令和２年度卒業論文

深層学習における自然物の認識特性の調査

宮崎大学工学部　情報システム工学科

佐枝　礼都

指導教員　椋木雅之

目次

1. はじめに

近年、大規模な畳み込みニューラルネットワークによる深層学習の手法は、コンピュータビジョン分野において大きな成果を上げている。画像認識問題の１つに画像に写っている対象のクラス名称とその対象が画像のどこに写っているかをピクセルレベルで判別するセマンティックセグメンテーション[1]がある。この問題では、車や飛行機などある程度形状が決まっている人工物を対象とすることが多く、空や海のような形状が決まっていない自然物を対象とすることが少ない。人工物は、ある程度形状が決まっていることから、特徴抽出の際に形状を重視していると仮定できる。しかし、形状が定まっていない自然物の特徴抽出は、人工物とは異なると考えられる。本研究では、セマンティックセグメンテーションのための手法の１つであるDeepLab-v3+[2]を用いて深層学習における自然物の認識特性を調査する。

2. 深層学習による物体認識

2.1. セマンティックセグメンテーション

本研究では、ピクセル単位での画像内の物体の領域を抽出するセマンティックセグメンテーション[1]を扱う。セマンティックセグメンテーションは画像単位のクラス分類やバウンディングボックスレベルの物体検出といった視覚認識タスクよりも厳密に位置を検出する必要があり、物体認識の分野において難易度の高い問題と考えられていたが、深層学習による手法が提案され、性能が向上してきている。セマンティックセグメンテーションでの認識は形状、色、配置といった特徴量を抽出していると考えられるが、何に注目しているかは自明ではない。

2.2.DeepLab-v3+

本研究では、セマンティックセグメンテーションを行う手段としてDeepLab-v3+を用いる。図 1にDeepLab-v3+のネットワーク構造を示す。DeepLab-v3+[2]はDeepLab-v3[3]を拡張したものである。シンプルかつ効果的なデコーダモジュールが追加されており、特に物体の境界付近のセグメンテーション処理の精度が高い。さらに、Atrous空間ピラミッドプーリングとデコーダモジュールの両方に対し、深さ方向に畳み込みを分解するDepthwise Separable Convolution[4]を適用することにより、高速で強力なセマンティックセグメンテーション用エンコーダデコーダネットワークを実現している。

学習済みのDeeLab-v3+に入力画像を与えると、各画素がどの認識対象物の一部であるか色分けされた画像（ラベル画像）が出力される。

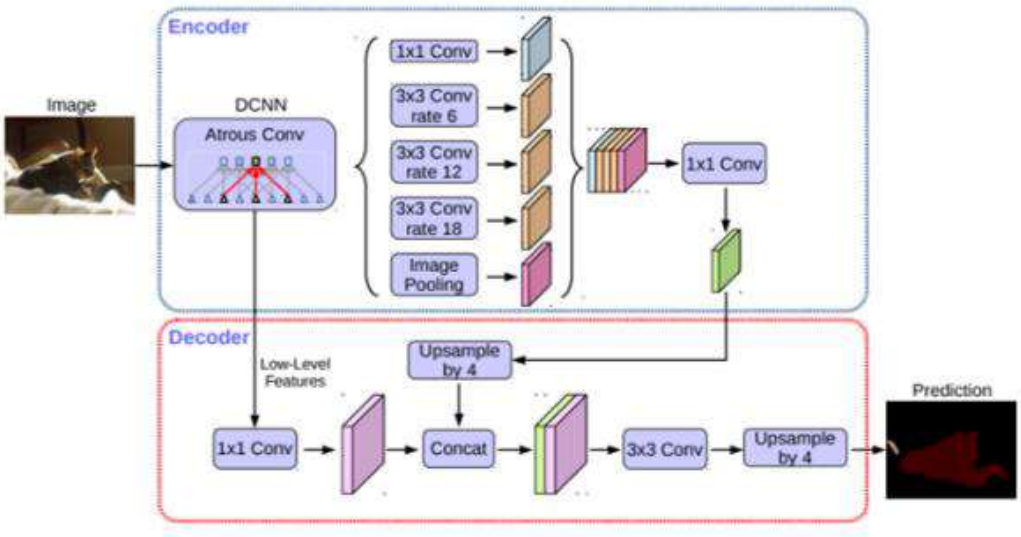

図 1　DeepLab-v3+のネットワーク構造[2]

