

A Blended Data Approach to Measuring Monthly Housing Starts:

Satellite Imagery, Survey Data and More!

Nicole Czaplicki¹, Colin Shevlin², Hector Ferronato², Aidan Smith¹,

Dwarakh Nayam², Lei Peng², Scott Springer¹, Doren Walker²

¹U.S. Census Bureau¹ ²Reveal Global Consulting

Abstract:

As part of the comprehensive Construction Re-engineering Initiative at the U.S. Census Bureau, alternative data sources are being considered to supplement or replace current data collection methods. For the Survey of Construction (SOC), which measures new residential construction, this includes observing housing starts from satellite imagery in place of the current interviews for housing starts conducted by field representatives. Satellite images are obtained monthly for a subset of places in the SOC sample. Convolutional neural network models are then applied to images to predict likely new residential construction projects, with the current focus being single-family housing starts. Several post prediction processing steps are applied including exclusions based on intersections with known buildings or roads, treatments for missing data due to cloud cover, and adjustments for the length of time between consecutive images, to ultimately produce place level estimates of housing starts. These place level estimates are then combined

¹ Any opinions and conclusions expressed herein are those of the authors and do not reflect the views of the U.S. Census Bureau. The Census Bureau has reviewed this product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data (Disclosure Review Board (DRB) approval number: CBDRB-FY26-ESMD009-001).

with the existing building permit level survey data to produce estimates of West South Central division level housing starts, an experimental data product from the Census Bureau.

1. Introduction

As part of the comprehensive Construction Re-engineering Initiative at the U.S. Census Bureau, alternative data sources are being considered to supplement or replace current data collection methods. In a time of rising collection costs, seeking out alternative sources of data is crucial to continue to produce accurate and timely estimates of monthly housing starts, a highly monitored economic indicator published in the New Residential Construction report. For the Survey of Construction (SOC), which measures new residential construction, this includes observing housing starts from satellite imagery in place of the current interviews for housing starts conducted by field representatives. As a proof of concept, the U.S. Census Bureau designed an experimental data product for monthly West South Central division-level single-family housing starts which combined survey data with data obtained from satellite imagery in the final estimates. Satellite images were obtained monthly for a subset of places in the SOC sample. Convolutional Neural Network (CNN) models were then applied to the images to predict likely new single-family housing starts. This was followed by several post prediction processing steps including exclusions based on intersections with known buildings or roads, treatments for missing data due to cloud cover, and adjustments for the length of time between consecutive images, to ultimately produce place-level estimates of housing starts. These place-level estimates were then combined with the existing building permit level SOC survey data as well as data from the Building Permits Survey to produce the final estimates.

Incorporating data from nontraditional data sources is a growing priority for producers of official statistics. Several advisory panels have convened to lay out a possible path forward within this new “Big Data” environment (ECOSOC 2015, Eurostat 2014, Japex et al. 2015, National Academies of Sciences, Engineering, and Medicine 2023). National statistics offices have already started down this path with several innovative projects combining data from multiple sources to produce estimates. Erciulescu et al. (2020) combined survey and auxiliary data to produce county level planted acreage predictions. At the U.S. Census Bureau, monthly state retail sales estimates (currently an experimental product) are produced by combining survey data from the Monthly Retail Trade Survey, the business register (a complete listing of all businesses in the United States), and third-party point of sale data (Hutchinson et al. 2023, Kaputa et al. 2024). Additionally, remote sensing data, such as from satellites, continues to be explored as a possible data source for the production of official statistics (Khan et al. 2021, Kitchin and Stehle 2021, Ryerson et al. 1983, Smith and Ferronato 2021, United Nations Task Team on Satellite Imagery and Geospatial Data 2017, USDA NASS).

This paper describes the process from image acquisition to estimate production of this experimental data product as well as highlighting some of the challenges that were encountered and solutions that were implemented while developing the procedure. While the experimental data product covered the West South Central Division only (consisting of Arkansas, Louisiana, Oklahoma, and Texas), efforts are underway to expand the geographic coverage of satellite data collection.

2. Background on the Survey of Construction

The Survey of Construction provides national and regional statistics on starts, completions, and characteristics of new, privately-owned single-family and multifamily housing units and on sales of new single-family houses. The survey is funded by both the U.S. Census Bureau and the Department of Housing and Urban Development and estimates are published as a joint release from the two sponsoring agencies. Estimates of the number of units authorized but not yet started, started, under construction, and completed are published monthly in the New Residential Construction report released on the 12th workday of the month. The report contains preliminary estimates for the most recent complete month, as well as first revision and second revision estimates for the prior two months. Additionally, estimates of single-family sales are published on the 17th workday of the month in the New Residential Sales report. SOC also publishes several less frequent reports including Quarterly Starts and Completions by Purpose and Design, and Annual Characteristics of New Housing. A full description of the SOC methodology is available at <https://www.census.gov/construction/soc/methodology.html>.

2.1 Sample design

The SOC is a multi-stage cluster sample of new residential buildings. The primary sampling unit (PSU) is a county or county equivalent selected from the Current Population Survey (CPS) sample. The CPS monthly data collection period complements that of the SOC allowing for the sharing of field representatives between the two surveys. PSUs were stratified by division and size with a self-representing stratum and several non-self-representing strata. One PSU per non-self-representing stratum was selected in the sample with methods to maximize overlap with the previous sample to retain existing trained field representatives.

From the sample of PSUs, in areas where a building permit is required, a stratified systematic sample of approximately 900 building permit offices (BPOs) was selected. BPOs with a large number of permits were included in the sample with certainty (probability = 1). There was no intentional process to overlap the sample of BPOs with the prior sample because it is assumed that a field representative in the PSU could visit any BPO within that PSU.

The final stage of the sample selection for areas where a building permit is required is a monthly sample of building permits. Every month, field representatives list all in-scope permits authorized in the prior month in the sampling instrument, and a systematic sample of permits is selected with a targeted overall sampling rate of 1 in 50 for permitted buildings containing between one and four housing units. Permits for buildings with five or more housing units are sampled with certainty.

For areas where a building permit is not required, a stratified sample of 80 land areas (groups of Census blocks) was selected. This is referred to as the non-permit sample. Areas were stratified based on housing unit population counts from the 2000 Census. For non-permit areas, field representatives canvass the area looking for new construction. All identified in-scope construction is included in the sample.

2.2 Data collection

Current data collection operations are conducted by field representatives who try to contact the builder or owner (respondent) associated with a sampled permit or non-permit project by phone to collect data for the sampled unit. If contact is not made with the respondent by phone, the field rep will make a site visit. For housing starts, the field representative will contact the respondent monthly to ask if the building has started construction in the prior month, leading to multiple interviews if the building does not start right after permit issuance. However, not every interview

is exclusively to collect a start date. SOC also collects monthly data on completions and new single-family sales as well as housing characteristics which are published annually. Although not every interview could be replaced by an alternative data collection method, such as satellite imagery, any reduction in interviews would reduce both field representatives' workload and respondent burden.

2.3 Estimation

SOC estimates of single-family housing starts for a given month t , \hat{S}_t , are computed as the sum of the estimated starts for non-permit areas ($\hat{S}_{non-permit,t}$) and the estimated starts from the permit survey ($\hat{S}_{permit,t}$).

$$\hat{S}_t^{SOC,production} = \hat{S}_{non-permit,t} + \hat{S}_{permit,t}$$

The monthly non-permit estimate for month t is simply the sum of the number of observed starts in the non-permit area in month t multiplied by the non-permit area's sampling weight

$$\hat{S}_{NP,t} = \sum_{q \in NP} w_q s_{q,t}$$

where $s_{q,t}$ is the number of single-family starts observed in non-permit area q in month t and w_q is the combined sampling weight for non-permit land area q .

