Catalogue no. 11-522-x
ISSN: 1709-8211

Proceedings of Statistics Canada Symposium 2021
Adopting Data Science in Official Statistics to Meet Society's Emerging Needs

Modernizing Construction Indicators Through Machine Learning and Satellite Imagery

by Aidan D. Smith and Hector Ferronato

Release date: October 15, 2021

Proceedings of Statistics Canada Symposium 2021
Adopting Data Science in Official Statistics to Meet Society's Emerging Needs

Modernizing Construction Indicators Through Machine Learning and Satellite Imagery

Aidan D. Smith, Hector Ferronato[1]

Abstract

Official statistical agencies must continually seek new methods and techniques that can increase both program efficiency and product relevance. The U.S. Census Bureau's measurement of construction activity is currently a resource-intensive endeavor, relying heavily on monthly survey response via questionnaires and extensive field data collection. While our data users continually require more timely and granular data products, the traditional survey approach and associated collection cost and respondent burden limits our ability to meet that need. In 2019, we began research on whether the application of machine learning techniques to satellite imagery could accurately estimate housing starts and completions while meeting existing monthly indicator timelines at a cost equal to or less than existing methods. Using historical Census construction survey data in combination with targeted satellite imagery, the team trained, tested, and validated convolutional neural networks capable of classifying images by their stage of construction demonstrating the viability of a data science-based approach to producing official measures of construction activity.

Key Words: Official Statistics; Housing Starts, Machine Learning, Satellite Imagery

1. Introduction

1.1 Overview of Census Bureau Construction Indicators

The U.S. Census Bureau produces thirteen principal federal economic indicators which serve as the most timely and leading official measures of U.S. economic activity. Three of these indicators focus exclusively on the construction sector and work in concert to report monthly measures of construction activity, housing inventories, and other key inputs to quarterly Gross Domestic Product (Bureau of Economic Analysis, 2020). New Residential Construction (NRC) provides several key measures of residential construction activity and inventory including building permits, housing starts, completions, and under construction estimates[2]. In addition to their direct insights, these stage-of-construction estimates also serve as critical inputs to our other two construction indicators. The New Residential Sales (NRS) indicator provides a variety of measures related specifically to sales of single-family homes including national and regional estimates of new houses sold and for sale as well as national average and median sales prices. Concurrent with the NRS release, the Census Bureau releases a monthly price index of new houses under construction allowing data users to derive a constant dollar series from the current dollar single-family construction spending series, and a quarterly price index that serves as a measure of inflation in new houses built for sale. And finally, the Construction Spending indicator provides the national total dollar value spent on private and public residential and non-residential construction by sector. Of note, the private residential component of this indicator, which relies heavily on the housing starts estimates from NRC and price information from NRS, accounts for roughly 40% of all construction spending and is a key input to the Private Fixed Investment component of Gross Domestic Product.

1.2 Survey of Construction

---

[1] Aidan D. Smith, Economic Directorate, U.S. Census Bureau, 4600 Silver Hill Road, Suitland, MD, USA, 20742, aidan.d.smith@census.gov; Hector Ferronato, Reveal Global Consulting, 8115 Maple Lawn Blvd, Suite 375, Fulton, MD, USA, 20759, hector.ferronato@revealgc.com

[2] The Census Bureau defines a house as started when excavation begins for the footings or foundation. A house is defined as completed when all finished flooring has been installed or at the time of occupancy, whichever occurs first. For buildings with two or more units, all units in the building are counted as completed when 50 percent or more of the units are occupied or available for occupancy.

Since 1959, the Census Bureau has conducted the Survey of Construction (SOC) to collect data on new housing starts, completions, sales, prices, and other residential construction characteristics[3]. The information collected through the SOC is used in conjunction with data from our Building Permits Survey (BPS) to produce many of the key estimates within our three construction indicators described in section 1.1.

The current SOC design employs a multistage process whereby a sample of issued permits is selected each month from a fixed sample of approximately 900 out of the roughly 20,000 permit-issuing jurisdictions within the U.S. For each of the 900 permit offices, field representatives manually list all permits for new residential construction, and a sample is drawn by selecting permits for buildings with one to four units at an overall rate of one in fifty and selecting permits for buildings with five or more units with certainty. To account for areas of the country where building permits are not required, field representatives canvas a sample of non-permit land areas looking for visual evidence of housing units started, all of which are included in the monthly sample. Data collection is entirely reliant on field representatives conducting personal interviews with the builders or owners of sampled projects each month to determine the stage of construction, sale information, and other characteristics of the units. In its current state, the SOC provides us with the data needed to produce monthly stage-of-construction estimates – housing starts, under construction, and completions – both at the national level and by Census Region.

