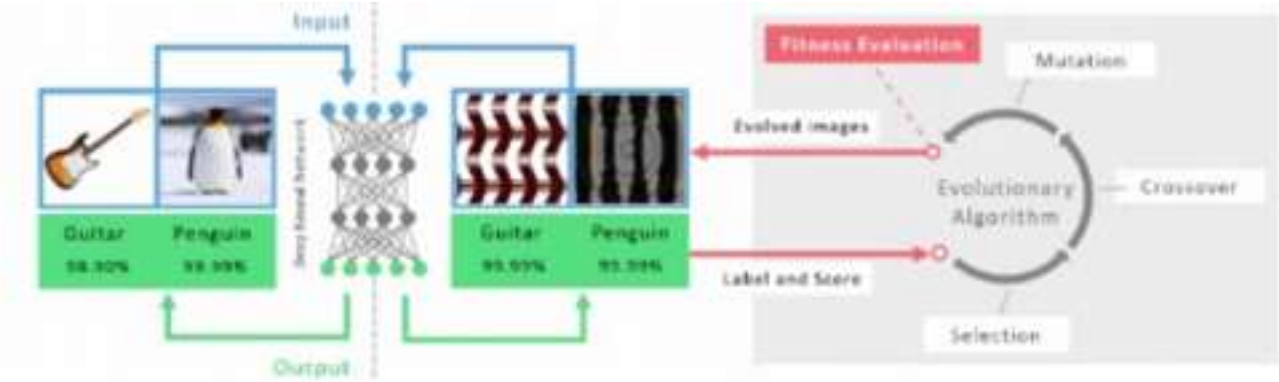


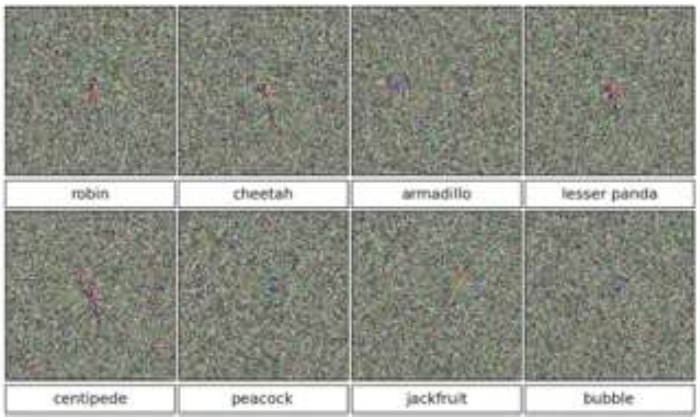
(b) Training Architecture



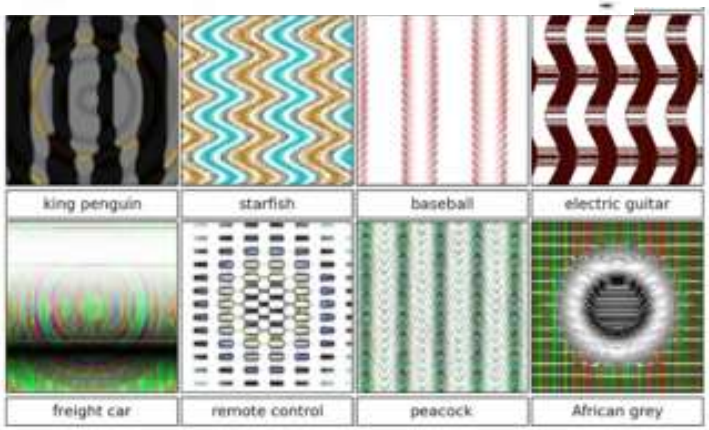
DNNを騙す画像の生成

- Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images(15’ A.Nguyen)
  - 元ある識別器を騙す画像を進化論的アルゴリズムにより作成
  - 識別精度が高くなるように画像のピクセルをランダムに変化させる





Direct encode



Indirect encode

終わりに

- まとめ
  - 深層生成モデルの中のVAEについての説明
  - 画像関連の生成モデル論文の紹介

- 印象
  - 生成モデルはここ1年が凄まじい
    - » VAE・GANの原論文以外はほぼ去年・今年の論文
    - » DeepMind, OpenAIが参入
    - » 静止画の生成モデルはかなり行われている印象
  - 動画像生成とかはまだまだ

参考文献

- Introduction to variational autoencoders
  - URL: https://home.zhaw.ch/~dueo/bbs/files/vae.pdf
  - VAEのスライド

- Deep Advances in Generative Modeling
  - URL: https://www.youtube.com/watch?v=KeJINHjyzOU
  - Youtubeでの深層生成モデルの解説

- Digit Fantasies by a Deep Generative Model
  - URL: http://www.dpkingma.com/sgvb_mnist_demo/demo.html
  - VAEのデモ

- LAPGANの解説(スライド)
  - 他のGANの話も載っているのでオススメ
  - URL: http://www.slideshare.net/hamadakoichi/laplacian-pyramid-of-generative-adversarial-networks-lapgan-nips2015-reading-nipsyomi