GLCICを用いた機械学習による マスク下の顔自動補完

日本大学文理学部情報科学科　谷聖一研究室

5417056 高田駿

目次

1. はじめに
   1.1 動機
   1.2 先行研究
   1.3 演習目的

2. 準備
   2.1 機械学習
   2.2 モデル
   2.3 機械学習における学習
   2.4 教示あり学習
   2.5 ニューラルネットワーク
   2.6 GAN
   2.7 GLCIC

3. 演習
   3.1 環境構築
   3.2 学習データセットの準備
   3.3 補完モデルの作成

4. 評価

5. まとめ

目次

1. はじめに
   1.1 動機
   1.2 先行研究
   1.3 演習目的

2. 準備
   2.1 機械学習
   2.2 モデル
   2.3 機械学習における学習
   2.4 教示あり学習
   2.5 ニューラルネットワーク
   2.6 GAN
   2.7 GLCIC

3. 演習
   3.1 環境構築
   3.2 学習データセットの準備
   3.3 補完モデルの作成

4. 評価

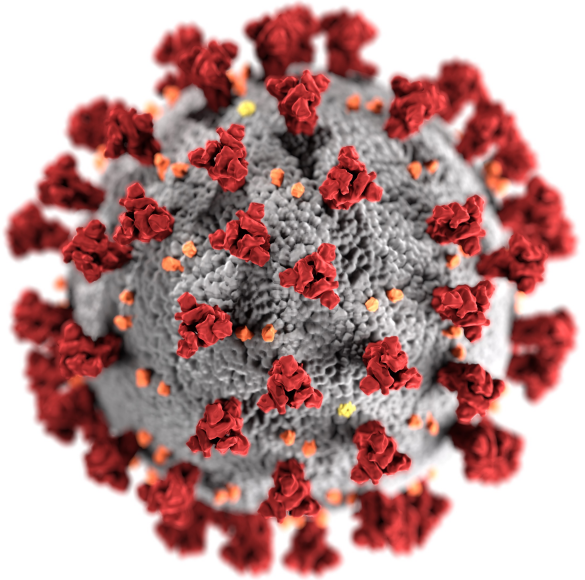
5. まとめ

1.1 動機

新型コロナウイルス感染症(COVID-19)

世界保健機関(WHO)による国際正式名称をCOVID-19といい，SARSコロナウイルス2がヒトに感染することによって発症する気道感染症.

2019年に「原因不明のウイルス性肺炎」と確認されて以降，世界的に感染が拡大し昨今も余談を許さない状況である.



1.1 動機

新型コロナウイルス感染症の流行

↓

マスク着用の常態化

↓

素顔を知らない知人の増加

↓

マスクをつけた人の顔画像から
マスク下の顔を補完したい！

1.2 先行研究

Masked Face Recognition Using Deep Learning: A Review
Electronics October 2021

Ahmad Alzu’bi, Firas Albalas,
Tawfik AL-Hadhrami, Lojin Bani Younis

機械学習を用いたマスク付きの顔認識技術についてのサーベイ論文.

機械学習を用いた
マスク付きの顔認識技術

→

既に200以上の研究が
存在する！

補完に関する研究だけで複数存在する！

1.2 先行研究

Globally and locally consistent image completion.
SIGGRAPH July 2017

SATOSHI IIZUKA, EDGAR SIMO-SERRA,
HIROSHI ISHIKAWA

GLCIC(Globally and locally consistent image completion)という画像に
ランダムに発生させた欠損の補完を行う手法を提案．GANという2つのニューラルネットワークが競い合いながら学習を進める手法を元にしている．

1.3 演習目的

GLCICの学習法

・ランダムに欠損を発生させて学習

動機：マスクをつけた人の顔画像からマスク下の顔を補完したい！

ランダムに欠損を発生させて学習 → 口元だけ欠損を発生させて学習

補完の質に差は生まれるのか？

目次

1. はじめに
   1.1 動機
   1.2 先行研究
   1.3 演習目的

2. 準備
   2.1 機械学習
   2.2 モデル
   2.3 機械学習における学習
   2.4 教示あり学習
   2.5 ニューラルネットワーク
   2.6 GAN
   2.7 GLCIC

3. 演習
   3.1 環境構築
   3.2 学習データセットの準備
   3.3 補完モデルの作成

4. 評価

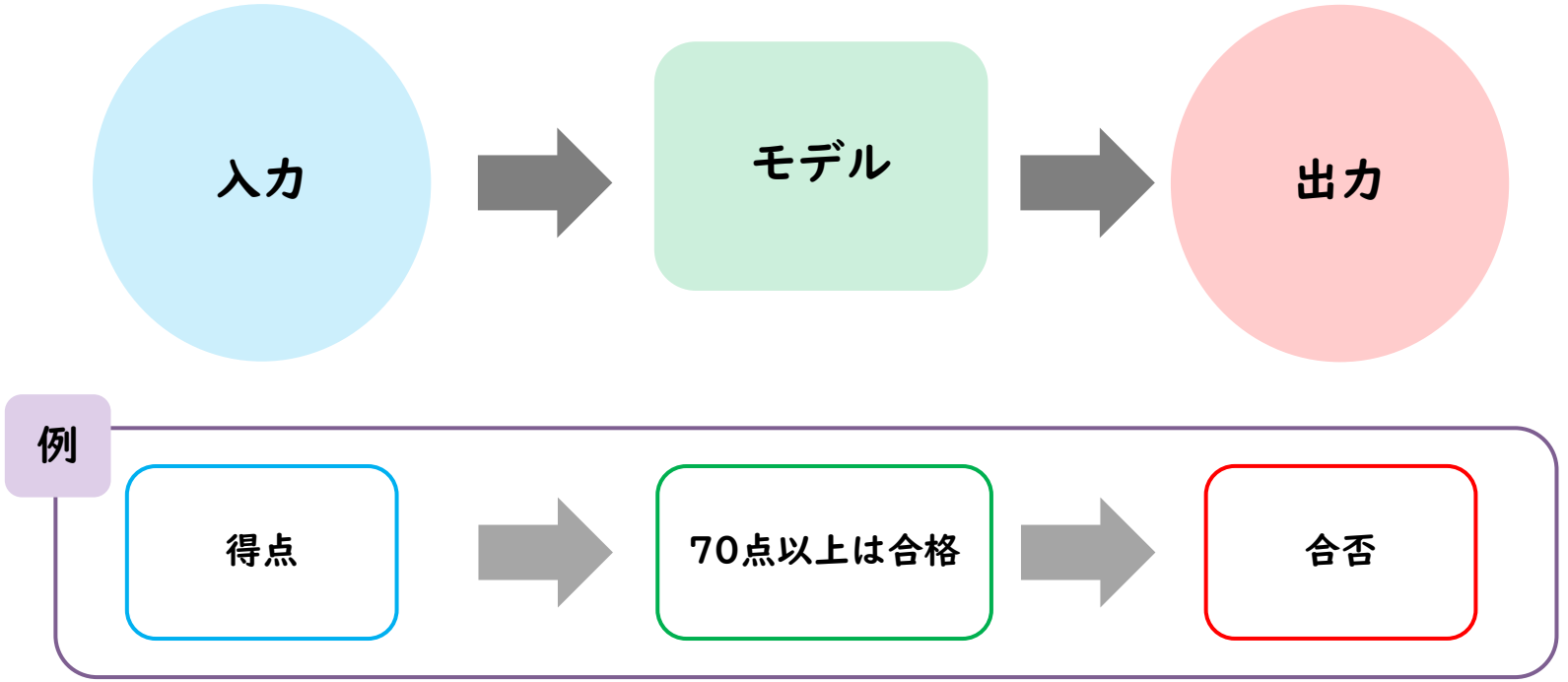
5. まとめ

2.1 機械学習

機械学習とは
明示的にプログラムしなくても学習する能力を
コンピュータに与える研究分野.
(Arthur Lee Samuel 1959)

2.2 モデル

入力から出力を得る仕組み．
評価や未来の予測，データの自動生成に活用されることが多い．



2.3 機械学習における学習

機械学習では，ニューラルネットワークなどでモデルを表現し，
何らかの学習方法でモデルを目的に適合させようとすることが多い．

教師あり学習

学習データに正解などの
情報を加えた状態で
学習を行う機械学習手法

教師無し学習

学習データに正解などの
情報を加えない状態で
学習を行う機械学習手法

強化学習

一連の行動をとった結果
ごとに報酬を設定し，
その報酬を最大化するよ
うに学習を行う機械学習
手法

2.4 教師あり学習

学習データに正解などの情報を加えた状態で学習を行う機械学習手法.