1.3 The Motivation to Modernize

As the needs and scope of the data user community continue to evolve, official statistical programs must pursue innovative methods, techniques, and data sources to remain relevant (Jarmin, 2019). The Census Bureau's construction indicators are not immune to this dynamic. While our data users continually demand construction data that are more timely, more accurate, and more granular, the existing survey designs on which these indicators are built impose direct limitations on our ability to meet those needs.

The current SOC data collection operation is entirely supported by field interviews and manual canvasing of sampled land areas both of which present considerable operational costs. Furthermore, obtaining data through personal interviews with builders and project owners has an associated respondent burden. The Paperwork Reduction Act (PRA) of 1995 gives the Office of Management and Budget (OMB) authority over the collection of information by Federal agencies with the goal of minimizing both respondent burden and cost of collection while simultaneously maximizing the utility of information collected (United States Office of Personnel Management, 2011). Within the context of the current SOC design, the dual constraints of collection cost and respondent burden limit the number of jurisdictions and residential projects we can include in the sample each month. This in turn results in a sample that can only support estimates with a level of granularity and accuracy that fall short of our data users' evolving needs. If our construction indicators are to remain relevant, we must explore new data sources and methods that can help decouple us from the collection cost and respondent burden constraints associated with the existing SOC.

2. Deep-Learning Status of Construction Via Remote-Sensing (DSCOVR)

2.1 Project Vision

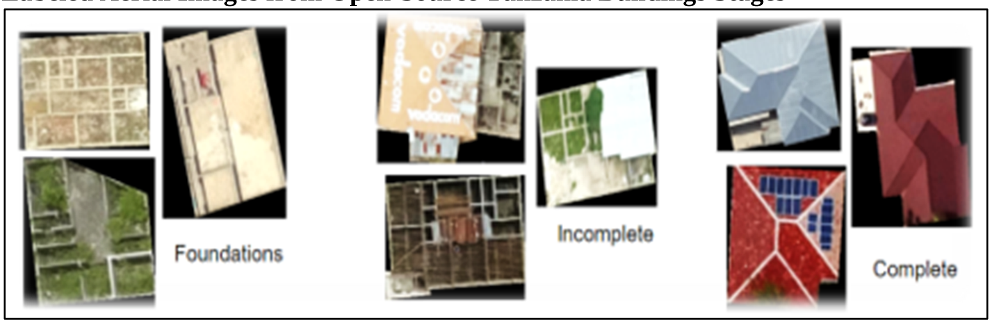
In 2019, the Census Bureau began researching the use of machine learning techniques with satellite imagery to produce two key construction estimates: housing starts and completions. Our vision of project success centers around a demonstration of this modern approach that can produce accurate starts and completions estimates while meeting existing monthly indicator timelines at a cost equal to or less than existing methods. If the project can satisfy these success criteria, we will have demonstrated an alternative approach to construction measurement that is decoupled from traditional survey collection cost and respondent burden – a critical first step to expanding our scope of collection to support more accurate and more granular estimates.

[3] Partial funding for the SOC is provided by the U.S. Department of Housing and Urban Development.

2.2 Initial Research (IC0)

After collaborating with Statistics Canada and further academic research, the team selected Convolutional Neural Network (CNN) as the best performing Deep Learning (DL) models for image classification (O'Mahony et al., 2019) and utilized the Fast.ai[4] Python library to train 3,000 open-source aerial images[5] labeled into three classes: foundation, incomplete, and complete. For the purposes of our initial research, we treated 'foundation' images as a proxy for a construction starts and 'complete' images – which show the presence of a completed roof – as a proxy for construction completions. While these did not align precisely with our existing start and completion definitions, they provided a reasonable baseline for the proof of concept.

Figure 2.2-1
Labeled Aerial Images from Open-Source Tanzania Buildings Stages



The labeled images were divided into two sets – 80% for training and 20% for validation. The classification results from the initial model (IC0) showed 88.5% accuracy after 29 epochs of training.

2.3 Proof of Concept Model (IC1)

The initial model showed significant potential for using CNN models to classify images into construction stages, even when using open-source and pre-labeled images. The team decided to improve the overall performance by training another model using newly sourced satellite images based on permit information and historical construction progress patterns[6]. Leveraging a database containing specific permit locations and issue dates, we automatically searched and downloaded three images for each permit location – one reflecting the pre-construction period, one representing the construction start, and one for construction completion. The image search for pre-constructions selected imagery for each permit location from three months before the associated permit issue date. The construction starts search selected imagery from between one month before and four months after the permit date. Completion images were selected from a period nine months beyond the permit date.

