

Semantic Segmentation for Aerial Imagery with Convolutional Neural Networks

March 2016

Shunta Saito

主 論 文 要 旨

報告番号	甲 第	号	氏 名	齋藤 俊太
主論文題目： Semantic Segmentation for Aerial Imagery with Convolutional Neural Networks (畳み込みニューラルネットワークを用いた航空画像の意味論的領域分割)				
(内容の要旨) 航空画像は、地図作成や災害規模推定、都市計画など、多岐にわたる応用において重要な情報源として利用されている。しかし多くの応用において専門家の目視による建物や道路の抽出という前処理が必要となり、これには多大な時間とコストが必要であった。そこで、この自動化に向けた研究が数多く行われてきたが、対象地域ごとに個別の画像特徴量設計が必要とされたり、抽出精度が十分でないなどの問題があった。しかし近年、大量の航空画像と地物マスク画像を含む大規模なデータセットを用いて畳み込みニューラルネットワークを訓練することで、特徴量設計の必要がなくなかつ旧来手法よりも高精度な意味論的領域分割による建物・道路の自動抽出が実現可能であることが示されている。ただし、これらの手法は地物の種別ごとに別々に訓練されたネットワークを必要とするため、複数種類の地物を同時に抽出することはできない。そこで本論文では、単一の畳み込みニューラルネットワークを用いて複数種類の地物抽出を同時に行うための意味論的領域分割手法を提案し、さらに分類問題において着目クラス以外のクラスとしてまとめられる「背景」クラスの特異性に注目することで、ネットワークの性能を向上させる新しい出力関数を提案する。また、航空画像からの建物抽出自動化の応用として、シームライン決定と呼ばれる複数の航空画像を繋ぎ合わせるための理想的な繋ぎ目の同定作業が高精度に実現できることを示す。第1章では、研究背景として、航空画像を活用した広範な応用事例の一部と、それらにおいて必要とされる地物抽出のための従来研究が持つ問題点について述べ、本論文でこれを解決するために用いる方法を俯瞰し、本研究の目的を明らかにする。第2章では、本論文と最も関連の深い先行研究について詳説し、本論文との関係を整理した上で本論文において提案される手法の位置づけを明らかにする。第3章では、提案手法について詳説する。本論文では、航空画像および建物・道路のマスク画像を含む大規模なデータセットを用いて、背景クラスの特異性に注目した新しい出力関数を持つ畳み込みニューラルネットワークを訓練することにより、建物・道路双方のマスク画像を同時に、かつ高精度に自動抽出する手法を提案する。第4章では、提案手法の有効性を公開データセットを用いて評価し、従来手法との定量的な比較を行う。第5章では、第4章の評価結果について考察する。また、学習済みネットワークのパラメータ解析および、データセットの規模についての検討を行う。第6章では、建物抽出の自動化によって航空画像のモザイクにおける理想的なシームラインの決定という課題が容易に自動化可能であることを示す。また、従来手法との比較を通し、より望ましいシームラインの決定が可能となることを示す。第7章では、本論文において提案された手法をまとめ、本論文により明らかになった最も重要な知見について俯瞰する。また、今後の課題と展望を示す。				

SUMMARY OF Ph.D. DISSERTATION

School Integrated Design Engineering	Student Identification Number	SURNAME, Given name SAITO, Shunta
Title Semantic Segmentation for Aerial Imagery with Convolutional Neural Networks		
Abstract <p>Aerial imagery has been utilized as important input for many applications such as map making, disaster relief, real estate management, and urban planning. In those applications, extracting ground objects such as buildings and roads from aerial imagery is necessary as a pre-processing. This task is to give an object class label for each pixel in an aerial image, so that it is highly attentional task for human and very costly and time-consuming process. Therefore, various methods have been proposed for automating this task. However, conventional methods usually need hand-crafted image features and specially designed classifiers for each target area. Furthermore, the extraction accuracy is not enough to be alternative to human experts. On the other hand, some recent studies show that this task can be performed automatically with relatively high accuracy and no feature engineering by utilizing a convolutional neural network (CNN) trained on a large scale dataset that includes many pairs of aerial imagery and ground object mask images. However, those recent achievements still require independent networks for each target object, so that simultaneous multiple object extraction using a CNN has not been evaluated.</p> <p>In this thesis, a method to extract multiple objects simultaneously with a single network is proposed. To train a single network efficiently for multiple object extraction task, a novel output function called channel-wise inhibited softmax that focuses on the specialty of the background class that is a class to represent all the other classes other than classes of interest, e.g., buildings and roads, is proposed. Then, how to improve seamline determination by utilizing the result of semantic segmentation for aerial imagery is shown.</p> <p>In Chapter 1, example applications of aerial imagery interpretation are introduced. Then problems in conventional methods for automated ground object extraction are presented, and our approaches to address the problems are overviewed to clarify the research goal.</p> <p>In Chapter 2, the details of the most related work and the relationship to our approach are presented to show the difference between them.</p> <p>In Chapter 3, the details of our methods are presented. A method to extract buildings and roads simultaneously and accurately from aerial imagery by training a single convolutional neural network with a novel output function on a large scale dataset that has aerial images and corresponding building and road mask images is proposed. The output function is designed by focusing on the specialty of the background class.</p> <p>In Chapter 4, empirical evaluation of the proposed method is performed on the large scale dataset. Then the effectiveness of the proposed method is compared to the conventional method.</p> <p>In Chapter 5, the results of the evaluation performed in Chapter 4 are discussed in details by visualizing the middle layer outputs and analyzing the trained parameters. Then the size of dataset that is required to train a convolutional neural network for this task is discussed.</p> <p>In Chapter 6, the fact that automatic seamline determination for aerial imagery mosaicking can be performed with optimal results by utilizing building extraction results that are acquired by semantic segmentation is presented. Then evaluation of the resulting seamlines by comparing to the conventional method based on non-semantic segmentation is performed.</p> <p>In Chapter 7, the proposed methods are summarized and the most important findings are clarified. Then conclusion and future work of this research are shown.</p>		