

Administrative Inventory Drift: Audit of "Ghost Improvements" in Cook County

UIC Research: Machine Learning as a One-State Inference for Modernizing Property Tax Baselines

Research Conducted By: Jarrett Williams

University of Illinois Chicago (UIC) College of Urban Planning and Public Affairs

Date: March 6, 2026

Summary Table from Demolitions/Structure removal prior to 2023

Township Name	2024 BOR Cert. Assessed building values	# of improvements	# of parcels
Bremen	\$ 1,918,624.00	234	256
West Chicago	\$ 951,116.00	53	54
Lake	\$ 2,349,810.00	312	317
Hyde Park	\$ 668,895.00	74	74
Thornton	\$ 3,726,470.00	552	583
Worth	\$ 386,498.00	82	83
Calumet	\$ 77,169.00	9	9
Rich	\$ 129,933.00	12	12
Lyons	\$ 67,517.00	4	5
Jefferson	\$ 113,439.00	8	8
South Chicago	\$ 39,257.00	2	3
Bloom	\$ 1,211,823.00	163	166
Proviso	\$ 292,580.00	16	16
Maine	\$ 18,000.00	1	1
Northfield	\$ 147,376.00	5	5
Berwyn	\$ 44,182.00	2	2
Wheeling	\$ 18,976.00	1	1
Riverside	\$ 25,156.00	1	1
Cicero	\$ 115,296.00	6	7
Leyden	\$ 17,981.00	1	1
Niles	\$ 44,774.00	2	2
Hanover	\$ 33,151.00	1	1
Barrington	\$ 56,000.00	1	1
Palos	\$ 27,888.00	1	1
Stickney	\$ 18,374.00	1	1

I. Executive Summary

This study investigates the phenomenon of "**Administrative Inventory Drift**," specifically the persistence of absent structural improvements within the Cook County property assessment rolls. By cross-referencing longitudinal assessment data with historical aerial imagery and tax sale records, this research quantifies the extent of "**Ghost Improvements**": major residential and commercial structures that exist in the administrative record but have been physically removed. Machine-learning models were used exclusively to identify candidate parcels for review; all reported findings reflect manually verified cases. The findings suggest a systemic decoupling of physical property reality from the tax certification process.

II. Methodology and Data Integration

The research employed a multi-vector data reconciliation strategy to identify and verify discrepancies in property characteristics:

- **Public Ortho-imagery:** Cook County 6-inch RGB ortho aerial imagery from 2024 and 2025 was utilized to train machine models.
- **Machine Learning Architecture:** Two Convolutional Neural Network (CNN) models **Swin** and **ResNet** were trained on 60,000 Cook County images depicting both improved and vacant parcels. A high-precision **YOLO** model, trained on 6,705 hand-labeled images, provided a secondary layer for detecting houses and detached garages.
- **Spatial Verification:** Longitudinal analysis of aerial imagery (1999–present) via Google Earth Pro and Vexcel established a "Ground Truth" for structural presence and the estimated year of demolition.
- **Permit Forensics:** A programmatic query of the Cook County Assessor's permit API was conducted using specific demolition identifiers ("wreck," "demo," "demolition").
- **Tax Sale Reconciliation:** An inner-join was performed between the identified "Ghost Improvement" inventory and the Treasurer's and County Clerk's Delinquency datasets for Annual and Scavenger sales.
- All machine-identified parcels were treated as preliminary candidates. No parcel was classified as a confirmed "Ghost Improvement" unless it was manually verified through aerial review and cross-referenced against current assessment records. The final reported figures reflect manually confirmed cases only, which account for prorations and parcel mismatches.

III. Technical Observations and Model Performance

The initial parcel candidates resulted from the convergence of two distinct architectures:

Model	Training Accuracy	Validation Accuracy	Final Validation Loss
ResNet18	99.52%	97.64%	0.1157
Swin-T	99.01%	97.26%	0.1000

While ResNet18 reached a slightly higher peak validation accuracy earlier in the training, the Swin-T model exhibited higher stability and a lower final validation loss, suggesting more consistent generalization across the dataset.

Early iterations were hampered by "noise" from tree canopies and "roof bleeds" from neighboring parcels. To resolve this, the models were retrained using **iterative labeling and active learning**, specifically targeting ambiguous bands and "Foundation Scars" vacant parcels where soil compaction from long-demolished structures provides a false signal for improvements.