学習データ

イヌ

ネコ

トリ

ニューラルネットワークを用いて学習

入力

モデル

出力

ネコ

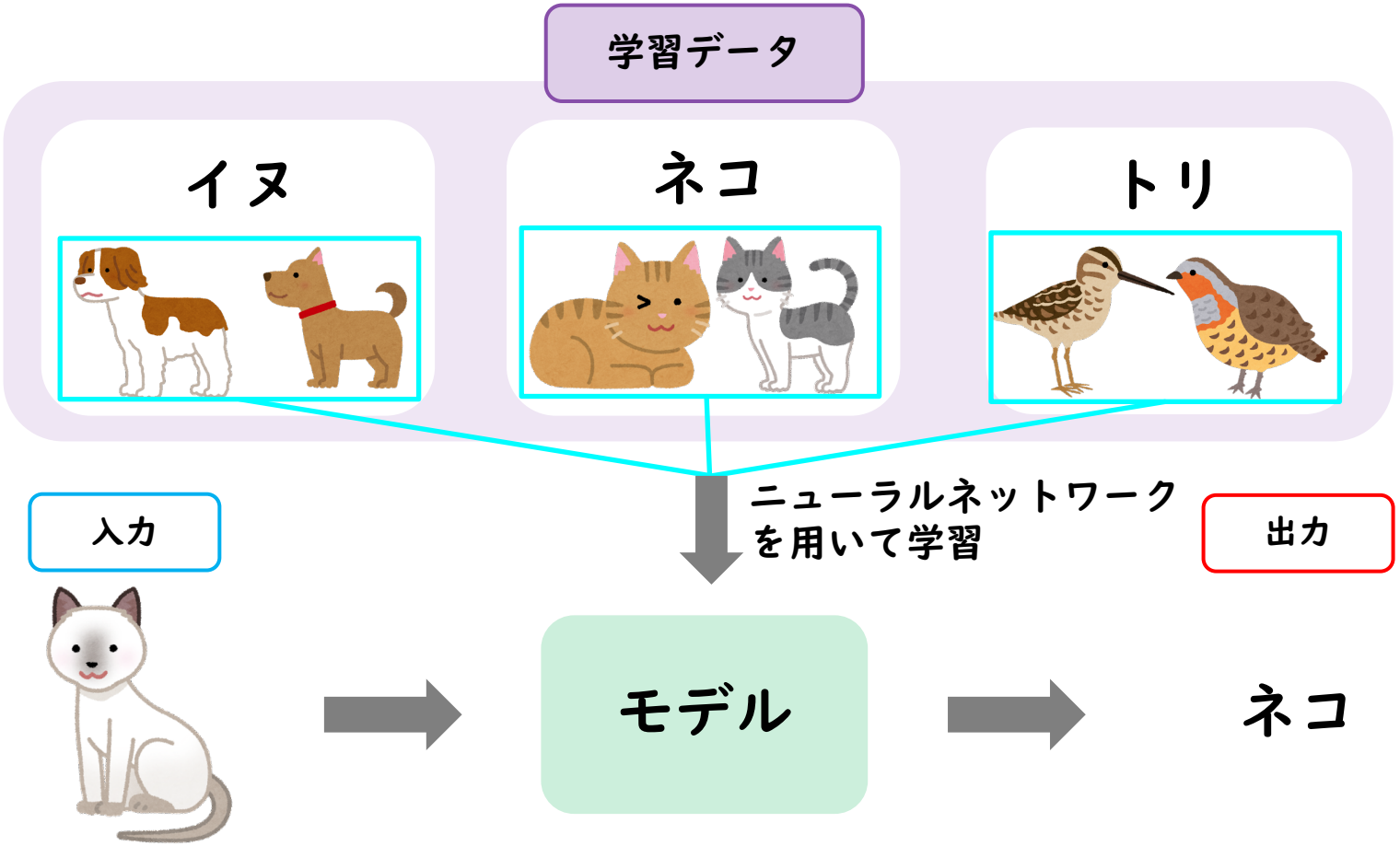
ラベル

データがどのカテゴリに属するかを表した値.

例：イヌ→0
ネコ→1
トリ→2

2.4 教師あり学習

学習データに正解などの情報を加えた状態で学習を行う機械学習手法.



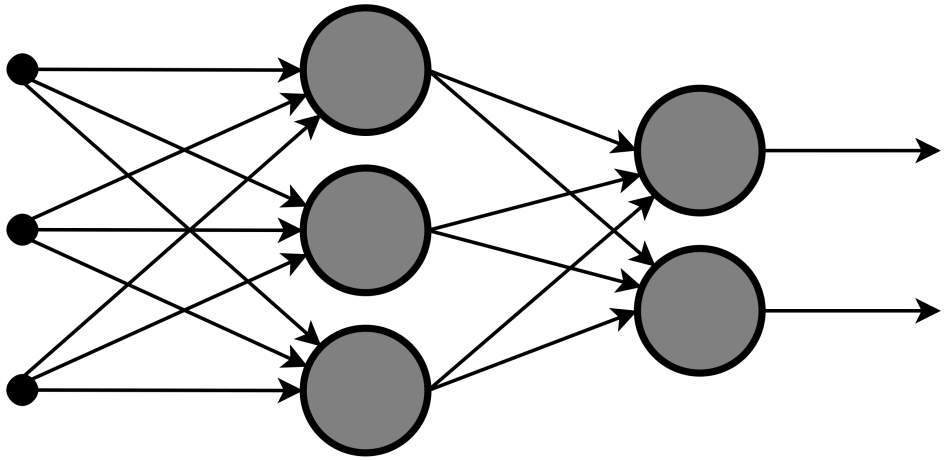
特徴量

データの特徴が数値化されたもの
(例：高さ，幅，RGB値)

ラベルと特徴量を持つ
学習データを受け取り
モデルを作成することが目的

2.5 ニューラルネットワーク

モデルの表現法のひとつ.
脳の神経回路を模した数理モデル.
誤差逆伝播法などの学習する機構を組み込める.

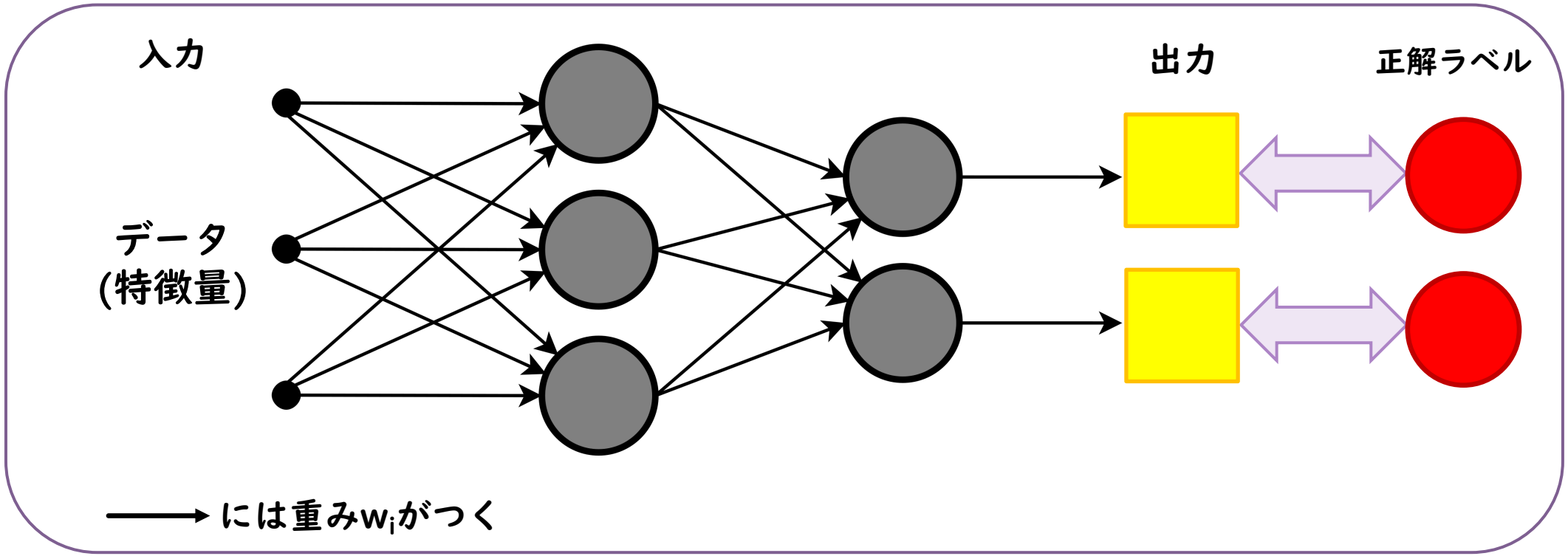


2.5 ニューラルネットワーク

ニューラルネットワークの学習

誤差逆伝播法

出力と正解ラベルの誤差を計算し，出力が正解ラベルに近づくように重みw_iを調整する

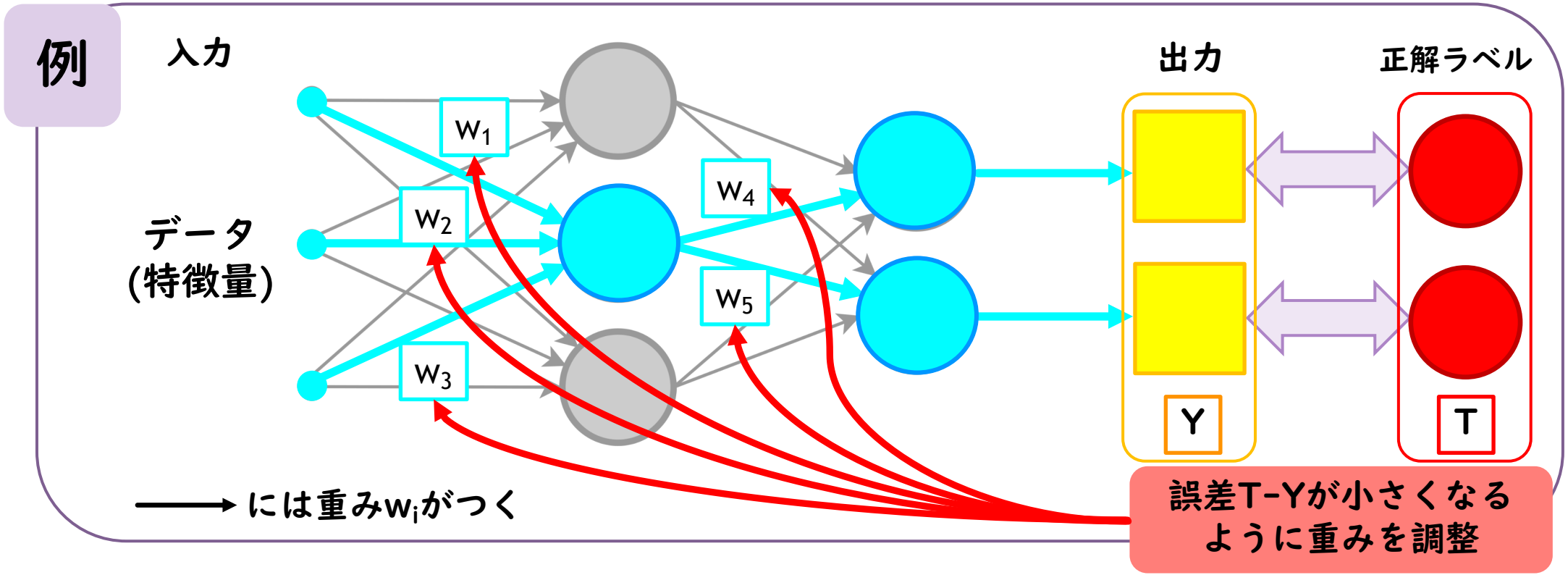


2.5 ニューラルネットワーク

ニューラルネットワークの学習

誤差逆伝播法

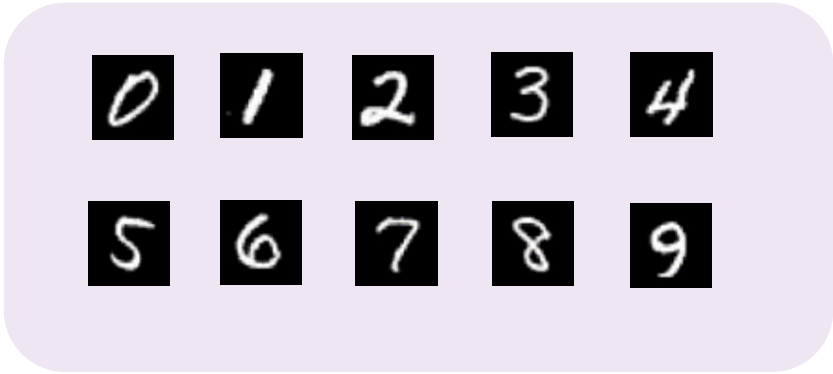
出力と正解ラベルの誤差を計算し，出力が正解ラベルに近づくように重みw_iを調整する