The permit survey estimate is computed using a separate ratio estimator, as shown in (1), which combines 14 ratio estimates of single-family housing starts, each corresponding to distinct permit issuing time periods. Each ratio estimate is of the form $\hat{G}_{t-i,t} \frac{Y_{t-i}}{\hat{H}_{t-i}}$. First, let w_j be the final SOC sampling weight (product of the CPS PSU sampling weight, SOC PSU sampling weight,

building permit office sampling weight and permit sampling weight) for permit j , $A_{j,t-i} = 1$ if permit j was authorized in month $t - i$ and zero otherwise, and $I_{j,t} = 1$ if permit j started in month t and zero otherwise. Then $\hat{G}_{t-i,t}$, given in (2), is the survey estimate of housing units authorized in month $t - i$ that started in month t , and \hat{H}_{t-i} , given in (3), is the survey estimate of housing units authorized in month $t - i$. Also, let Y_{t-i} be the value of housing units authorized in month $t - i$ from the Building Permits Survey (BPS). Since 2022, the BPS has included all building permit offices with more than six permits issued annually in its monthly data collection, making it a very reliable measure of total permits authorized. Separate monthly ratio estimates are computed for the current month, t , and the 11 months preceding it. Months $t - 12$ to $t - 17$ are combined into a single estimate as are months $t - 18$ to $t - 59$. This estimator assumes that all housing units start construction in the month the permit is authorized or in a subsequent month and that no units start construction before the permit authorization month.

$$\hat{S}_{permit,t} = \left[\sum_{i=0}^{11} \hat{G}_{t-i,t} \frac{Y_{t-i}}{\hat{H}_{t-i}} \right] + \left[\sum_{i=12}^{17} \hat{G}_{t-i,t} \left[\frac{\sum_{i=12}^{17} Y_{t-i}}{\sum_{i=12}^{17} \hat{H}_{t-i}} \right] \right] + \left[\sum_{i=18}^{59} \hat{G}_{t-i,t} \left[\frac{\sum_{i=18}^{59} Y_{t-i}}{\sum_{i=18}^{59} \hat{H}_{t-i}} \right] \right] \quad (1)$$

$$\hat{G}_{t-i,t} = \sum w_j A_{j,t-i} I_{j,t} \quad (2)$$

$$\hat{H}_{t-i} = \sum w_j A_{j,t-i} \quad (3)$$

2.4 Reliability of the estimates

As with all survey estimates, estimates of housing starts are subject to both sampling and nonsampling errors. Nonsampling errors include errors due to nonresponse, collection errors, coverage errors, processing errors and all other errors not related to sampling. These could arise from houses that started before the permit was authorized, missing permits from the monthly

permit listing, late reported housing starts, housing starts identified in the incorrect month, among other sources of nonsampling error. These are not directly measured but quality control measures are applied to minimize these errors. Sampling error is estimated using the modified half-sample method, a variation on balanced repeated replication (Rao and Shao, 1996; Thompson, 1998). These are reported in terms of the relative standard error (RSE), also called the coefficient of variation (CV), defined as the standard error divided by the estimate. The average RSE over the last six months is published alongside the most recent monthly estimates. Estimates of West South Central division single family housing starts are not currently published and so their RSEs are not publicly available. The most comparable published estimate is single family housing starts for the South region which in the most recent publication (January 2026) had an RSE of 8%.

3. Satellite Data Collection Project Overview

Satellite data collection offers distinct advantages over current field data collection methods. Since it is cost prohibitive to send field representatives to every construction site in an area, a sample of new single-family homes are selected for SOC from a listing of newly authorized building permits. However, with satellite imagery, all new construction sites for the area can be observed, increasing the sample size of new homes and precision of the subsequent estimates. Additionally, with current data collection methods multiple interviews with the builder may be required to collect a start date if the home does not start in the month of permit issuance. Satellite data collection reduces respondent burden by eliminating the need for start interviews (although interviews to collect other data items would still be needed). Furthermore, satellite imagery can reduce the incidence of late reported starts that occur when 1) a new permit was not included in the monthly permit listing (and therefore not available for the current month's sample), 2) a builder cannot be reached for an

interview about a sampled permit and a site visit was unable to be completed or 3) a house starts before permit issuance.

3.1 Geographic scope

The BPOs selected for satellite collection were self-representing (sampled with probability = 1) BPOs from self-representing PSUs in the current SOC sample for the West South Central Division. Therefore, these units combined PSU and BPO sampling weights are equal to one. SOC survey data was used for the remaining sampled BPOs as well as the non-permit areas in the West South Central Division to compute the resulting monthly estimate of division-level single-family housing starts. This division-level estimate provides a more geographically granular tabulation than the current production estimates which go down only to the region level. Further details on the estimation methodology are provided in Section 7.

3.2 Residential building type

For the initial effort at satellite data collection, the focus was on single-family homes because each observed start typically corresponds to one housing unit. Single-family attached homes (i.e. townhomes) are an exception, as they are included in single-family estimates but may have multiple units corresponding to an individual satellite observed start when all of the attached units are on the same stage of construction.

Expanding into multi-family starts will require a process to both identify housing starts and then to predict the number of housing units associated with the building, which may be difficult at the time of start when only the excavation or foundation are in place. Although building permit data does include the number of housing units for the building, most addresses for new construction cannot be geocoded (located on a map) in real time. In other words, even if the building permit

data for a new multi-family building were available, it could not be located on the satellite image during the production process in order to assign the appropriate number of housing units to the observation.

Furthermore, single-family homes are typically built in areas with other single-family homes. These single-family residential neighborhoods are easily identified on satellite imagery and are used in the building type model described in section 5.2 to limit the counting of starts to areas that are likely to contain single-family homes. However, multi-family buildings are more likely to be found near city centers and may be hard to distinguish from non-residential structures, such as an office building or hospital, at the start of construction. Estimating multi-family housing starts from satellite imagery is an area for future research.

3.3 Timing

The project was intended to be a proof of concept of how satellite imagery could be used to replace at least some of the data collected by field representatives and used in a production environment. Therefore, it adhered closely to the current New Residential Construction report production timeline, with the experimental data product publication occurring on the business day following the New Residential Construction publication.

4. Image Collection and Pre-Processing

4.1 Defining the areas of interest (AOIs)

BPOs with significant single-family home construction activity were selected as the areas of interest (AOIs) for satellite data collection. These AOIs include incorporated places (cities,

towns, and villages), census-designated places, counties, and unincorporated areas. Places refer to specific, bounded areas that are either incorporated under state law with their own local government or, in the case of census-designated places, unincorporated but identified for statistical purposes. Counties, on the other hand, are larger administrative units that encompass multiple places as well as unincorporated areas. To define the exact boundaries of these AOIs, Census TIGER/Lines files were used. County geometries were derived from the U.S. county file, while geometries for places come from state-specific place files. Unincorporated areas of a county were determined by combining county geometries with incorporated places, identified through a spatial join using BPS data and place TIGER/Lines files, leaving the unincorporated portion as the new AOI for that county's unincorporated area.

Once the geometries were established, they were sent to an image vendor for augmentation to meet satellite tasking requirements. The vendor simplifies the AOIs by reducing the number of vertices (corners) to ensure the total for the project remains below a defined limit. Additionally, if an AOI is less than 50 square kilometers, it may be combined with a larger nearby AOI. In some cases, the vendor may remove areas such as airports, state and federal parks, military bases, and bodies of water, where residential construction is not possible. The modified AOIs were reviewed with the vendor to confirm that any changes were appropriate. These augmentations do not affect the original AOIs used for data reporting. Finally, the vendor and image provider assessed whether the areas could be captured during the required timeframe before tasking began.

The list of AOIs were merged to form a few large polygons which encompass the total area of interest for satellite image tasking. At delivery, a collection of images were returned, each representing a loose grid tile. The satellites make multiple passes over the area and captured many images that were split into tile grids. Each tile is around 18432 by 18432 pixels for 50 centimeter resolution or around 9.2 kilometers by 9.2 kilometers. Often, these tiles from different passes can overlap. Steps are taken during the blob search portion of the process detailed in Section 6.1 to remove duplicates that arise from image overlap.

4.2 Image collection window

Ideally, to mirror current data collection procedures, satellite imagery would be captured on the first day of the month following the reference month. However, the satellites do not pass over each AOI every day so a multi-day collection window is needed to ensure the satellite can make a collection attempt for each AOI. Additionally, weather conditions can render an AOI or portions of an AOI unobservable by satellite. A longer image collection window allows for multiple collection attempts and a greater chance of obtaining usable satellite imagery. As a result, the satellite tasking window was defined as the 25th of the reference month to the 5th of the following month.

4.3 Image quality requirements

Satellite-based estimates of housing starts are contingent on quality satellite images. Images that are too dark, too blurry, or obscured by clouds cannot be correctly classified and are, therefore, unusable in producing satellite-based estimates of housing starts. We provided the satellite vendor with thresholds for the maximum cloud cover of 50% and a maximum incidence angle of 30%. The satellite makes multiple passes over an AOI during the image collection window and images are returned when the quality requirements are met. Therefore, the final set of images for

an AOI for a given month may come from multiple days within the collection window.

Furthermore, if the image quality requirements are not met for an AOI or part of the AOI during the collection window, images are not returned resulting in missing satellite data.

5. Convolutional Neural Networks for Image Classification

5.1 Introduction to CNNs

A CNN is a type of deep-learning model that is especially good at understanding images (Zhao et al. 2024). In simple terms, in the context of image recognition, a CNN learns by sliding small “filters” over an image and looking for simple patterns like edges, corners, and textures. CNNs are typically composed of several layers that work together to identify patterns of interest. Early layers detect very basic shapes, and as the image goes through more layers, the network combines those simple shapes into bigger, more meaningful features. Each layer keeps the most useful information and gradually ignores noise, so the model becomes better at recognizing what matters for the task. For segmentation, the CNN does this pattern-learning at every pixel, allowing it to decide which class each pixel belongs to based on both local details and the broader scene.

