

セマンティックセグメンテーションと
敵対的生成ネットワークを用いたイラスト自動生成システム
—Illustration Automatic Generation System Using Semantic Segmentation and GAN—

武蔵野大学 データサイエンス学部 データサイエンス学科
中舘滯男 福原義久

目次

1. 目的と背景

- ・ 本研究の目的
- ・ 本研究の活用
- ・ 先行研究
- ・ 本研究の特徴

2. 提案手法

- ・ 提案システム
- ・ 利用技術
- ・ パーツ画像の抽出
- ・ 新規パーツ画像の生成
- ・ 新規イラスト画像の生成

3. 実験結果

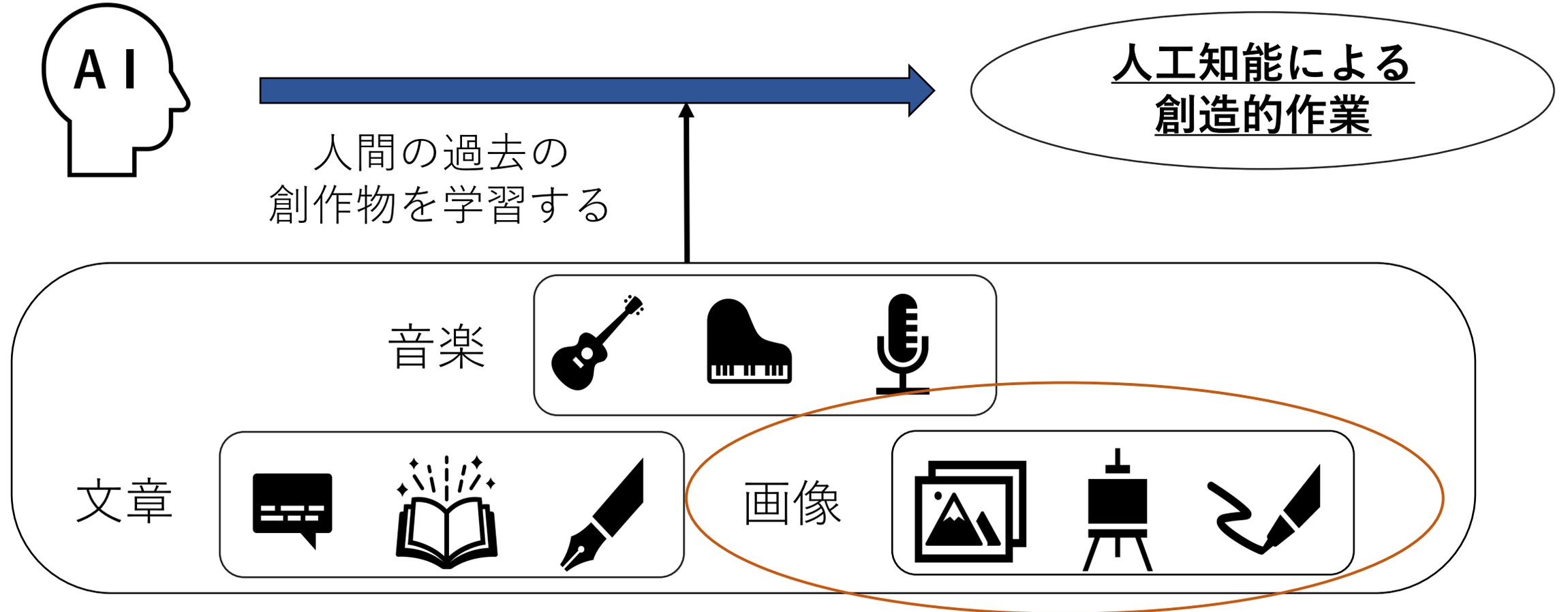
- ・ 「未知の果物」

4. 総括

- ・ 今後の展望

5. 参考文献

本研究の目的

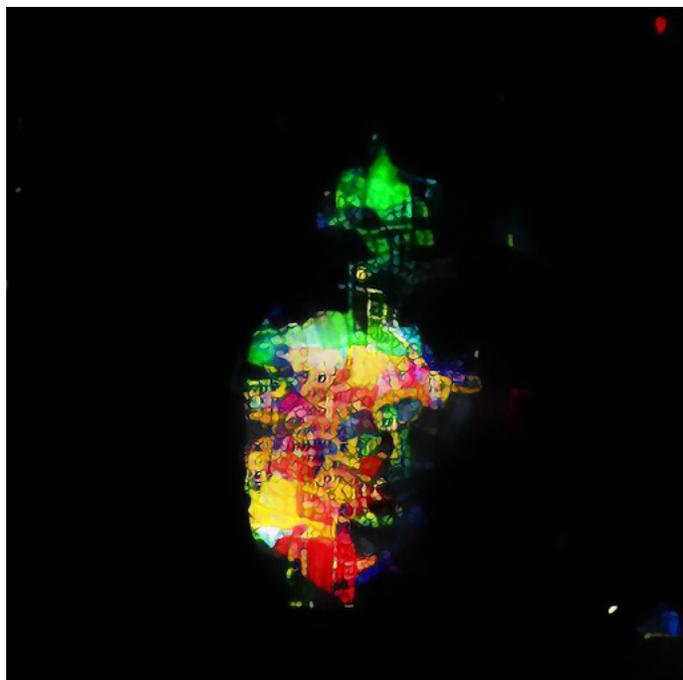


- 人間の過去の創作物をもとに人工知能による創造的な作業を行う
→そうした作業により**人間に新しい閃きを与え創作の支援**をする
- 本研究では**画像分野**に着目する

本研究の活用

出力画像を加工したり、鮮明化させて利用する

→出力される画像は架空のイラスト画像のため、漫画やゲームなどフィクション作品で利用しやすいと考える



本研究の結果（鮮明化加工済み）

使用例

- ・漫画で登場する果物
- ・ゲーム内のアイテム
- ・果物など品種改良を行う会社のロゴマーク

先行研究：人工知能による音楽・文章生成

1. 作詞



『電☆アドベンチャー』の歌詞 (抜粋)
作詞・A | 仮面 作曲・谷内翔太

夜、抱きしめる！ 月を祈りに うさぎを招いて笑おう
ここにこうぼうばブルーベリー
月明かりを 包みこめ
夜、追いこそう！！詩的な自由
輝く川がこみ上げる 助け合って月を染める
いろんな明日
挑戦、めちゃくちゃ！Yeah!

クリーブラッツ提供

- ・ 単語のイメージと色の関係を学習
- ・ 小説など64万の文書をもとに生成

[1]. 坂本真樹「作詞 AI による人の創造力と想像力増幅の試み：電☆アドベンチャー作詞の経緯と今後の可能性 (小特集 創造性・芸術性における AI の可能性)」, 『電子情報 通信学会』, 102(3), pp.234-239. 2019

2. 画像から文章生成

Concept-Set: a collection of objects/actions.

dog, frisbee, catch, throw



Generative Commonsense Reasoning

Expected Output: everyday scenarios covering all given concepts.

- A dog leaps to catch a thrown frisbee. [Humans]
- The dog catches the frisbee when the boy throws it.
- A man throws away his dog's favorite frisbee expecting him to catch it in the air.

