

Pretext task を用いた植物画像からの分けつ数の推定

内海ゆづ子[†] 中村浩一朗[†] 岩村 雅一[†] 黄瀬 浩一[†]

[†] 大阪府立大学大学院工学研究科 〒599-8531 大阪府堺市中区学園町 1-1

E-mail: {yuzuko, masa, kise}@cs.osakafu-u.ac.jp, cn346ann0@yahoo.co.jp

あらまし 農学や生物学において、植物の生育状態の把握や植物の特徴を知るため、植物の形態形質（形状やサイズなどの見た目の特徴）の計測が必要不可欠である。しかし、形質によっては計測に非常に手間がかかることから、計測の自動化が求められている。本稿では、植物の形質の中でも、イネなどの単子葉植物の枝分かれの様子を表す分けつに焦点をあて、画像から分けつ数を自動的に推定する手法の提案をする。現在、画像認識の分野では、Deep Neural Network (DNN) に基づく手法が圧倒的な性能を発揮しており、植物計測に対しても DNN の利用が有効であると考えられる。しかし、DNN では大量の学習データが必要とされるため、分けつのような計測に手間がかかるものに対して、大量の正解ラベル付き学習データを用意することが難しい。そこで、本研究では、分けつ数が不明な大量の植物画像を用いて、自動的に正解ラベルを付与できる pretext task によりネットワークを学習し、これを分けつ数の推定に転用する。このことにより、小規模なラベル付きデータしか用意できない分けつ数の推定に対して、DNN を適用可能となる。実験の結果、pretext task を用いた一部の推定器が、従来の計測値に基づいて構築した分けつ数推定モデルよりもよりよい推定結果を示すことが明らかとなった。

キーワード 植物形態形質計測, 分けつ, Deep Neural Network (DNN), Pretext task

1. はじめに

植物の草丈や形状といった見た目の特徴（形態形質）を計測することは、生育状態の把握や、植物の特性などを知るのに非常に重要な作業である。例えば、草丈を観測することで、植物の生育状態を把握したり、草の拡がりやを把握することで、その木の倒れやすさや、光合成のしやすさなどの特性を知ることができる。しかし、植物の形状を計測する作業は困難な事が多く、作業に多くの手間がかかることが多い。そのため、破壊を伴わない植物の形状形態の自動推定が必要とされている。

本研究では、植物の形質のうち、イネなどの単子葉植物の生育状態を把握する際に用いられる分けつを対象とし、その自動推定を行う。分けつは、図 1 に示すように、単子葉植物の枝分かれを示し、生育の初期段階から追従して葉の数の増加を調べることで、分けつ数を数えることができる。また、生育の初期段階が分からない場合は、破壊調査をすることで分けつ数を知ることができる。このように分けつを数えるには非常に手間がかかったり、破壊を伴う調査が必要になることから、画像から自動的に推定できれば、作業の省力化に貢献できる上、非破壊の調査が可能となることから、継続した植物の観測が可能となる。

従来、分けつの推定は、衛星から観測された可視域の赤色光の反射率と近赤外域の反射率の比である Normalized Difference

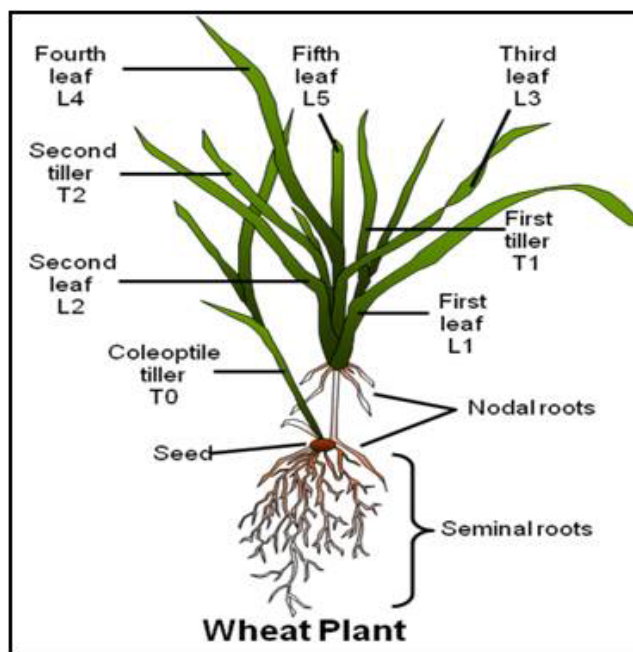


図 1: 麦の分けつの例^(注1).