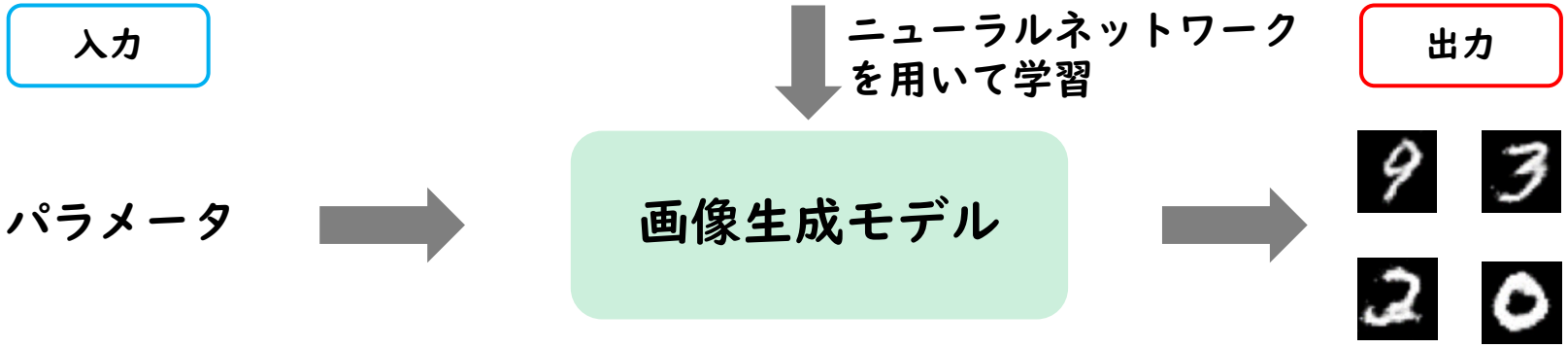


2.6 GAN

GAN(Generative Adversarial Networks)とは

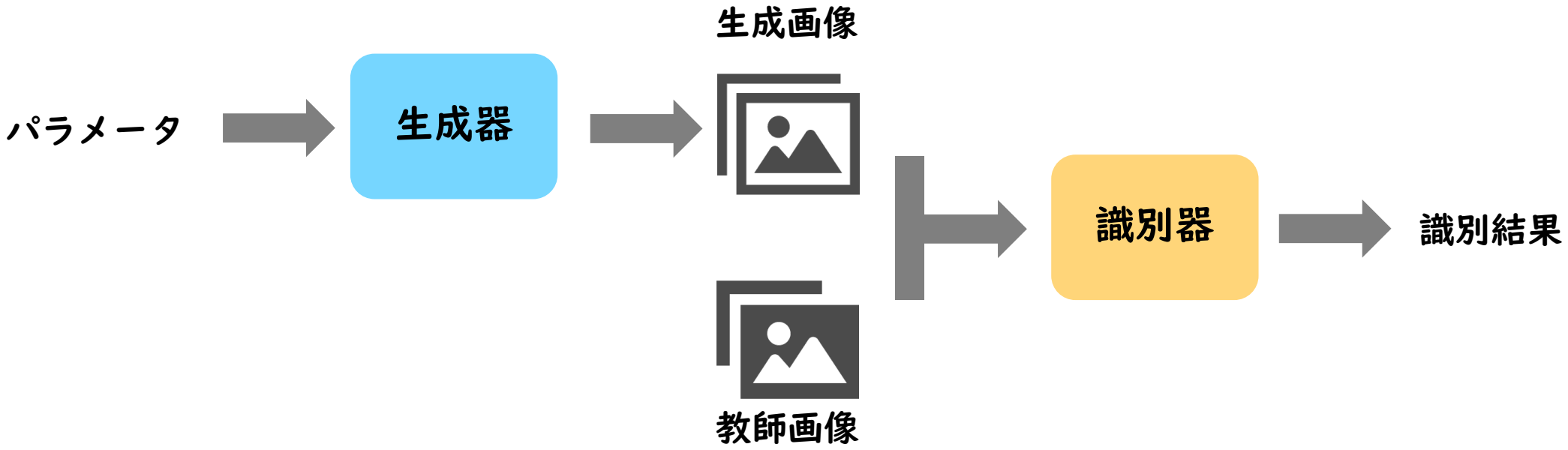
正解ラベルのついた学習データの集合(教師画像)から学習を行い，パラメータを受け取り画像を生成して出力する画像生成モデルを，2つのニューラルネットワークで構築する手法.





2.6 GAN

GANは教師画像から生成器と識別器という
2つのニューラルネットワークを競わせながら学習を進める.

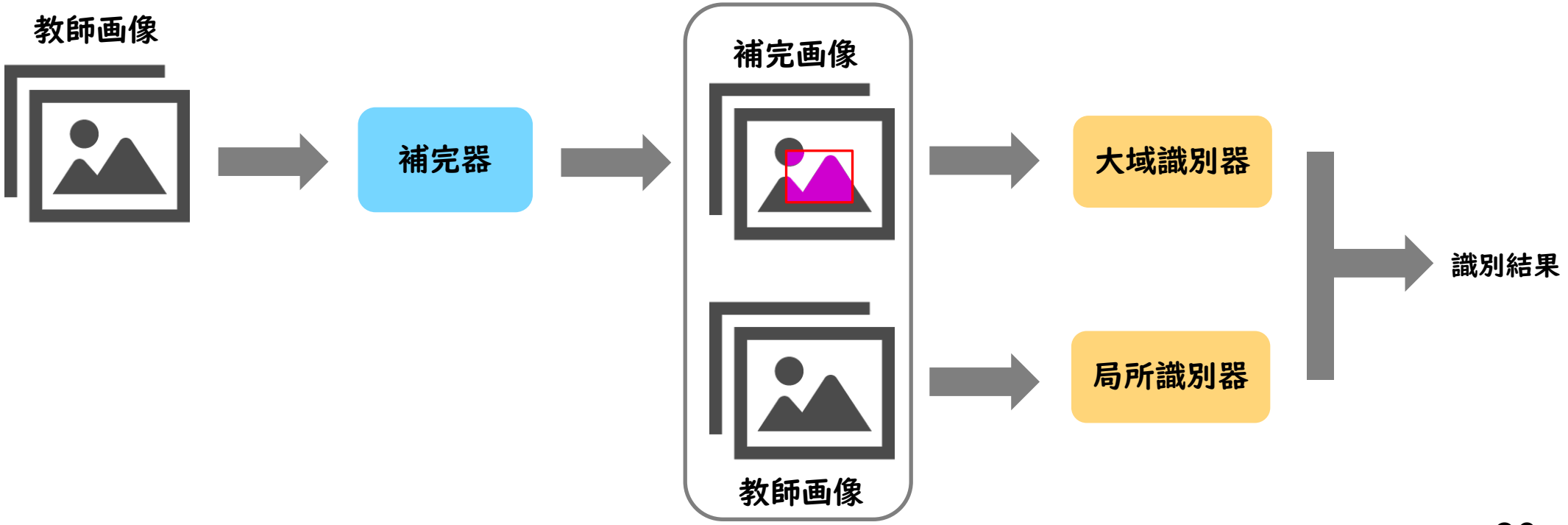


2.7 GLCIC

GLCIC(Globally and Locally Consistent Image Completion)とは，
正解ラベルのついた学習データの集合(教師画像)から学習を行い，画像における欠損の補完を行う画像補完モデルを，3つのニューラルネットワークから構築する手法.

2.7 GLCIC

GLCICは教師画像から，補完器というニューラルネットワークと，大域識別器と局所識別器という2つのニューラルネットワークを競わせながら学習を進めていく．



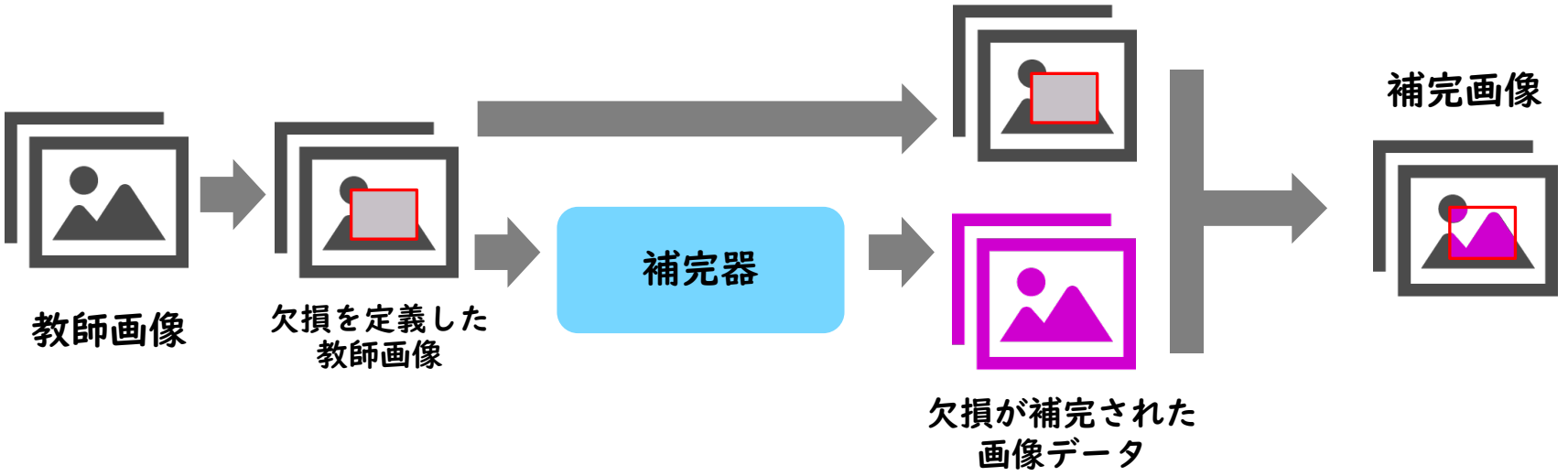
2.7 GLCIC

補完器

欠損を定義した教師画像を入力として受け取り，欠損が補完された画像データを生成.

入力：欠損を定義した教師画像
出力：欠損が補完された画像データ

補完した欠損部分以外を教師画像に置き換えた画像データを補完画像とする.