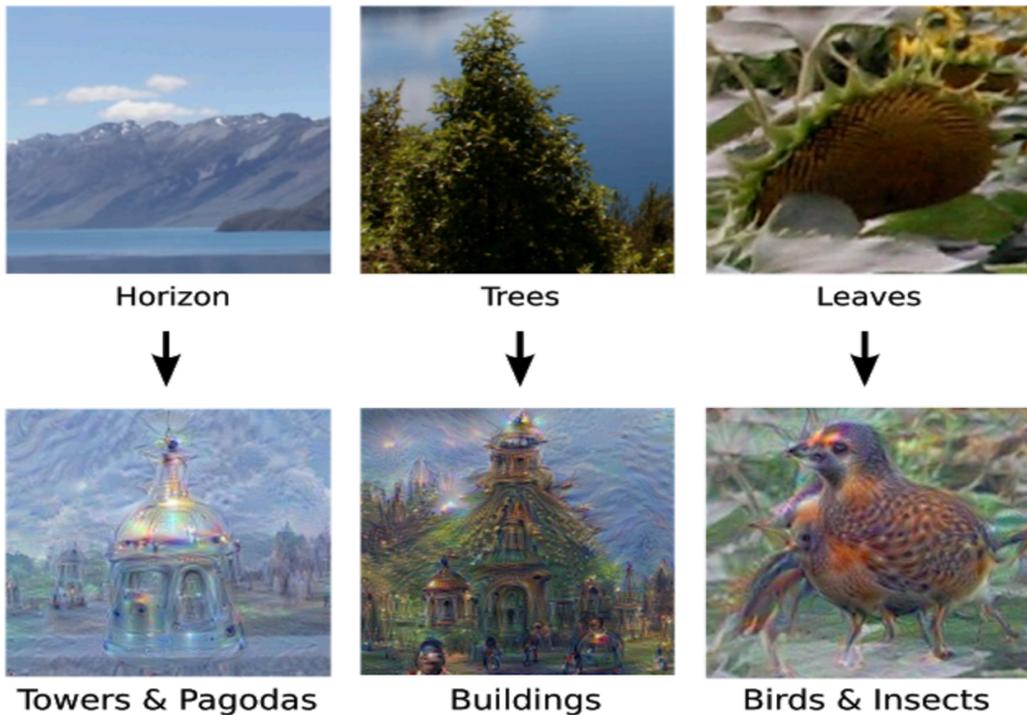
GPT2: A dog throws a frisbee at a football player. [Machines]
UniLM: Two dogs are throwing frisbees at each other.
BART: A dog throws a frisbee and a dog catches it.
T5: dog catches a frisbee and throws it to a dog

- ・ 3万5千の概念と7万7千の文を学習

[2]. Bill Yuchen Lin et al., "CommonGen: A Constrained Text Generation Challenge for Generative Commonsense Reasoning", EMNLP-Findings'20, 2020

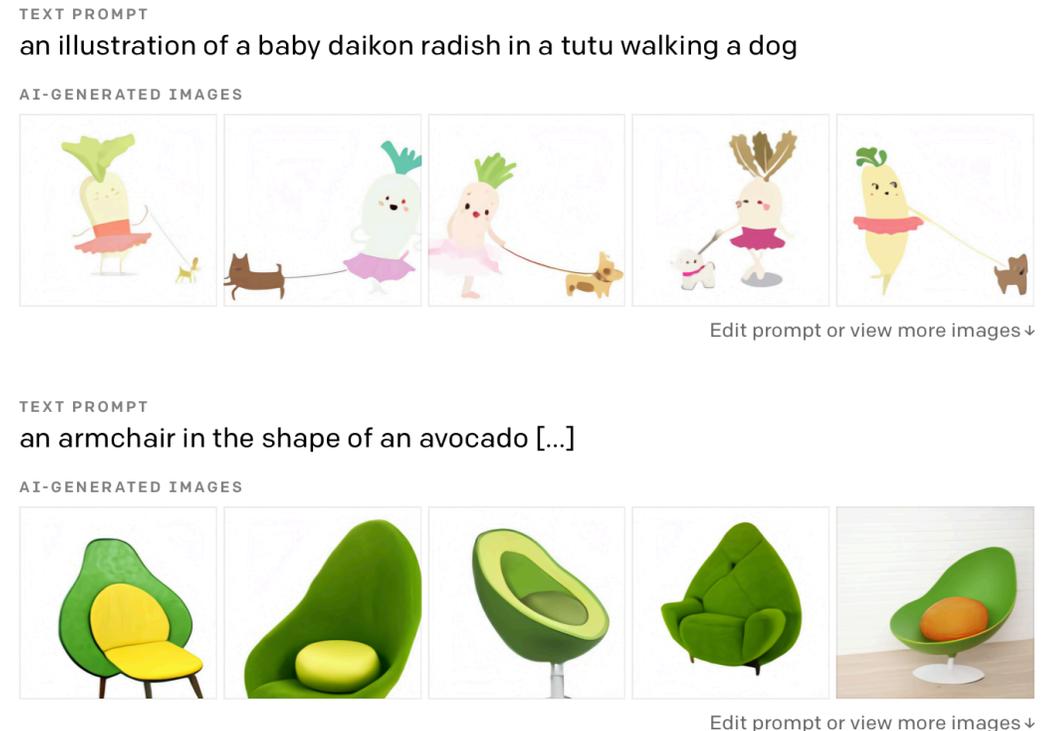
先行研究：人工知能による画像生成

3. 類似画像に変換



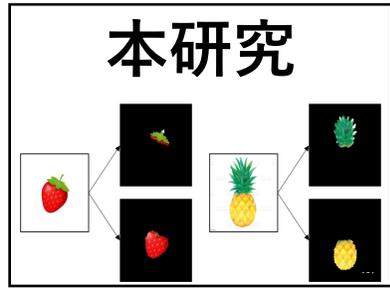
- ・ 入力画像と学習した概念との類似しているものを認識し変換する (Deep dream)

4. テキストから画像生成



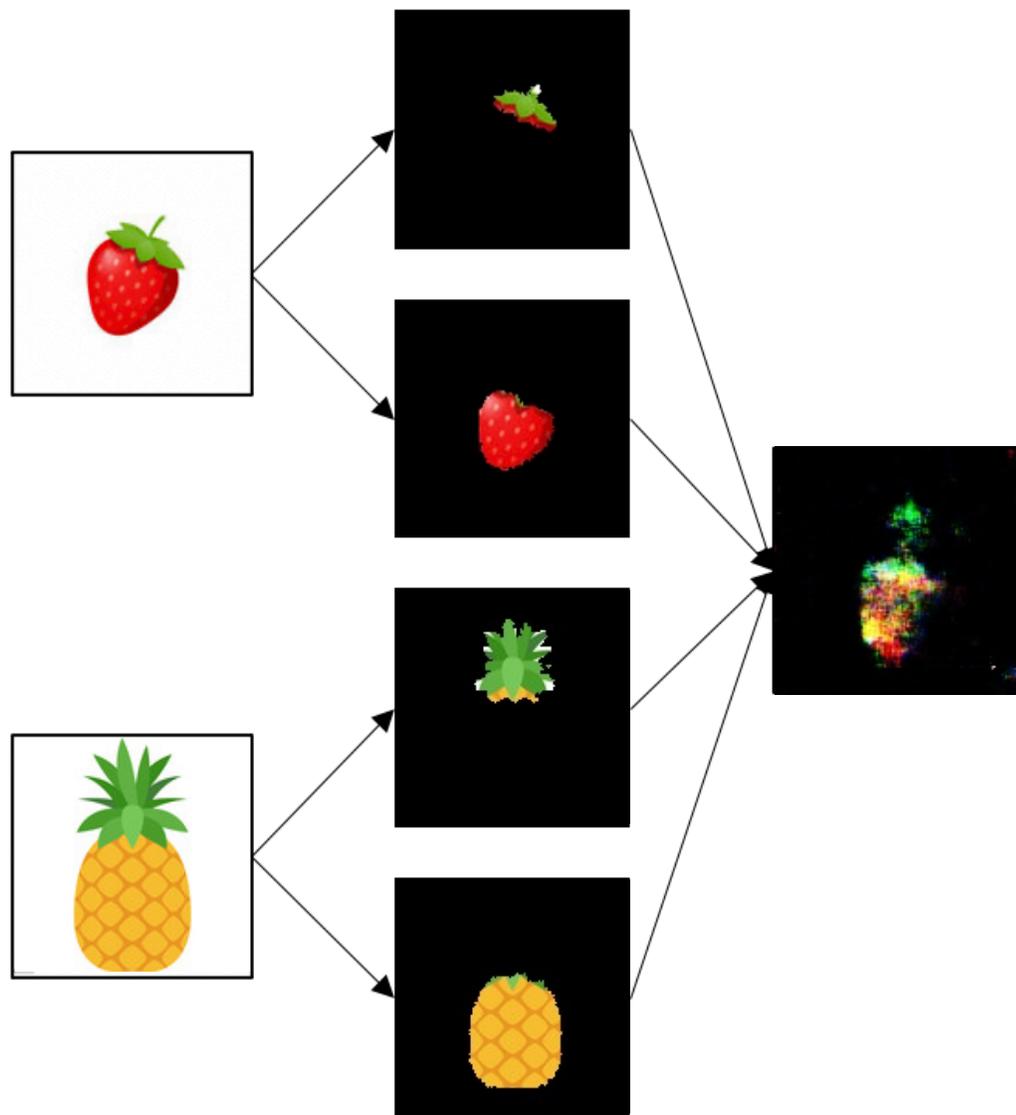
- ・ GPT-3を用いてテキストをもとに画像を生成する (DALL·E)

