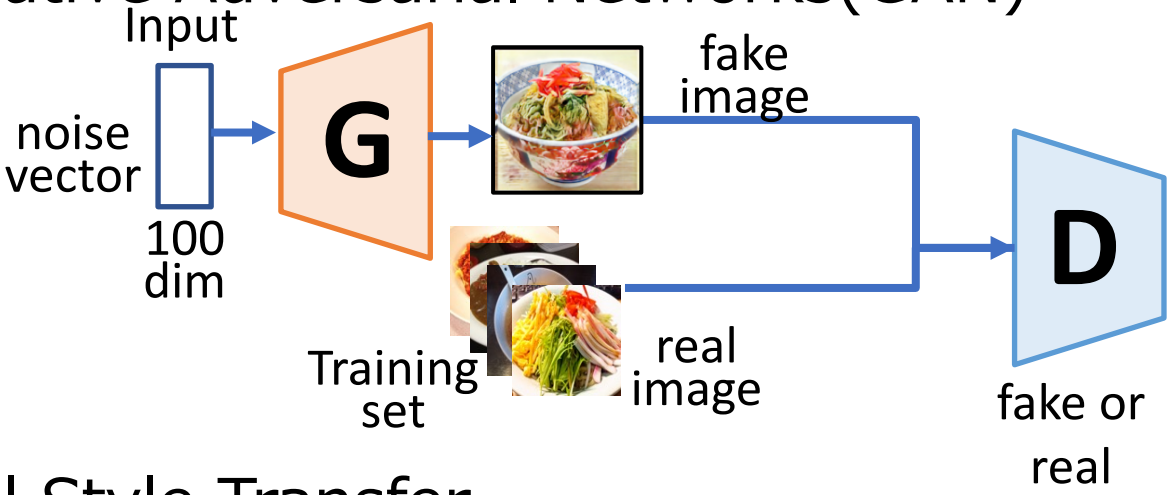
Conditional GANによる食事写真の属性操作

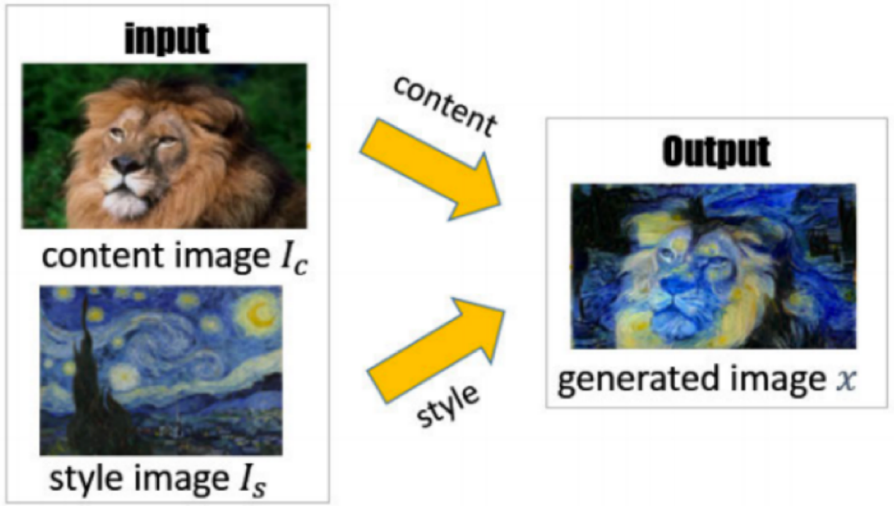
成冨志優† 堀田大地††
丹野良介††† 下田和††† 柳井啓司†††
電気通信大学 †情報理工学域1類 メディア情報学プログラム
†† 情報理工学域3類 機械システムプログラム
††† 大学院情報理工学研究科 情報学専攻

はじめに

- 深層学習を用いた画像生成・変換
  - Generative Adversarial Networks(GAN)

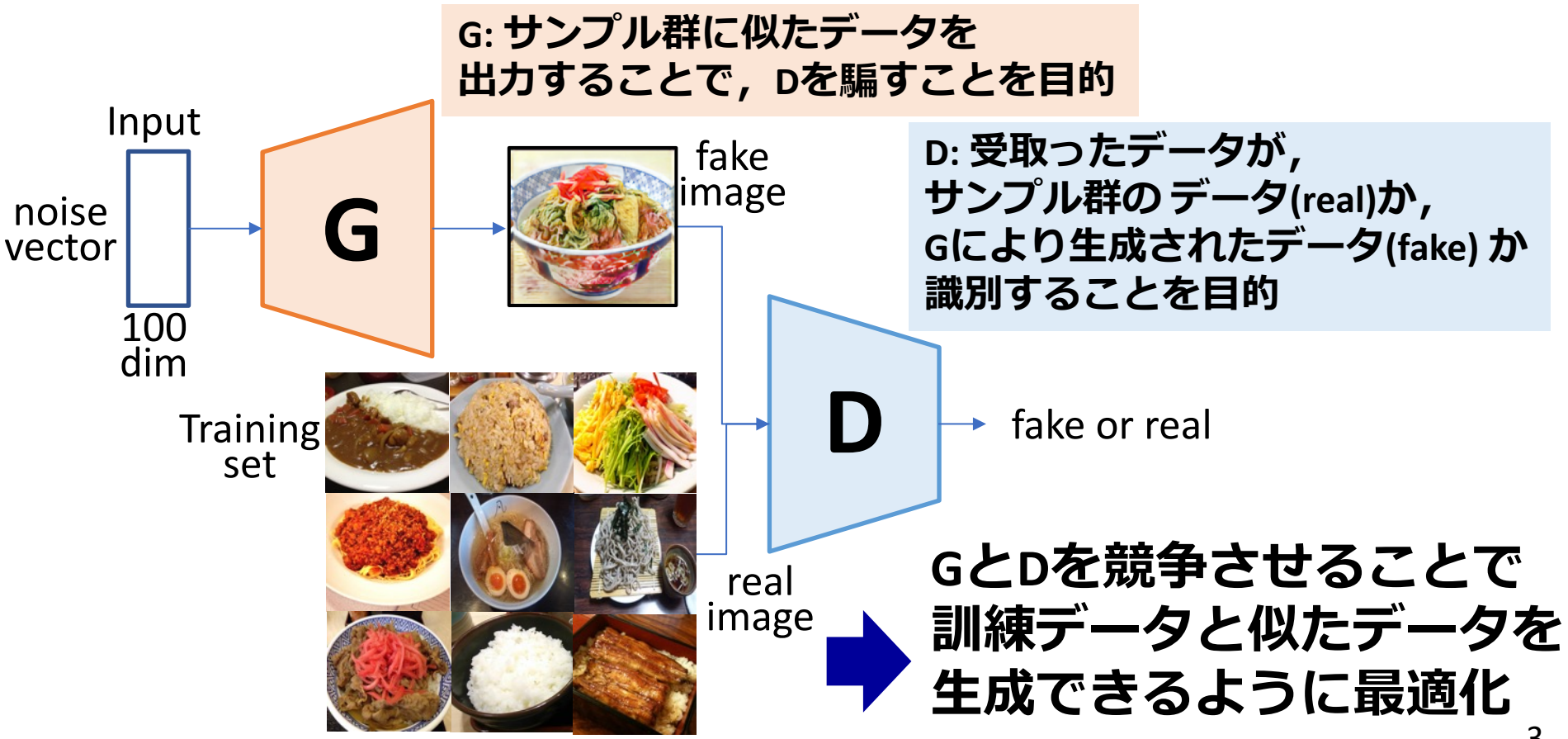


  - Neural Style Transfer



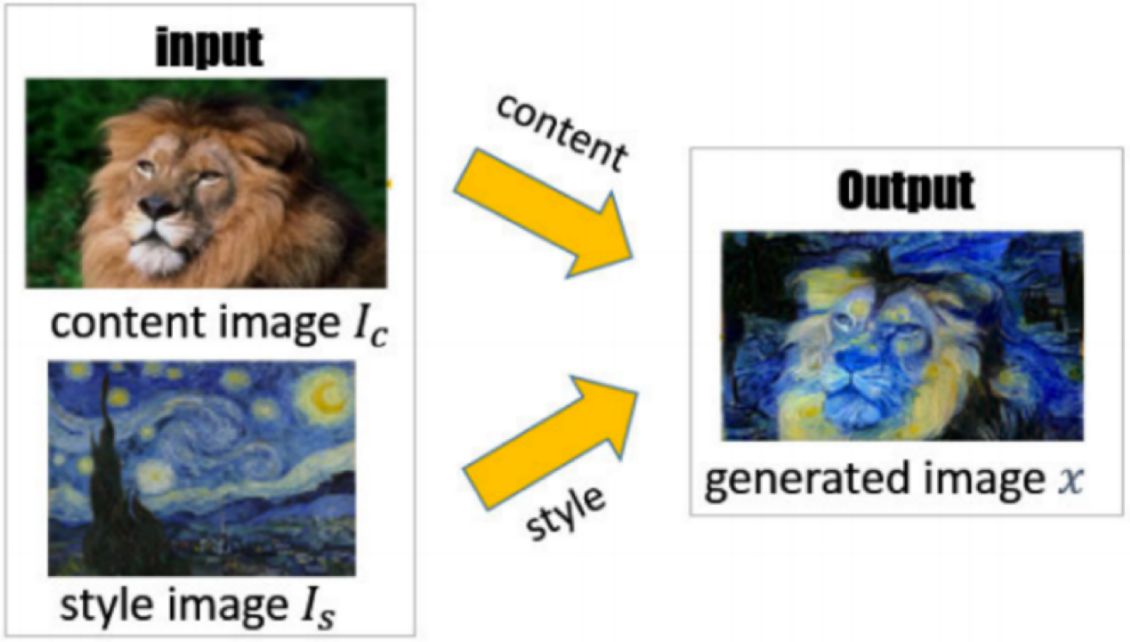
GANとは

- 生成モデルと深層学習を組合せた深層生成モデル
  - Generative Adversarial Network (GAN) [Goodfellow+ NIPS-14]



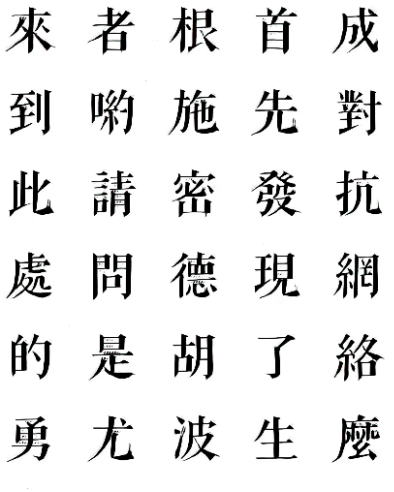
Style Transferとは

- 画風を変換するアルゴリズム(2015年8月に公開)
  - 2枚の画像を入力として，片方をコンテンツ画像，片方をスタイル画像とする
  - コンテンツ画像に書かれた物体の配置をそのままにして，画風をスタイル画像に変換した画像を生成
  - AIによる芸術画像の生成として研究が盛んに