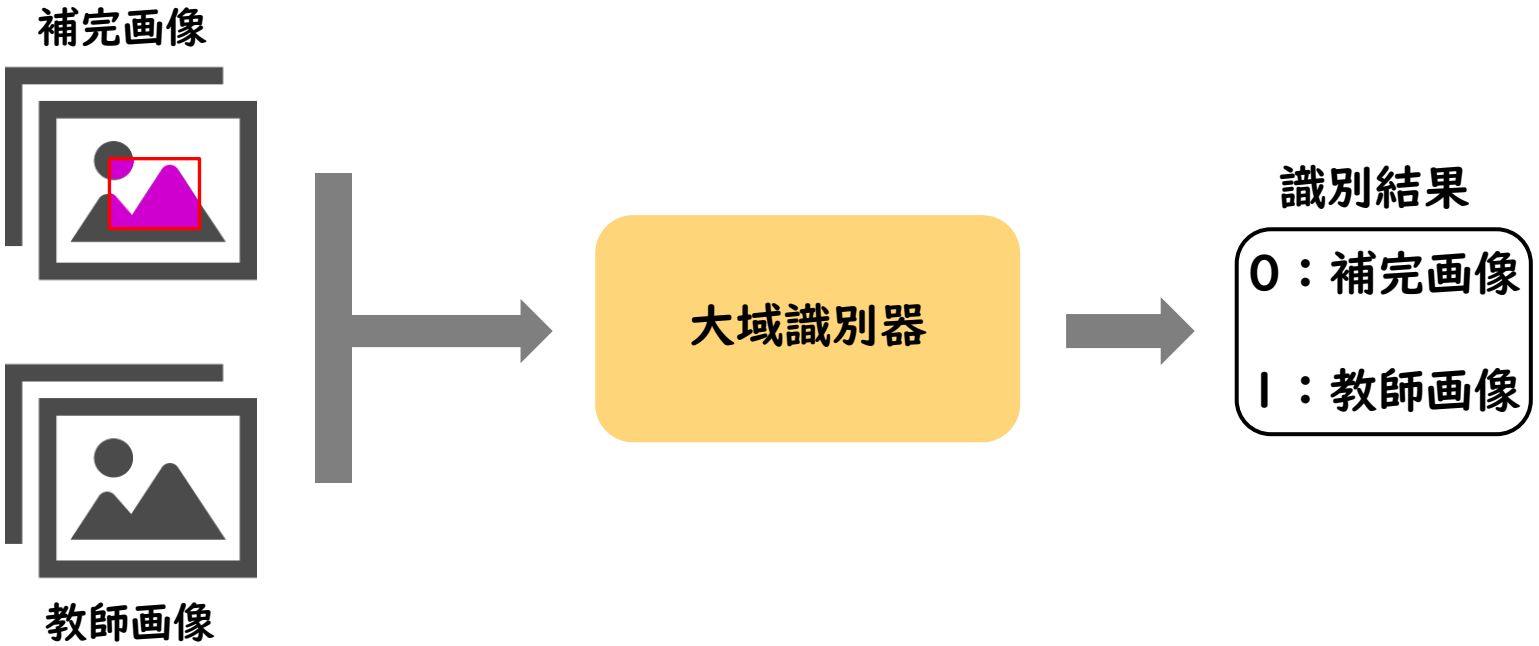
2.7 GLCIC

大域識別器

画像全体が自然な画像になっているかを評価

画像データを入力として受け取り，識別結果を出力

入力：正誤のラベルのついていない画像データ
出力：入力が教師画像であると判断した場合は1，そうでない場合は0



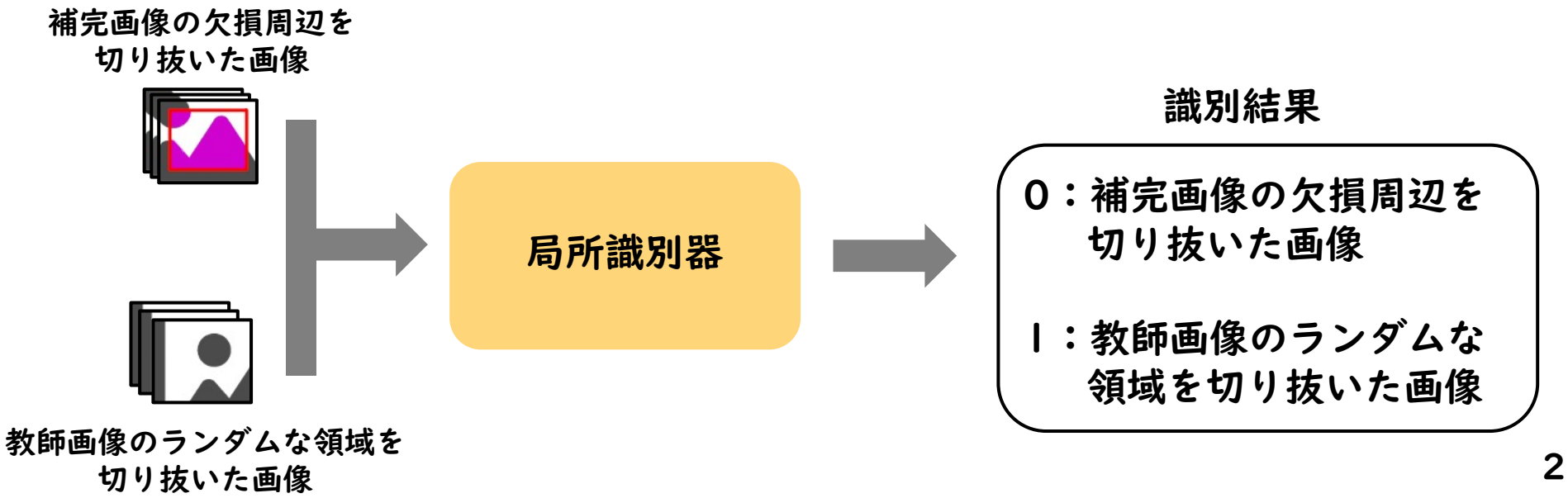
2.7 GLCIC

局所識別器

詳細な整合性によって
画像を評価

補完画像の補完を行った欠損周辺を切り抜いた画像，または教師画像のランダムな領域を
切り抜いた画像を受け取り，識別結果を出力．

入力：正誤のラベルのついていない画像データ
出力：入力が教師画像のランダムな領域を切り抜いた画像であると判断した場合は1，
　　そうでない場合は0



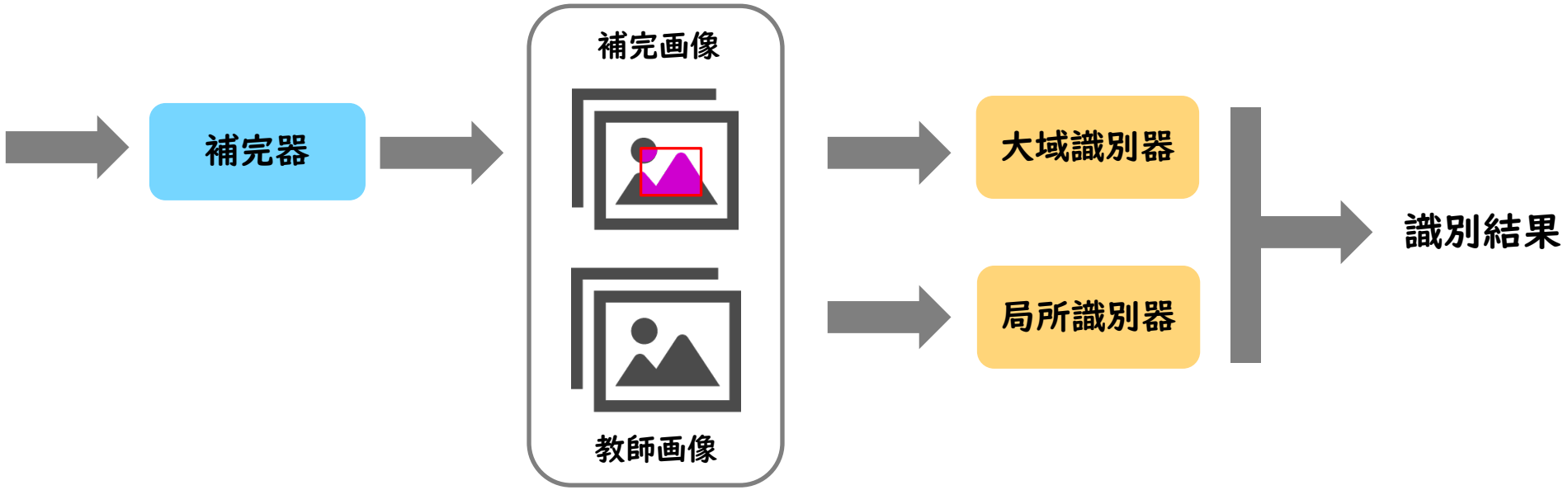
2.7 GLCIC

学習の流れ

手順１，２を交互に繰り返すことで学習を進める．
学習には誤差逆伝播法を用いる．

１．大域識別器と局所識別器を学習

２．補完器を学習



2.7 GLCIC

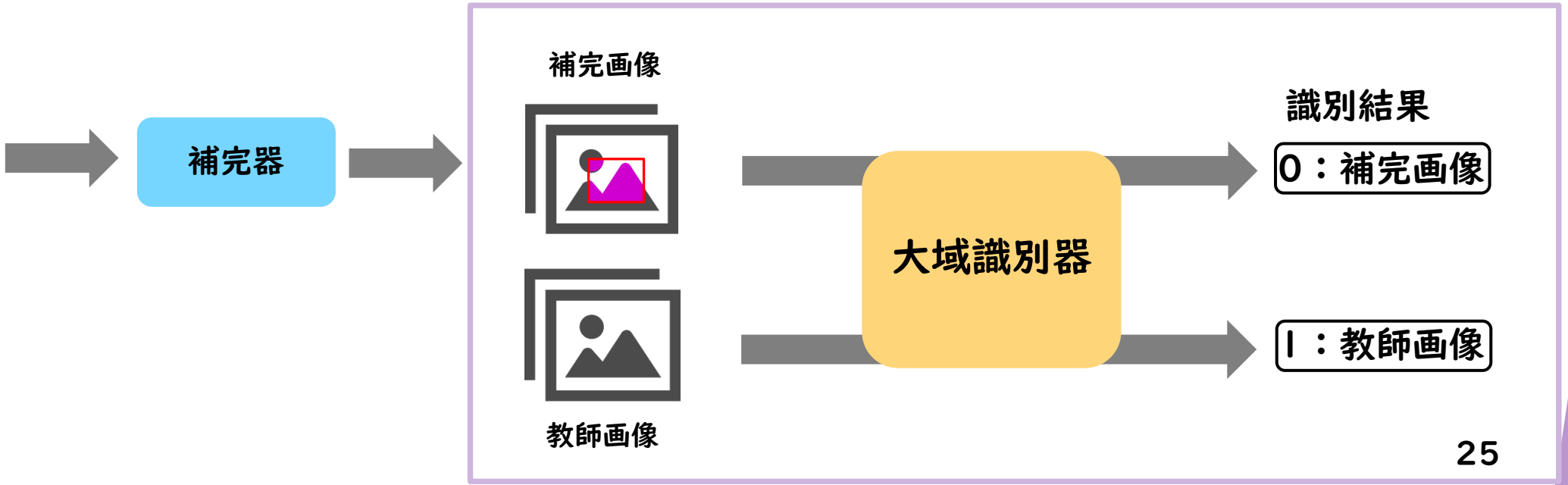
学習の流れ

手順１，２を交互に繰り返すことで学習を進める．
学習には誤差逆伝播法を用いる．

1. 大域識別器と局所識別器を学習

2. 補完器を学習

補完画像は0，教師画像は1と出力することを目指して学習



2.7 GLCIC

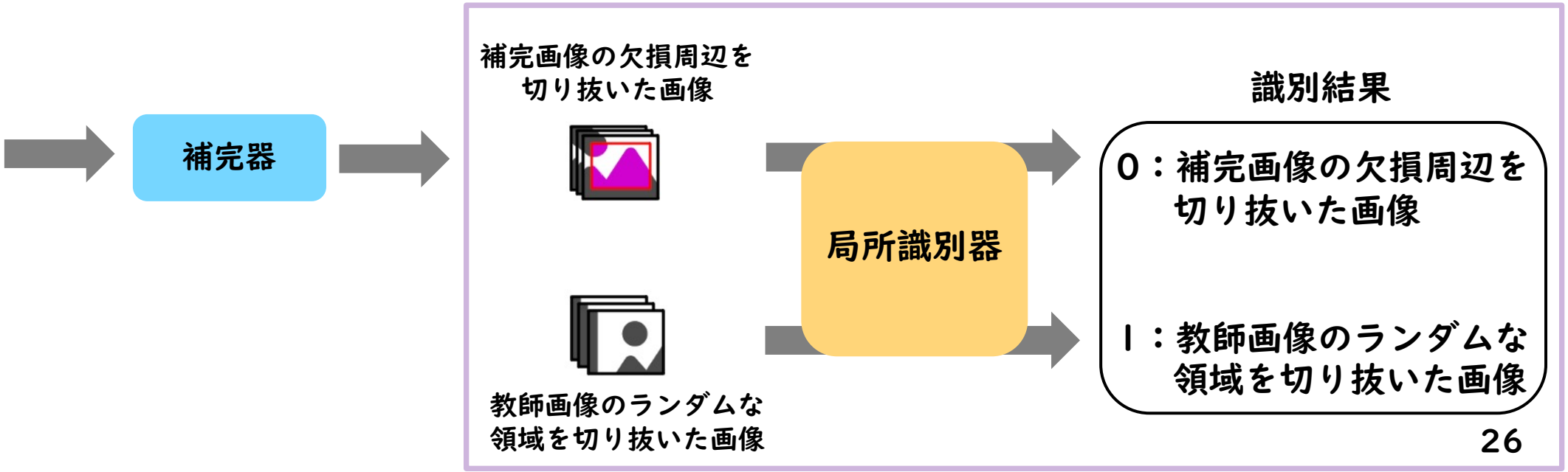
学習の流れ

手順１，２を交互に繰り返すことで学習を進める．
学習には誤差逆伝播法を用いる．

１．大域識別器と局所識別器を学習

２．補完器を学習

教師画像のランダムな領域を切り抜いた画像は1，
そうでない場合0と出力することを目指して学習



2.7 GLCIC

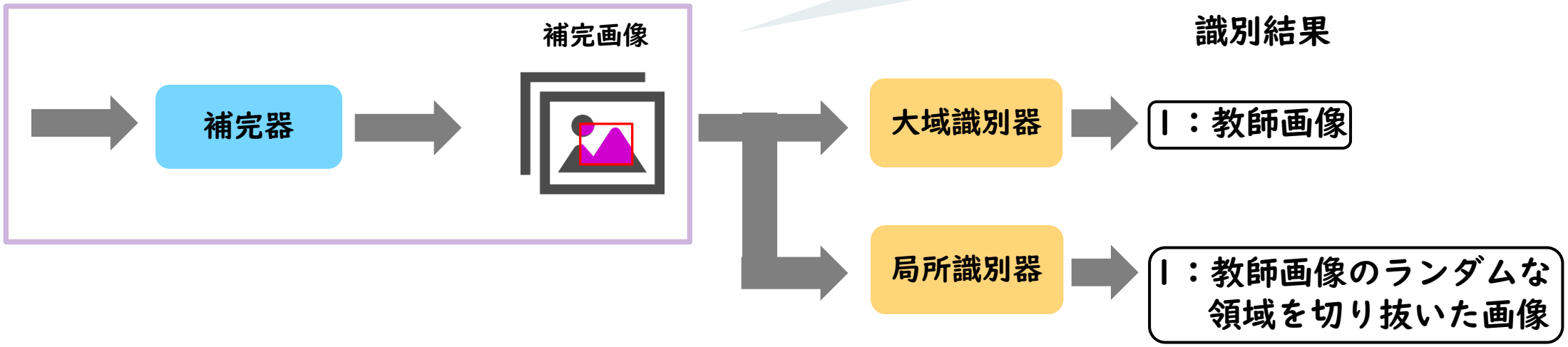
学習の流れ

手順１，２を交互に繰り返すことで学習を進める．
学習には誤差逆伝播法を用いる．