Vegetation Index (NDVI) を用いて、圃場の単位面積あたりに含まれる植物の分けつ数をおおまかに推定するものがほとんどであった [1], [2], [3]. そのため、従来の多くの手法では、個体の分けつ数を推定することは不可能であった。これに対して、個体の分けつ数を推定する手法として、画像処理を用いて推定するものが提案されている [4], [5]. これらの手法は、研究者の

(注1) : <http://losangelesbreadbakers.blogspot.com/2013/01/more-on-wheat-tillering.html>.

知見に基づいて作られた特徴量を基に、線形回帰を用いて分けつ数を推定している。そのため、画像中の一部の限られた特徴のみが分けつ数の推定に反映されておらず、画像から分かる見えた目の特徴を全てを活用できていない。

画像認識の分野では、Deep Neural Network (DNN) に基づく手法が圧倒的な性能を発揮している。DNN では、画像から直接特徴量を学習するため、見えた目の特徴を十分に反映させた特徴が得られることから、分けつ数の推定に対しても、DNN を用いることが有効であると予測できる。しかし、DNN の性能を最大限発揮するには、大量の学習データが必要である。上で述べた通り、分けつ数を数えることは非常に手間がかかることから、大量のラベル付学習データを用意することが困難であるため、DNN を直接利用することは難しい。

コンピュータビジョンの分野でも、大量の学習データはDNNを利用する際の障害となっており、学習データが十分に用意できない場合でもDNNの性能を発揮できるような手法が提案されている。例えば、本来解くべき問題の学習データとは異なる学習データを用いてネットワークを学習し、得られたモデルを用いて識別を行う transfer learning [6]、一部のみラベル付けがされた学習データを利用する semi-supervised learning [7]、自動的にラベルを付与してネットワークを学習する self-supervised learning [8]、が提案されている。Self-supervised learningのうち、自動的に付与されるラベルを用いて本来解きたい問題とは異なるタスクを解くことで特徴量を学習し、その特徴量を用いて本来解くタスクを解く手法が提案されており、従来の本来のタスクを直接解く手法と同等の性能を示している [8], [9], [10]。この、本来解きたい問題とは異なるタスクを pretext task と呼び、大半にラベルが付与されていない学習データに対しても適用できる。

本稿では、大量のラベル付き学習データを用意できない分けつ数の推定に対して、DNN を用いた手法を提案する。我々の知る限り、分けつ数の推定にDNNを用いたのは本稿が初めてである。提案手法では、ラベル付きの学習データが大量に用意できないことから、pre-trained model や pretext task を用いることで、DNN を用いた分けつ数の推定を実現する。

2. 関連研究

本章では、分けつ数の推定と pretext task に焦点を当て、関連研究について詳しく述べる。

2.1 分けつ数の推定

分けつ数は麦、米といった単子葉植物の生育を把握するための指標として用いられている。数えるのに手間がかかったり、破壊を伴う調査が必要な場合もあることから、自動でかつ非破壊に計測をする手法が提案されてきた。生育の把握を目的とした場合、リモートセンシングを用いて広大な圃場を計測し、分けつ数を推定する手法が主に提案されている ([1] の Table 1 を参照)。その中で、NDVI は麦の分けつ数推定のために最も頻繁に利用されている指標である [1], [2], [3]。リモートセンシングを用いた推定は、広い領域にある植物のおおまかな分けつ数を調査するには向いているものの、個体の分けつ数を推定