本研究の特徴：先行研究との比較

<div style="text-align: center;">研究</div> <div style="text-align: center;">評価基準</div>	<div style="text-align: center;">1.作詞</div> 	<div style="text-align: center;">2.画像から文章生成</div> 	<div style="text-align: center;">3.類似画像に変換</div> 	<div style="text-align: center;">4.テキストから画像生成</div> 	<div style="text-align: center;">本研究</div> 
創造性	○	×	△	○	○
利便性	△	○	×	○	○
使用データ種類	色/文章	画像/文章	画像	画像/文章	画像

本研究は、**画像データのみで利便性の高い架空の画像を生成する**

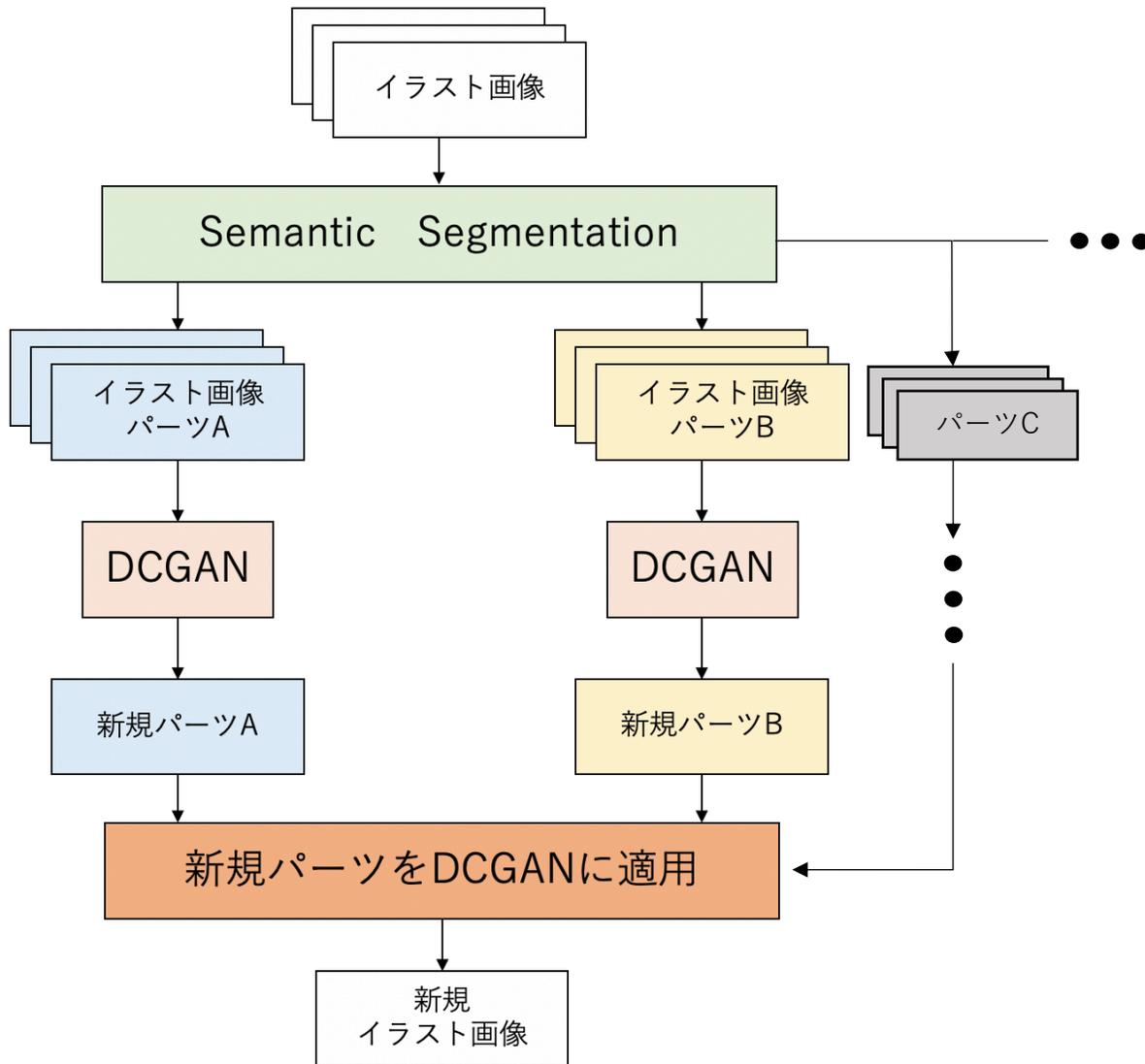
本研究の特徴：提案手法の概要



画像データの組み合わせ

- 必要なパーツ画像を選択して組み合わせると架空のイラスト画像ができる
- パーツ画像に分けることで各特徴を効果的に学習できると仮定

提案システム



パーツ画像抽出

- ・ Semantic Segmentationを用いて、イラスト画像の各パーツ画像を抽出

新規パーツ画像生成

- ・ 抽出したパーツ画像に DCGANを適用し、新規パーツ画像を生成

新規イラスト画像生成

- ・ 生成した新規パーツ画像に再び DCGANを適用し、新規イラスト画像を生成

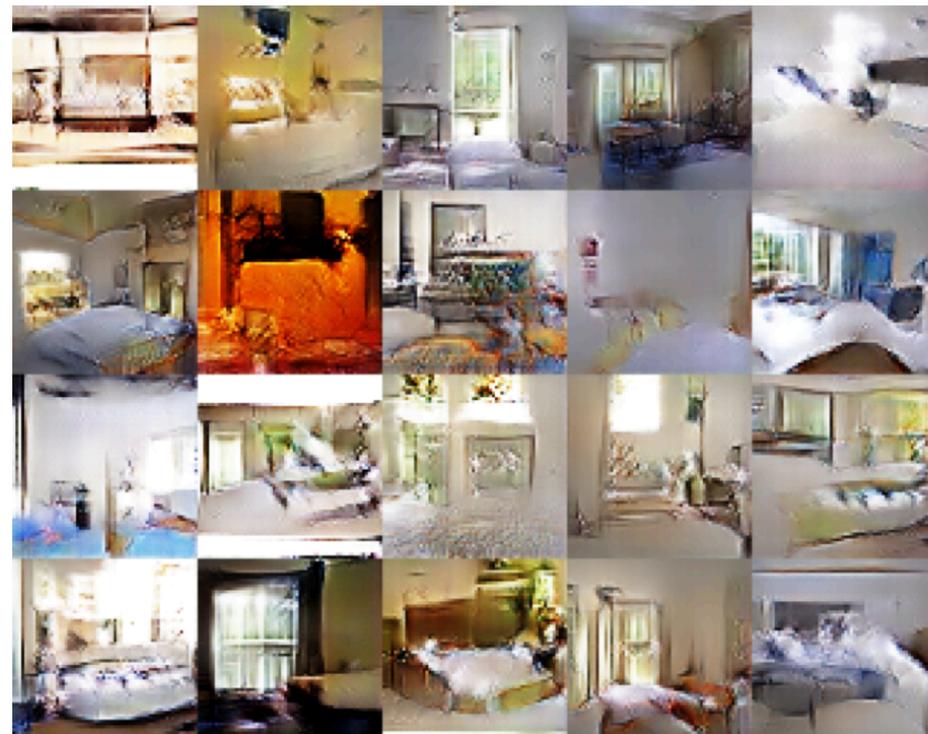
Semantic Segmentation



- ・ 深層学習を用いて画像を画素ごとにカテゴリ分けする技術

[5]. Liang-Chien Chen et al., "Rethinking Atrous Convolution for Semantic Image Segmentation", 2017