１．大域識別器と局所識別器を学習

２．補完器を学習

大域識別器と局所識別器が補完画像を１と出力することを目指して学習



目次

1. はじめに
   1.1 動機
   1.2 先行研究
   1.3 演習目的

2. 準備
   2.1 機械学習
   2.2 モデル
   2.3 機械学習における学習
   2.4 教示あり学習
   2.5 ニューラルネットワーク
   2.6 GAN
   2.7 GLCIC

3. 演習
   3.1 環境構築
   3.2 学習データセットの準備
   3.3 補完モデルの作成

4. 評価

5. まとめ

3.1 環境構築

開発環境：Google Colab

使用言語：Python

Google Colabとは？

Googleの提供しているサービス

- ブラウザからPythonを実行することができる
- 環境構築が殆ど必要無い
- 高スペックなPCを要求されない
- 安価で高機能なGPUを活用することができる

Pythonとは？

データ解析や機械学習の分野で広く使われているプログラミング言語

- 可読性が高い
- ライブラリが豊富
- Google Colabの主要言語

3.2 学習データセットの準備

画像データの収集

Google画像検索から
Webスクレイピングで人物画像を収集

検索キーワード
「日本人」「日本人 顔」「日本人 女性」「日本人 男性」
「顔」「芸能人」「女優」「俳優」「男」「女」

計1000枚の画像を収集

3.2 学習データセットの準備

画像データの絞り込み

以下の条件から主観で画像を絞り込む(手作業)

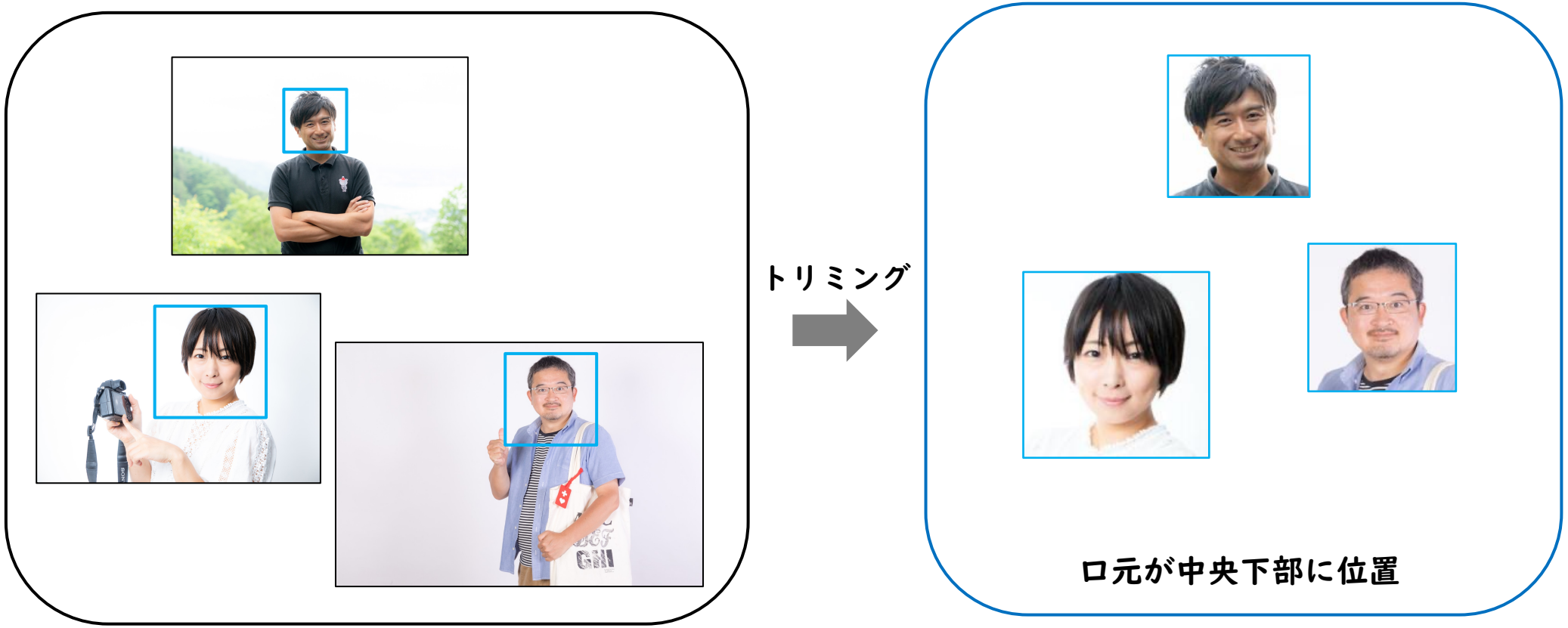
- 口元の写っていない画像を除去
- 同一人物が対象とならないように画像を除去
- サイズの小さい画像を除去
  (顔部分が128×128ピクセル以下)

収集した画像1000枚→528枚

3.2 学習データセットの準備

画像データの正規化

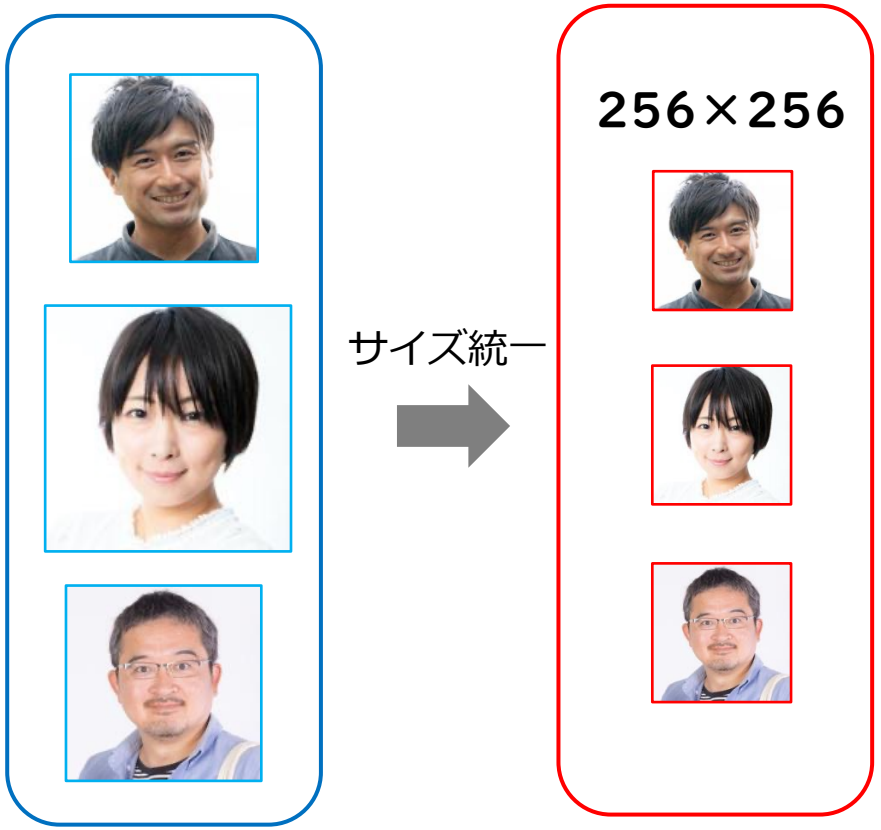
画像に対して口元が中央下部に位置するようトリミング(手作業)