することは困難である。

近年、植物の遺伝子型データ (ジェノタイプ) から、表現型データ (フェノタイプ) を予測するため、植物から形態形質を計測するフェノタイピングの研究が盛んに行われている。フェノタイピングでは、生育状態の把握とは異なり、植物個体の形態形質の計測が必要となる。形態形質の1つである分けつ数でも、フェノタイピングを目的として、個体からの推定が行われるようになってきた。Fahlgren らは、横方向から撮影したエノコログサの画像を用いて、画像中の植物の縦横比や、画像中に占める植物の領域を説明変数として、分けつ数を推定する線形回帰モデルを提案している [4]。画像中の縦横比や面積はオープンソースソフトウェア PlantCV^(注2)を用いて計算しているため、画像を撮影するだけで、自動的に分けつ数の推定が可能である。また、Boyle らは、Frangi filter [11] を用いて麦の画像から特徴量を抽出し、この特徴量を説明変数として線形回帰モデルにより分けつ数を推定している [5]。いずれの手法もヒューリスティックな特徴量を用いているため、これらの特徴量が分けつ数の推定に最も適しているかは不明である。加えて、これらの特徴量は画像の一部の特徴のみを反映させたものであり、画像から分かりうる全ての見えた目の特徴を分けつ数推定に利用できていない。

2.2 Pretext task

ラベル付けされたデータを大量の集めることが困難な問題に対してDNNを適用するのに、pretext task に基づくネットワークの学習の手法が提案されている。通常、pretext task は本来解きたい目的のタスクとは関係のないタスクであり、pretext task で用いられる学習データには、自動的に付与できるラベルを用いる。Pretext task を用いて特徴量を学習し、その特徴量を用いて本来解きたい目的の問題を解くことで、DNN を十分に学習できるだけの学習データがなくても、学習データが充分にある場合と同等の精度を達成することが可能である。

これまで、pretext task としてさまざまなタスクが提案されてきた。画像のカラー化 [12] や、ジグソーパズルを解く [13]、画像の回転角の予測 [8]、画像中の対象物体の数を数える [10] などのタスクが提案されている。これらのタスクにより学習された特徴量は、画像認識やセグメンテーション、物体認識に用いられ、目的の問題を直接解いた時と同等の精度を達成している。提案手法では、いくつかの pretext task を設定し、得られた特徴量をもとに分けつ数の推定を行う。

3. DNN に基づく分けつ数の推定

提案手法は、特徴抽出をする CNN モデルと、分けつ数の推定をする回帰モデルからなる DNN に基づく分けつ数の推定手法を提案する。本稿で特徴抽出に用いる CNN モデルは以下の通りである。

- 一般物体認識用に学習された pre-trained モデル
- Pretext task を解いた モデル

また、用いる回帰モデルは以下の通りである。

(注2) : <https://plantcv.danforthcenter.org>

- サポートベクタ回帰 (Support Vector Regression: SVR)
- 重回帰分析 (Multiple Linear Regression: MLR)

本章では、特徴抽出に用いた2つのモデルと、提案手法での分けつ数の推定方法について説明する。

3.1 Pre-trained モデル

ラベルが付与された学習データが少ない場合、DNN は性能を十分に発揮できない。この、ラベル付き学習データが少ない場合の対象方法の1つに、予め学習されたモデルを用いることが挙げられる。一般的に、目的のタスクに近いタスクで学習されたモデルは、高精度な結果をもたらす特徴量を抽出できると考えられている。このことから、植物を撮影した画像から分けつ数を推定する我々の問題では、自然画像を用いたタスクを解いた pre-trained モデルを特徴抽出に用いるのがふさわしい。そこで、本稿では、ImageNet dataset [14] で学習した VGG 16 の pre-trained モデル [15] を用いる。

3.2 Pretext task で学習されたモデル

分けつ数推定の精度向上のため、pre-trained モデルだけでなく、pretext task を解いたモデルを用いて特徴抽出をする。本稿では、VGG 16 [15] を用いて、画像処理を用いて自動で算出可能な植物の見た目の特徴量を予測する問題を解くことで、特徴量の学習する。提案手法では、画像中の植物の面積と縦横比を予測する pretext task を用いて特徴量を学習することにした。植物の面積と縦横比は、[4] での分けつ数の推定のための回帰モデルの中で説明変数として用いられている。