The model for this application was trained by showing it many satellite images paired with their matching human applied labels, described in Section 5.3. During training, the model made a prediction for each image, compared it to the true mask, and computed the error according to the loss function (Formula Box 1). It then calculated the mean of intersection over union (Formula Box 2) to determine how to correct the predictions. The training algorithm then adjusted the

model's internal weights to reduce that error, repeating this process over many batches and epochs until performance stabilized.

To validate the results, we combined quantitative metrics with manual validation. During the manual validation, a random sample of predicted masks was overlaid on the original images and reviewed by human analysts. This visual inspection confirmed whether the predicted construction stages matched the observed structures and identified any new construction projects which were missed by the model predictions.

Formula Box 1: Loss Functions

The total loss function used in this application is a combination of the dice loss function and the focal loss function to both maximize overall region correctness and the detection of more detailed features. For class $c \in \{1, \dots, C\}$ and pixels $i \in \{1, \dots, N\}$, define:

- $p_{i,c}$: predicted probability for class c at pixel i
- $y_{i,c}$: ground-truth one-hot label (0 or 1)
- ϵ is a small constant for numerical stability (equal to 10^{-6} for our application)

Dice Loss

Dice coefficient per class is defined as

$$\text{Dice}_c = \frac{2 \sum_{i=1}^N p_{i,c} y_{i,c} + \epsilon}{\sum_{i=1}^N p_{i,c} + \sum_{i=1}^N y_{i,c} + \epsilon}$$

The mean dice coefficient over all classes is

$$\text{Dice} = \frac{1}{C} \sum_{c=1}^C \text{Dice}_c$$

Then the dice loss function is given by

$$\mathcal{L}_{\text{dice}} = 1 - \text{Dice}$$

Focal Loss

The cross-entropy function is the predicted probability of the true class of pixel i, p_i^+

$$p_i^+ = \sum_{c=1}^C y_{i,c} p_{i,c}$$

Then the focal loss function is given by

$$\mathcal{L}_{\text{focal}} = -\frac{1}{N} \sum_{c=1}^C \sum_{i=1}^N \alpha_c (1 - p_{i,c})^\gamma \log(p_i^+)$$

- α_c : class weighting factor
- γ : focusing parameter (equal to 2 for our application)

Total Loss

$$\mathcal{L}_{\text{total}} = 0.4\mathcal{L}_{\text{dice}} + 0.6\mathcal{L}_{\text{focal}}$$

$$\mathcal{L}_{\text{total}} = 0.4 \left(1 - \frac{1}{C} \sum_{c=1}^C \frac{2 \sum_i p_{i,c} y_{i,c} + \epsilon}{\sum_i p_{i,c} + \sum_i y_{i,c} + \epsilon} \right) + 0.6 \left(-\frac{1}{N} \sum_{i=1}^N \alpha (1 - p_i^+)^\gamma \log(p_i^+) \right)$$

Formula Box 2: Intersection over Union (IoU), also called Jaccard's Index

For a given class c :

- TP_c : pixels correctly predicted as class c
- FP_c : pixels predicted as class c but actually not c
- FN_c : pixels that are class c in truth but predicted as other classes

$$\text{IoU}_c = \frac{TP_c}{TP_c + FP_c + FN_c}$$

Then the Mean IoU is given by

$$\text{mIoU} = \frac{1}{C} \sum_{c=1}^C \text{IoU}_c$$

5.2 General structure of the models used

The CNN models used in this project have an encoder-decoder architecture. This architecture is highly beneficial for tasks like semantic segmentation because it balances feature extraction and spatial resolution restoration. The encoder focuses on extracting high-level, abstract features from the input image, such as shapes and patterns by progressively downsampling the spatial dimensions and applying deep convolutional layers. Then the decoder restores the spatial details, ensuring fine-grained predictions by upsampling the features back to the original resolution. This

combination allows the model to capture global context and preserve local details simultaneously.

One of the key advantages of this architecture is its ability to learn features at multiple scales. The encoder compresses the image, enabling the model to extract global information, while the decoder combines this with local features to produce detailed segmentation outputs. This multi-scale learning is particularly useful for handling objects of varying sizes within an image. Additionally, skip connections and feature concatenation between the encoder and decoder facilitate the retention of fine-grained spatial details, resulting in sharper boundaries and more accurate predictions.

The encoder-decoder structure is also computationally efficient. By reducing the spatial dimensions in the encoder stage, the model can focus on learning complex patterns while requiring less computational power. The decoder then uses this compact representation to reconstruct meaningful spatial information. Another significant benefit is its robustness to noise and small datasets. The bottleneck structure in the encoder forces the model to learn essential features, which helps reduce overfitting on noisy data. Furthermore, pretrained encoders can be leveraged to transfer knowledge from large datasets, enabling efficient training even with limited labeled data. This architecture can also incorporate attention mechanisms, enhancing its ability to focus on the most important regions of the image, thus improving segmentation quality.

5.3 Construction stages model

The construction stages model was designed to classify observed pixels into one of eight classes corresponding to the stages observed for new residential construction. Three of these classes, land, vegetation, and background, are considered pre-construction and indicate that neither a construction site nor an existing structure is present. The remaining five classes follow the stages of construction from start to completion. They are excavation, foundation, framing, unfinished roof, and finished roof.

This model was trained on 4,550 images labeled with these eight classes. Following a segmentation approach, the model predicted the most likely class for each pixel in the image. The resulting output is referred to as a prediction mask, a pixel-by-pixel construction stage prediction.

5.3.1 Model architecture

Figure 1 illustrates an Attention U-Net architecture designed for semantic segmentation tasks. The architecture takes an input image of dimensions $256 \times 256 \times 3$ and processes it through an encoder-decoder structure, enhanced by attention gates to focus on relevant regions. The encoder consists of convolutional layers (3x3 convolutions with ReLU activation) and max-pooling layers (2x2) to progressively downsample the input and extract hierarchical features. Skip connections between the encoder and decoder preserve spatial details by directly transferring features from the encoder to corresponding layers in the decoder.

The decoder upsamples the feature maps using transpose convolution (2x2) and refines them with convolutional layers, progressively reconstructing the segmentation map to match the input resolution. Attention gates (AG) are introduced along the skip connections to emphasize the most

relevant features while suppressing background noise. These gates dynamically learn to focus on key regions by multiplying incoming feature maps with attention weights computed based on contextual importance.

At the output, a 1x1 convolution layer maps the refined features into a segmentation mask of $256 \times 256 \times 7$, where each pixel corresponds to one of the output classes. The use of attention gates ensures precise segmentation by enhancing the model's focus on critical regions while reducing interference from irrelevant parts of the image, making this architecture particularly effective for tasks requiring fine-grained and accurate pixel-wise predictions. For additional details on Attention U-Net see Jing (2019).

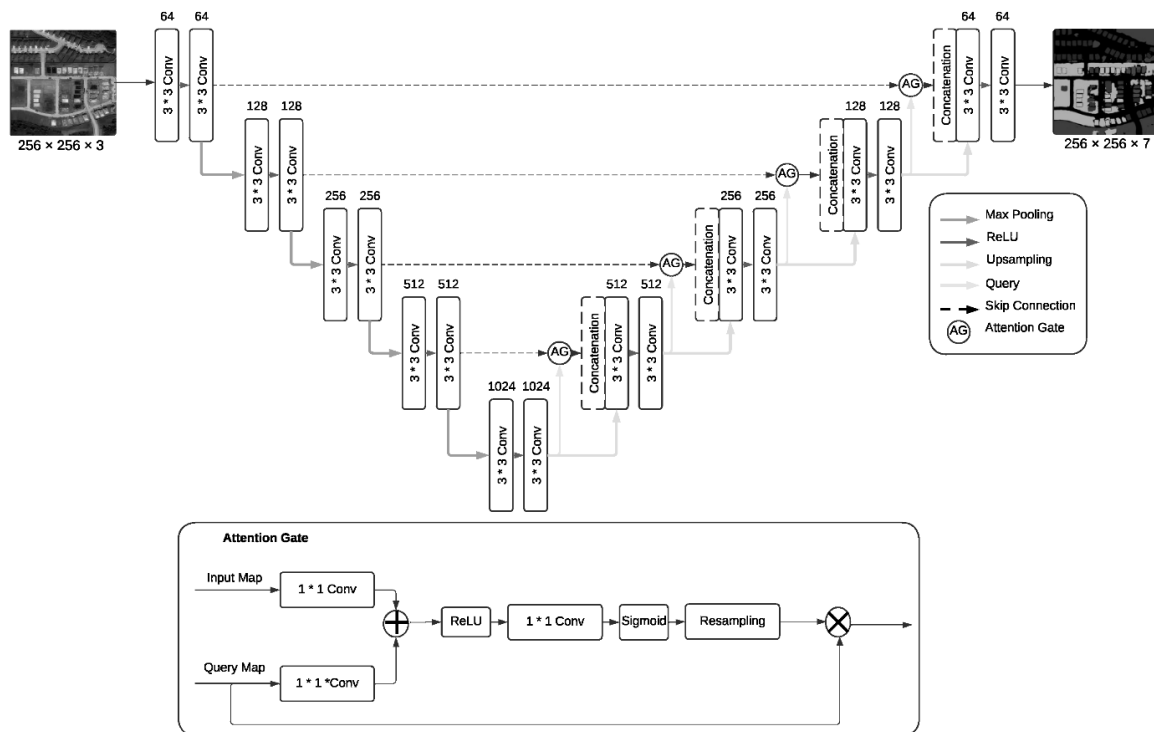


Figure 1. Construction stage model using Attention U-Net architecture.