3. 自然物の認識

3.1. 自然物と人工物の定義

本研究では、空や山のように物体としての輪郭が存在せず形状が一律に決まっていない物体を自然物と定義する。本研究で定義した自然物の画像（以後自然物画像と表記する）の例を図２に示す。自然物画像の特徴として、空と海は似たような色をしていたり、山と海、空と海の境界はある程度はっきりしていたりする特徴がある。一方で、車や飛行機等、物体としての輪郭が存在し、形状がある程度一律に決まっている物体を人工物と定義する。本研究で定義した人工物の画像（以後人工物画像と表記する）の例を図３に示す。ビン、コンピュータのようにそれぞれ単体の物体としての輪郭が決まっている。

図 2　自然物画像の例

図 3　人工物画像の例

3.2. 認識特性の調査

人工物は、物体としての輪郭が存在していることから、深層学習による特徴抽出では形状を重視していると仮定できる。一方、自然物は、物体としての輪郭が存在せず、形状も一律でないことから、色や模様を重視していると仮定できる。自然物を認識対象とする場合、このような性質の違いが影響し、従来認識対象とされていた人工物と異なるアプローチが必要となる可能性がある。本研究ではこのような仮定の下、深層学習によるセマンティックセグメンテーションにおいて、人工物と自然物で認識特性がどのように異なるか調査する。

4. 実験

4.1. 実験１. 転移学習を用いた調査

深層学習では多数の画像を畳み込みニューラルネットワークで学習することで、ネットワーク内の重みとして特徴抽出器が形成される。人工物画像を学習した場合、形状を重視した特徴抽出器が形成されると考えられる。本実験では人工物画像で学習した特徴抽出器が、異なる性質をもつ自然物画像の認識に有効であるか調査する。

一般に転移学習とは、あるタスクについて得られている知識を、関連する新規のタスクに利用する手法の総称である。本実験では人工物画像で学習済みの重みを特徴抽出器として利用し、自然物画像の学習に利用する転移学習を行う。

実験では、転移学習あり、なしで認識精度にどのように影響するか調査する。一般に、転移学習は認識精度の向上に有効とされているが、自然物と人工物に認識処理上の特性の違いがあれば、転移学習がそれ程有効に働かないと考えられる。

4.1.1. 実験手順

DeepLab-v3+の学習・評価には、入力画像とその正解のセマンティックセグメンテーション結果の画像（正解画像）の組からなる訓練用データ・評価用データが必要である。

自然物画像の学習、評価のために SUN2012 データセット[5]を用いる。このデータセットは 16873 枚の画像があり、4919 種類のクラスが含まれている。本実験では「山」「海」「空」を認識対象物とする。認識対象物のそれぞれの領域に対して、注釈情報としてその領域の位置とクラスが与えられている。注釈情報から正解画像を作成する。正解画像は、画像中の各画素がどの認識対象物に属するか色分けされたラベル画像となる。入力画像とそれに対する正解画像の例を図４に示す。入力画像と正解画像 8336 組を訓練用データとする。また、100 組を評価用データとする。

実験では転移学習ありと転移学習なしの２通りを行う。転移学習ありでは、PASCAL VOC 2012 Semantic Segmentation Dataset[6]によって事前に学習済みの重みを使用する。このデータセットは、人工物を主とした 20 クラスの認識対象物を含んでいる。データセットから 1464 組を訓練用データとして用いて事前に学習を行い、得られた