関連研究

- 画像生成、変換の分野では文字画像、顔画像、部屋画像などの研究が行われている。



食事画像ではやられていない

食事画像にも応用したい…！！

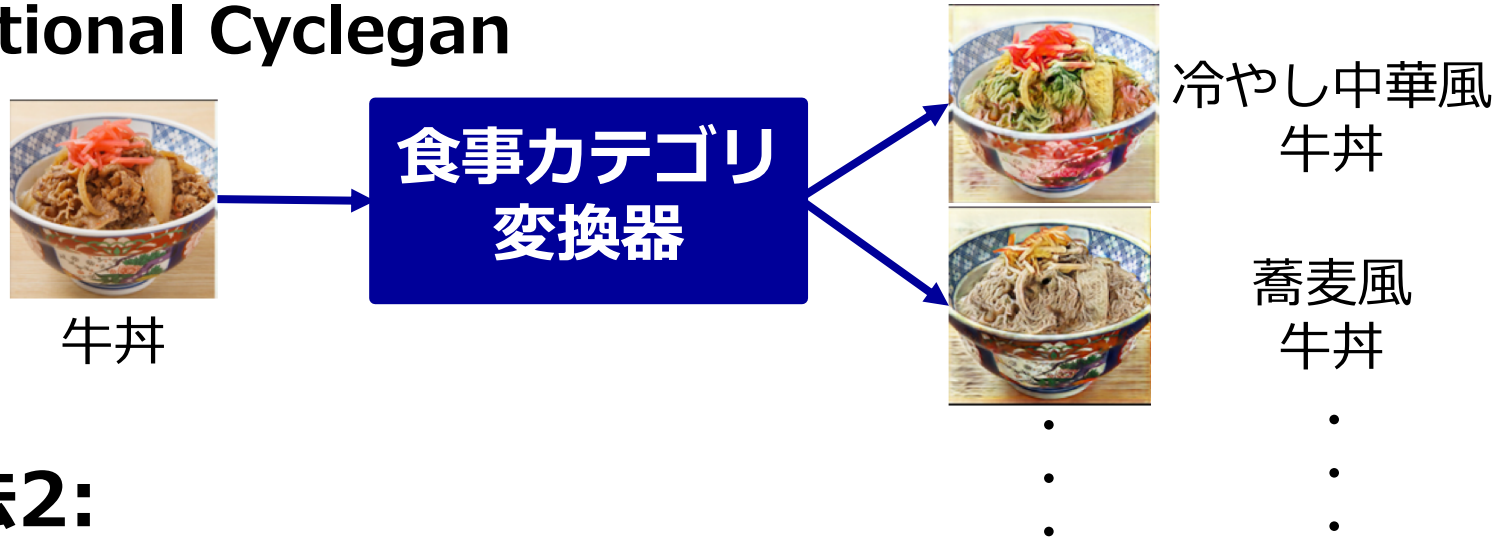
研究目的

食事画像の生成・変換において

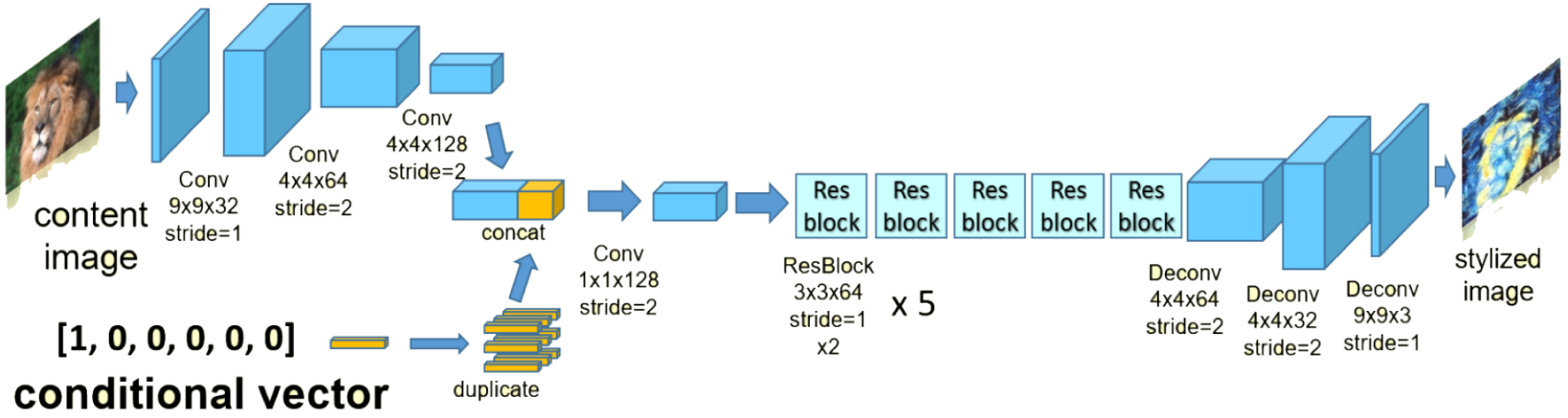
最適な手法の検討

手法概要

- **変換手法1:**
  - **Conditional Cyclegan**

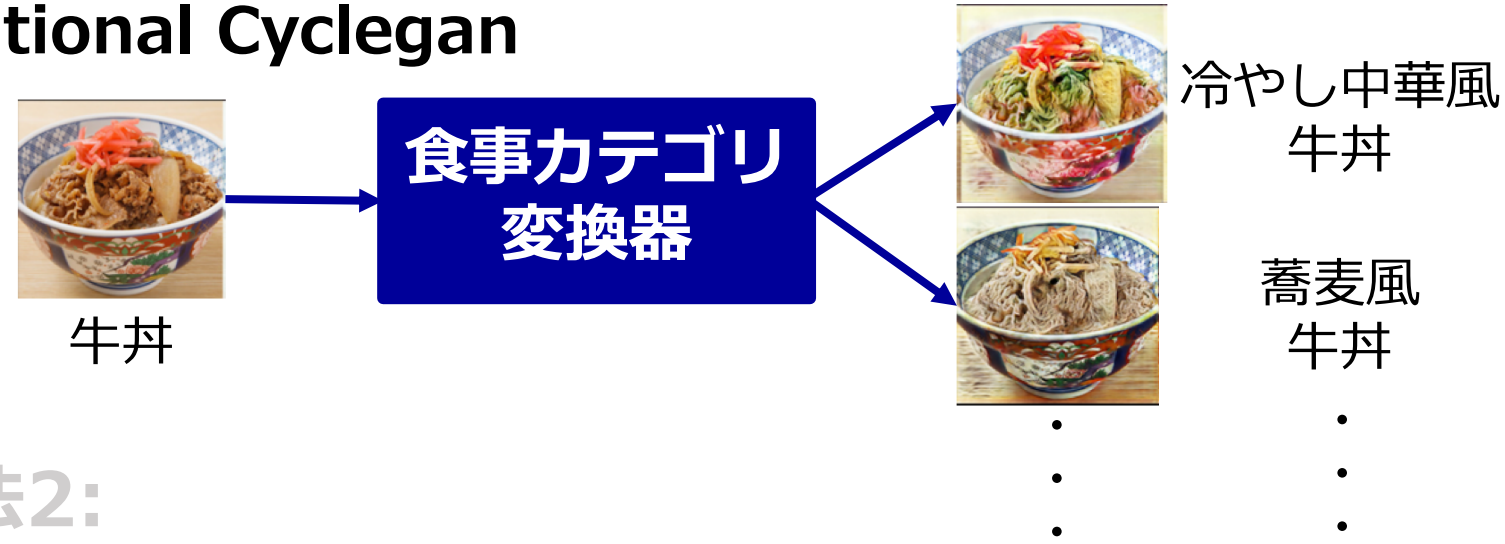


- **変換手法2:**
  - **Multi Style Transfer** [Tanno+ MMM2017]

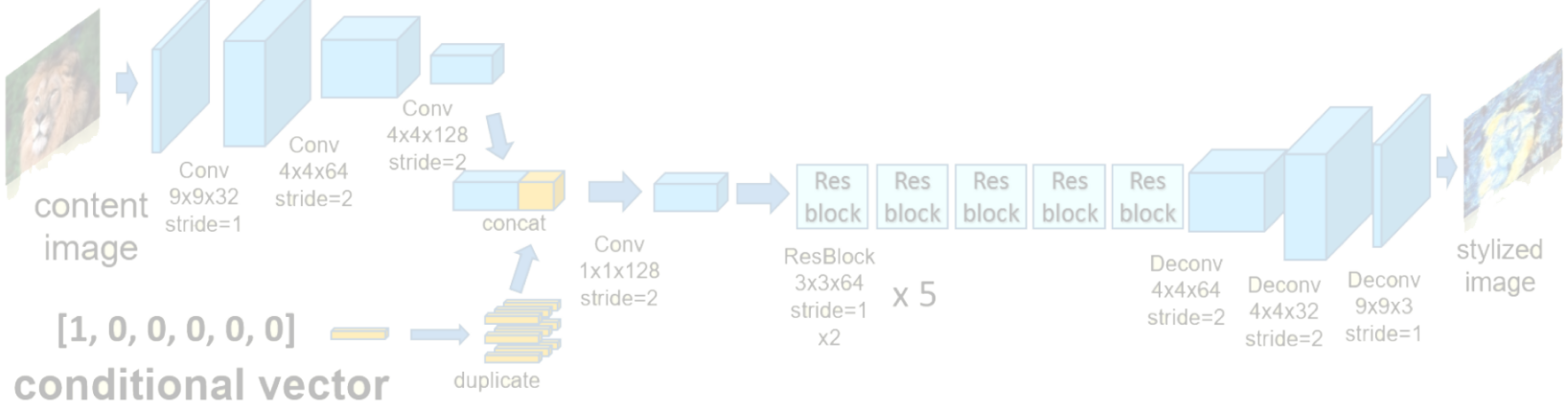


手法概要

- **変換手法1:**
  - **Conditional Cyclegan**

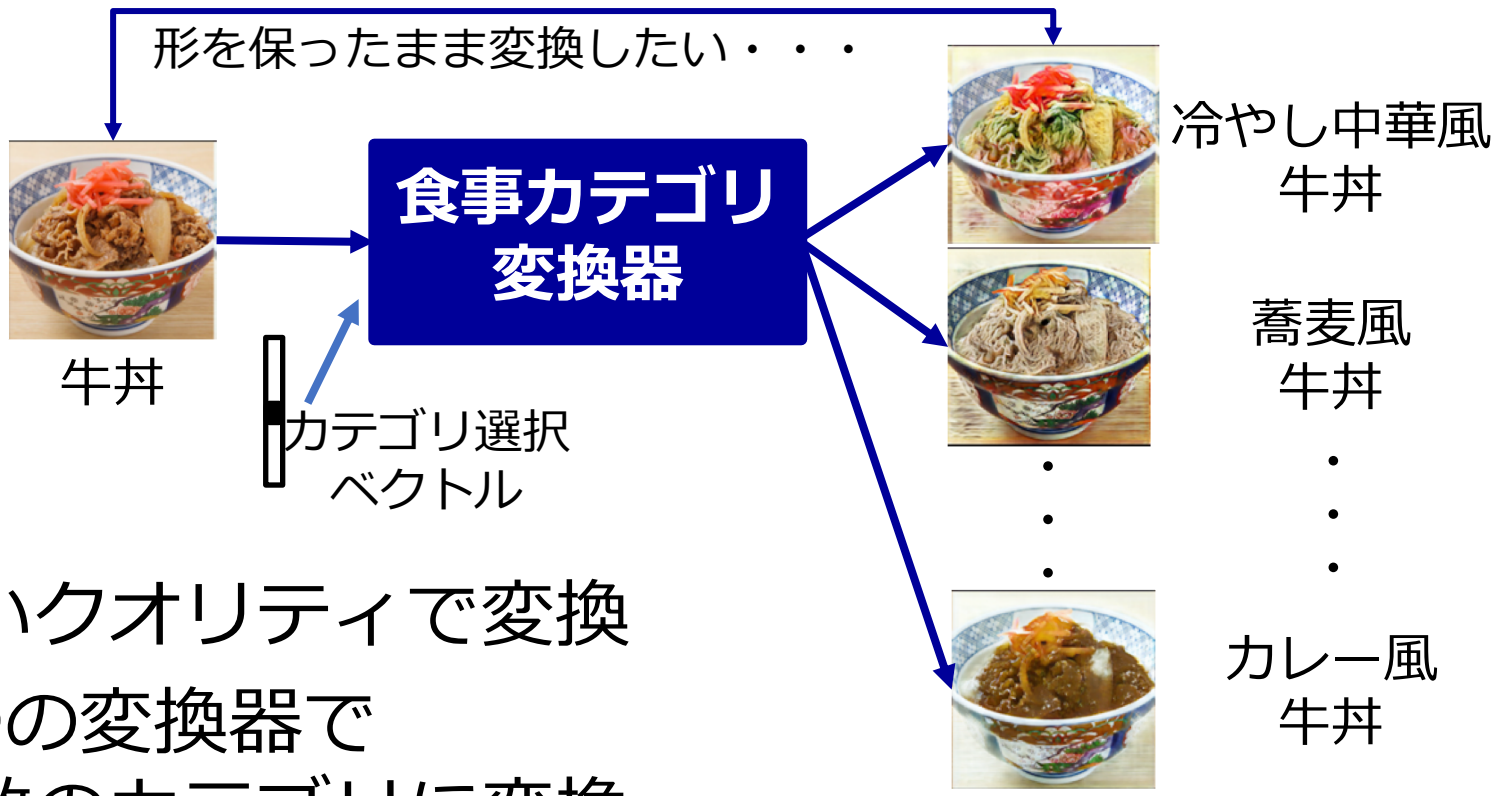


- **変換手法2:**
  - **Multi Style Transfer** [Tanno+ MMM2017]