5.3.2 Model training and evaluation

In addition to the Attention U-Net architecture, several additional architectures were tested for the construction stages model including DeepLab + ResNet34, U-Net, HRNet, and DeepLabV3+ResNet50. The U-Net, DeepLabV3+ResNet50, and Attention U-Net emerged as the top performers. These models demonstrated superior accuracy and robustness in segmenting distinct construction stages, even with constrained data availability. U-Net excelled due to its efficient encoder-decoder structure, it uses a simple, efficient ‘compress then rebuild’ design that keeps important image details. DeepLabV3+ResNet50 leveraged multi-scale context effectively through atrous convolutions, looking at the ‘big picture’ to understand the surrounding area. Attention U-Net improved results by adding a ‘focus’ step that helps the model pay attention to the most useful parts of the image and ignore background noise.

Our approach to model training and testing utilized normalization and the CLAHE color function for image processing, effectively improving feature visibility in various lighting conditions. The AdamW optimizer with weight decay and the Categorical Focal Crossentropy loss function were employed, optimizing generalization and handling class imbalances by focusing on difficult-to-classify examples. These tailored strategies collectively ensured that the top three models were particularly well-suited to our construction stage segmentation project, delivering clear and precise segmentation critical for monitoring progress.

The three model architectures were evaluated on accuracy, validation loss, and mean intersection of union (MioU, see Formula Box 2) with the results presented in Table 1. The models were validated on a dataset of 240 images. Accuracy is the proportion of correctly predicted pixels out

of the total number of pixels. Validation loss assesses how well the model generalizes to unseen data, with lower values indicating higher performance. Higher values of MIoU indicate better model performance.

	U-net	Attention-Unet	DeepLab_V3
Validation loss	0.03	0.03	0.02
Accuracy	0.94	0.93	0.94
mIoU	0.86	0.84	0.85

Table 1. Construction Stage model architecture evaluation statistics based on a dataset of 240 images. Source: Reveal Global Consulting. Training and testing images from March 2023 to August 2024.

5.4 Building type model

The building type model was designed to distinguish single-family residential areas from other areas on the satellite image. New construction for a store or other non-residential structure can look very similar to new single-family residential construction, especially in its early stages.

Early in the project we considered using parcel data, which would include zoning information indicating whether the land was zoned for residential use. However, we quickly discovered that publicly available parcel data is not updated as frequently as needed for the monthly estimates we sought to produce and would often result in the incorrect exclusion of in-scope new construction. A common example we encountered was a large parcel that was previously zoned agricultural was now being developed into new single-family homes but the zoning on the publicly available parcel data was not yet updated to reflect the change from agricultural to residential.

Therefore, we developed the building type model to determine which areas on the image were likely to be single-family residential, and therefore in scope to our estimates. There were five classes defined for the building type model: background, single-family residential detached, single-family residential attached, multi-family residential, and non-residential.

5.4.1 Model architecture

Figure 2 illustrates the semantic segmentation architecture used in the building type model to process an input image through an encoder-decoder framework. The process begins with the input image being fed into the encoder, which uses a ResNet50 backbone for feature extraction. The encoder captures hierarchical feature representations, progressing from low-level details like edges and textures to high-level semantic features. To enhance multi-scale feature learning, the architecture incorporates an Atrous Spatial Pyramid Pooling (ASPP) module within the encoder. This module consists of 1x1 convolutions for point-wise feature extraction, 3x3 convolutions with varying dilation rates (6, 12, and 18) to capture features at different spatial scales, and global image pooling to encode context across the entire image. The outputs of these operations are concatenated and passed through a 1x1 convolution to ensure compact and meaningful feature representations.

The decoder complements the encoder by restoring spatial resolution and refining predictions. Low-level features from earlier layers of the encoder are integrated into the decoder through skip connections, ensuring that fine-grained spatial details are preserved. These low-level features are first processed using a 1x1 convolution to reduce their channel dimensions, making them compatible with the high-level features from the ASPP module. The decoder upsamples the high-level features by a factor of four, combining them with the low-level features through

concatenation. Additional refinement is achieved through a series of 3x3 convolutions and further upsampling, which ensures that the output segmentation mask matches the resolution of the original input image. Finally, the model outputs a pixel-wise segmentation mask, which predicts the class of each pixel in the input image. For more on semantic segmentation see Kou et al. (2022).

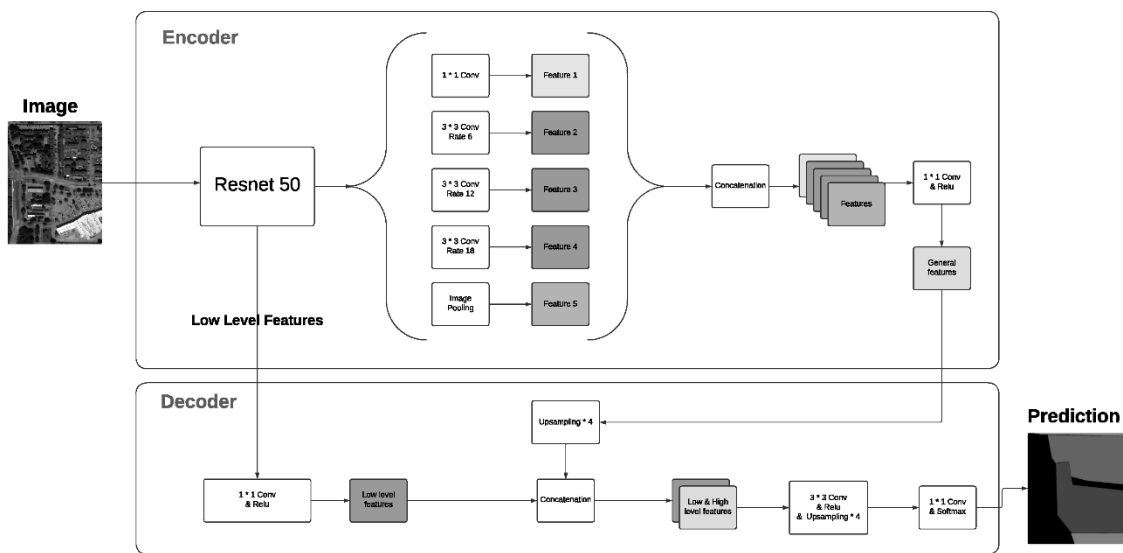


Figure 2. Illustration of semantic segmentation architecture used in the building type model.

5.4.2 Model training and evaluation

We evaluated the performance of DeepLabV3+, U-Net, and Attention U-Net models, training on 32,691 images and validating on 1,721 images. The results of this analysis are shown in Table 2.

The validation loss indicated that the U-Net model was not generalized enough to handle unseen data so this architecture was abandoned early in the development of the building type model and we did not compute the mIoU at the time. Instead, we focused our evaluation on the other two methods. The results indicate that both models performed adequately with DeepLabV3 demonstrating slightly better performance in accuracy. However, the validation losses and mIoU suggest room for improvement, especially in generalizing to unseen data. We continue to work on improving the building type model by incrementally adding more training data, especially for less common classes like multi-family and single-family attached, which is expected to boost model accuracy, reduce loss, and enhance segmentation quality across all building types.

	U-net	Attention-Unet	DeepLab_V3
Validation loss	0.57	0.08	0.06
Accuracy	0.86	0.81	0.89
mIoU	na	0.38	0.54

Table 2. Building model architecture evaluation statistics based on a dataset of 1,721 images.

Source: Reveal Global Consulting. Training and testing images from March 2023 to August 2024.