学習済みの重みを初期値として、上記の自然物画像を学習する。転移学習なしでは、初期値をランダムとする。

転移学習あり、なしともに DeepLab-v3+の損失関数が収束するように 12000 回学習を行う。学習終了時には、損失関数は十分収束していた。

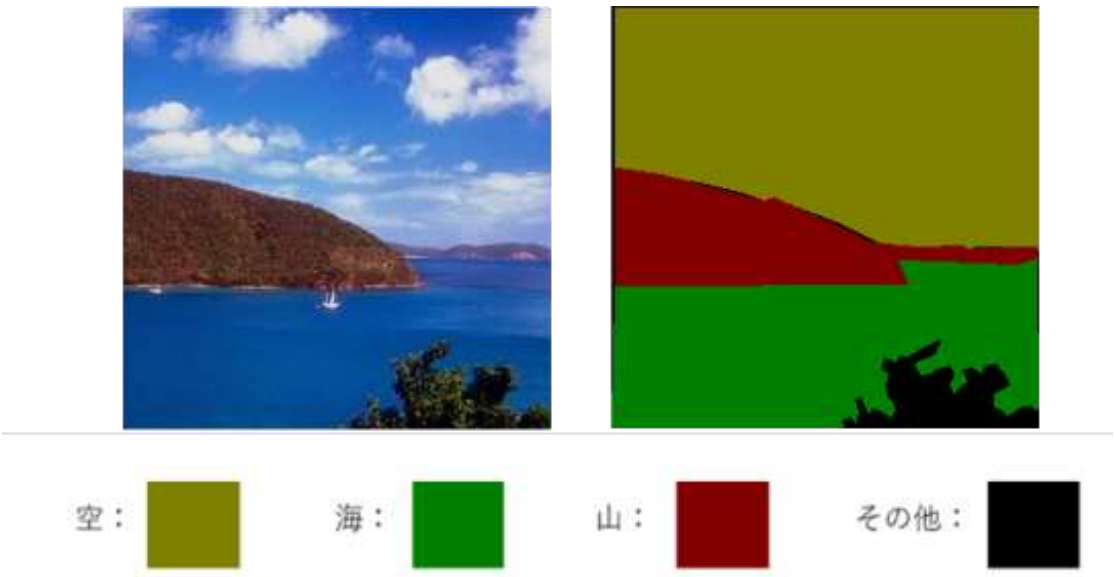

図 4　入力画像とその正解画像の例

4.1.2. 実験結果

実験結果を表１に示す。表中の値は、評価用データに対する認識対象物の IOU の平均値である。IOU は下式で示されるように、正解画像中の認識対象物の領域（正解領域）と認識結果画像中の認識対象物の領域（予測領域）の重なり度合いを表す指標である。正解領域と予測領域の重なりが大きいほど値が大きくなる。

$$IoU = \frac{\text{正解領域} \cap \text{予測領域}}{\text{正解領域} \cup \text{予測領域}}$$

表１より、「山」「海」に関しては、転移学習なしよりも転移学習ありの方が約 0.2 良い結果となっている。「空」に関しては、転移学習なしとの差が約 0.02 とわずかであったが、転移学習ありの方が良い結果となっている

これより、自然物画像では、人工物画像を事前に学習した重みから生成された特徴抽出器が有効であり、深層学習による自然物と人工物の認識特性には、この実験からは違いが見られなかった。

表 1　実験１の結果（平均 IOU）

	転移学習なし	転移学習あり
山	0.556	0.769
海	0.525	0.727
空	0.879	0.902

4.2. 実験２. 回転画像を用いた評価

人工物に対して深層学習による認識が形状を重視しているのであれば、入力画像を回転させて与えた場合、形状が変わるので認識精度は下がると考えられる。一方、自然物に対しては色や模様を重視しているのであれば、回転画像を与えても認識精度にはそれほど影響を与えないと考えられる。