Conditional CycleGANの概要

- 大前提:変換前の元の形はなるべく維持

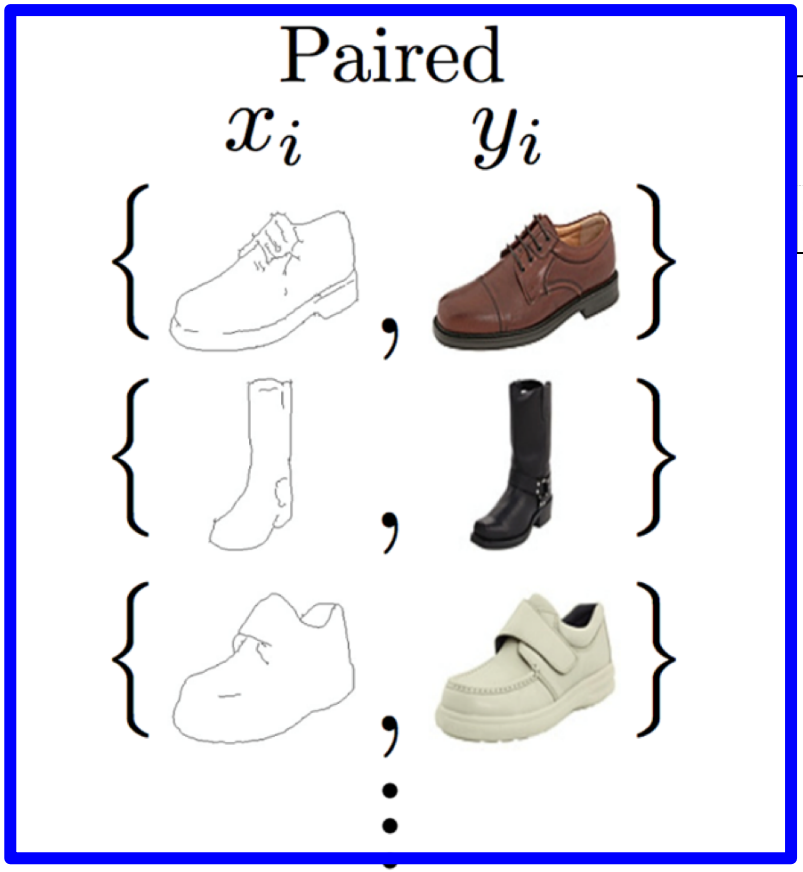


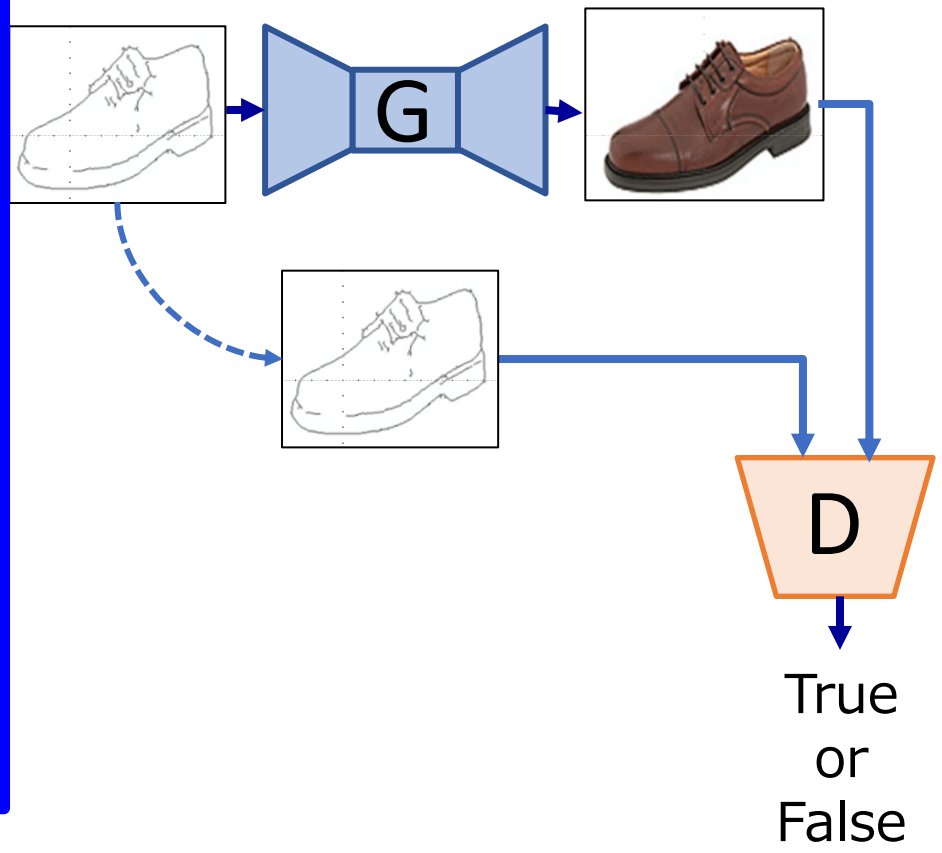
- 課題：
  1. 高いクオリティで変換
  2. 1つの変換器で
     複数のカテゴリに変換

Image-to-Imageの変換が可能な
CycleGANを拡張

CycleGANとは

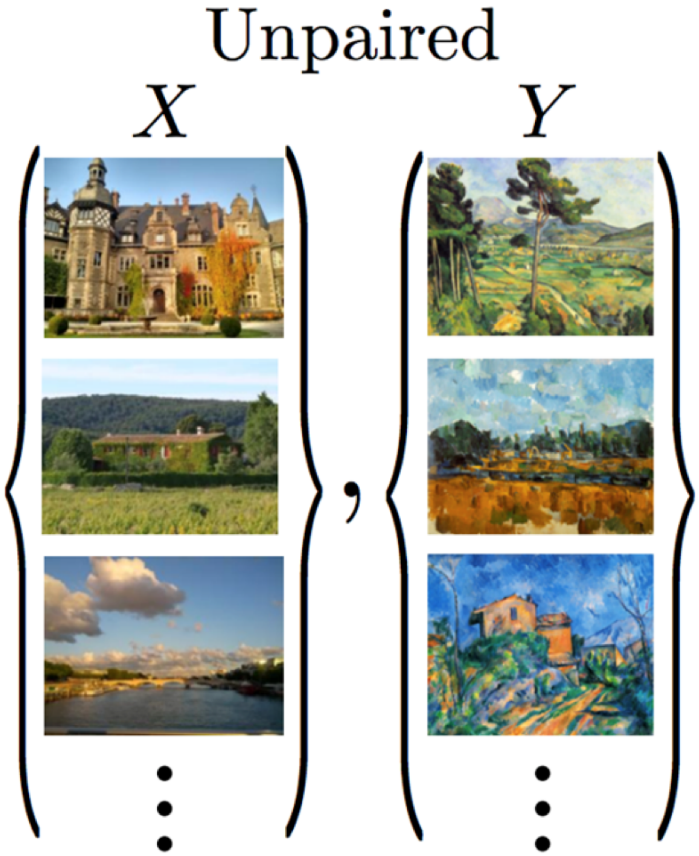
- 従来手法(pix2pix[Isola+ CVPR-17])
  - 元画像と変換先画像のペアが必要