6. Post Processing

6.1 Blob search

Both the construction stage model and building type model's prediction mask output is a grayscale Geo-TIF image file which consists of different grayscale colors for each pixel representing different classes of interest. The next step is to extract useful information from this

mask. We initially tried using parcel geometry data to crop the mask output for a single property and determine the state of construction using the pixel distribution available. However, the available parcel data can be out of date, especially for new construction zones. This is primarily due to processing time by each county in updating publicly available parcel data. The delay could range from a couple of months to over a year. Parcel data would be better suited for tracking existing structures and is not well suited to this application.

The solution we developed was to extract the data through a process called Blob Search. This process extracts clustered or connected pixels of the same classification, hereafter known as blobs. Each blob represents one of the following: a roof, a particular stage of construction of a building, land, or vegetation. This provides a clear outcome of the property's stage of construction without the need for third party data. Additional processing is done to remove duplicate blobs that exist due to overlapping images due to tasking. This ensures that there is only one blob per building per month.

Figure 3 provides an example of a satellite image and the corresponding construction stage prediction mask.

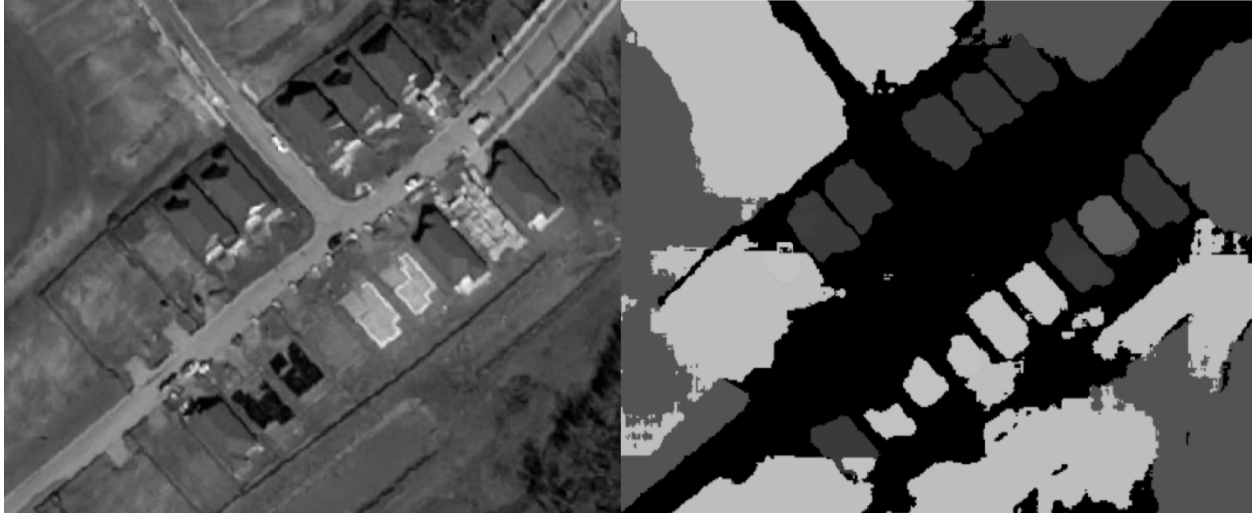


Figure 3. Satellite image and construction stage prediction mask. Image from February 2026.

Prediction Mask Source: Reveal Global Consulting using construction stage model trained on February 18, 2026.

6.2 Footprint exclusion

To reduce noise in the blob search output, a process was developed that utilizes open-source building footprint data. Bing Maps released a dataset containing 1.4 billion building footprint polygons detected on imagery between 2014 and 2024. Start blobs identified through blob search are overlaid on top of building footprint shapes and any overlaps are excluded as probable noise. Care was taken to ensure building shapes were extracted from imagery taken prior to the reference month image. Additionally, blobs were overlaid with known roads using the 2023 TIGER/Lines Road Data to exclude blobs (excluding land and vegetation) that overlap with roads. Testing indicated a large reduction in false starts with minimal loss of true starts. In time, model improvements should eliminate the need for this post-processing step. Continuous training of the model will improve accuracy and reduce the dependency on additional data sources.

6.3 Business classification

Specific business rules were applied to interpret construction progress by analyzing transitions between predictions from the previous month to the current month. These rules help determine the construction status for each property site based on the logic defined in Table 3.

For example, if a site was predicted to be in the LAND phase in the previous month and the current month shows EXCAVATION, the business class START is assigned, indicating that construction has progressed from land preparation to active development. Similarly, a transition from EXCAVATION to FOUNDATION would indicate further progress, thus the business class IN PROGRESS would be assigned.

The previous month's blobs are matched with the current month's blobs using geographic metadata and matching algorithms, accounting for any size differences between the blobs across months. Once the business rules are applied, predictions are aggregated across the defined place. This aggregation provides an estimate of the number of single-family housing starts in the AOI.

The model may predict some noise because of cloud cover, poor visibility, or limited sunlight during winter months. These cases can produce discrepancies in what should be observed. A post process step is taken to adjust these mis-matched areas accordingly. For example, if an area has been predicted to have a roof for 3 months, the likelihood of it becoming vegetation is highly unlikely. Therefore, we modify the result to be roof to roof which takes the MISSCLASSIFIED BACKWARDS to an EXIST category.

Changing seasons are additional source of noise in the model predictions. Houses with trees nearby may be classified as VEGETATION when leaves are present but classified as ROOF after the leaves fall. This results in START COMPLETION categorization. From our microvalidation efforts, we determined that the vast majority of blobs identified as START COMPLETION were false starts. However, while the progression from a preconstruction stage (BACKGROUND, LAND, or VEGETATION) to a roof within a month is uncommon, it does occur, most often in areas with many construction projects occurring close together, such as a new neighborhood, and so cannot be completely ignored. Consequently, START COMPLETION blobs that lie within the same Geohash (see the imputation section below for more details on Geohashes) as a blob categorized as START are included in the starts estimate. All other START COMPLETION blobs are excluded. We note that the “COMPLETION” business class does not correspond to the SOC definition of a completed house, which is defined by the presence of finished flooring. However, since the focus of this effort was housing starts it seemed sufficient for identifying a house that has reached the finished roof stage. A method for identifying completed units that more closely aligns with the survey definition is one of many areas of future research.

	Current Month Prediction							
Previous Prediction	Background	Land	Vegetation	Excavation	Foundation	Framing	Unfinished Roof	Roof
Background	Unknown	Land	Land	Start	Start	Start	Start	Start Completion
Land	Land	Land	Land	Start	Start	Start	Start	Start Completion
Vegetation	Land	Land	Land	Start	Start	Start	Start	Start Completion
Excavation	Image Error	Land	Land	In Progress	In Progress	In Progress	In Progress	Completion
Foundation	Image Error	Misclassified Backwards	Misclassified Backwards	In Progress	In Progress	In Progress	In Progress	Completion
Framing	Image Error	Misclassified Backwards	Misclassified Backwards	In Progress	In Progress	In Progress	In Progress	Completion
Unfinished Roof	Image Error	Misclassified Backwards	Misclassified Backwards	In Progress	In Progress	In Progress	In Progress	Completion
Roof	Image Error	Misclassified Backwards	Misclassified Backwards	Start Reconstruction	Start Reconstruction	Start Reconstruction	Rehab	Exist

Table 3. Business classification rules based on construction stage predictions.

6.4 Imputation

The image quality requirements discussed in Section 4.3 define the minimally acceptable satellite image that can be classified by the models presented in Section 5. However, when images meeting these requirements are not obtained in the defined image capture window for a given area, there is missing data. The entire AOI could be missing, or more frequently, only part of an AOI is missing. Furthermore, there could be consecutive months of missing data for a given area due to cloud cover.

New residential construction is not typically evenly distributed across an AOI, which could be as large as an entire city or county. It is typically localized to a small number of areas with many new homes being built close together. When a partial AOI is observed, rather than applying a missing data mechanism at the AOI level, we took a localized approach using Geohashes to define the area for imputation. Geohash is a public domain geocoding system to define geographic areas in a hierarchical spatial structure (Wikipedia, 2025). The number of characters in the hash represents the size, a 6-character Geohash is larger in area compared to 7-character Geohash. For imputation, we divide an AOI into its component 7-character Geohashes (153m x 153m). At this more granular geographic level, both the assumption of evenly distribution construction activity in the current month and an assumed correlation in activity in consecutive months are more plausible than at the AOI level.