本稿では、植物の面積や縦横比を予測するのに、2つの方法を用いた。1つは、面積は縦横比を n 段階の離散値に量子化し、入力画像の値がどれにあたるかを推定する。これは、識別問題を解いているとみなすことができる。もう1つの方法は、値そのものを推定した。これは、回帰を用いた推定である。

Pretext task の利用による特徴量の学習を定式化すると、以下のように表現できる。 X を植物画像、 $g(\cdot)$ を画像から植物の縦横比や面積の計算を行うオペレータであると仮定すると、画像から抽出される縦横比や面積 v は以下のように表現できる。

$$v = g(X). \quad (1)$$

v を回帰用いて推定する場合、VGG 16 モデル $F(\cdot)$ はモデルパラメータ θ を用いて、以下の通り v を推定する。

$$F(X|\theta) = \hat{v} \quad (2)$$

今、 N 組の学習データ $\{(X_i, v_i) | i = 1, 2, \dots, N\}$ が与えられたとすると、VGG 16 モデルは以下の問題を解くこととなる。

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N (v_i - F(X_i|\theta))^2 \quad (3)$$

続いて、 v の推定を離散値を用いて認識問題とみなして解く場合を考える。 v を n 段階に量子化し、 $q \in \{0, 1, \dots, n-1\}$ を量子化インデックスとすると、VGG 16 モデル $F(\cdot)$ は v を直接推定するのではなく、量子化インデックス q を推定する。量子化インデックスの推定は、モデルパラメータ θ と、画像 X の値 v が j 番の量子化インデックスとなる確率を表現するベク

トル $\hat{q} = (\hat{q}_0, \hat{q}_1, \dots, \hat{q}_{n-1})^\top$ を用いて

$$F(X|\theta) = \hat{q}. \quad (4)$$

と表す。 N 組の学習データ $\{(X_i, \mathbf{q}_i) | i = 1, 2, \dots, N\}$, $\mathbf{q}_i = (q_{i,0}, q_{i,1}, \dots, q_{i,n-1})^\top$ が与えられた時、VGG 16 モデルは以下の問題を解くこととなる。

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \text{loss}(X_i, \theta) \quad (5)$$

ここで、loss function $\text{loss}(\cdot)$ は

$$\text{loss}(X_i, \theta) = -\frac{1}{n} \sum_{j=0}^{n-1} q_{ij} \log(F^j(X_i|\theta)) \quad (6)$$

となり、 $F^j(X_i|\theta)$ は出力ベクトルの j 番目の値である。

3.3 回帰モデル

予め学習された CNN モデルや、pretext task により学習されたモデルを用いて特徴抽出を行い、得られた特徴量に対して回帰モデルを適用し、分けつ数を推定する。本稿では、SVR と MLR を用いて推定をする。

SVR はサポートベクタマシン (Support Vector Machine: SVM) を回帰に応用したものである。SVR の最大の利点は、カーネルトリックを用いて非線形問題を線形問題として扱うことが可能となる点である。SVR では、カーネル関数を用いて特徴空間を高次元の空間に写像する。この際、カーネル関数が非線形であれば、SVR は非線形回帰問題を扱うことが可能となる。これに加えて、SVR は、DNN で用いるよりもはるかに少ない学習データで学習が可能である。

SVR に加えて、本稿では MLR を用いた。MLR は最も単純な回帰モデルで、隠れ層がない全結合のネットワークと同等である。特徴抽出に CNN モデルを用いているので、特徴抽出をしたネットワークに全結合層を加えるだけで簡単に実装ができることから、回帰モデルとして用いた。