3.2 学習データセットの準備

画像データの正規化

画像処理ライブラリPillowを用いて画像サイズを256×256ピクセルへ統一



3.2 学習データセットの準備

収集枚数	正規化後枚数
1000枚	528枚

学習データ	テストデータ
428枚	100枚

3.3 画像補完モデルの作成

使用ライブラリ

OpenCV	画像解析ライブラリ
Keras	機械学習ライブラリ
PyTorch	機械学習ライブラリ
numpy	データ分析ライブラリ
Pillow	画像処理ライブラリ
selenium	Webブラウジングライブラリ

3.3 画像補完モデルの作成

転移学習

学習済みのモデルを利用して学習を行う手法.
ある領域で学習させたモデルを、別の領域に適応させることが可能.

Google Colabには連続稼働時間に制限がある

一度の稼働で約80000回の学習まで

転移学習を用いて前回の学習を持ち越すことで連続稼働時間制限の問題を回避

3.3 画像補完モデルの作成

画像補完モデルの構築

口元を補完する画像補完モデル

入力：補完したい画像
出力：欠損を補完した画像
学習回数：480000回(80000回×6)

以下を参照

[1] SATOSHI IIZUKA, EDGAR SIMO-SERRA,　HIROSHI ISHIKAWA
Globally and locally consistent image completion. SIGGRAPH July 2017

[2] GLCICで無かったことにしたいアレコレ（GANを使った画像生成を Globally and Locally Consistent Image Completion で理解してみる）
https://qiita.com/shtmr/items/ce4bb7fc4f719c8c6737

3.3 画像補完モデルの作成

学習方法

ランダムに欠損を発生させて学習(ランダム学習)

口元だけ欠損を発生させて学習(口元学習)

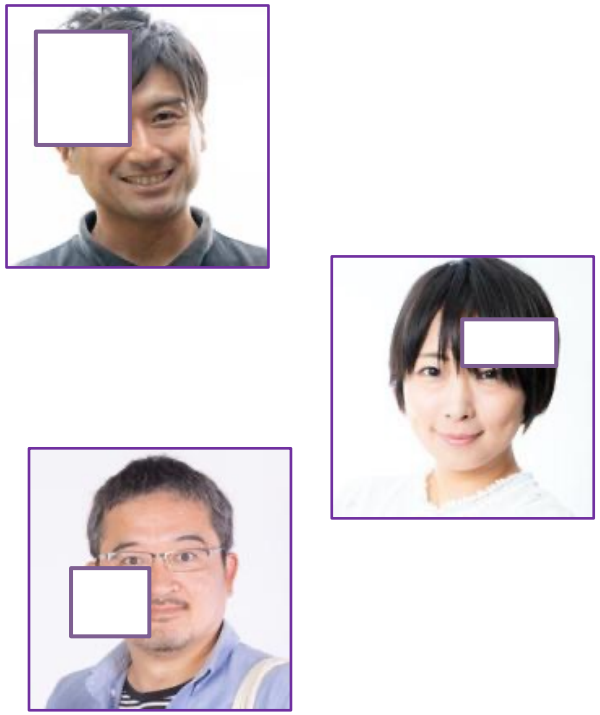
異なる学習法で2つの画像補完モデルの作成.
補完の質に差は生まれるのか？

3.3 画像補完モデルの作成

ランダムに欠損を発生させて学習(ランダム学習)

学習データセット

ランダムに欠損を発生



欠損を発生させる領域
座標：(5, 5)～(145,145)
幅　：48～96
高さ：48～96

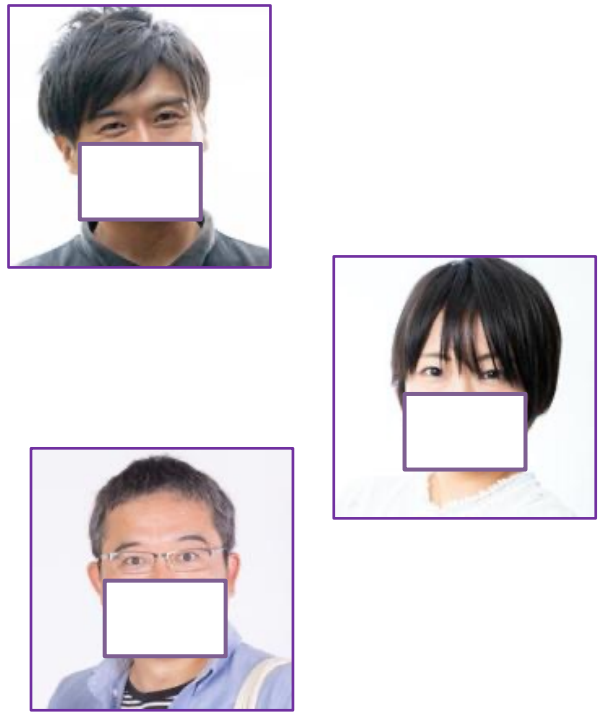
画像補完モデルの
学習に使用

3.3 画像補完モデルの作成

口元だけ欠損を発生させて学習(口元学習)

学習データセット

口元だけ欠損を発生



欠損を発生させる領域
座標：(56, 128)
幅　：144
高さ：96

画像補完モデルの
学習に使用

目次

1. はじめに
   1.1 動機
   1.2 先行研究
   1.3 演習目的

2. 準備
   2.1 機械学習
   2.2 モデル
   2.3 機械学習における学習
   2.4 教示あり学習
   2.5 ニューラルネットワーク
   2.6 GAN
   2.7 GLCIC

3. 演習
   3.1 環境構築
   3.2 学習データセットの準備
   3.3 補完モデルの作成

4. 評価

5. まとめ

4. 評価

先行研究では

	先行研究	今回の演習
学習データセットの用意	CelebAデータセットの画像	Webスクレイピングで収集した人物画像
学習データセットの枚数	200000枚	528枚
学習期間	2ヶ月	1週間
学習環境	4台のK80 GPUを搭載した1台のマシン	1台のK80 GPUを搭載した1台のマシン

CelebAデータセット(CelebFaces Attributes Dataset)とは
有名人の顔画像をカラーで202599枚集めたデータセット

(先行研究と比べて)十分な学習を行えていない.

4. 評価

生成結果

作成した2つの画像補完モデルで画像補完

補完したい画像

口元を欠損

ランダム学習

口元学習

2種類の観点から評価

4. 評価

主観的評価

アンケート

・補完元の画像
・ランダム学習によって作成した画像補完モデルで口元を補完した画像
・口元学習によって作成した画像補完モデルで口元を補完した画像

どの画像補完モデルで補完したか伏せた状態で3枚の画像を見せ，
補完の質の高い画像を選んでもらう．

4. 評価

主観的評価

問題数

10問

選択肢(アンケートを行う際は画像を上下に表示している)

1.上の方が似ている
2.上の方がどちらかというと似ている
3.どちらも同じくらいの補完である
4.下の方がどちらかというと似ている
5.下の方が似ている

アンケート対象

10代から60代までの男女10名

実施回数

学習回数320000回時点と400000回時点の2回実施

4. 評価

主観的評価

アンケート結果を集計

評価の優れている問題数

	学習回数320000回	学習回数400000回
ランダム学習	2	3
口元学習	7	7
同等	1	0

口元学習で作成された画像補完モデルの方が評価が高い

4. 評価

客観的評価

特徴点マッチング

画像解析ライブラリOpenCVを用いて異なる画像間で角や輪郭など特徴を指す固有の点を対応付ける手法.
ターゲットと比較対象の画像を渡すことで類似度を表す値を返す．類似度は値が小さいほど2枚の画像の差が小さい，似ていることを示す.

実行結果の例

```
TARGET_FILE: target.jpg
b.jpg 66.49019607843137
a.jpg 59.78921568627451
```