CycleGANとは

- CycleGAN([Zhu+ ICCV-17])
  - ペア画像がなくても学習可能に

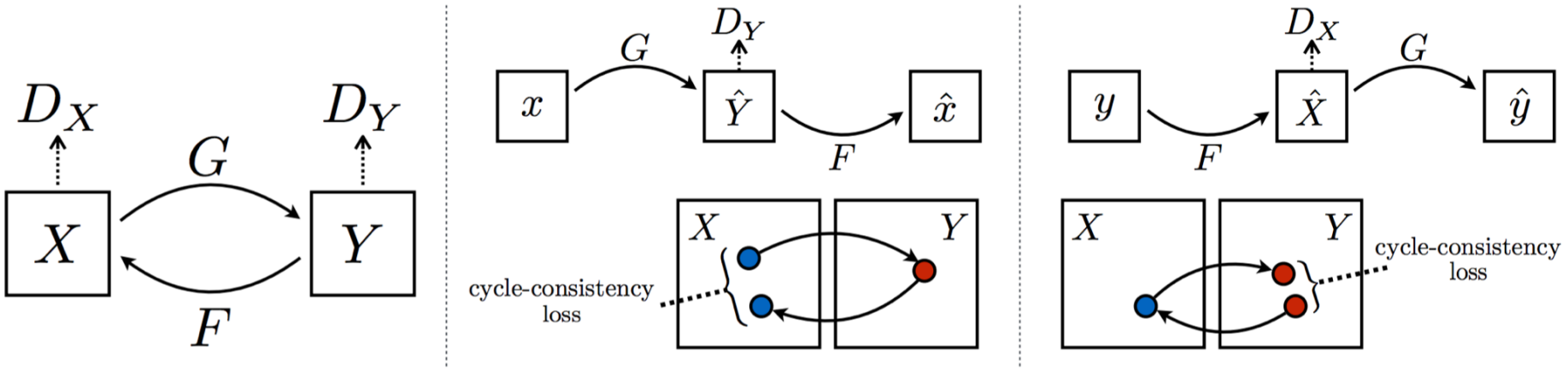


CycleGANとは

- CycleGAN([Zhu+ ICCV-17])
  - 「元の画像」と、その画像を「変換＋逆変換」して元に戻したものの間の誤差で学習

➡ ペア画像が無くても変換可能に

Cycle Consistency Lossを導入

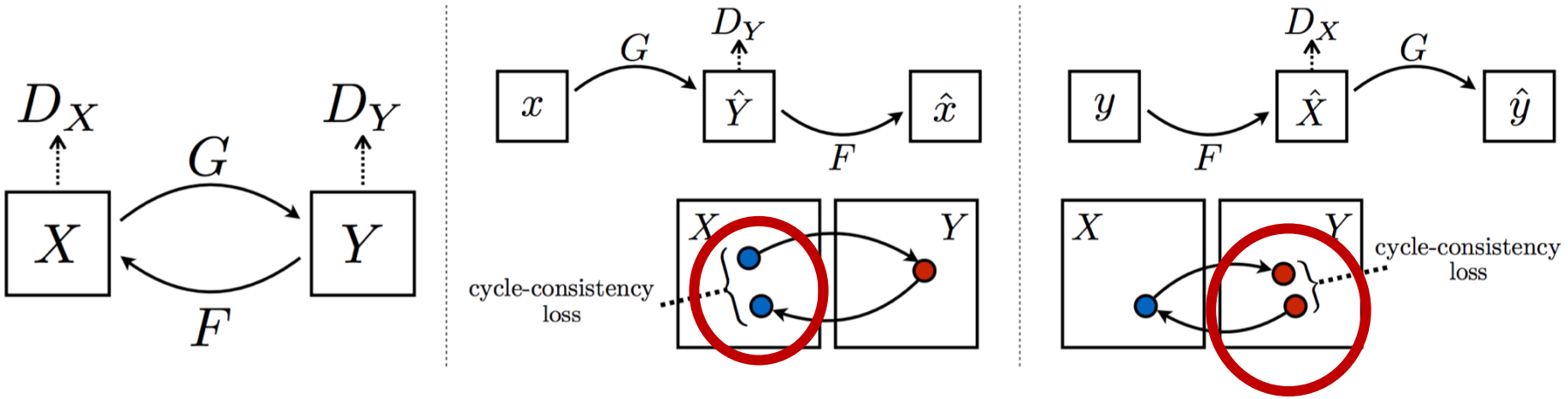


CycleGANとは

- CycleGAN([Zhu+ ICCV-17])
  - 「元の画像」と、その画像を「変換＋逆変換」して元に戻したものの間の誤差で学習

➡ **ペア画像が無くても変換可能に**

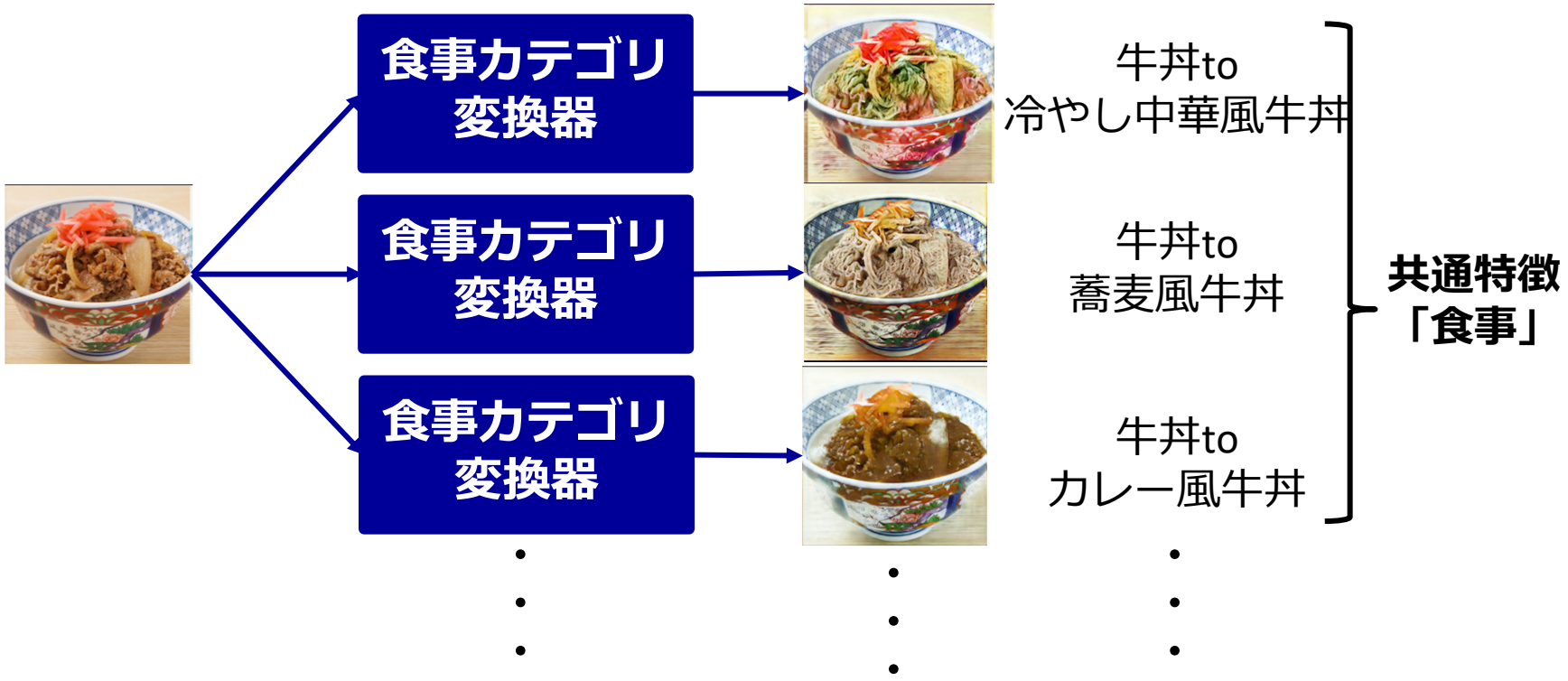
Cycle Consistency Lossを導入



CycleGANの課題

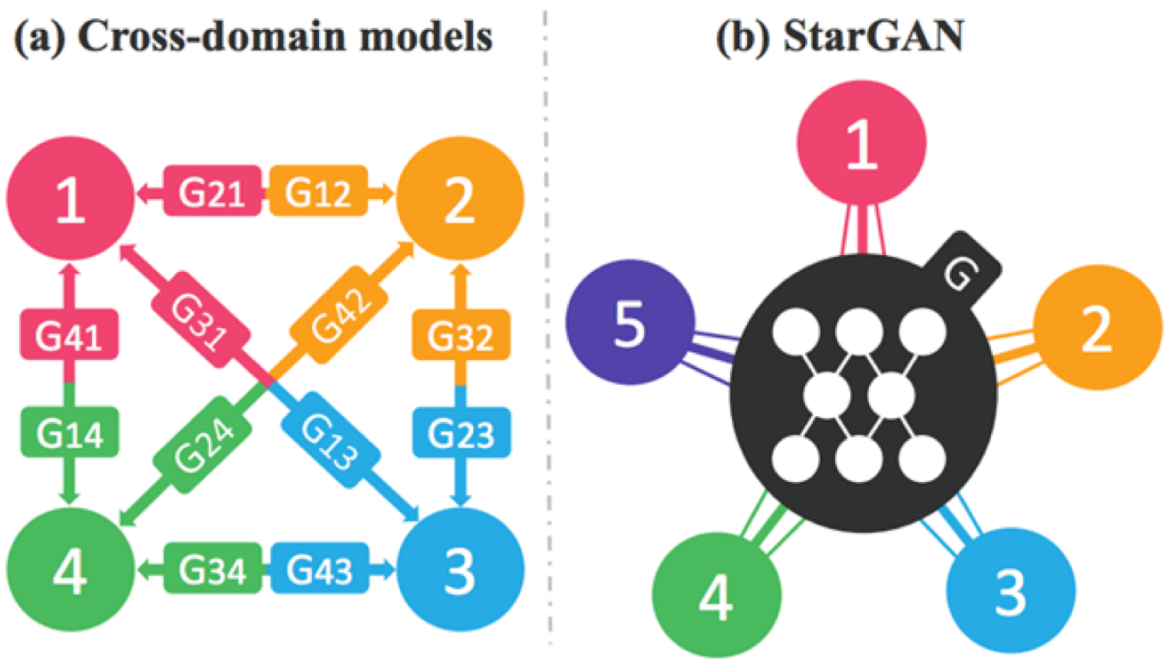
1. 1:1対応の変換しかできない
2. 変換器が独立しているので，
   他の共通する特徴を活かせない

➡ 1:n 変換に拡張



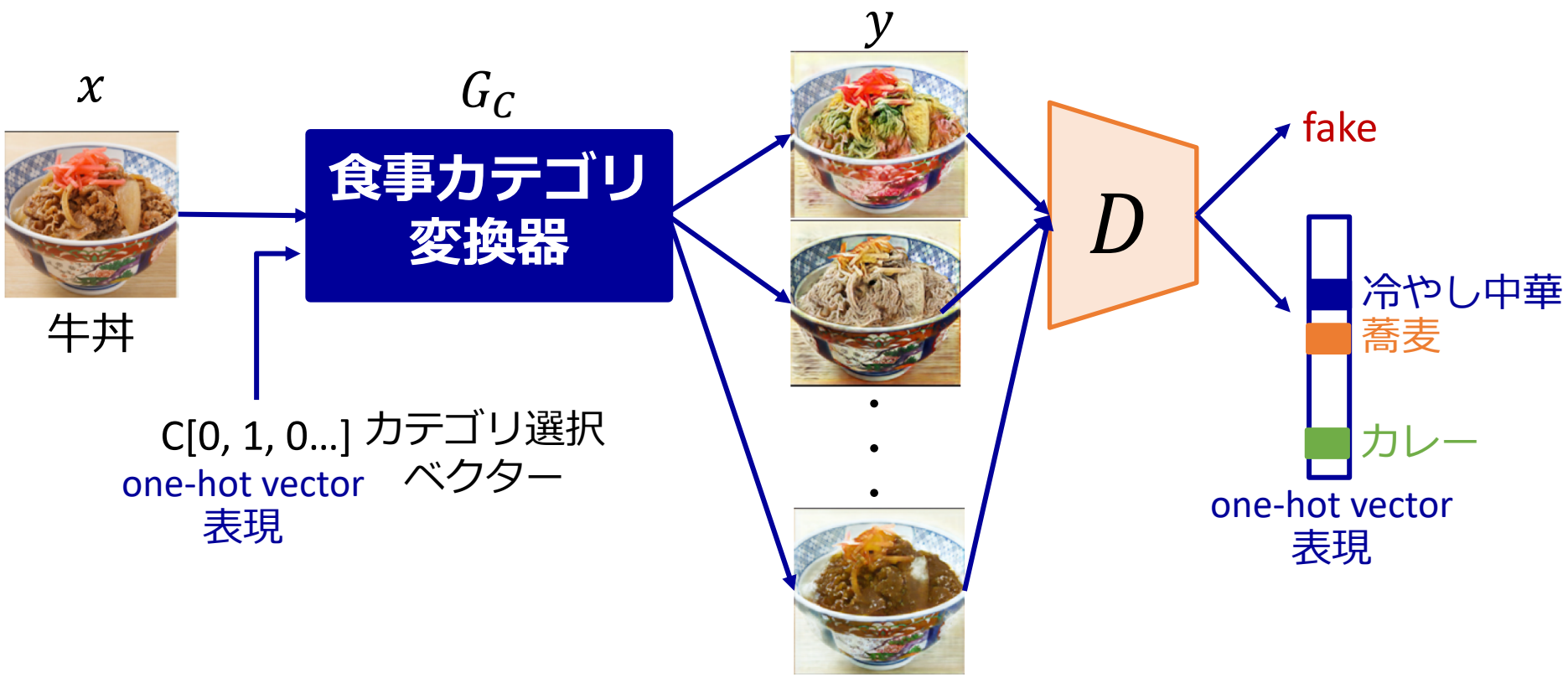
CycleGANの拡張

- StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation [Choi+ CVPR-18]
  - 1:1から1:nへ拡張
  - CycleGAN+**AC-GAN**の組合せで実現
  - 発想がシンプル



