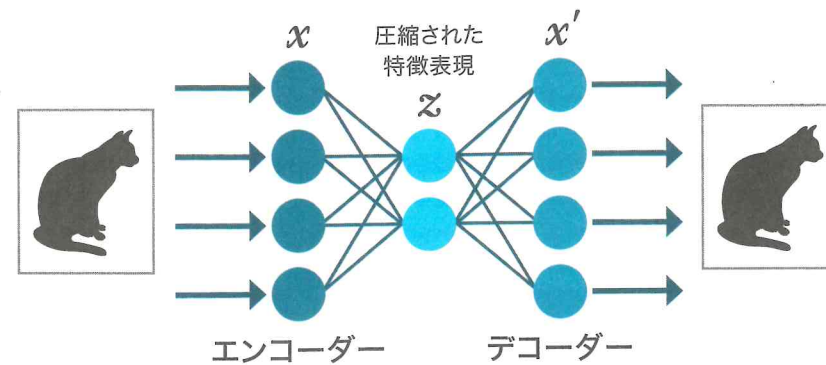


■図2-2-31 オートエンコーダー



なお、デコーダーに着目すると「圧縮された特徴表現」から、出力情報を生成していることになるのでGeneratorと呼ばれることがある。

### VAE (Variational Autoencoder)

このオートエンコーダーの中間層の圧縮された特徴表現の数値を確率分布に従うように変更し、さらにエンコーダーの出力とデコーダー (Generator) の入力をこれに合わせて変更したものがVAE (Variational Autoencoder)である。確率分布に従うようにすることで、圧縮された特徴表現の数値を変動させ、デコーダー (Generator) の所期の性能 (いろいろなネコの画像の生成) を達成できる。例えばFrey Faceデータセット<sup>※66</sup>の顔のデータで学習することで、怒った顔に対する特徴表現と笑った顔に対する特徴表現を得て、これらの特徴表現の間で徐々に変化させるよう中間層の変数値を制御することにより、怒った顔から笑った顔へ徐々に変化するデータを生成することができる<sup>※67</sup>。

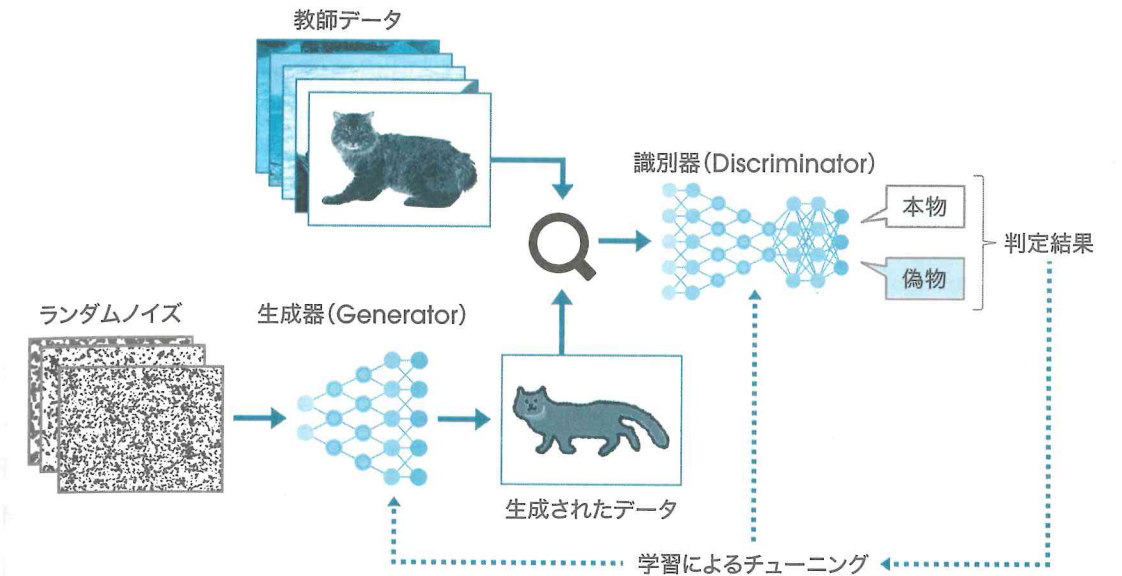
### GAN

生成モデルのもう一つの手法であるGANは、ランダムノイズベクトルから入力データのレプリカをつくり出そうとする生成ネットワーク (Generator) と、生成されたデータが、「教師データ」か「生成ネットワークの出力物」かを識別する識別ネットワーク (Discriminator) の2つのネットワークを組み合わせ、互いを欺くようにトレーニングを行う (図2-2-32)。例えば、図の識別器は2枚の画像の違いを見分けられるための特徴を抽出しようとし、生成器はその特徴について差をなくすように生成を行う。トレーニングに利用したデータに類似した出力を生成できるため、希少データの水増しや、作風をまねた絵画や楽曲の生成など、様々な応用が進められている。

※66 Brendon Frey氏の顔の様々な表現を約2000枚集めたデータセット。

※67 sam roweis : data <<https://cs.nyu.edu/~roweis/data.html>>

■図2-2-32 GANの構成



### (3) 最新技術動向

#### GANの様々なバリエーション

高い品質での再現が可能なGANは、生成ネットワークと識別ネットワークを回路のように組み合わせ、様々な工夫を行うことにより多数のバリエーションを生み出しつつある。

例えば、Cycle-GAN [4] では、無関係の2つの画像を使って、お互いに変換しあうようにサイクル状にネットワークを構成し、画像の変換を行ってみせた。「形や様式の見分けがつくか」が判断できる。風景写真と名画を組み合わせることで「モネ風」、「ゴッホ風」、「セザンヌ風」の風景絵画を生成できた。そのほかにも風景写真の季節や時刻 (昼間と夕暮れ) を入れ替えるといった例が紹介されている。

AttnGanは、マイクロソフトが開発した「キャプション→画像」の生成を可能にしたGANである。いくつかの情報を基に、足りない情報を自分で補って構成し直すというプロセスを経る。そのため、自然言語処理向け深層ニューラルネットに着目されているアテンション機構を導入し、「黄色の羽」、「赤い腹部」などの複雑な条件でも、画像中の細かく区分けされた領域のうち特定の単語と最も関連性の強い領域を描画することで複雑な画像を生成できるという [5]。

StyleGANは、ノイズベクトルを入力層から与えるという従来の生成ネットワーク (Generator) の構造を大きく変更し、まずノイズベクトルを別の空間にマッピングし、そこから得られた情報を生成ネットワークの各層に入力していくことで生成画像の大局的な構造から詳細構造までを柔軟に制御することを実現している。

BigGAN [6] は、ImageNetのような極めてバリエーションが多く複雑なデータセットを使い、ラベルからの画像生成を過去最大規模で学習させることに成功している。BigGAN発表直前におけるInception<sup>※68</sup>最高スコアは52.52であったが、BigGANは166.3を達成し、大幅な改善となった。

GANでは、出力を高精細にしようとすると、ノード数が増えるため学習時間も増えてしまう

※68 Inceptionスコアとは、生成画像の「良さ」を以下2つの観点から点数付けしたもの。1) 画像識別器が識別しやすい、2) 物体クラスのバリエーションが豊富。Inceptionスコアが大きいほど、性能が高い。