When we encountered missing data for an entire Geohash or part of a Geohash, we drew from the most recent observed data for that Geohash to develop an imputed value. When less than 50% of a Geohash was observed in the current month, the Geohash's starts value was set to the

maximum of the Geohash starts observed in the target month and the Geohash's starts from the prior (or last observed) month. When at least 50% of the Geohash was observed in the current month, the current month's starts are adjusted by dividing the observed start by the proportion of the Geohash that is observed. For example, if 75% of the Geohash was observed and three starts were identified in the observed area, the final starts for the Geohash would be calculated as $3/0.75 = 4$ starts.

When the missing part of the Geohash is next observed, a rectification process was applied to ensure that housing starts would not be double counted. For example, assume a Geohash was completely covered by clouds in the prior month and had a value of four imputed starts based on its previous month's fully observed image. In the current month, the Geohash is fully observed and compared to the last observed image from two months ago to determine the business classification for the observed blobs. The resulting comparison results in the identification of six housing starts. However, only two starts would be recorded in the current month for that Geohash because four of the six starts were accounted for in the imputation in the prior month.

6.5 Capture window adjustment

Capture window adjustment is used to adjust the post imputed counts for the difference in days between image dates and the number of days in the target month. Due to the long image capture window, the number of days between observations could vary between 20 days to 42 days. We adjust the estimated starts for a given area by the ratio of the number of days in the target month to the number of days between observations.

For example, if the image for May was taken on June 1st and the image for June was taken on July 4th, there would be 33 days between images when the June estimates were calculated and there are 30 days in June. Therefore, the June estimate would be multiplied by 30/33 to correct for the length of time between observations. This simple solution assumes that construction activity is the same for every day of the week and makes no attempt to adjust for differences in activity based on the day of the week or the presence of holidays. A more nuanced capture window adjustment that does attempt to capture these differences is an area for future research.

6.6 Place level validation

After these processes were complete, the final monthly estimate of housing starts for the AOI was computed by summing the final estimate of starts (including imputation and capture window adjustment) for each Geohash in the AOI. These estimates were validated by BPO level estimates (for the corresponding AOI) computed from the SOC data. This served as a quality control measure on the satellite AOI estimates. The SOC sample was not designed to produce estimates below the region level and consequently, direct BPO level estimates from the SOC data are highly variable. Instead, we computed simple model-based estimates of BPO level starts from the SOC data, conceptually mirroring the existing SOC estimator. We estimated division level ratios of starts in reference month t , authorized in month $t - i$, to the number of permits authorized in $t - i$ for $i = 0$ to 11 since nearly all single-family houses start within one year of permit authorization (i.e. $\frac{\hat{G}_{t-i,t}}{\hat{H}_{t-i}}$ defined in equations (2) and (3)). These ratios were multiplied by the number of permits authorized for the BPO for the corresponding month and summed to obtain the BPO level estimated starts for month t .

$$\hat{S}_{BPO,t} = \left[\sum_{i=0}^{11} Y_{BPO,t-i} \frac{\hat{G}_{t-i,t}}{\hat{H}_{t-i}} \right]$$

7. Estimation methodology

As discussed in Section 2.3, the current SOC housing starts estimates are computed as the sum of the estimates from both permit and non-permit issuing areas. The permit estimate makes use of the permit authorization date as an auxiliary variable that can be used to create a ratio adjustment to the known monthly values of permits authorized from the BPS. As such, the permit authorization date for each housing structure is a necessary component of the current estimation procedure for permit issuing areas.

However, the estimates obtained from the satellite imagery are at the BPO level and thus cannot be directly incorporated into the current estimator based on permit level data. Moreover, we are unable to match building permits in real time to individual observed housing starts on the satellite imagery due to geocoding constraints (addresses often do not yet exist in any existing geocoding database at the time of a housing start and cannot be located on the satellite image). In other words, we cannot assign a permit authorization month to observed housing starts from satellite imagery.

Therefore, a blended data estimator was used, combining the BPO level estimates obtained from the satellite imagery with a modified version of the currently used separate ratio estimator for the non-satellite permit places, essentially estimating the satellite portion separately from the non-satellite portion. It also includes a PSU calibration factor to add stability to the ratio estimates from the non-satellite sample data. The non-permit portion of the estimate is unchanged.

7.1 PSU calibration factor

The self-representing places that were chosen for satellite data collection generally have many sampled permits in SOC, all with relatively low sampling weights. When these permits were removed from the SOC sample to avoid double counting with the satellite data, the resulting dataset consisted of a smaller sample of permits, often with larger sampling weights. The resulting ratios used in the separate ratio estimator are less stable than their full SOC sample counterparts, resulting in high variances for estimates of housing starts.

With the BPS data, we can compute the number of permits authorized in a given month across any geographic domain. Therefore, we computed BPS totals at the PSU level and used those totals to construct a PSU calibration weight to add stability to the non-satellite SOC sample ratio estimates.

Divide the SOC permit sample of places into those selected to be collected by satellite (Q_S) and those that will not (Q_N). All calculations using SOC survey data detailed in Sections 7.1 and 7.2 only use records coming from places in Q_N .

Let w_j be the final weight for permit j (product of CPS sampling, SOC PSU sampling, BPO sampling, and permit sampling weights) and w_k^{PSU} be the sampling weight for PSU k (product of CPS sampling, SOC PSU sampling). Thus, the BPO-permit sampling weight for permit j from place p in PSU k is $w_j^{BPO-permit} = w_j / w_k^{PSU}$.

Define the number of building permits authorized in month $t - i$ from BPO p recorded in BPS as $y_{p,t-i}$ and let $F_p = 0$ if BPO p is collected by satellite and 1 otherwise. Also, let $A_{j,t-i} = 1$ if permit j was authorized in month $t - i$ and 0 otherwise. The PSU calibration factor, defined in

(4), for PSU_k in month $t - i$ is the ratio of the BPS value of permits authorized in month $t - i$ for non-satellite places in PSU_k , defined in (5), to the SOC estimate of permits authorized in month $t - i$ for non-satellite places in PSU_k , shown in (6).

$$w_{PSU_k,t-i}^{cal} = \frac{Y_{PSU_k,t-i}}{\hat{H}_{PSU_k,t-i}} \quad (4)$$

$$Y_{PSU_k,t-i} = \sum_{p \in PSU_k} F_p y_{p,t-i} \quad (5)$$

$$\hat{H}_{PSU_k,t-i} = \sum_{j \in Q_N \cap PSU_k} A_{j,t-i} w_j^{BPO-permit} \quad (6)$$

The final weight, w_j , for permit j belonging to PSU k , authorized in month $t - i$ is multiplied by the corresponding PSU calibration factor to define the calibrated final weight ($w_{j,t-i}^*$) that will be used for the non-satellite portion of the estimate.

$$w_{j,t-i}^* = w_j w_{PSU_k,t-i}^{cal}$$

7.2 Blended data estimator

The non-satellite permit sample estimate of housing starts uses a modified version of the separate ratio estimator currently used for SOC permit sample estimation.

Define $I_{j,t} = 1$ if permit j was a start in month t , and 0 otherwise. Then the non-satellite permit area estimate of units authorized in month $t - i$ and started in month t is given in (7). Likewise, the non-satellite permit estimate for total permits authorized in month $t - i$ by non-satellite collected places is defined in (8).

$$\hat{G}_{t-i,t} = \sum_{j \in Q_N} w_{j,t-i}^* A_{j,t-i} I_{j,t} \quad (7)$$

$$\hat{H}_{t-i} = \sum_{j \in Q_N} w_{j,t-i}^* A_{j,t-i} \quad (8)$$

Since all satellite collected places are in the BPS, we can remove their data from the BPS total. As mentioned in section 4.1, all satellite collected places are self-representing BPOS from self-representing PSUs, so their SOC place-level sampling weights are equal to one. Therefore, we can subtract their unweighted building permits for each month, $t - i$, from the BPS total number of permits for that month (Y_{t-i}). Define the adjusted BPS total as

$$\tilde{Y}_{t-i} = Y_{t-i} - \sum_p (1 - F_p) y_{p,t-i}$$

The total starts estimate for the satellite collected places (all with sampling weight=1) is simply the sum of the place-level estimates obtained from the satellite imagery for month t .

$$S_{SAT,t} = \sum_{p \in Q_S} S_{SAT,p,t}$$

The blended data estimator for total single-family starts is defined in (9) with the non-satellite permit issuing portion of the estimate presented in (10).

$$\hat{S}_t^{Blended, PSU\ cal} = \hat{S}_{permit,t}^{NS} + \hat{S}_{non-permit,t} + S_{SAT,t} \quad (9)$$

$$\hat{S}_{permit,t}^{NS} = \left[\sum_{i=0}^{11} \hat{G}_{t-i,t} \frac{\tilde{Y}_{t-i}}{\hat{H}_{t-i}} \right] + \left[\sum_{i=12}^{17} \hat{G}_{t-i,t} \left[\frac{\sum_{i=12}^{17} \tilde{Y}_{t-i}}{\sum_{i=12}^{17} \hat{H}_{t-i}} \right] \right] + \left[\sum_{i=18}^{59} \hat{G}_{t-i,t} \left[\frac{\sum_{i=18}^{59} \tilde{Y}_{t-i}}{\sum_{i=18}^{59} \hat{H}_{t-i}} \right] \right] \quad (10)$$