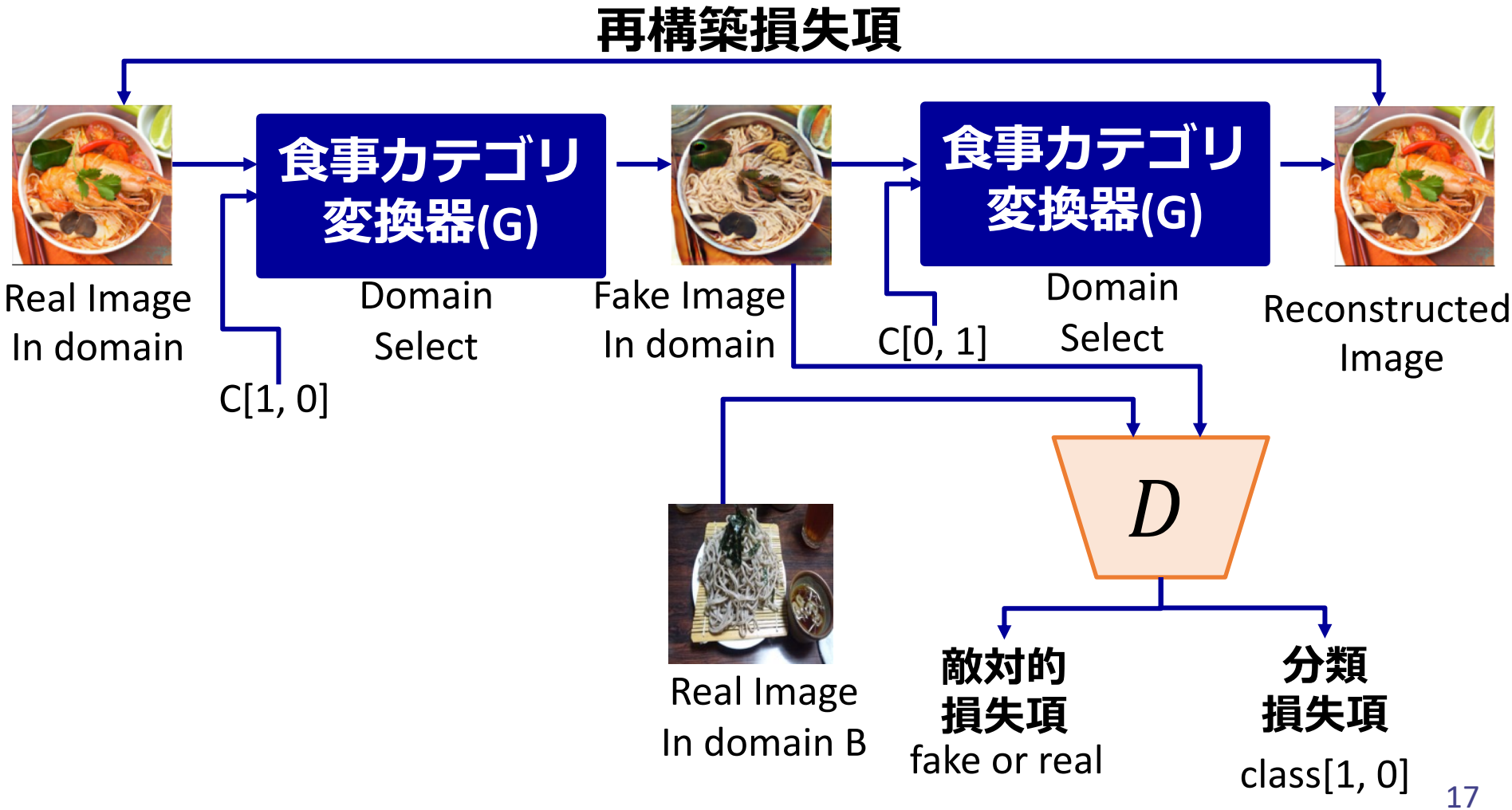
AC-GAN[Odena+ ICML-17]とは

- Conditional Image Synthesis With Auxiliary Classifier GANs
  - 単に生成するだけではなく，**クラスの識別をさせる補助的なタスクを識別器(D)に追加**



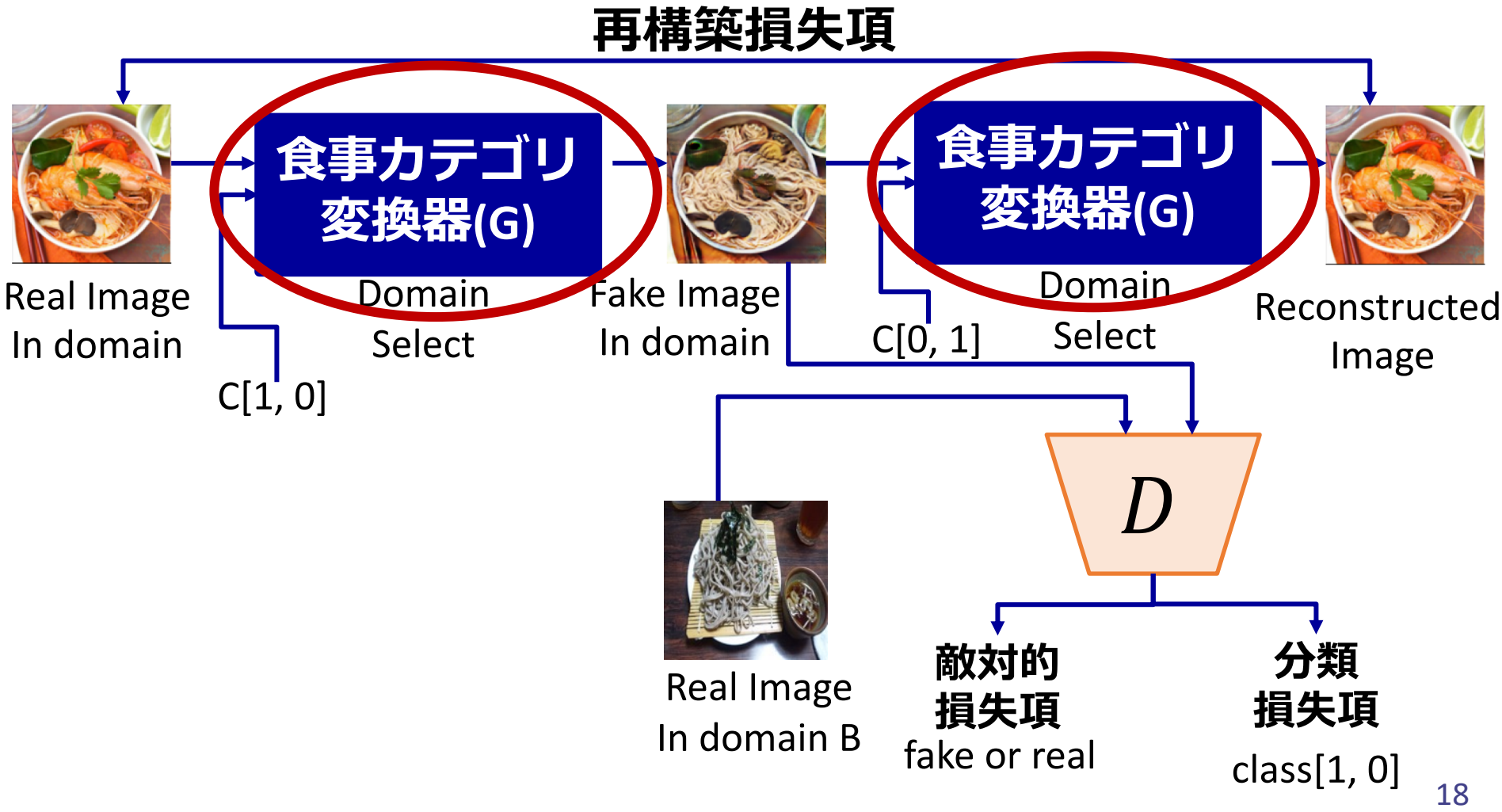
ネットワーク全体の構成

- 構成要素は3つ
  - 敵対的損失、再構築損失、分類損失の組み合わせ



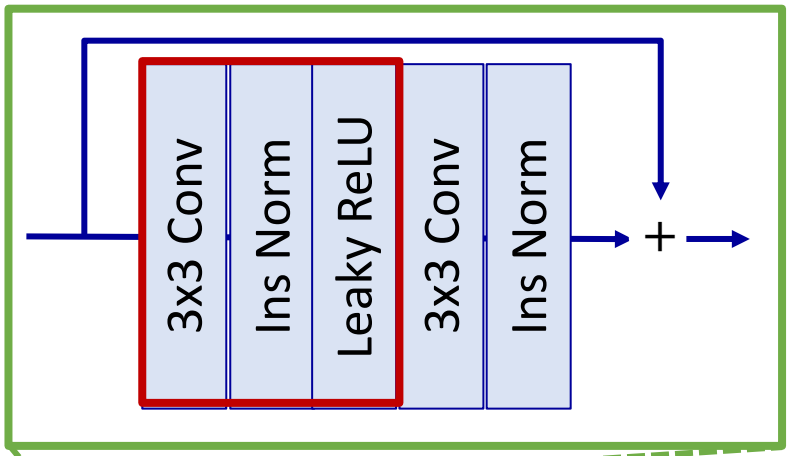
ネットワーク全体の構成

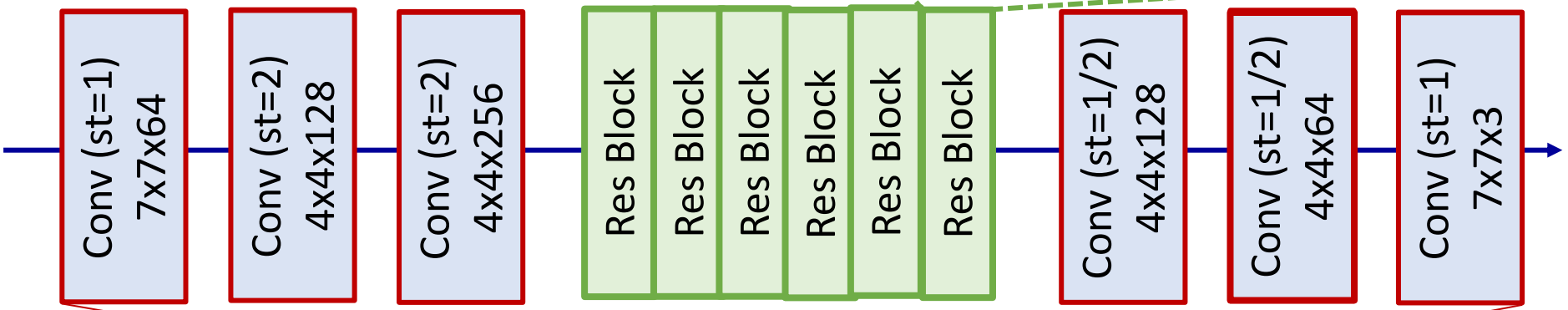
- 構成要素は3つ
  - 敵対的損失、再構築損失、分類損失の組み合わせ



食事カテゴリ変換器のネットワーク詳細

- Perceptual Losses for Real-Time Style Transfer and Super-Resolution [Johnson+ ECCV-16]で提案されたネットワークを参考
  - ConvDeconvの中間層にResBlockを積層
  - 変換の表現力が高くなる