4. 実 験

提案手法を評価するため、実験を行った。本稿では、単子葉植物のうち、エノコログサの分けつ数を推定し、推定結果を従来手法である [4] と比較した。

4.1 実験環境

実験には、[4] で用いられたデータセット^(注3)を利用した。データセットの一部を図2に示す。このデータセットは、屋内の照明条件が一定の環境で撮影された約80,000枚の鉢植えのエノコログサの画像で構成されており、データセットのうち、846枚の画像に手で数えた分けつ数が付与されている。画像はRGBカラーで、解像度は2,454×2,056 pixelsである。実験では、植物の横方向から撮影され、分けつ数が付与されていない画像のうち、10,000枚をpretext taskの学習データとし、2,000枚を評価用データとした。分けつ数が付与されたデータ

(注3) : https://plantcv.danforthcenter.org/pages/data-sets/2013/setaria_burnin2.html

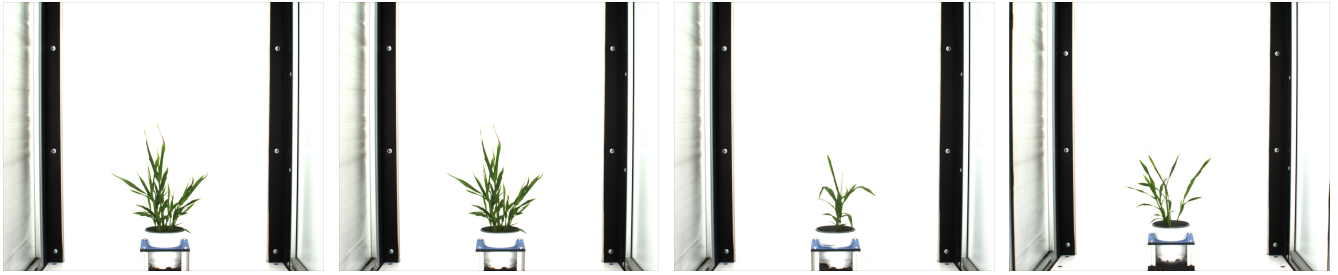


図 2: 実験に用いた画像例

から、577 枚を回帰に用いた。すべての画像は、画像の縦横比を保ったまま、解像度を 224×224 pixels に正規化して用いた。

Pretext task で推定した画像に占める植物の面積や縦横比は、PlantCV を用いて計算した。面積や縦横比を量子化する際の量子化レベル n は 4, 8 とした。分けつ数の精度評価には、6-fold cross validation を用い、それぞれの平均絶対誤差 (Mean Absolute Error: MAE) を用いて評価をした。Pretext task を用いた特徴量の学習はそれぞれ 4 回行い、それぞれで 6-fold cross validation を行って精度を評価した。

Pre-trained モデルによる特徴抽出, pretext task による特徴量と学習と特徴抽出, MLR の実装は Keras を用いて行い, SVR の実装は scikit-learn を用いた。SVR では、3 次の多項式カーネルを用い、ソフトマージンを制御するパラメータ C を 0.001 とし、これ以外のパラメータをデフォルト値として実験を行った。

提案手法の性能評価のため、従来手法 [4] と精度の比較をした。従来手法では、分けつ数 TC を、植物の画像上での縦横比 HW 、新鮮重量 M_{fw} を説明変数として、以下のように表現する。

$$TC = 0.220M_{fw} - 2.19HW + 5.26, \quad (7)$$

ただし、新鮮重量は画像中の植物が占める面積 A_{sv} を用いて以下のように計算される。

$$M_{fw} = 3.755 \times 10^{-5} A_{sv} + 0.2704. \quad (8)$$

4.2 結果

表 1, 2 に回帰に SVR, MLR を使った場合のそれぞれの MAE を示す。Pretext task では、1 つのタスクに対して 4 回づつ学習をし、得られたモデルで特徴抽出をしたため、1 つのタスクにつき MAE が 4 つ表示されている。ただし、pre-trained モデルはモデルが 1 つであるため、SVR での評価は 1 度のみである。また、MLR では、全結合層の重みの初期値によって分けつの推定精度が異なるため、pre-trained モデルでも分けつ推定の評価を行っている。それぞれの表の最終行は、従来手法での MAE である。面積を回帰で求める pretext task では学習に失敗したため、分けつ推定は行わなかった。

表 1, 2 から、提案手法の一部では、従来手法と比較してより高精度な分けつ数の推定が可能となった。MLR を用いた場合、pre-trained モデル、量子化数が 8 のときの面積の推定の pretext task で学習したモデルが従来手法を上回る精度を示した。この中でも、Pre-trained モデルが最もよい精度を示した。SVR を

表 1: SVR を用いた時の推定結果.