4. 評価

客観的評価

主観的評価のアンケートの際に使用した10枚の画像を使用

ターゲット画像

補完元の画像

比較対象

・ランダム学習によって作成した画像補完モデルで口元を補完した画像
・口元学習によって作成した画像補完モデルで口元を補完した画像

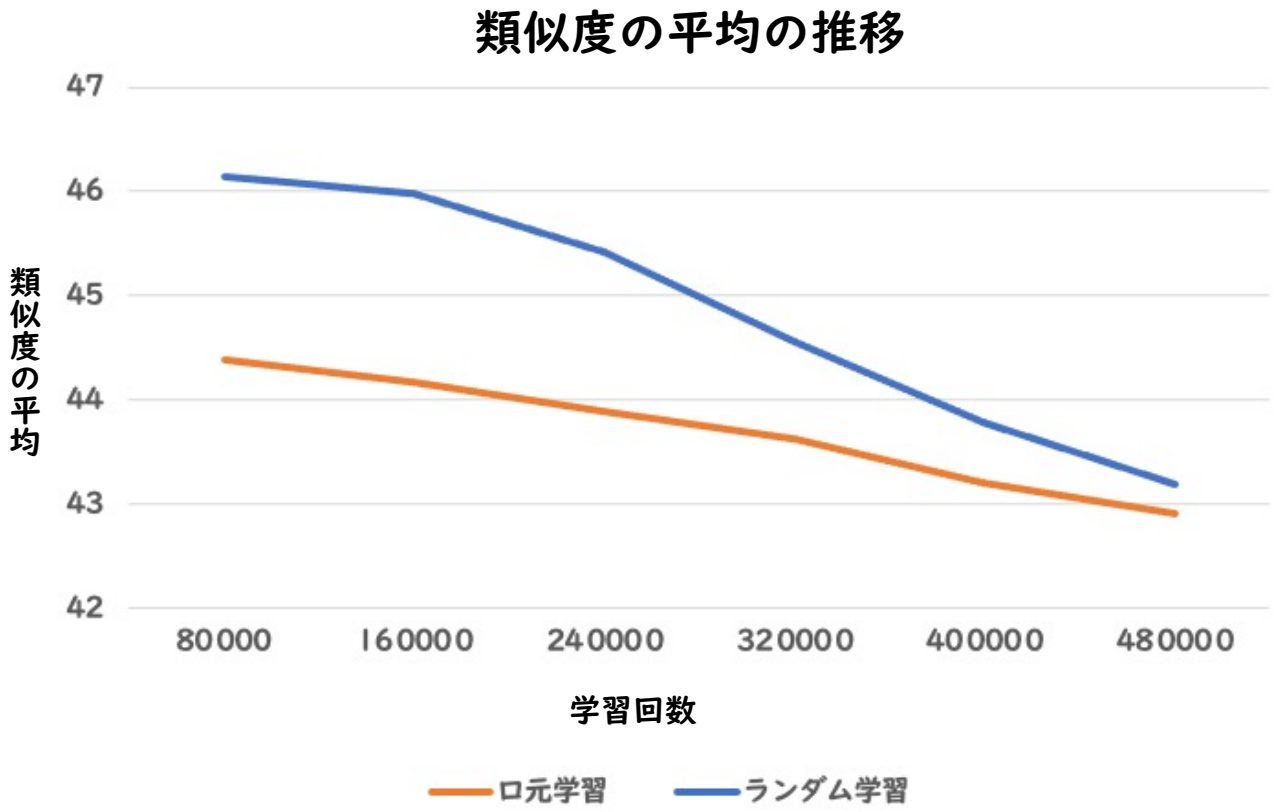
学習回数80000回ごとに類似度を比較

4. 評価

客観的評価

ロ元学習で作成された画像補完モデルの方が，類似度が優れている傾向にある．

ランダム学習で作成された画像補完モデルの方が，今後類似度の更新が見込めると推察される．



目次

1. はじめに
   1.1 動機
   1.2 先行研究
   1.3 演習目的

2. 準備
   2.1 機械学習
   2.2 モデル
   2.3 機械学習における学習
   2.4 教示あり学習
   2.5 ニューラルネットワーク
   2.6 GAN
   2.7 GLCIC

3. 演習
   3.1 環境構築
   3.2 学習データセットの準備
   3.3 補完モデルの作成

4. 評価

5. まとめ

5. まとめ

考察

ランダムに欠損を発生させて学習 → 口元だけ欠損を発生させて学習

補完の質に差は生まれるのか？

5. まとめ

考察

今回作成した2種類の画像補完モデルはどちらの質も低く，差は僅かであった．

ロ元学習で作成した画像補完モデルはランダム学習に比べて主観的評価，客観的評価どちらにおいても画像補完の質が高いと評価された．

客観的評価における類似度の推移より，ロ元学習よりもランダム学習の方が類似度の更新が見込めると推察される．

5. まとめ

今後の課題

先行研究との比較

	先行研究	今回の演習
学習データセットの用意	CelebAデータセットの画像	Webスクレイピングで収集した人物画像
学習データセットの枚数	200000枚	528枚
学習期間	2ヶ月	1週間
学習環境	4台のK80 GPUを搭載した1台のマシン	1台のK80 GPUを搭載した1台のマシン

5. まとめ

今後の課題

・学習データセットの増量

・学習が本演習時点より進んだ際の質の観察

・収集した画像データの偏りを無くす

Ex. 評価

主観的評価

集計結果(学習回数320000回時点)

回答者No	画像1	画像2	画像3	画像4	画像5	画像6	画像7	画像8	画像9	画像10
1	3	4	4	4	2	4	3	4	4	4
2	2	4	3	3	2	4	2	4	2	3
3	4	4	2	2	2	4	3	4	3	4
4	1	3	4	3	4	4	2	4	3	3
5	2	4	2	3	2	3	4	5	2	3
6	5	5	1	4	3	3	4	4	4	4
7	5	5	5	3	3	4	3	3	4	3
8	5	5	3	4	2	3	3	2	1	3
9	5	4	5	3	1	1	2	5	5	3
10	1	5	1	4	2	5	3	4	3	3
平均	3.3	4.3	3	3.3	2.3	3.5	2.9	3.9	3.1	3.3
	口元	口元	五分	口元	ランダム	口元	ランダム	口元	口元	口元

五分	1
ランダム	2
口元	7

スコアの平均が3より高いほど，口元学習によって作成された
画像補完モデルの方が補完の質が高いことを示している．

Ex. 評価

主観的評価

集計結果(学習回数400000回時点)

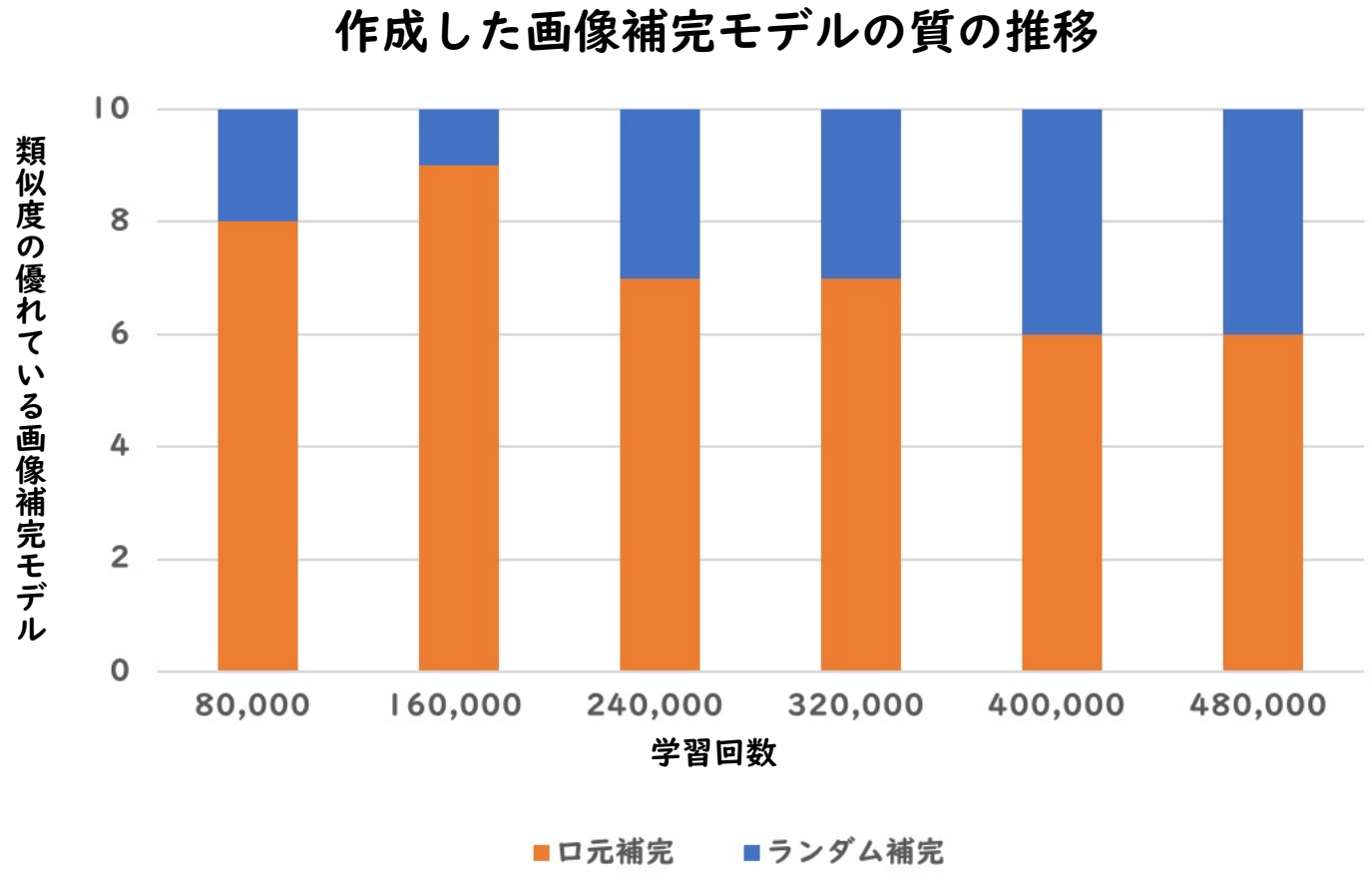
回答者No	画像1	画像2	画像3	画像4	画像5	画像6	画像7	画像8	画像9	画像10
1	4	4	3	3	2	4	2	4	2	3
2	2	4	2	2	2	4	3	4	3	4
3	5	3	4	3	4	4	2	4	3	3
4	4	4	2	3	2	3	4	5	2	3
5	1	5	1	4	3	3	4	4	4	4
6	3	4	4	4	2	4	3	4	4	4
7	4	5	4	5	4	5	3	1	1	3
8	1	5	3	3	1	5	3	5	5	3
9	1	5	5	3	2	3	3	3	4	5
10	4	5	5	3	2	4	2	4	3	2
平均	2.9	4.4	3.3	3.3	2.4	3.9	2.9	3.8	3.1	3.4
	ランダム	口元	口元	口元	ランダム	口元	ランダム	口元	口元	口元

ランダム	3
口元	7

スコアの平均が3より高いほど，口元学習によって作成された
画像補完モデルの方が補完の質が高いことを示している．

Ex. 評価

客観的評価


作成した画像補完モデルの質の推移
類似度の優れている画像補完モデル
10
8
6
4
2
0
80,000
160,000
240,000
320,000
400,000
480,000
学習回数
口元補完
ランダム補完