食事カテゴリ変換器

目的関数

- 敵対的損失＋再構築損失＋分類損失で構成
  - 損失重みは$\lambda_{classifier}=10, \lambda_{cicle}=1$に設定

識別器(D)の目的関数

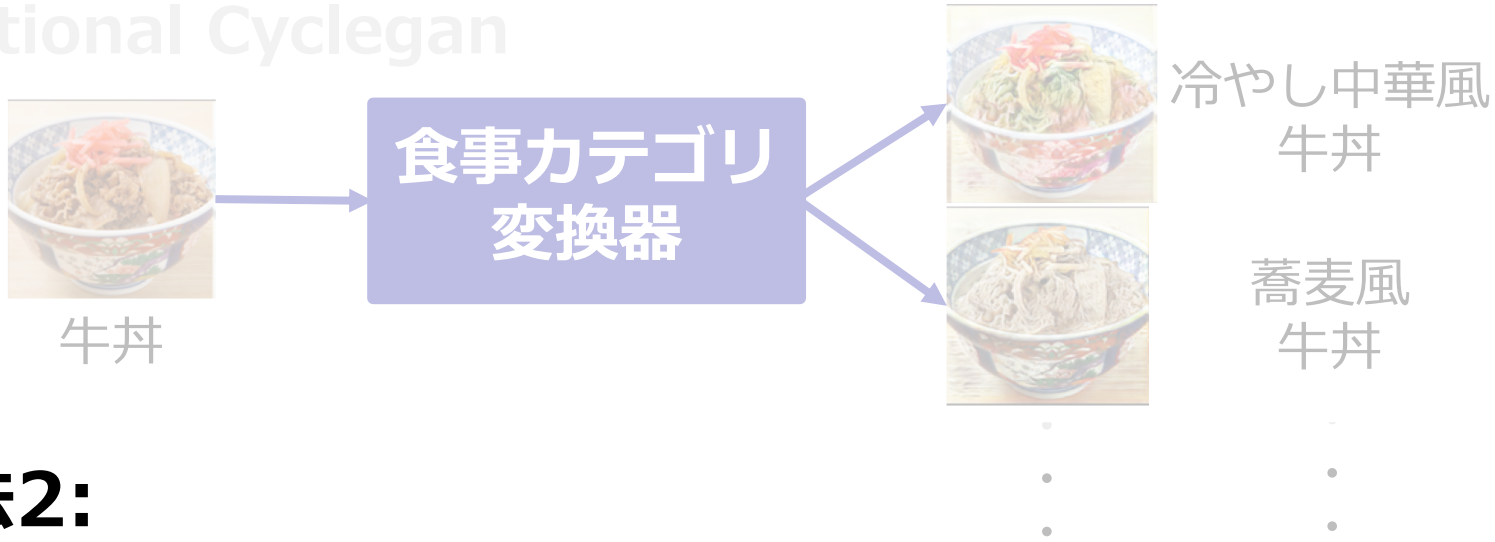
$$L_D = \underbrace{L_{adversarial}}_{\text{敵対的損失}} + \underbrace{\lambda_{classifier} L_{classifier}}_{\text{分類損失}}$$

変換器(G)の目的関数

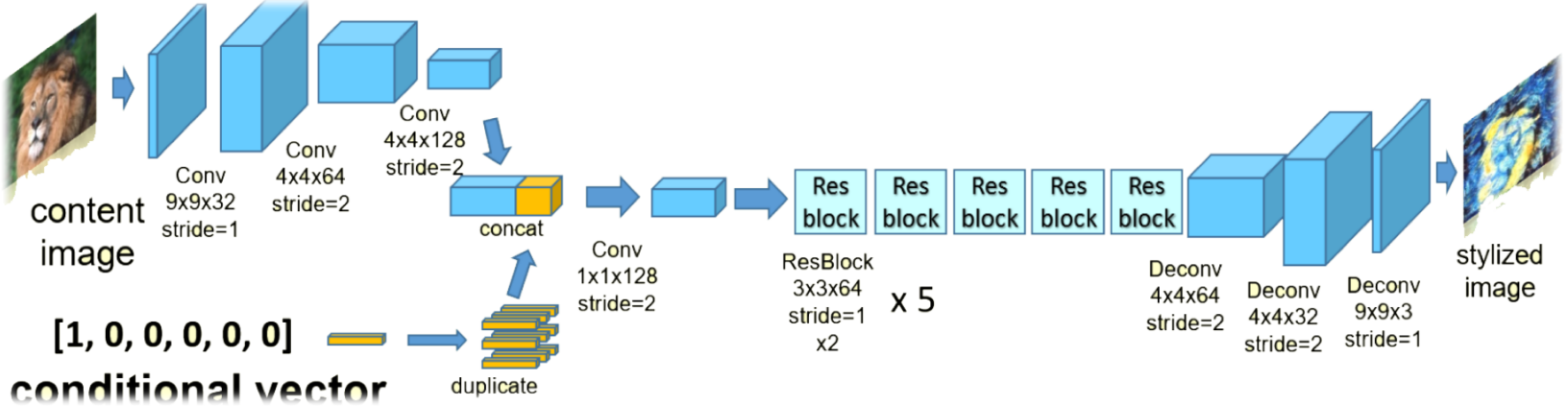
$$L_G = \underbrace{L_{adversarial}}_{\text{敵対的損失}} + \underbrace{\lambda_{classifier} L_{classifier}}_{\text{分類損失}} + \underbrace{\lambda_{cycle} L_{cycle}}_{\text{再構築損失}}$$

手法概要

- 変換手法1:
  - Conditional Cyclegan

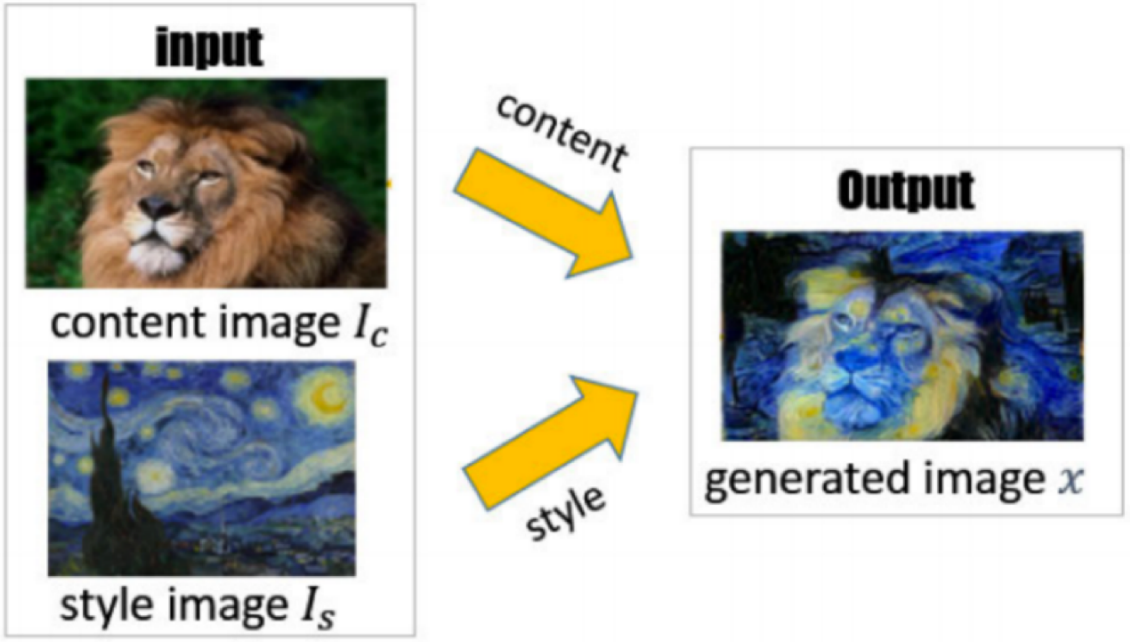


- **変換手法2:**
  - **Multi Style Transfer** [Tanno+ MMM2017]



Style Transferとは(再掲)

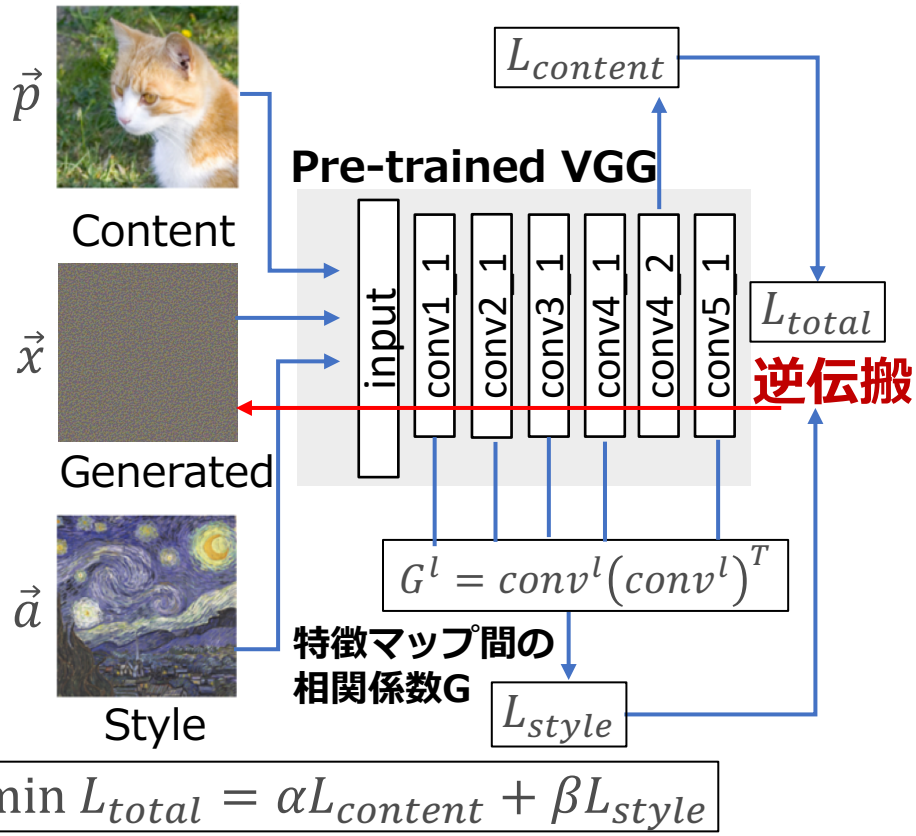
- 画風を変換するアルゴリズム(2015年8月に公開)
  - 2枚の画像を入力として，片方をコンテンツ画像，片方をスタイル画像とする
  - コンテンツ画像に書かれた物体の配置をそのままにして，画風をスタイル画像に変換した画像を生成
  - AIによる芸術画像の生成として研究が盛んに



オリジナル手法(Gatysらの手法)

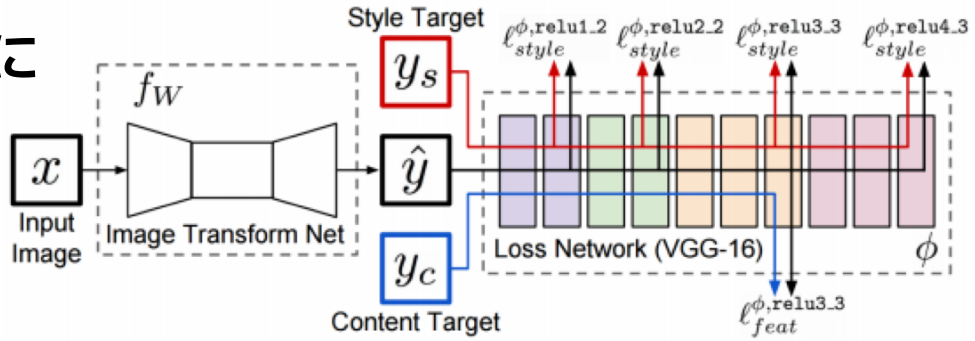
- Neural Style Transferを提案 [Gatys+ CVPR2016]
  - 損失関数に学習済みCNNを利用
  - ContentとStyleのズレを相関行列により ロスとして定義
  - **順伝搬＆逆伝搬**を反復

- **問題点**
  - **順伝搬と逆伝搬を繰り返すため生成に時間がかかる**
  - **GPU利用で数十秒程度**

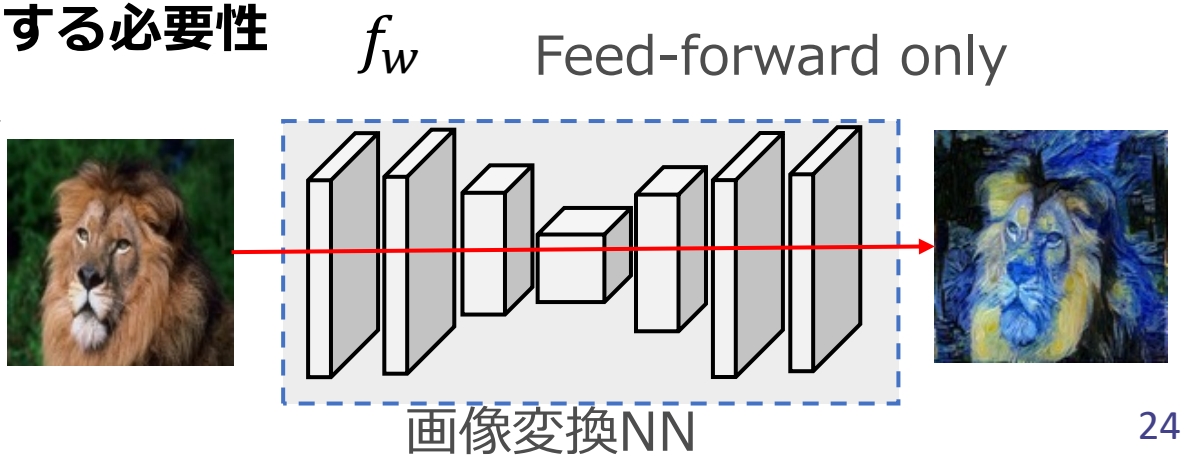


高速化手法 (Johnsonらの手法)