Feature extraction methods	MAE			
Pre-trained VGG 16 model	0.81			
Pretext task (area, 4 levels)	1.02	0.93	0.93	0.90
Pretext task (area, 8 levels)	1.09	1.02	1.02	1.06
Pretext task (area, regression)	-			
Pretext task (aspect ratio, 4 levels)	0.70	0.87	0.83	0.81
Pretext task (aspect ratio, 8 levels)	1.01	1.07	1.12	0.98
Pretext task (aspect ratio, regression)	0.66	0.69	0.71	0.70
Fahlgren et al. [4]	0.98			

表 2: MLR を用いた時の推定結果.

Feature extraction methods	MAE			
Pre-trained VGG 16 model	0.55	1.79	1.81	1.71
Pretext task (area, 4 levels)	1.79	1.76	2.38	1.70
Pretext task (area, 8 levels)	1.73	2.45	0.57	1.74
Pretext task (area, regression)	-			
Pretext task (aspect ratio, 4 levels)	2.93	2.36	2.39	1.21
Pretext task (aspect ratio, 8 levels)	1.79	2.45	1.84	1.72
Pretext task (aspect ratio, regression)	1.16	2.33	1.22	2.95
Fahlgren et al. [4]	0.98			

用いた場合は、pre-trained モデルや、量子化数が 8 の時の面積の推定、回帰を用いた面積の推定の pretext task で学習したモデル、量子化数が 4 の縦横比の推定、回帰による縦横比の推定で学習したモデルが従来手法を上回る精度を示した。以上から、分けつ数の推定に DNN を用いたことで、推定精度が向上したことが明らかとなった。

表 2 から、MLR を用いた場合、分けつ数の推定精度に大きくばらつきが見られた。例えば、pre-trained モデルを用いた場合、1 回目の評価では、本稿の中で最もよい推定精度を示しているが、それ以外の精度は従来手法を下回っている。一方で、SVR を用いた推定の精度は安定している。よって、MLR 結果にばらつきがあるのは、学習に初期値が大きく影響しているためであると考えられる。

続いて、実験結果のさらなる考察のため、各サンプルの推定結果をグラフに表示した。そのグラフを図 3 に示す。比較のため、従来手法のグラフも図 3 に示している。グラフの水平方向の軸は分けつ数の真値で、水平方向は分けつ数の推定値を示す。また、青または黒の直線は、分けつ数の真値と推定値が一致し