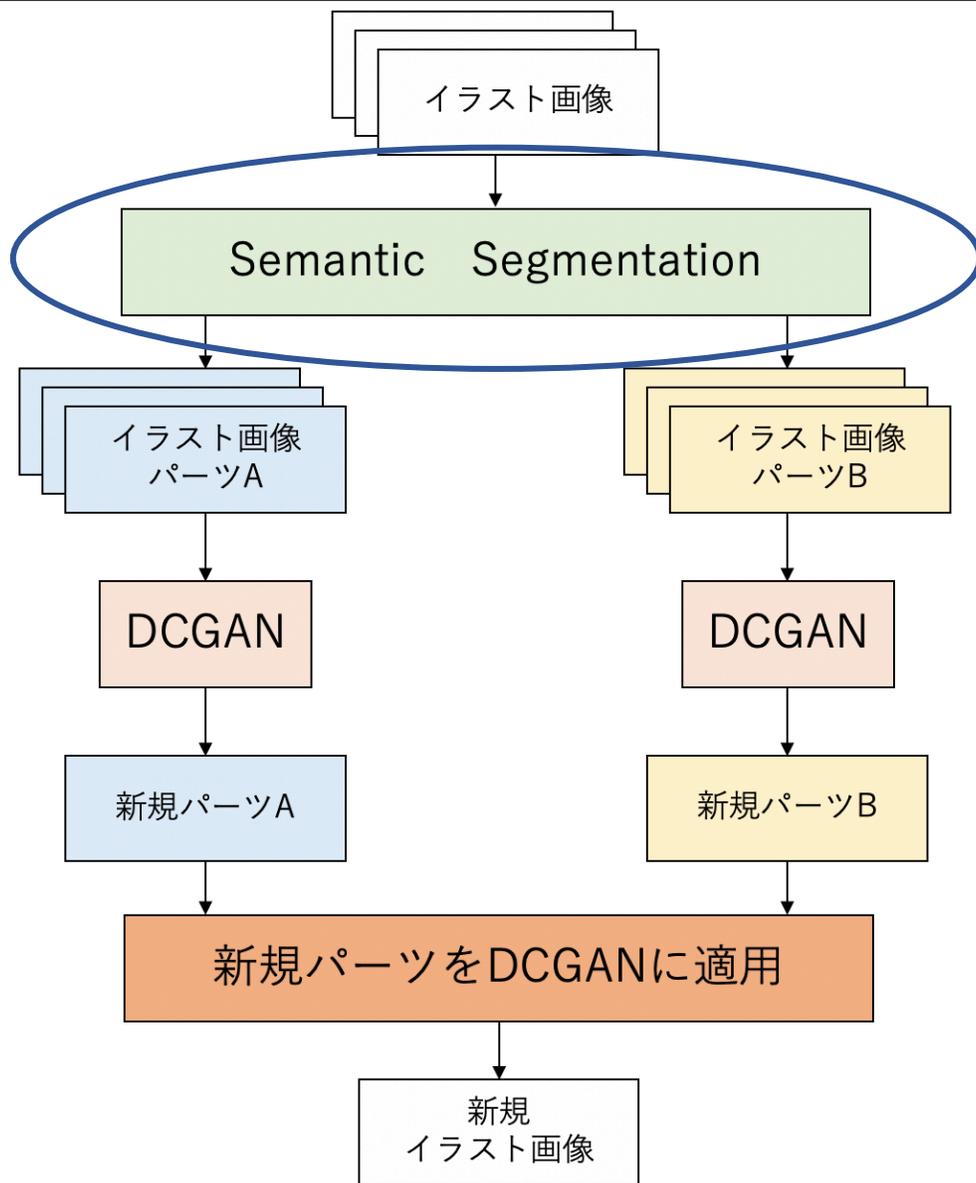
DCGAN



- ・ 学習データと類似した架空の画像の生成、画像の変換ができる

[6]. Alec Radford et al., "UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS", Under review as a conference paper at ICLR 2016