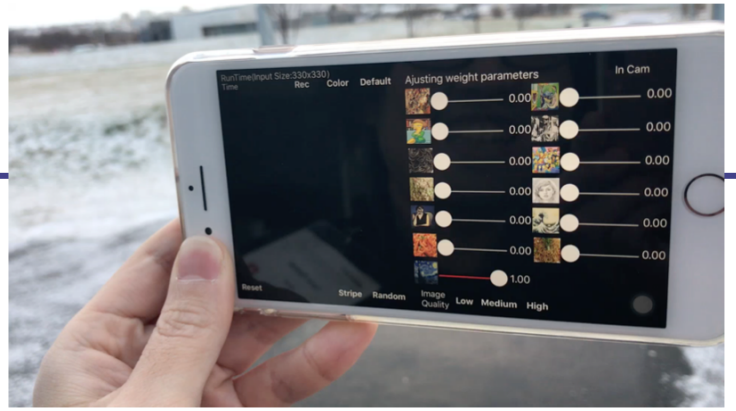
- **特定スタイルを変換可能なネットワークを学習**
  - １回のFeed-forwardでスタイル変換可能
  - 順伝搬でスタイル変換可能
    - **非常に高速に画像生成可能に**
    - **Gatysらの手法の1000倍**



- **問題点**
  - **1つのモデルで1つのスタイル表現**
    - **スタイル毎に学習する必要性**
    - **消費メモリの増大**



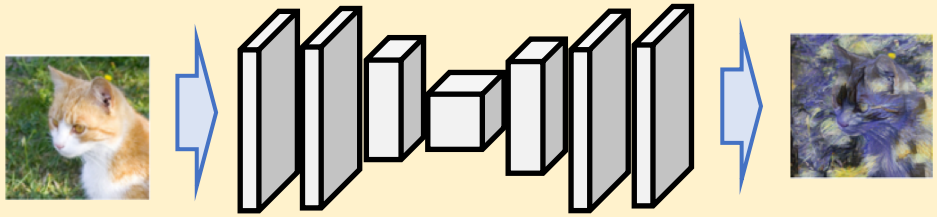
Multi Style Transfer



- Johnsonらの手法を拡張
  - スタイル選択の入力信号を追加
  - **同時に最大30種類程度のスタイルを学習可能**

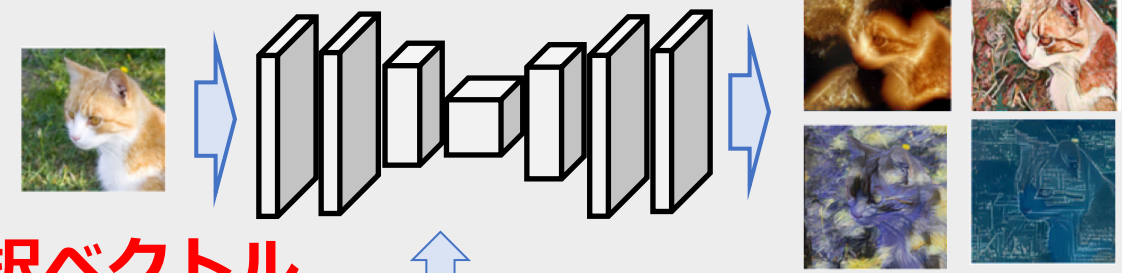
[Tanno+ MMM2017]

従来手法



1対**1**

提案手法



1対
複数

スタイル選択ベクトル
(1,0,0,0,…0)

実験

- 変換手法
  - 手法 1 : Conditional CycleGAN
  - 手法 2 : Multi Style Transfer

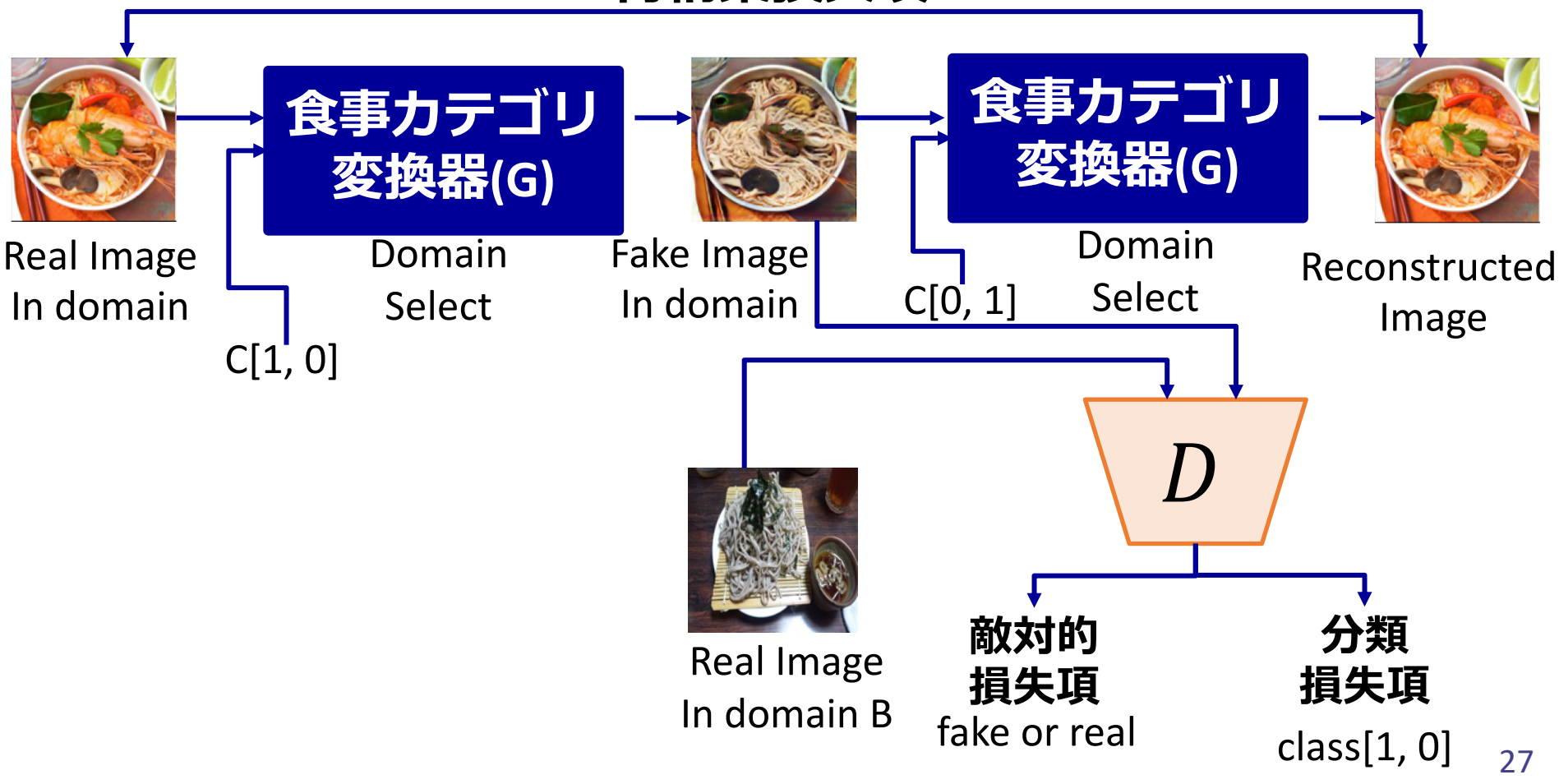
- 学習環境
  - Nvidia Quadro P6000

- 学習データ
  - Conditional CycleGAN: 10カテゴリ合計約23万枚
  - Multi Style Transfer: MS-COCO約8万枚(content画像)
    +10カテゴリ(style画像)

学習データ(手法１：cCycleGAN)

- **変換先の形が変わりすぎると学習が難しい**
  - 多様性が有りすぎると変換は困難

形状が似ていれば
学習が上手くいく可能性



学習データ(手法１：cCycleGAN)

- 変換元と変換先で形状が似ていれば学習がうまくいく可能性

➡ 「丼」という制約を設けて１０個の食事カテゴリを選出


カレー


炒飯


牛丼


冷やし中華


ミートスパ


ラーメン


白飯


蕎麦


うなぎ


焼きそば

学習データの選別(手法１：cCycleGAN)

- 選出した食事カテゴリについて画像をTwitterから収集
  - **単にキーワード検索「牛丼」だとノイズ画像が多数**

食事認識エンジン

認識率が高い順にリランキング

形状が似た画像群が得られる

学習データ枚数(手法１：cCycleGAN)

- 合計約２３万枚を利用
  - 学習用：９割
  - テスト用：１割

食事カテゴリ	画像枚数
ラーメン	74007
カレー	34216
炒飯	27854
焼きそば	24760
白米	21324
牛丼	18396
冷やし中華	13499
ミートスパゲティ	7138
うな重	5329
そば	3530
合計	**230053**

実験データ(手法2:Multi Style Transfer)

- 各カテゴリの中から代表style画像を主観で1枚選出

  主観ではなく、工夫してスタイル画像の選択を行いたい...