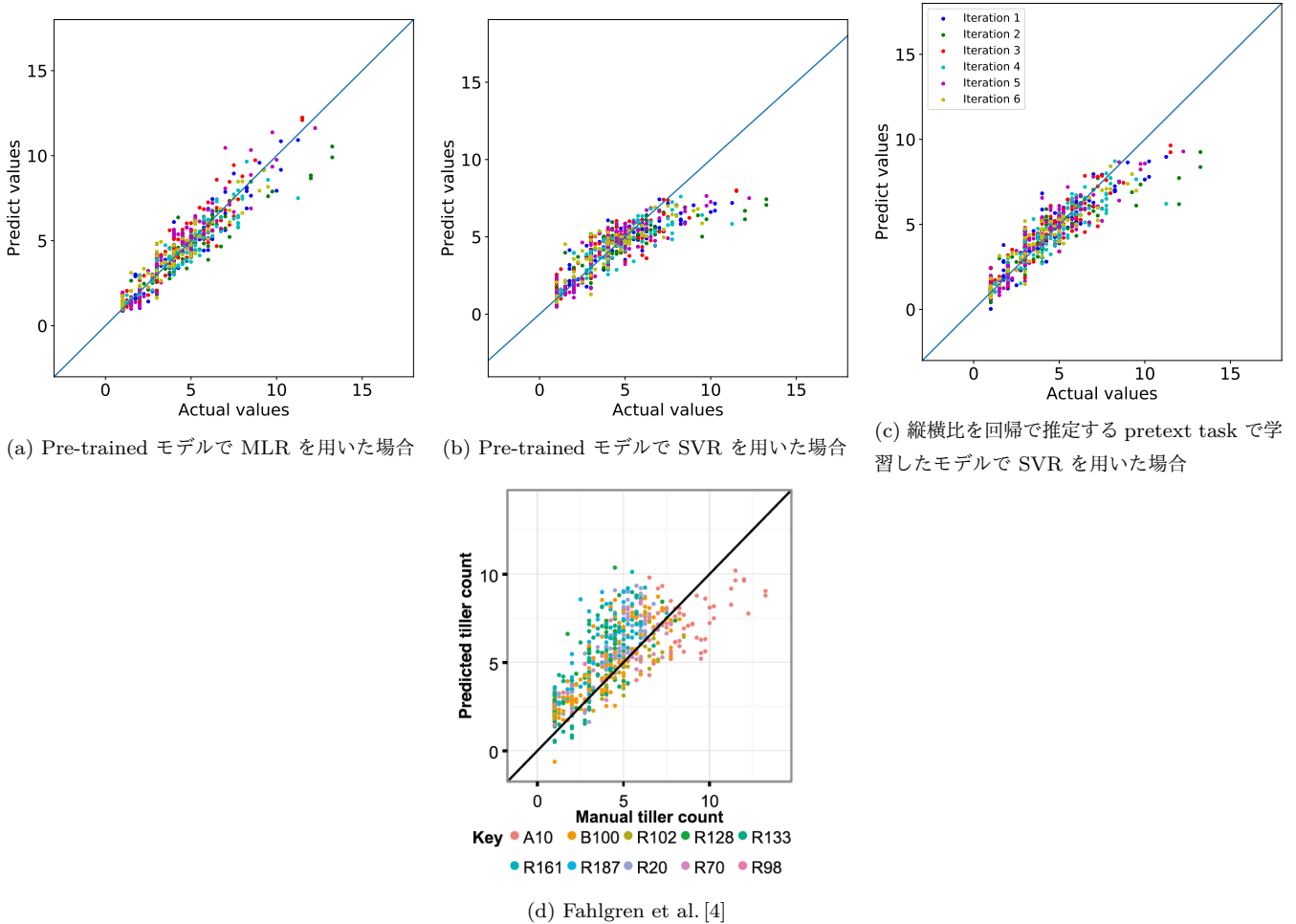


図 3: 実験結果の真値と推定値を示すグラフ。

た時を示す。つまり、直線から点までの距離は、そのサンプルの分けつ数の推定精度を表す。図 3(a) は、pre-trained モデルを用いて MLR で推定をした結果であり、最もよい MAE を示した結果である。図 3(b) と図 3(c) は pre-trained model と SVR を用いた場合、植物の画像の縦横比を回帰ですべて低する pretext task で学習されたモデルと SVR を用いた結果である。図 3(a), 3(b), 3(c) と図 3(d) を比較すると、提案手法の方がより推定精度が高いことがわかる。図 3(a) と図 3(b) を比較すると、特に分けつ数の真値が 7 以上の時に、MLR を用いた方が SVR を用いた時よりもより精度がよい。また、図 3(b), 3(c) を比較すると、pretext task を用いた方が、pre-trained モデルよりも、真値が 7 以上の時に推定精度が改善していることが分かる。よって、pretext task で問題に合わせて CNN を学習することで、精度の改善を図ることが可能であるといえる。