After curating 3,000 new images into the three classes, the team manually filtered and transferred images from one category to another for any images that appeared to be mislabeled. This manual review and relabeling was required since our automated image selection and labeling process relied on typical permit-to-start and permit-to-completion durations, and at least some of the selected projects would be expected to deviate from these average durations. The new image set was then used to train model IC1, leveraging the same CNN library and ResNET architecture. The validation set classifications yielded 92% accuracy, and F1-scores[7] of 96%, 92%, and 84% for completions, pre-

---

[4] Fast.ai is an open-source deep learning library which provides practitioners with high-level components, including the pre-trained model templates used in this project. Pre-trained models are already trainings on millions of generalized images and only require the training of last layers for specialized learning. Source: https://github.com/fastai/fastai

[5] This dataset contains computer generated building footprints derived using Bing Maps algorithms on satellite imagery. Source: https://github.com/microsoft/Uganda-Tanzania-Building-Footprints

[6] The Census Bureau produces annual statistics on the length of time from permit authorization to start and from start to completion, available here: https://www.census.gov/construction/nrc/lengthoftime.html

[7] The F1 score metric is 2*((precision*recall)/(precision+recall)). It conveys the balance between the precision and recall of the model.

constructions, and starts classes respectively. The lower accuracy for the starts class reflects the relative rarity of starts images within the overall image set. This phenomenon is somewhat intuitive given that the window of opportunity to observe an actual start is narrower than the observation windows for pre-construction and completion. Nonetheless, the confusion matrix for IC1 classifications presented in Table 2.3-1 shows the model performs well when classifying the images into the three construction stages.

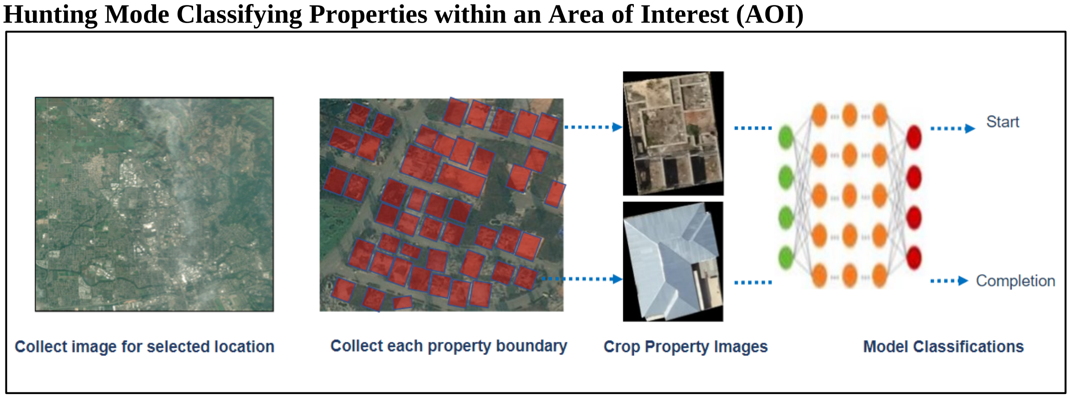
Table 2.3-1
Confusion Matrix for IC1 Classifications

		Predicted		
		COMPLETE	PRE-CONSTRUCTION	STARTS
Actual	COMPLETE	**362**	3	5
	PRE-CONSTRUCTION	2	**236**	10
	STARTS	17	17	**156**

2.4 Production Application

The solution architecture for applying the trained CNN model for satellite imagery construction classification was designed in two modes – Hunting and Tracking. On a monthly basis, an overall image covering a specified Area of Interest (AOI) will be downloaded and cropped to each individual property boundary within that AOI. Each individual property image is classified by the model, and properties classified as starts are tagged for continued tracking over the following months.

Figure 2.4-1
Hunting Mode Classifying Properties within an Area of Interest (AOI)



Tracking mode ensures each location classified as start will be provided as input for additional classification in subsequent months until the property image is classified as a completion. The results include the location, estimated construction start date based on when the model classified the property image as start, and estimated completion date corresponding to when the model classified the property image as a completion.

2.5 Validation Framework

Our team is exploring several strategies as part of an overall model validation framework, including classification progression consistency review, permit cross-referencing, manual visual inspection, and pixel activation review. As a first step, we run the model on new imagery outside the initial curated set and covering a period of multiple distinct months. This allows us to review classification progression across multiple months for logical consistency. For example, if the model produces a classification progression for a construction project where a completion precedes a start, this likely indicates suboptimal model performance based on our real world understanding of typical construction progression. We are also able to layer permit location data on top of the bounded properties within the AOI allowing us to compare model predictions for these locations against expected start and completion status based on permit issuance. Additionally, for each set of classifications, a random 5% sample is selected for manual visual inspection of image and classification pair. A more recent addition to the validation framework is the review of pixel activation from specific layers of the CNN. The goal of this review is to validate if the model is 'looking at' the meaningful pixels when making a classification. When the model classifies an image as a completion, we expect it to have arrived at that classification via the pixels that form the completed building. Figure 2.5-1 shows a successful example of this with the original image on the left, an initial CNN layer in the middle, and the final layer – where a completion was detected – on the right. The purple colors indicate the areas of the image where the model is focusing at each iteration.