7.3 Nonresponse adjustment

The current nonresponse adjustment method for SOC accounts for late reported starts that are expected to be received up to twelve months late by multiplying the permit area housing starts

estimate by a nonresponse undercoverage adjustment factor (NUAF). The NUAF for the preliminary housing starts estimate is determined by examining the ratios of the final estimate to the preliminary estimate of housing starts for the most recent 12 months. The asymmetric fences outlier detection method (Thompson, 1999) with parameter $k=1.5$ is applied to the set of 12 ratios and extreme values are excluded. If at least six ratios remain, the average ratio is calculated from the remaining ratios and used as the NUAF for preliminary housing starts. If fewer than six ratios remain after outlier detection, the median ratio is used as the NUAF instead. NUAF factors for single-family starts are currently calculated at the region level. For the experimental data product, the same methodology was applied at the division level and excluded satellite places from all estimates used to define the NUAF. The calculated NUAF was applied only to the non-satellite permit portion of the estimate.

7.4 Reliability of the blended data estimates

The modified half-sample variance estimation method (Thompson, 1998) that is currently used for SOC was used to estimate sampling error for the non-satellite portion of the estimate. The satellite portion of the estimate has no sampling error because all the selected satellite places are self-representing and satellite data collection is attempted for all in-scope areas for the selected places.

There are non-sampling errors in the satellite-based estimates that were not quantified or included in the variance estimates for the experimental product. The sources of non-sampling error for the satellite estimates are analogous to the non-sampling errors of the survey data described in Section 2.4. The tasking AOI may not fully match the jurisdiction covered by building permit office and may include (exclude) areas where the jurisdiction does not (does) issue permits which could lead to consistent over (under) estimation for the AOI; a type of

coverage error. Misclassifications in the construction stage model or building type model are types of processing errors that can occur in the satellite estimation procedure as are missed exclusions based on existing building or road footprints. Additional processing errors can occur when a single blob covers more than one unit, leading to underestimation of the total units. Homes constructed with unique architectural styles or construction techniques may also go undetected by the satellite process if they substantively differ from the training images and are not appropriately classified by the construction stage or building type model. Furthermore, when the image capture dates are anything other than the last day of month, measurement errors arise because the period between images does not exactly correspond to the reference month. Finally, errors can occur from missing data due to cloud cover, similar to errors due to nonresponse. Throughout the satellite process, quality control measures (imputation, capture window adjustment, and survey estimate comparisons, etc.) were implemented to minimize these errors where possible. Although some of these errors cannot be quantified, a topic for future research is developing measures of uncertainty for the construction stage and building type models that can help quantify some of the uncertainty in the satellite estimates.

8. Experimental Data Product

These methods were used to produce an experimental data product, blending satellite data with survey data, of monthly single-family housing starts for the West South Central Division of the United States published by the U.S. Census Bureau available at <https://www.census.gov/construction/nrc/data/satellitestarts.html>. At the product's initial release in December 2024, estimates for July 2024 - November 2024 were published. Additional monthly data releases followed on the business day following the release of the New Residential Construction report, keeping essentially the same timeline as the current production indicator.

The final publication of this experimental product was released in September 2025, resulting in a series of 14 estimates from July 2024-August 2025 shown in Figure 4 with the accompanying 90% confidence interval.

8.1 Comparison with estimates from survey data

We compared these blended data estimates to those obtained using the survey data alone. It should be noted that the Census Bureau does not currently publish estimates at the division level nor is the survey specifically designed to produce estimates at the division level. However, we considered the sample size and reliability of the survey West South Central Division estimates to be sufficient to be used as a “ballpark” estimate for comparison.

First, we computed estimates from the survey data only using the PSU calibration factor from the experimental product. This analysis isolates the change from survey data collection to satellite data collection, leaving the estimation methodology for the survey data portion of the estimate unchanged. We examined the ratios of the blended estimates incorporating satellite data to those using survey data only to gauge how the point estimates compared. We also looked at the rate at which the blended estimates fell within the 90% confidence interval of the survey data estimates. This analysis was repeated using the existing production estimation methodology (without the PSU calibration factor) for the survey data. The results of both comparisons are given in Table 4.

The point estimates incorporating satellite data fell within the confidence intervals of the survey estimates over 70% of the time, indicating that they are not statistically different than what would have been obtained from the survey data alone for those months at the 90% confidence level. Furthermore, the satellite estimates neither consistently overestimated nor underestimated total housing starts relative to the survey estimates, with the average ratios being close to one.

Based on these results and other internal validation efforts, we concluded that the blended estimates incorporating satellite data were plausible.

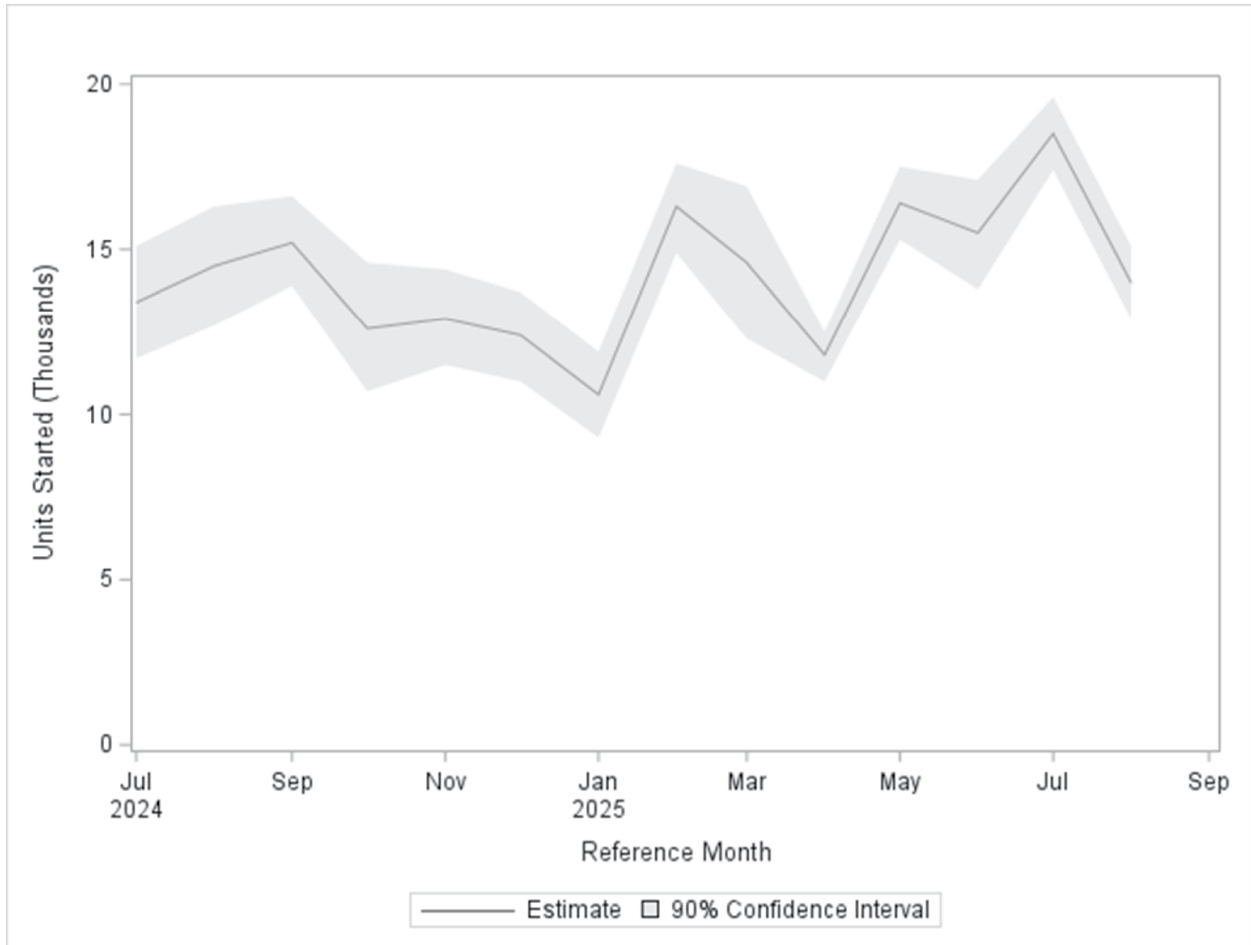


Figure 4. Blended data estimate incorporating satellite data with SOC survey data for new privately-owned single-family housing units started in the West South Central Division with 90% confidence interval, July 2024 - August 2025. Source: U.S. Census Bureau. Disclosure Review Board (DRB) approval number: CBDRB-FY26-ESMD009-001.