そこで、人工物画像、自然物画像での深層学習が、回転画像での評価に影響を与えるかどうかを調査する。回転画像は0°、90°、180°回転させた画像を用いる。

4.2.1. 実験手順

学習は自然物画像、人工物画像ともに実験 1 と同じ人工物画像で学習済みの重みを用いて転移学習を行う。

自然物画像の認識実験では SUN2012 データセット[5]から 8336 組の画像を訓練用データとし、300 組を評価用データとする。人工物画像の認識実験では PASCAL VOC 2012 Semantic Segmentation Dataset[6]から画像 8364 組を訓練用データとし、300 組を評価用データとする。自然物画像、人工物画像それぞれに対して DeepLabv3+ の損失関数が十分収束するように 12000 回学習を行う。訓練用データは回転させずに与える。学習済みのネットワークに対し、自然物画像、人工物画像の評価用データの入力画像として 0°、90°、180°回転させた画像を与え、ラベル画像を出力する。自然物画像の評価対象は「山」「海」「空」とし、人工物画像の評価対象は「テーブル」「ビン」「コンピュータ」とする。出力されたラベル画像と評価用データの正解画像を比較することで結果を評価する。

4.2.2. 実験結果

実験結果を表２、表３に、出力されたラベル画像の例を図５、図６、図７に示す。

まず、自然物画像の評価に関して、表２より、「山」「海」「空」ともに 0° の結果よりも 90° と 180° の IOU が低い。特に「海」に関して、90° で 0.030 と「山」「空」と比較しても IOU が低い結果となった。自然物画像では回転画像を与えても深層学習での認識に影響を与えないと考えられたが、実際には影響を受けていることが分かる。また、図５より、「空」、「海」に関して、180° では、「空」である場所を「海」と認識し、「海」である場所を「空」と認識していることが分かる。これより、自然物画像の深層学習では、位置の情報を利用していると考えられる。

次に人工物画像の評価に関して、表３より、「ビン」「テーブル」は 0° と比較して 90° 、180° で IOU が低い。また、図６より、特に「テーブル」のラベル画像では 90° 、180° で認識ができていないことが明らかである。一方、「コンピュータ」は 90° 、180° の IOU 結果が約 0.85 であり、「ビン」「テーブル」と比較して IOU が高い。図７からも、それぞれの回転画像で認識に差があまり見られない。これより、

人工物画像では回転画像での認識精度は下がると考えられたが、「ビン」「テーブル」のように下がるものもあれば、「コンピュータ」のように下がらないものもあることが分かる。