パーツ画像の抽出

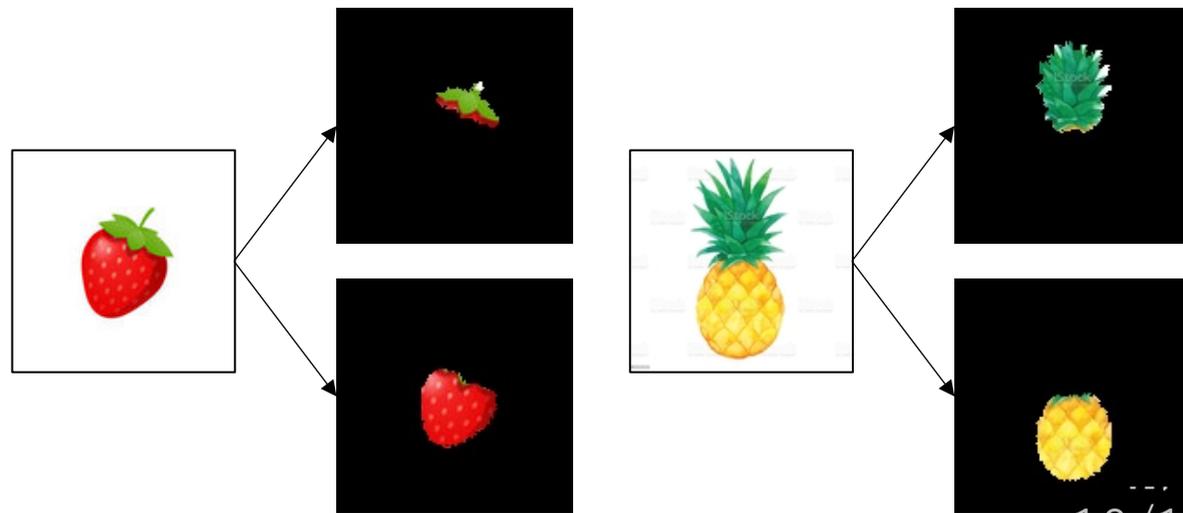


入力したイラスト画像から複数に分解したパーツ画像を抽出する

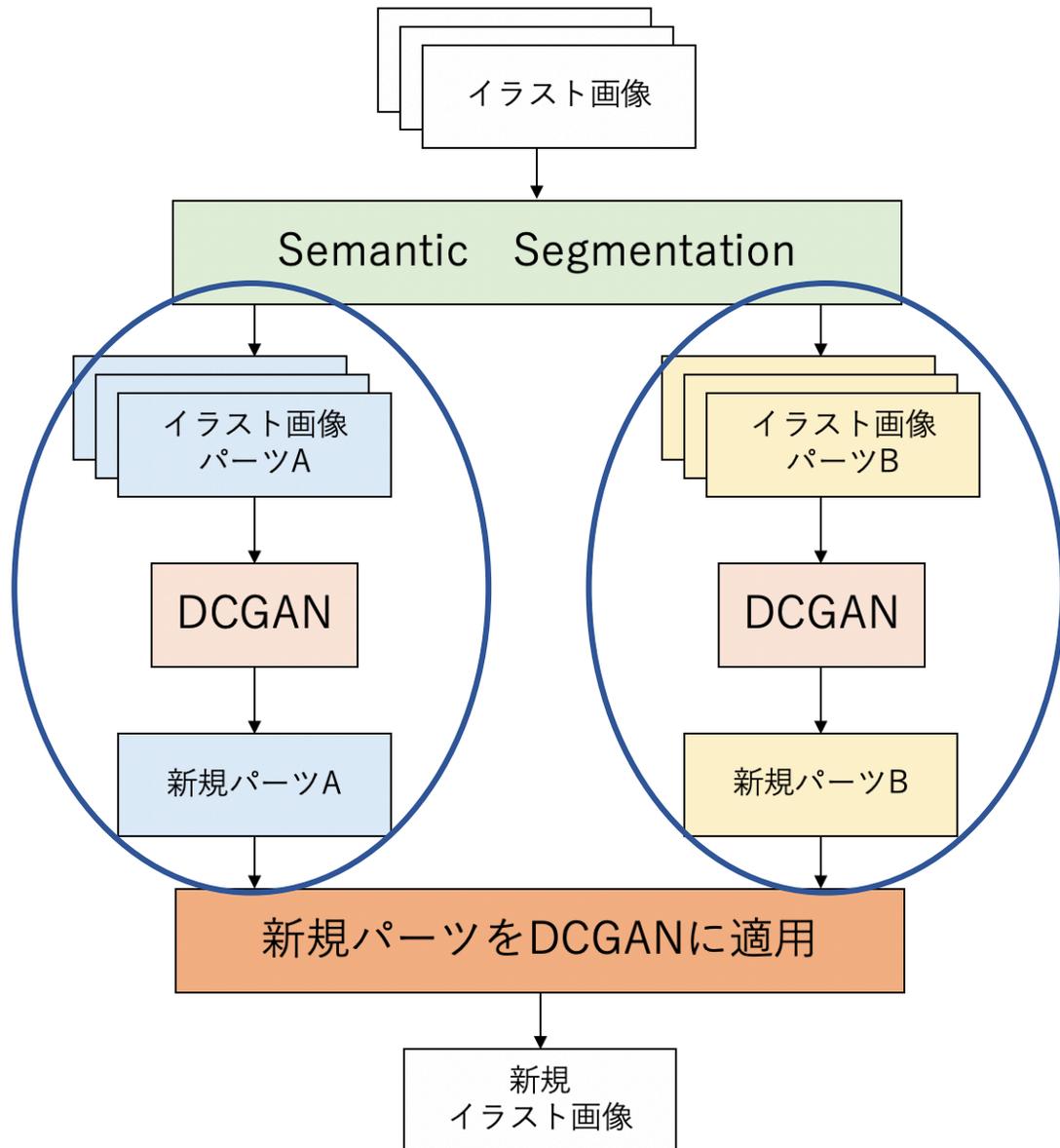
詳細情報

- ・ アノテーションデータ数
→ イチゴ：92枚、パイナップル：71枚
- ・ 入力画像、出力画像ともに128×128px
- ・ 手法はDeep Lab v3+を用いる

[7]. Jonathan Long et al., "Fully Convolutional Networks for Semantic Segmentation", 2015