5. まとめと今後の課題

本稿では、単子葉植物で生育度合いを計るのに用いられる分けつ数の推定に DNN を適用する手法を提案した。分けつ数のデータを収集するのは労働集約的で困難であるため、正解ラベルの付いた学習データが大量に集めることができない。そこで、transfer learning や pretext task を用いることで、大量のラベル付きデータが入手困難な場合でも、DNN を適用することを

可能にした。実験の結果、提案手法の一部で従来手法よりもより精度の高い分けつ数の推定が可能となった。

今後の課題として、他の pretext task を試してみることで、このタスクにはどの pretext task が有効であるかを明らかにすることが挙げられる。また、本稿で用いたエノコログサ以外の単子葉植物に対しても、提案手法が有効であるか評価する必要がある。

謝辞 本研究は JSPS 科研費 若手研究 JP18K18074 の補助による。

文 献

- [1] I.M. Scotford and P.C. Miller, “Estimating tiller density and leaf area index of winter wheat using spectral reflectance and ultrasonic sensing techniques,” *Biosystems Engineering*, vol.89, no.4, pp.395–408, 2004.
- [2] P. Dampney, S. Quegan, and P. Meadows, “Advanced radar for measuring green area index (GAI), biomass and shoot numbers in wheat (Radwheat),” *Technical Report 252, HGCA Project Report*, 2001.
- [3] J.C. Taylor, G.A. Wood, J.P. Welsh, and S. Knight, “Exploring management strategies for precision farming of cereals assisted by remote sensing,” *Aspects of Applied Biology*, vol.N/A, no.60, pp.53–60, 2000.
- [4] N. Fahlgren, M. Feldman, M.A. Gehan, M.S. Wilson, C. Shyu, D.W. Bryant, S.T. Hill, C.J. McEntee, S.N. Warnasooriya, I. Kumar, T. Ficor, S. Turnipseed, K.B. Gilbert, T.P. Brutnell, J.C. Carrington, T.C. Mockler, and I. Bax-

- ter, “A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in setaria,” *Molecular Plant*, vol.8, no.10, pp.1520–1535, 2015. <http://dx.doi.org/10.1016/j.molp.2015.06.005>
- [5] R.D. Boyle, F.M. Corke, and J.H. Doonan, “Automated estimation of tiller number in wheat by ribbon detection,” *Machine Vision and Applications*, vol.27, no.5, pp.637–646, 2016.
- [6] Y. Huang, Y. Cheng, D. Chen, H. Lee, J. Ngiam, Q.V. Le, and Z. Chen, “Gpipe: Efficient training of giant neural networks using pipeline parallelism,” *arXiv preprint arXiv:1811.06965*, 2018.
- [7] T. Miyato, S.-I. Maeda, S. Ishii, and M. Koyama, “Virtual adversarial training: A regularization method for supervised and semi-supervised learning,” *IEEE Trans. PAMI*, 2018.
- [8] S. Gidaris, P. Singh, and N. Komodakis, “Unsupervised representation learning by predicting image rotations,” *Proc. ICLR*, 2018.
- [9] M. Noroozi, A. Vinjimoor, P. Favaro, and H. Pirsiavash, “Boosting self-supervised learning via knowledge transfer,” *Proc. ACML*, 2018.
- [10] M. Noroozi, H. Pirsiavash, and P. Favaro, “Representation learning by learning to count,” *Proc. ICCV*, 2017.
- [11] A.F. Frangi, W.J. Niessen, K.L. Vincken, and M.A. Viergever, “Multiscale vessel enhancement filtering,” *Proc. Medical Image Computing and Computer-Assisted Intervention*, 1998.
- [12] R. Zhang, P. Isola, and A.A. Efros, “Colorful image colorization,” *Proc. ECCV*, 2016.
- [13] M. Noroozi and P. Favaro, “Unsupervised learning of visual representations by solving jigsaw puzzles,” *Proc. ECCV*, 2016.
- [14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” *Proc. CVPR*, 2009.
- [15] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *Proc. ICLR*, 2015. <http://arxiv.org/abs/1409.1556>