Figure 2.5-1
Pixel Activation Review – Construction Completion

As a practical exercise of the validation framework, Houston, TX imagery was sourced for validation and collection of accuracy and F1 scores. The AOI box in Houston measured 100sqkm, and one image for each of 4 months was collected for the validation experiment. After cropping, a total of 250,000 images were classified. Our validation strategies immediately indicated some issues and limitations of the training image set and classification categories that needed to be addressed before collecting metrics. We discovered a roof color bias wherein the training imagery skewed heavily toward one color of roofing negatively impacting model performance when running on an entirely new AOI. The number of false positives present in this exercise also suggested that we would need to refine our classification categories. This included the need to account for images completely unrelated to construction such as trees or bodies of water and addressing the fact that the pre-construction and starts categories appeared to be too similar. Based on the insights gathered from execution of the validation framework, the team began work on several improvements for the next model iteration, IC2.

2.6 Model Improvements (IC2)

Collecting more images from different locations to increase training set diversity and redefining the categories for classification improved the model IC2 substantially. The new images training set included images from four distinct locations, and the new categories facilitated identifying non-construction images which were considered false positives previously. Classification categories were reorganized into a tiered structure:

- Class A: Ongoing Construction Activity
    - Class A1: Excavation
    - Class A2: Foundation
    - Class A3: Framing

- Class B: Completed Structures
    - Class B1: Residential Buildings
    - Class B2: Commercial Buildings
    - Class B3: Streets/Parking Lots

- Class C: Land, Water, and Vegetation
    - Class C1: Trees
    - Class C2: Grass
    - Class C3: Clean Land

Model optimization was also performed to select the most efficient number of layers and model within Neural Network types. The most optimal training combination for model IC2 was of type Resnet50 with 128 for resizing parameter, which achieved 97.1% accuracy and 98% F1-Score on train/test validation.

This model was also tested outside the training images and compared to human labels from control group images. From approximately 160 images, 139 images were labeled as completions by humans, 21 as vegetation, and only 4 as starts. IC2 model classified correctly, when compared to human's labels, 95.7% of completions, 90.5% of vegetation, and 50% of starts. The results indicate the need for increasing the number of control group images labeled by humans, specifically for starts images. With only four images from the random 5% sample being of starts class, the 50% accuracy results are not significant, however the completions and vegetations classes' 96% and 91% accuracy can be validated due to higher number of samples observed and labeled. The team plans to target high-construction and permitting activity areas when further validating the model to confidently measure accuracy and F1 scores for the construction activity class specifically.

3. Conclusion

3.1 Summary and Next Steps

The results of our research and proof of concept to date are a promising first step in demonstrating the viability of a data science-based approach to producing official measures of construction activity. Based on the success of this initial research, the project will move forward with goals targeting scalability, incorporating sophisticated image classification methods, and ultimately integrating with our existing construction indicators. While our initial proof of concept focused on a few construction dense locations, we must now expand our scope of training and test imagery to include a wide variety of climates as well as a mix of both rural and urban environments. A typical completed roof in the rural southwest may look quite different than one from a metropolitan area in the northeast, and our methodology will need to reliably identify both. The team will continue to research the application of instance segmentation to extract more detailed information such as the construction project boundary, square footage, and multiple classifications within a single image (e.g. a start adjacent to a completion indicating an addition). Integration planning will need to address issues such as image frequency vs. resolution and reconciliation of the old and new methodologies and stage-of-construction definitions to ensure indicator time series integrity and comparability. While there is still work to be done as we move closer to production integration, the initial project success has demonstrated the potential to release our construction indicators from the traditional constraints of collection cost and respondent burden that have limited our ability to meet our data users' evolving needs.

References

Bureau of Economic Analysis (2020), "Concepts and Methods of the U.S. National Income and Product Accounts", Washington, DC, Retrieved from https://www.bea.gov/resources/methodologies/nipa-handbook/pdf/all-chapters.pdf (accessed 8 November 2021).

Jarmin, Ron S. (2019), "Evolving Measurement for an Evolving Economy: Thoughts on 21st Century US Economic Statistics", *Journal of Economic Perspectives*, 33 (1), pp 165-184.

O'Mahony, N., Campbell, S., Carvalho, A. et al. (2019), "Deep Learning vs. Traditional Computer Vision", *Advances in Computer Vision Proceedings of the 2019 Computer Vision Conference*, Springer Nature Switzerland AG, pp 128-144.

United States Office of Personnel Management (2011), "Paperwork Reduction Act (PRA) Guide Version 2.0", Washington, DC, Retrieved from https://www.opm.gov/about-us/open-government/digital-government-strategy/fitara/paperwork-reduction-act-guide.pdf (accessed 8 November 2021).