表 2　実験２:自然物画像の結果（平均 IOU）

	0°	90°	180°
山	0.769	0.462	0.419
海	0.727	0.030	0.220
空	0.902	0.684	0.214

表 3　実験２:人工物画像の結果（平均 IOU）

	0°	90°	180°
ビン	0.788	0.462	0.434
テーブル	0.963	0.282	0.229
コンピュータ	0.949	0.845	0.838

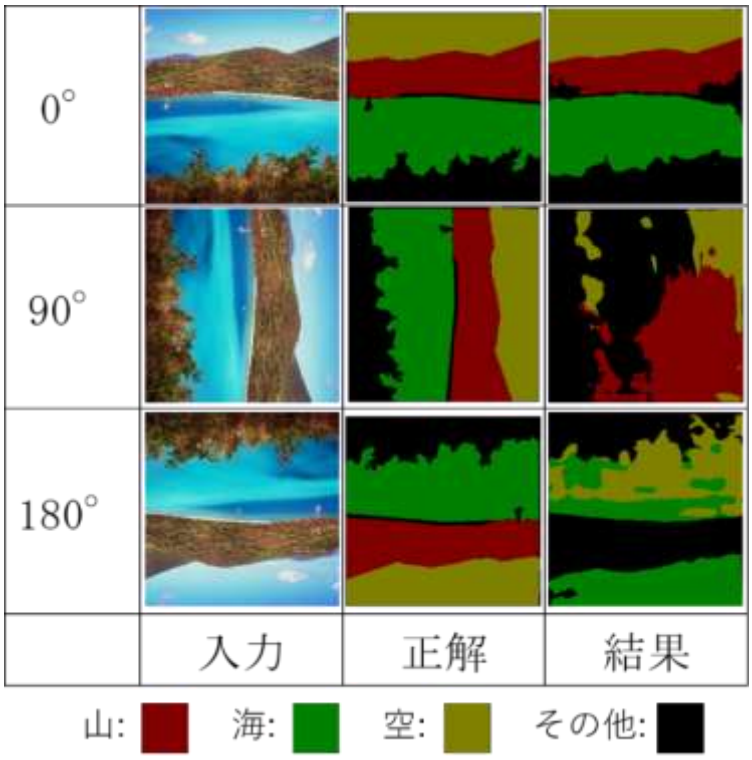

図 5　実験２:自然物画像の結果の例

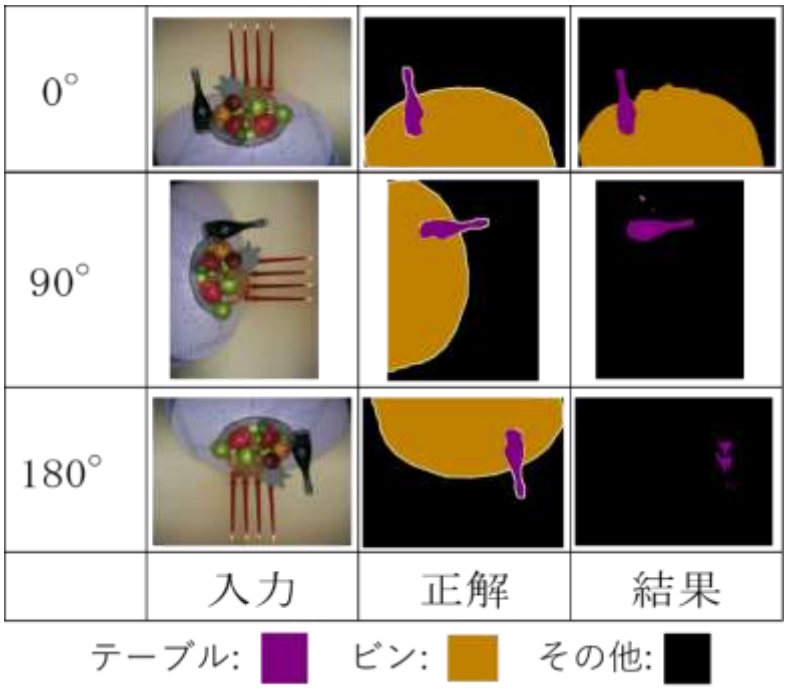

図 6　実験２:人工物画像の結果の例１

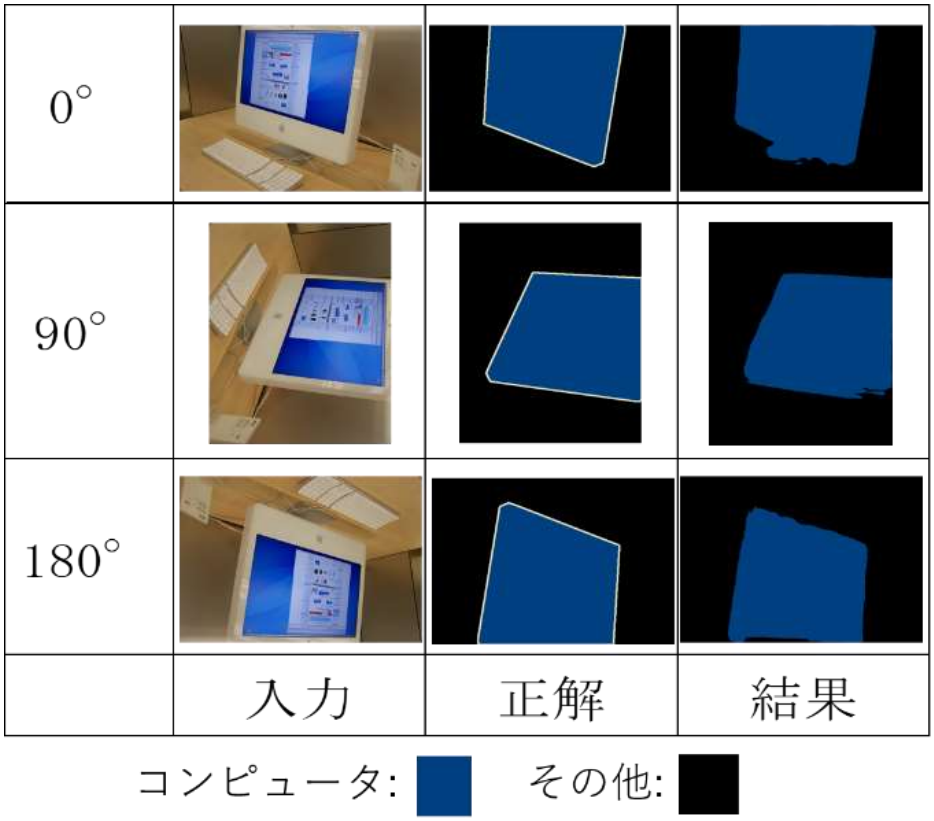

図 7　人工物画像の結果の例２

4.3. 実験３. 転移学習を用いた追加調査

実験１では、人工物画像での学習済みの重みを転移学習で利用して自然物画像を学習した場合と、自然物画像のみを学習した場合でセマンティックセグメンテーションの結果を比較した。この場合、前者は後者よりも多くの画像を利用して学習しており、認識結果は、その影響も受けて前者の方が良かった可能性もある。

本実験では、同じ訓練用データを使って、転移学習結果を比較することで、自然物と人工物の認識特性の違いについて追加調査する。具体的には、自然物画像、人工物画像の訓練用データを用意し、人工物画像で学習済みの重みを自然物画像の学習に利用した場合と、自然物画像で学習済みの重みを人工物画像の学習に利用した場合を比較する。

4.3.1. 実験手順