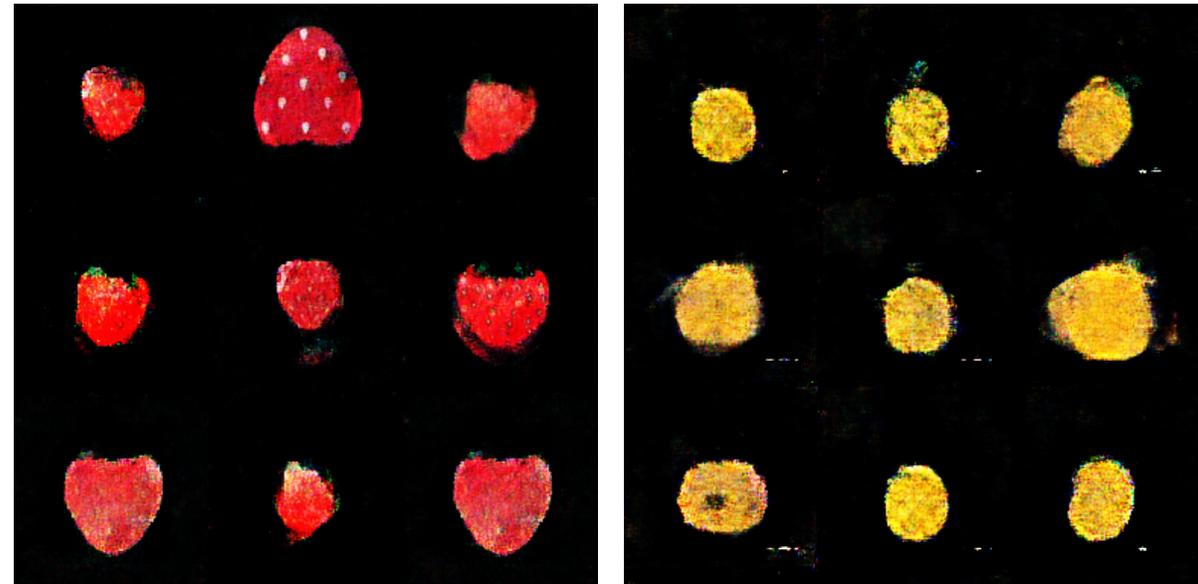
新規パーツ画像の生成



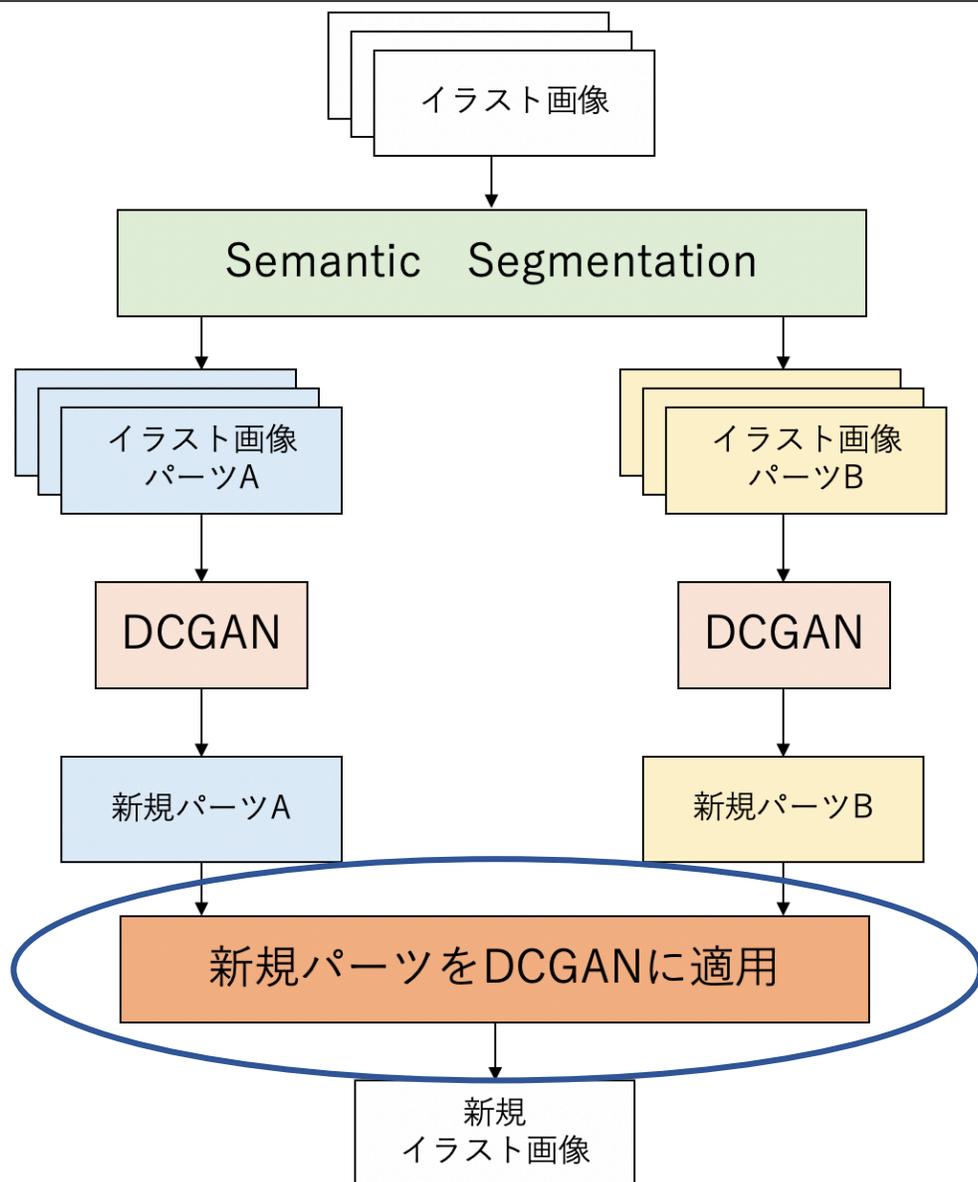
抽出されたパーツ画像をDCGANに適用し新規パーツ画像を生成する

詳細情報

- ・ 入力画像、出力画像ともに128×128px
- ・ 各パーツ画像2000epochまで生成



新規イラスト画像の生成



生成された新規パーツ画像を一つのデータセットに集約しDCGANに適用、各パーツ画像の特徴を持った新規イラスト画像を生成する

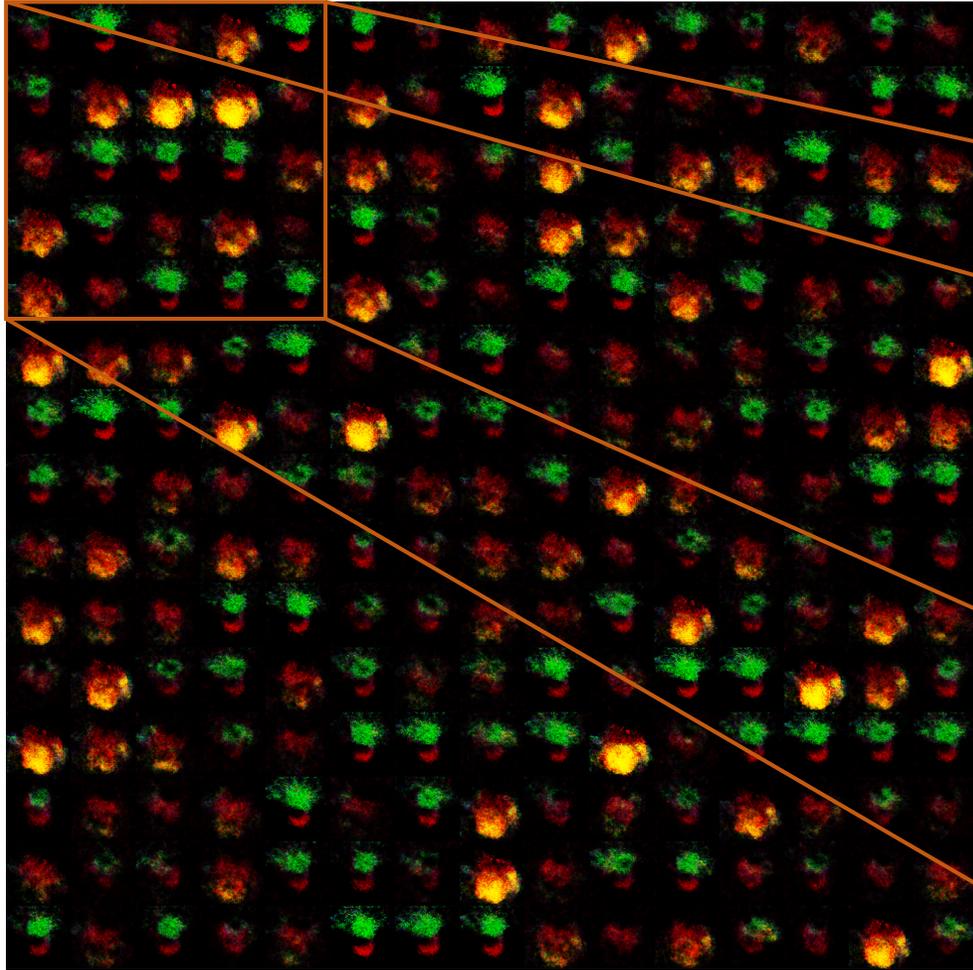
詳細情報

- ・ 入力画像、出力画像ともに128×128px
- ・ データセット内の各パーツ画像数は同一



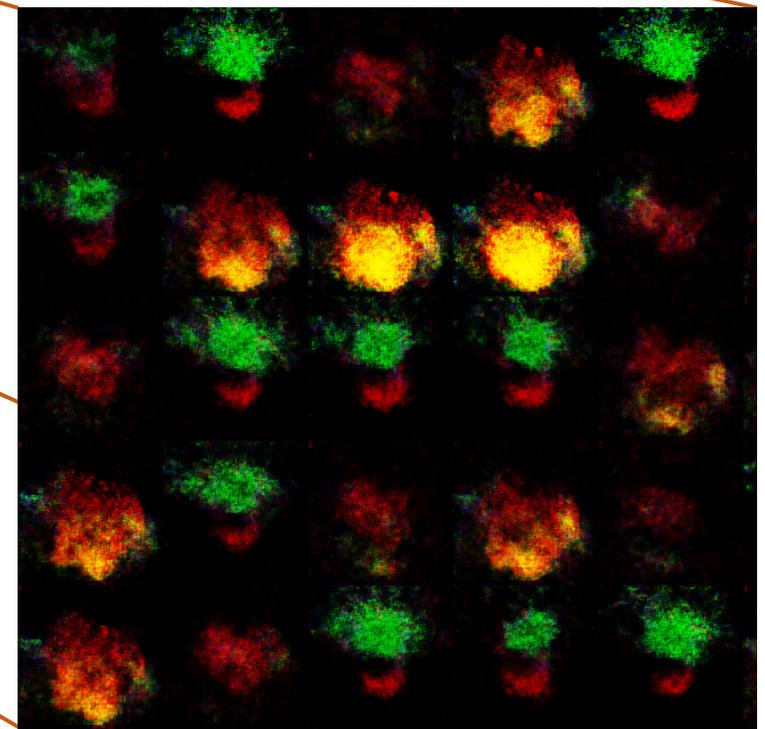
データセットの例

実験結果：イチゴ+パイナップル

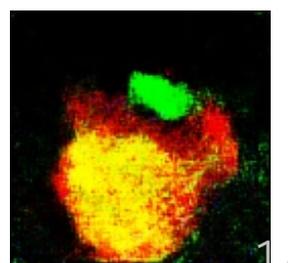
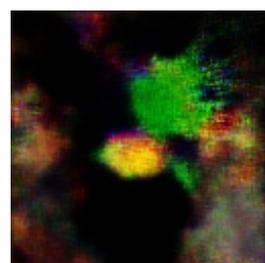
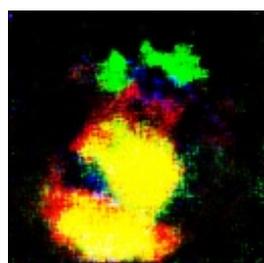
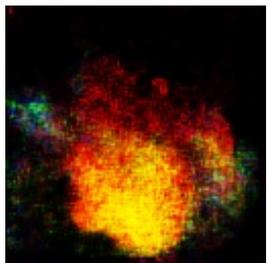
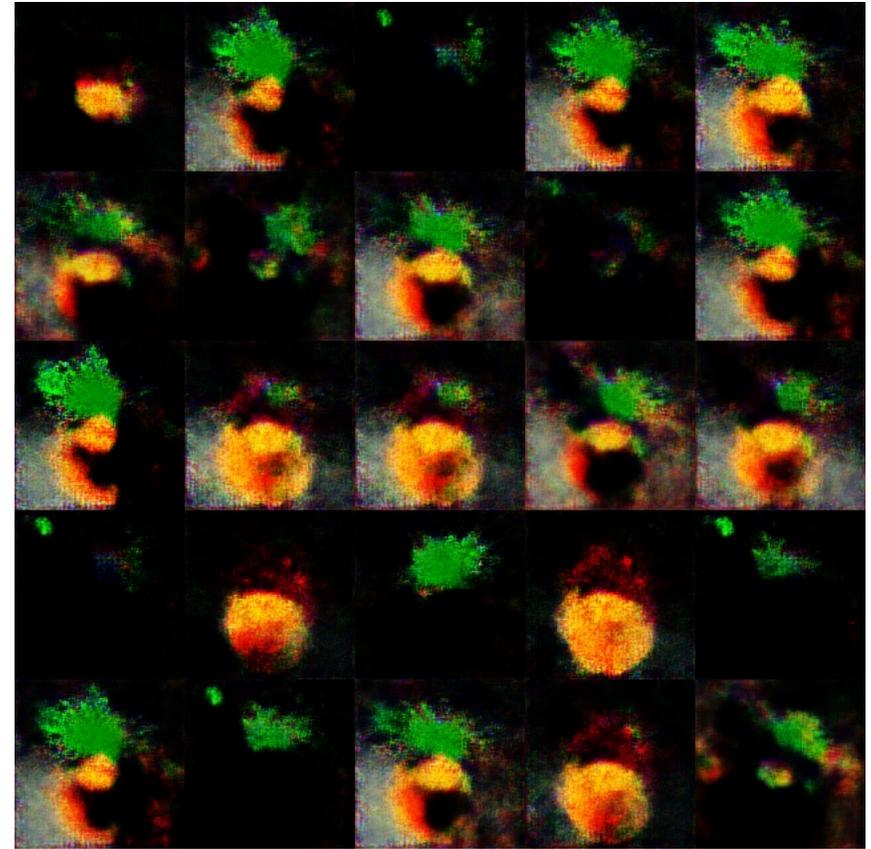
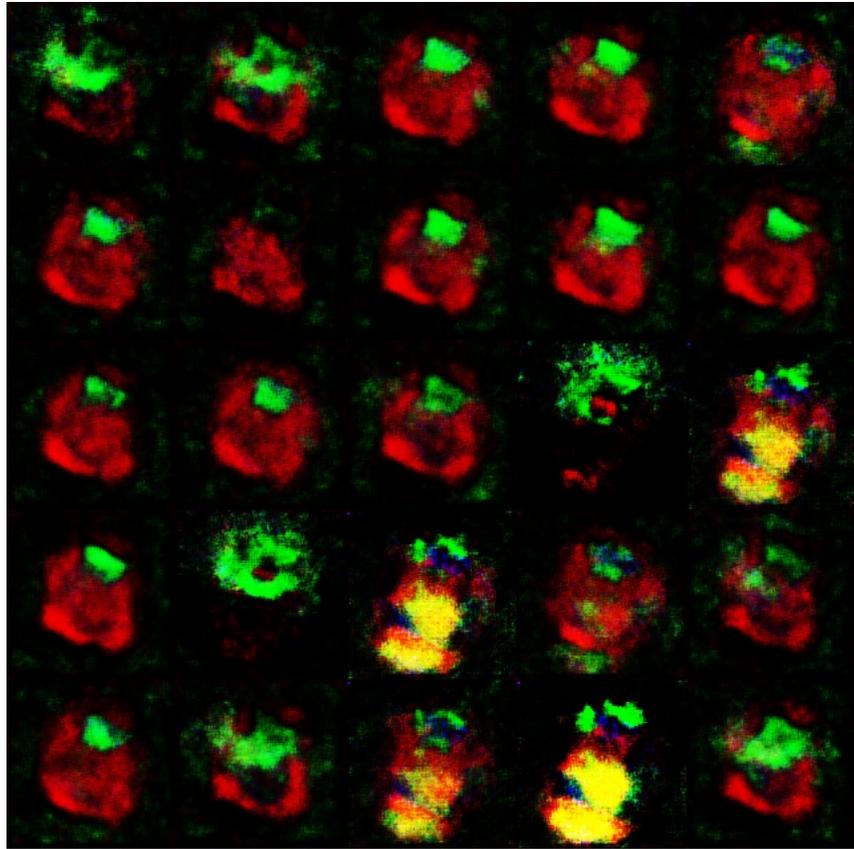