Comparison	Minimum	Maximum	Average	90% CI Inclusion Rate
$\frac{\hat{S}_t^{Blended,PSU\ cal}}{\hat{S}_t^{SOC,PSU\ cal}}$	0.89	1.33	1.02	78.6%
$\frac{\hat{S}_t^{Blended,PSU\ cal}}{\hat{S}_t^{SOC,production}}$	0.75	1.34	1.06	71.4%

Table 4. Summary statistics for ratios comparing blended data estimates to survey only estimates over the period, July 2024 - August 2025. Source: U.S. Census Bureau. Disclosure Review Board (DRB) approval number: CBDRB-FY26-ESMD009-001.

9. Conclusions and Next Steps

This experimental data product for monthly West South Central Division single-family housing starts represented a significant step forward in the exploration of alternative methods for measuring construction activity in the United States. The effort successfully demonstrated that modern computer vision techniques can be applied to high-resolution satellite imagery to reliably identify new construction activity and produce estimates consistent with legacy methods used for SOC.

The implementation yielded several notable accomplishments. It established that monthly detection of construction stages from satellite imagery is feasible across seasons and showed that

satellite-derived housing starts estimates could be incorporated into SOC estimation without changes to the underlying sample design. It also provided operational experience in coordinating satellite tasking and image processing on an indicator production timeline. As a result, the Census Bureau's technical capabilities in the use of satellite imagery were substantially expanded.

The project also highlighted key constraints that limit the future scalability of this approach. Attempting to expand toward regional or national coverage while remaining reliant on high-resolution imagery would impose substantial acquisition costs, and competition for tasking capacity may restrict the ability to ensure consistent monthly coverage. Additionally, expansion without decoupling from the SOC sample design would continue to require traditional data collection to supplement satellite-based observations.

With these constraints in mind, the experimental data product, in its current form, has been discontinued. This research has provided a strong foundation for the next phase of this project that is currently underway. We are exploring the feasibility of lower resolution imagery at both 1.5m and 4.7m resolution. To create training data for these types of images, we are leveraging the existing labeled 50cm resolution training data and downsampling the pixels to match the lower-resolution images. We have also decoupled satellite data collection from the SOC sample to expand satellite data collection beyond the existing sample places and are exploring model-based estimation methods utilizing the monthly BPS data to develop estimates based only on satellite observations. As satellite data collection expands geographically, additional validation will be required to ensure that the model is robust to a variety of geographic features, seasonal changes, and variations in architecture and construction techniques. If necessary, multiple models can be developed to accurately measure construction in different parts of the country. Overall, we

are hopeful this new approach could pave a more sustainable road toward broader geographic coverage and may allow for even more detailed geographic estimates.

Declaration of Conflicting Interests and acknowledgement of funding

The authors declare that there are no conflicts of interest and no funding sources to declare.

Acknowledgements

The authors thank Angela Delano, Brian Dumbacher, Katie Genadek, Joseph Gyourko, Bonnie Kegan, and Linda Young for their thoughtful comments on earlier versions of this manuscript.

References

- ECOSOC. 2015. "Report of the Global Working Group on Big data for official statistics." United National Economic and Social Council. 46th Statistical Commission. Available at: <http://unstats.un.org/unsd/statcom/doc15/2015-4-BigData.pdf> Accessed March 3, 2025.
- Erciulescu, Andreea L., Nathan B. Cruze, and Balgobin Nandram. 2020. "Statistical challenges in combining survey and auxiliary data to produce official statistics." *Journal of Official Statistics* 36 (1): 63-88.
- Eurostat. 2014. "Big data – an opportunity or a threat to official statistics?" 62nd plenary session at Conference of European Statisticians, Paris, April 9 – 11 2014. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2014/32-Eurostat-Big_Data.pdf Accessed March 3, 2025.

- Hutchinson, Rebecca, Scott Scheleur, and Deanna Weidenhamer. 2023. "Alternative Data Sources in the Census Bureau's Monthly State Retail Sales Data Product", In *Advances in Business Statistics, Methods and Data Collection*, edited by. Ger Snijkers, Mojca Bavdaž, Stefan Bender et al. John Wiley & Sons, Ltd.
- Japac, Lilli, Frauke Kreuter, Marcus Berg, et. al. 2015. "Big Data in Survey Research: AAPOR Task Force Report." *Public Opinion Quarterly* 79 (4): 839-80.
<https://doi.org/10.1093/poq/nfv039>.
- Jing, Hong. 2019. (2019, December 8). "Biomedical Image Segmentation – Attention U-Net." *Jinglescode*. December 8. Available at
<https://jinglescode.github.io/2019/12/08/biomedical-image-segmentation-u-net-attention/> Accessed December 2024.
- Kaputa, Stephen J., Darcy Steeg Morris, and Scott H. Holan. 2024. "Bayesian Multisource Hierarchical Models with Applications to the Monthly Retail Trade Survey." *Journal of Survey Statistics and Methodology*. 2024(00): 1-23.
<https://doi.org/10.1093/jssam/smae019>.
- Khan, Md Kamrul Hasan, Avishek Chakraborty, Giovanni Petris, and Barry T. Wilson. 2021. "Constrained Functional Regression of National Forest Inventory Data Over Time Using Remote Sensing Observations." *Journal of the American Statistical Association* 116(535): 1168–80. <https://doi.org/10.1080/01621459.2020.1860769>
- Kitchin, Rob, and Samuel Stehle. 2021. "Can Smart City Data be Used to Create New Official Statistics?" *Journal of Official Statistics* 37 (1): 121-47. <https://doi.org/10.2478/jos-2021-0006>.

Kou, Lei, Mykola Sysyn, Szabolcs Fischer, Jianxing Liu and Olga Nabochenko. 2022. "Optical Rail Surface Crack Detection Method Based on Semantic Segmentation Replacement for Magnetic Particle Inspection." *Sensors* 22(21): 8214. <https://doi.org/10.3390/s22218214>.

National Academies of Sciences, Engineering, and Medicine. 2023. "Toward a 21st Century National Data Infrastructure: Enhancing Survey Programs by Using Multiple Data Sources." The National Academies Press. <https://doi.org/10.17226/26804>.

Rao, J. N. K., and J. Shao. 1996. "On Balanced Half-Sample Variance Estimation in Stratified Random Sampling." *Journal of the American Statistical Association* 91(433): 343-48. <https://doi.org/10.1080/01621459.1996.10476694>.

Ryerson, R. A., J.-L. Tambay, R. J. Brown, L. A. Murphy, and B. McLaughlin. 1983. "A Timely and Accurate Potato Acreage Estimate from Landsat: Results of a Demonstration." *Survey Methodology* 9 (1): 119-138.

Smith, Aidan and Hector Ferronato. 2021. "Modernizing Construction Indicators Through Machine Learning and Satellite Imagery." *Proceedings of Statistics Canada Symposium 2021*. Available at <https://www150.statcan.gc.ca/n1/en/pub/11-522-x/2021001/article/00019-eng.pdf?st=vefDpbSy> Accessed March 3, 2025.

Thompson, Katherine J. 1998. "Evaluation of Modified Half-Sample Replication for Estimating Variances for the Survey of Construction (SOC)." *Economic Statistical Methods Report Series* ESM-9801.

Thompson, Katherine J. 1999. "Ratio Edit Tolerance Development Using Variations of Exploratory Data Analysis (EDA) Resistant Fences Methods." *Proceedings of the 1999 Federal Committee on Statistical Methodology Conference*.

United Nations Task Team on Satellite Imagery and Geospatial Data. 2017. “Earth Observations for Official Statistics: Satellite Imagery and Geospatial Data Task Team Report.” Available at https://unstats.un.org/bigdata/task-teams/earth-observation/UNGWG_Satellite_Task_Team_Report_WhiteCover.pdf Accessed March 3, 2025.

USDA NASS. n.d. “Cropland Data Layers – FAQs.” Available at: https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php#who Accessed March 3, 2025.

Wikipedia contributors. 2025. “Geohash.” Wikipedia, The Free Encyclopedia. Last modified December 16. 11:22 UTC. Available at: <https://en.wikipedia.org/w/index.php?title=Geohash&oldid=1327842327>.

Zhao, Xia, Limin Wang, Yufei Zhang, Xuming Han, Muhammet Devici, and Milan Parmar. 2024. “A Review of Convolutional Neural Networks in Computer Vision.” *Artificial Intelligence Review*. 57: 99. <https://doi.org/10.1007/s10462-024-10721-6>.