人工物画像で学習済みの重みを特徴抽出器として自然物画像の学習に利用する転移学習（転移学習①）と、自然物画像で学習済みの重みを特徴抽出器として人工物画像の学習に利用する転移学習（転移学習②）を行う。

まず、自然物画像、人工物画像の学習・評価のためのデータを用意する。自然物画像訓練用データとしては、SUN2012 データセット[5]から入力画像とその正解画像 8336 組を、自然物画像評価用データとしては、同じく 100 組を用いる。人工物画像訓練用データとしては、PASCAL VOC 2012 Semantic Segmentation Dataset[6]から 8364 組を、人工物画像評価用データとしては、同じく 100 組を用いる。

転移学習①では人工物画像訓練用データで事前に学習を行った上で、得られた重みを初期値として自然物画像訓練用データで学習する。逆に、転移学習②では、自然物画像訓練用データで学習した重みを初期値として人工物画像訓練用データを学習する。

転移学習①、転移学習②ともに DeepLab-v3+の損失関数が十分収束するように事前学習 12000 回、転移学習 12000 回学習を行う。自然物の評価対象は「山」「海」「空」とし、人工物の評価対象は「テー

ブル」「ビン」「コンピュータ」とする。

4.3.2. 実験結果

実験結果を表４、表５に示す。

表４の転移学習①の結果は、実験１の「転移学習あり」の結果より悪くなっている。実験１では、公開されている事前学習済みの重みを利用したのに対し、本実験では、人工物画像の事前学習も手元で行った。このため、違いが生じたと考えられる。表４と表５を比較すると、「テーブル」の結果がやや悪いが、全体としては同程度の平均 IOU となっている。