カレー


炒飯


牛丼


冷やし中華


ミートスパ


ラーメン


白米


そば


うな重


焼きそば

実験データまとめ

- 手法1: conditional CycleGAN
  - 10カテゴリ合計約23万枚

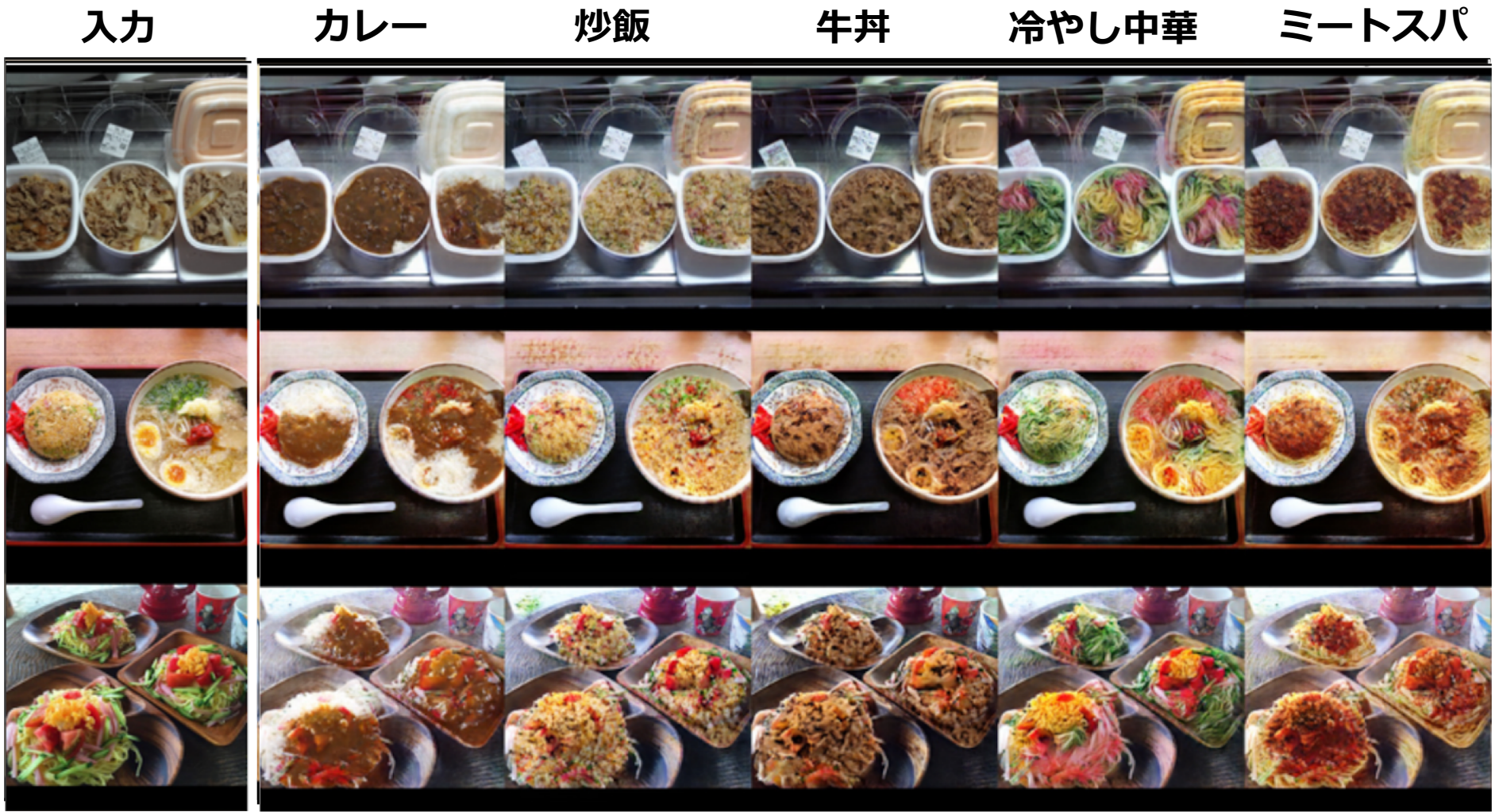
- 手法2: Multi Style Transfer
  - 各カテゴリの中から代表style画像を主観で1枚選出

結果(手法１：cCycleGAN)

- 画像中に食事が１つのみの場合

結果(手法１：cCycleGAN)

- 画像中に食事が複数ある場合

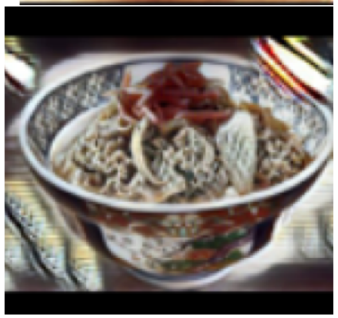


結果(手法２： Multi Style Transfer)

- 各カテゴリの中から代表style画像を主観で1枚選出

手法比較

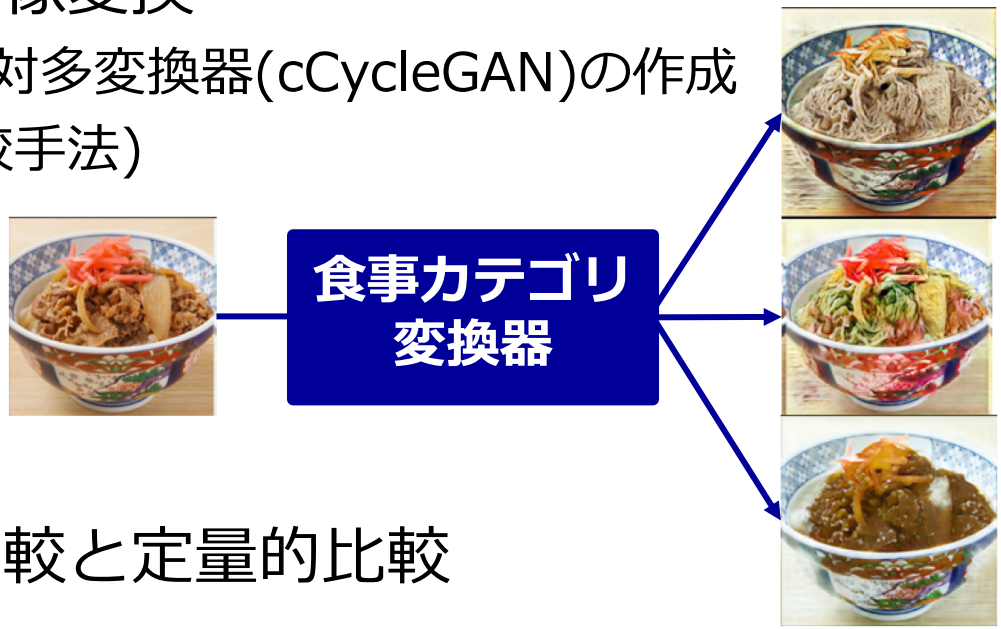
入力

	ミート スパゲティ	そば	冷やし中華
cCycleGAN			
Multi Style Transfer			

未学習データでの変換

まとめと今後の課題

- まとめ
  - 深層学習を用いた食事画像変換
    - CycleGANの拡張による1対多変換器(cCycleGAN)の作成
    - Multi Style Transfer(比較手法)



- 今後の課題
  - より多くの他手法との比較と定量的比較
  - 「丼」制約の緩和
  - 任意カテゴリ(学習済みカテゴリ以外)への拡張

ご清聴ありがとうございました

F1-3

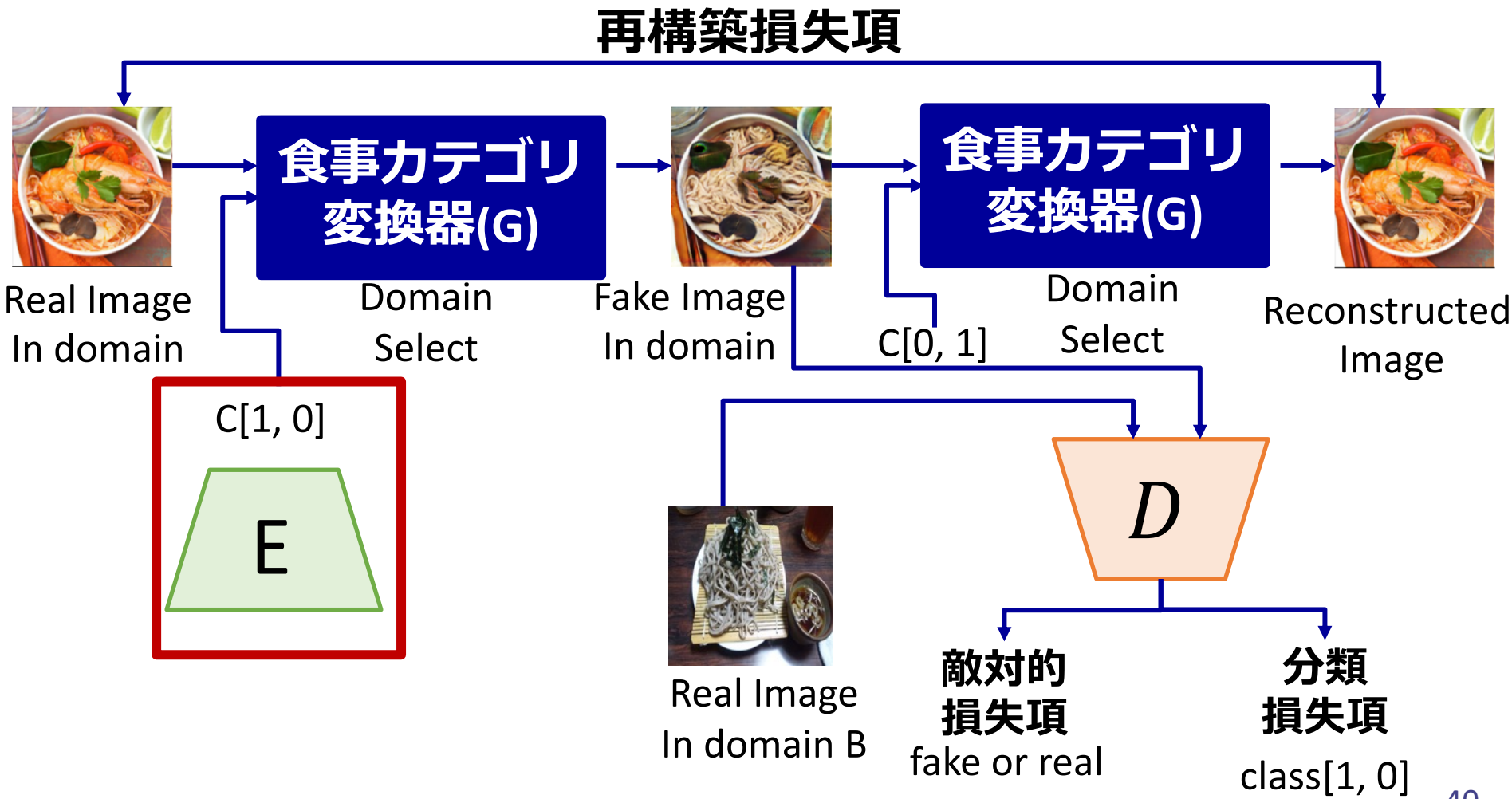
紫雲殿

ロビー

インタラクティブセッション1日目

D09

任意カテゴリへの変換


再構築損失項
食事カテゴリ
変換器(G)
食事カテゴリ
変換器(G)
Real Image
In domain
Domain
Select
Fake Image
In domain
C[0, 1]
Domain
Select
Reconstructed
Image
C[1, 0]
E
D
Real Image
In domain B
敵対的
損失項
fake or real
分類
損失項
class[1, 0]

質問

- 食事画像を変換する意味、モチベーション
  - インスタ映えする画像を取りたい
    - 難しい
  - 食事画像を魅力ある画像に変換したい
    - とりあえずカテゴリ変換から
  - 食事画像の変換に関する研究は行われていない。

質問

- ラーメンの枚数が多いようですが…
  - 一年間でつぶやかれる画像の中でラーメンが一番多い
  - 1対多変換なので、少ないデータの特徴を補える
  - ラーメン大好きだから

- 食事エンジンとは
  - AlexNetをベースにしたCNNをImageNet1000カテゴリでpre-train済みのモデルを弊大学の柳井研究室で提供している食事データセットでfine-tuneしたものを使用

質問

- 定量評価(pix2pixと同様)
  - 画像1枚を1秒だけ表示させて、本物か偽物かを人に判断させる
  - 被験者8名(GANを知る者)

一秒で消える

	画像変換
生成画像本物と誤認識した場合の割合	30%
本物を生成画像と誤認識した場合	32%

質問

- **なぜ，これまで食事画像の変換はされてこなかったのか**
  - 多様性がある学習データだと学習が難しいこと
  - 顔や服，文字などは公開データセットがあるが，食事はない
    - 顔はCelebAの20万枚，文字はMNISTなど
  - 今回，学習に用いたデータもTwitterから収集し， ノイズを除去したクリーンなデータセットを用意した
  - 変換がうまくいったので公開を予定

結果(手法１：cCycleGAN)

- 画像中に食事が１つのみの場合

結果(手法２： Multi Style Transfer)

- 各カテゴリの中から代表style画像を主観で1枚選出

結果(手法１：cCycleGAN)

- 画像中に食事が複数ある場合

実験データ(手法2:Multi Style Transfer)

- 複数のstyle画像の統合
  - 意味的に遠いスタイル画像を統合することでより多様性のあるスタイル表現が得られることを期待


カレー


炒飯


牛丼


冷やし中華


ミートスパ


ラーメン


白米


そば


やきそば


うな重

結果(手法２： Multi Style Transfer)

- 複数のstyle画像の統合

結果(手法２： Multi Style Transfer)

- 複数のstyle画像の統合