生成画像の詳細

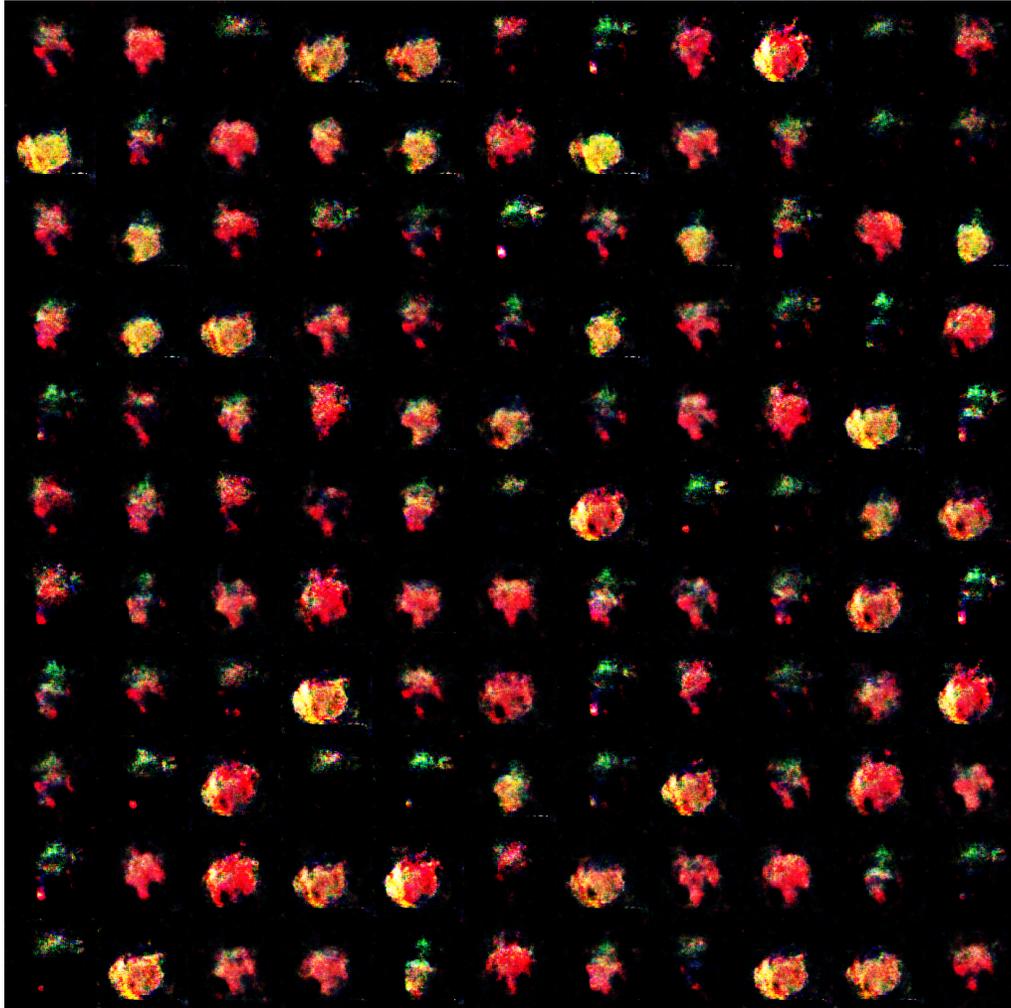
- ・各パーツ画像を144枚使用
- ・653epochまで生成
- ・画像サイズ128×128px



実験結果：イチゴ+パイナップル

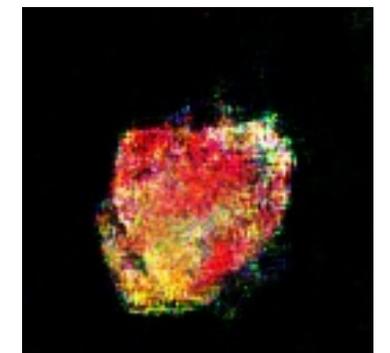
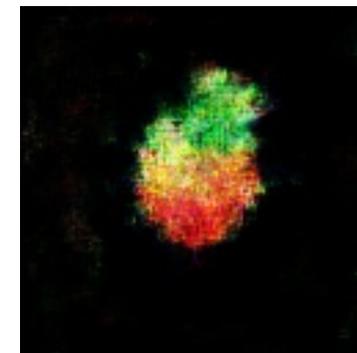
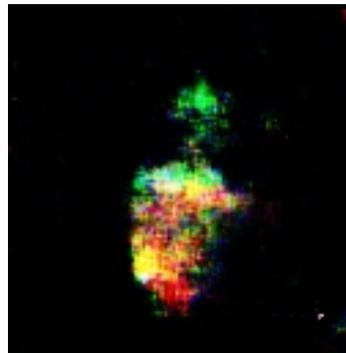