転移学習の効果を評価するために、事前学習時点での結果と比較する。表６、表７に事前学習時点での結果を示す。

自然物画像の結果に関して、表４、表７より、自然物画像のみでの学習よりも転移学習①の結果が「山」では 0.080、「海」では 0.195 良い。「空」に関しては転移学習した方が 0.004 悪い。実験１ほど大きな効果は見られなかったが、転移学習①については、人工物画像で事前学習したことが、自然物画像の学習に有効であったと言える。従って、実験１同様、転移学習①の実験からは、自然物と人工物の認識特性の違いは見いだせなかった。

人工物画像の結果に関して、表５、表６より、人工物のみでの学習

よりも転移学習②の結果が「ビン」では 0.04 良くなっているが、「テーブル」では 0.160、「コンピュータ」では 0.061 悪い。転移学習②では、事前学習のみの方が、全体的に結果が良く、自然物画像で事前学習したことが人工物画像の認識に悪影響を及ぼしたと言える。これは、自然物と人工物の認識特性の違いに起因することも考えられるが、転移学習①ではこのような影響は起こっていなかった。本実験に用いた人工物画像のデータセット（PASCAL VOC）は、画像自体は２０クラス以上の人工物を含んでおり、多様性が高かった。一方、自然物画像のデータセット（SUN2012）は風景画像のみで比較的類似した画像が多かった。事前学習では多様な画像を与える方が、より汎用的な特徴抽出器を学習できる。転移学習②の実験では、使用した画像の多様性の差が結果に表われたと考えられる。

表 4　実験３:転移学習①の結果（平均 IOU）

	転移学習①
山	0.636
海	0.720
空	0.875

表 5　実験３:転移学習②の結果（平均 IOU）

	転移学習②
ビン	0.738
テーブル	0.598
コンピュータ	0.835

表 6　実験３:転移学習①の事前学習の結果（平均 IOU）

	事前学習
ビン	0.694
テーブル	0.758
コンピュータ	0.896

表 7　実験３:転移学習②の事前学習の結果（平均 IOU）

	事前学習
山	0.556
海	0.525
空	0.879

5. おわりに

本研究では、深層学習における自然物の認識特性調査を行った。自然物は形状が決まっていないのに対し、人工物は形状がある程度決まっている。このことから、深層学習において自然物は色や模様を重視していると仮定し、人工物は形状を重視していると仮定したが、本研究での２つの実験では、自然物と人工物の認識特性に違いは見られなかった。よって、本研究では、自然物に対して人工物とは異なる特有の深層学習のアプローチが必要とは言えない。

実験２において、自然物画像の認識は位置による影響を受けていると考えられた。今後の課題として、位置の情報を使うことにより、認識率にどう影響するかを調査することが必要と考える。

謝辞

最後に、本研究を行うにあたり、ご指導いただいた椋木雅之教授に深く感謝致します。指導教員である椋木雅之教授には、お忙しい中研究のアイデアや実験、論文の記述方法に関する助言といったアドバイスなど沢山のご指導をいただきました。また、忙しいにも関わらず、研究に関するアドバイスをしてくださり、知見を共有できた椋木研究室の皆様にお礼申し上げます。

参考文献

[1] Martin Thoma, ”A Survey of Semantic Segmentation”, arXiv:1602.06541[cs.CV], 2016

[2] Liang-Chieh Chen, “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”,arXiv:1802.02611 [cs.CV], 2018

[3] Liang-Chieh Chen,“Rethinking Atrous Convolution for Semantic Image Segmentation”, arXiv:1706.05587 [cs.CV], 2017

[4] Chollet F, “Xception: Deep learning with depthwise separable convolutions” CVPR. 2017

[5]SUN Database,https://groups.csail.mit.edu/vision/SUN/ (2017/02参照)

[6]VOC2012, http://host.robots.ox.ac.uk/pascal/VOC/voc2012/ (2017/02 参照)