Ex. GAN

生成器

パラメータを入力として受け取り画像データを生成

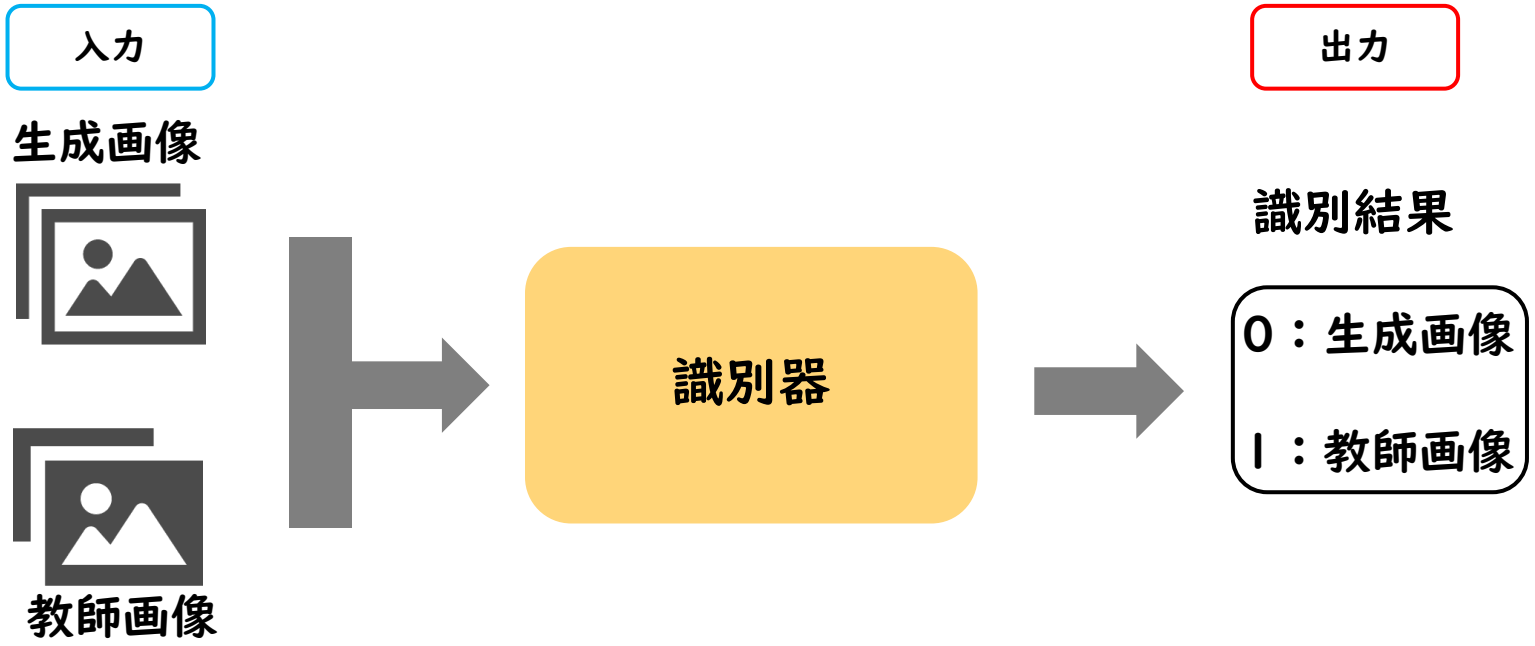
入力：パラメータ
出力：画像データ(生成画像とする)

Ex. GAN

識別器

画像データを入力として受け取り識別結果を出力

入力：正誤のラベルのついてない画像データ
出力：入力が教師画像であると判断した場合は1，そうでない場合は0



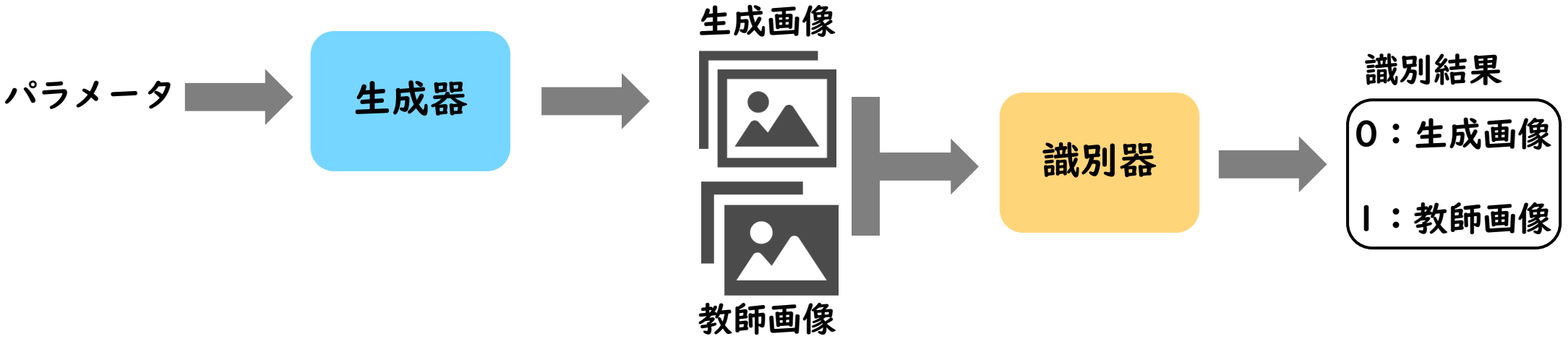
Ex. GAN

学習の流れ

手順１，２を交互に繰り返すことで学習を進める．
学習には誤差逆伝播法を用いる．

１．識別器を学習

２．生成器を学習



Ex. GAN

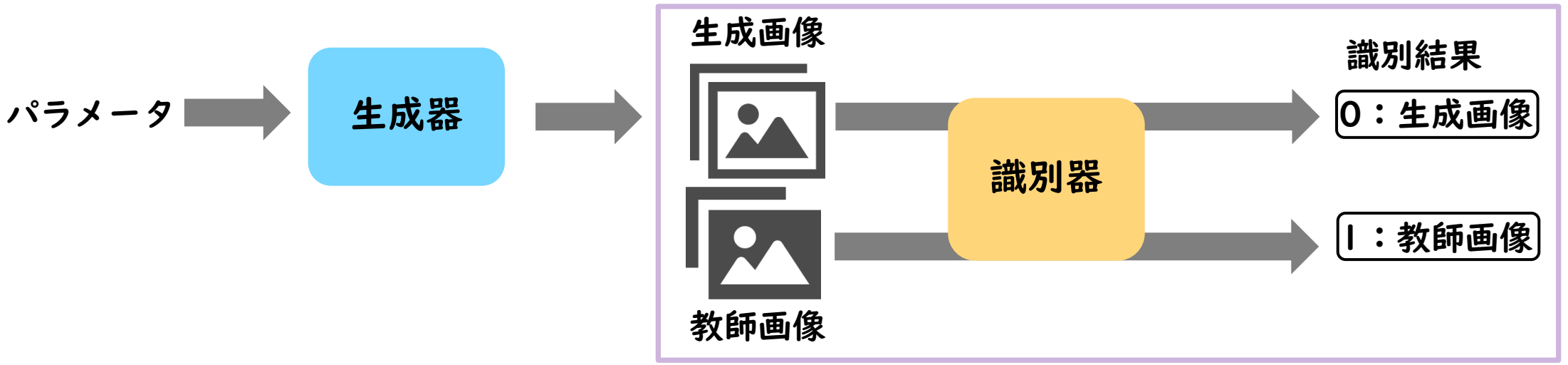
学習の流れ

手順１，２を交互に繰り返すことで学習を進める．
学習には誤差逆伝播法を用いる．

1. 識別器を学習

2. 生成器を学習

生成画像は0，教師画像は1と出力することを目指して学習



Ex. GAN

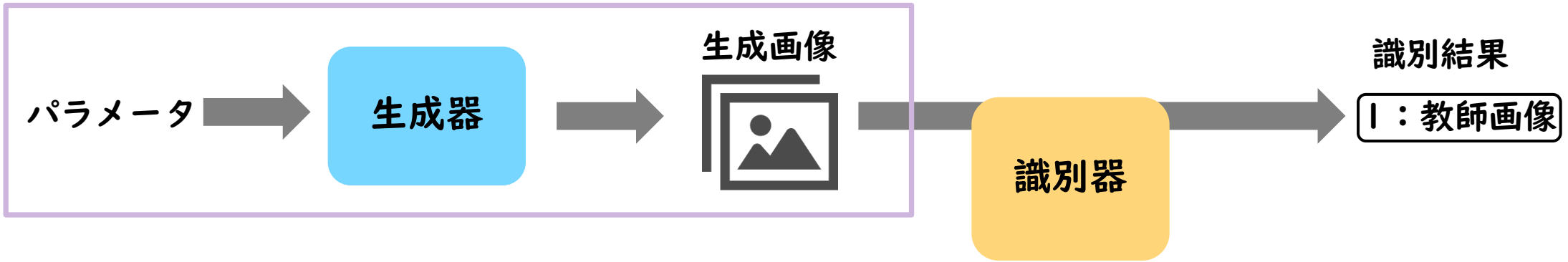
学習の流れ

手順１，２を交互に繰り返すことで学習を進める．
学習には誤差逆伝播法を用いる．

１．識別器を学習

２．生成器を学習

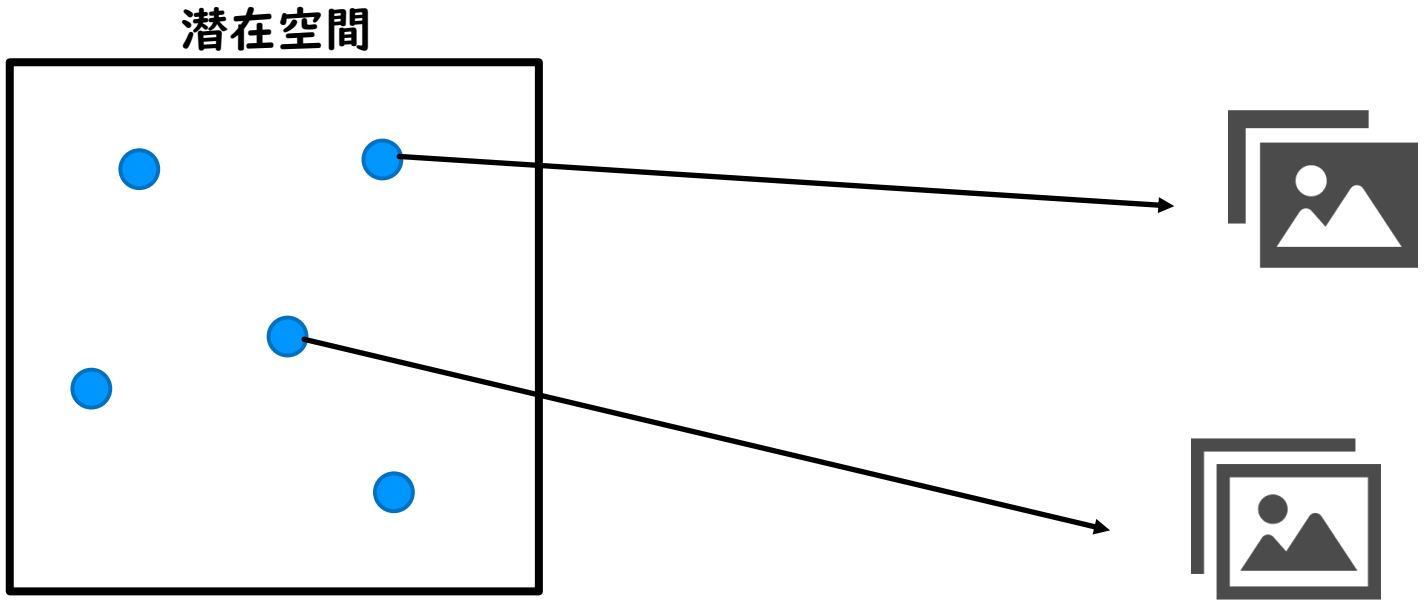
識別器が１と出力するような画像を生成することを目指して学習



Ex. 付録

パラメータとは？

ランダムにサンプリングされた多次元ベクトル

潜在空間

生成器の入力の多次元ベクトルが存在する空間
学習済みの生成器では潜在空間内の一点からは決まった画像が生成される