実験結果：イチゴ+パイナップル（改良）

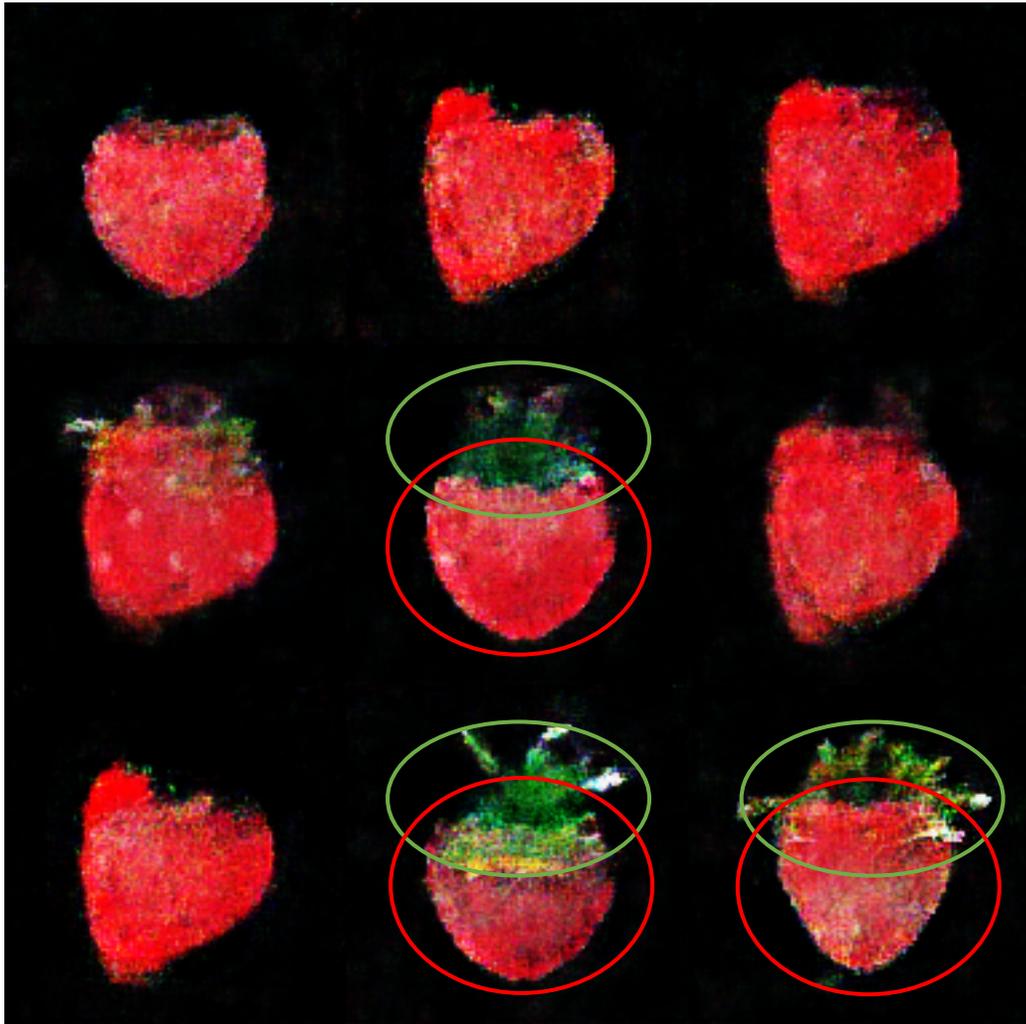


生成画像の詳細

- ・ 各パーツ画像を35枚使用
- ・ 3000epochまで生成
- ・ 画像サイズ128×128px

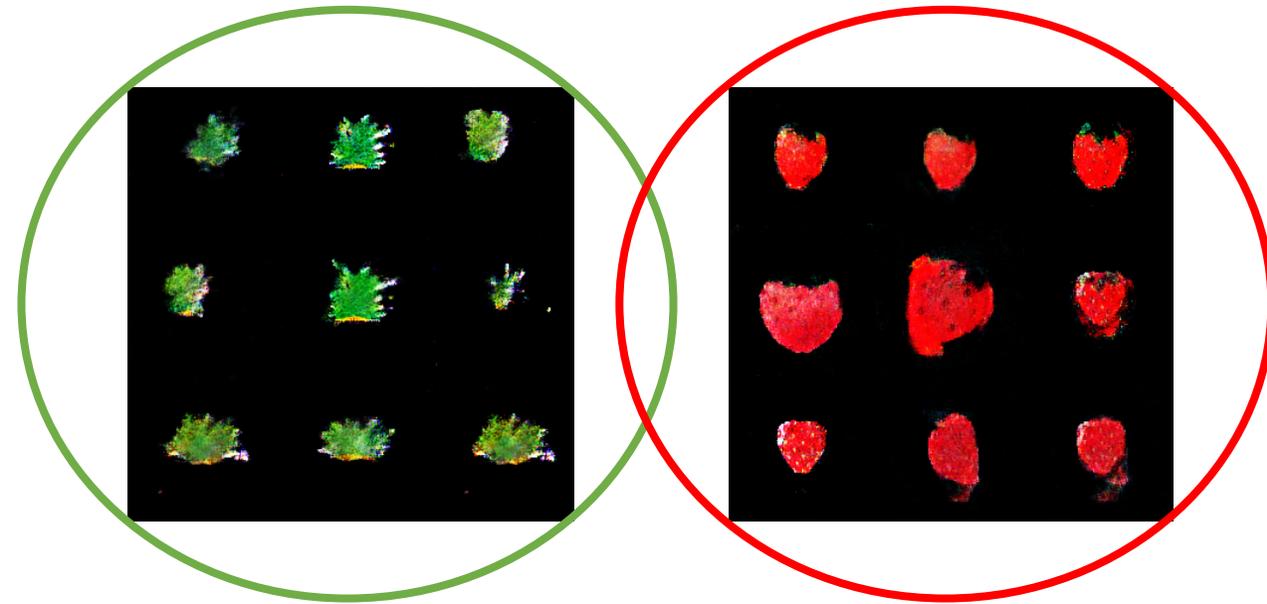


実験結果：パイナップルの冠芽+イチゴの実

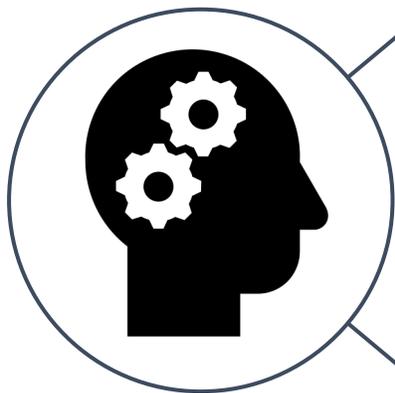


生成画像の詳細

- ・各パーツ画像を30枚使用
- ・3000epochまで生成
- ・画像サイズ128×128px

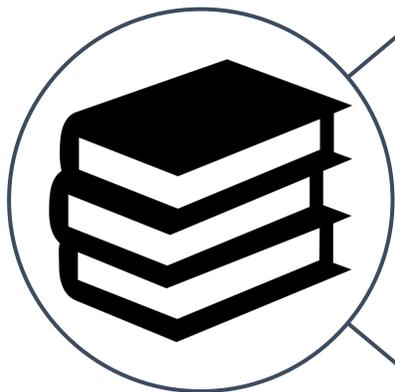


今後の展望



利便性を高める

- 被験者実験などを行い、一般的な感性を学習
→ より利用者の意図に沿ったものを生成



創造性を高める

- 本実験で使用した画像以外のパーツ画像を学習
→ 多様な出力画像を生成

参考文献

- [1]. 坂本真樹「作詞 AI による人の創造力と想像力増幅の試み：電☆アドベンチャー作詞 の経緯と今後の可能性 (小特集 創造性・芸術性における AI の可能性)」, 『電子情報 通信学会』, 102(3), pp.234-239. 2019
- [2]. Bill Yuchen Lin et al., “CommonGen: A Constrained Text Generation Challenge for Generative Commonsense Reasoning”, EMNLP-Findings’20, 2020
- [3]. Alexander Mordvintsev et al., “Inceptionism: Going Deeper into Neural Networks”, 2015
- [4]. DALL·E, “Creating Images from Text”, January 5, 2021
<https://openai.com/blog/dall-e/>
- [5]. Liang-Chien Chen et al., “ Rethinking Atrous Convolution for Semantic Image Segmentation”, 2017
- [6]. Alec Radford et al., “UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS”, Under review as a conference paper at ICLR 2016
- [7]. Jonathan Long et al., “Fully Convolutional Networks for Semantic Segmentation”, 2015
- [8]. Ian J. Goodfellow et al., “Generative Adversarial Nets”, 2014