IV. Findings and Geographic Distribution

Through automated identification and manual verification, the study isolated **1,544 "Ghost Improvements"** and **6,115 records** featuring unaccounted detached garages.

- **Localized Concentration:** The "Ghost Improvement" phenomenon is not a universal system glitch. It is geographically concentrated within the **Southern Triad of Cook County** and the **Southern Townships of the City Triad**.
- **Northern Control Group:** An exhaustive random scan of 80,000 improved parcels in the Northern and Northwest townships yielded **12** ghost improvements, reinforcing that the disconnect is tied to specific socio-economic and administrative pressures in the south.

V. Administrative and Fiscal Impact

1. The Failure of Reactive Appeal Channels

Of the 1,610 parcels analyzed, only **2.3%** initiated formal appeals post-demolition. This indicates that traditional grievance channels are insufficient for remediating inventory drift in regions with **orphaned properties** and no active steward to engage the process.

2. The "Administrative Mask"

"Ghost Improvements" distort surrounding property values by acting as erroneous comparables in mass-appraisal models. Furthermore, they hide new construction; if a new house is built on a parcel already carrying an old "Improved" record, the system is less likely to trigger reconciliation, leading to a permanent loss of data integrity.

VI. Proposed Remediation

The Assessor's Office does not have a duty of omniscience, and it is reasonable to expect cooperation from partner agencies. To bridge the "**Digital Divide**" in municipal record-keeping, the following proactive steps are proposed:

1. **Predictive "Economic Life" Flagging:** Periodically review records that have surpassed their structural life (e.g., a 500-sq-ft house built in 1889).
2. **Environmental Signal Integration:** Reconcile the master roll with **Cook County Department of Environment and Sustainability (DES)** paper trails, including 10-day demolition notices, Asbestos Abatement Permits, and 3D Ordinance debris reports.

Conclusion: Ultimately, these parcels represent the death of a property's economic life. Implementing an internal, proactive verification framework will ensure that the County's assessment data aligns with physical reality without requiring a new formalized mandate.

This study was initiated following a review of the Cook County Assessor’s Mass Closed permit event that occurred on 02/10/2023. Permits were closed in mass ahead of system rollover to Tyler Technologies IAS System.

While this review was not intended to evaluate institutional intent or overall permit reconciliation rates, it revealed recurring categories of drift that later served as the typological basis for county-wide machine learning models.

The observed discrepancy patterns informed the development of targeted detection logic for:

- Demolished improvements remaining actively assessed
- New construction absent from characteristic updates
- Detached garages not reflected in parcel characteristics

These observations ultimately led to the development of a broader parcel-scale reconciliation framework designed to identify inventory drift at jurisdictional scale.

Through the automated identification and manual verification of these specific Error Types, **6,115 records** featuring unaccounted detached garages and **1,544 “ghost” improvements** were discovered. The coexistence of both overstated improvements (demolished structures remaining on record) and understated improvements (existing garages coded as “0-car”) indicates that the reconciliation lag is structural rather than directional.

To ensure data integrity and avoid false positives, the garage sample was restricted to **single-improvement-PINs** where the primary structure and garage is contained entirely within the parcel boundaries. At the time of review, the official assessment record for each of these 6,115 parcels identified the garage characteristic as **'0-car'**. While it is possible that a garage record could be on an adjacent parcel under common ownership, the scope of this detection was intentionally narrowed to structures contained entirely within the legal boundaries of a single parcel. This constraint was applied to isolate the specific failure of the 'Improvement-to-PIN' reconciliation logic, regardless of the owner’s broader portfolio. The following results mentioned were all manually confirmed.

To identify detached garages, a dual-model detection framework was developed combining a ResNet18 binary classifier with a YOLO object-detection architecture.

The initial ResNet18 model was trained on a binary classification task distinguishing:

- Residential parcels containing only a primary structure (“House”)
- Residential parcels containing both a primary structure and detached garage (“House + Garage”)

Early training data was disproportionately composed of post–World War II Chicago alley-oriented residential layouts, where detached garages are typically positioned at the rear of the

parcel. While this produced strong localized performance, testing revealed significant canonical orientation bias within the model.

For example, identical parcel imagery subjected to rotational transforms produced materially different confidence outputs despite no physical change in parcel characteristics. This indicated the model had partially learned orientation-dependent spatial patterns rather than generalized garage morphology.

To mitigate this bias, the training dataset was expanded geographically to include broader suburban development patterns and randomized rotational augmentation was introduced during training. In parallel, a YOLO object-detection model was developed to independently identify primary structures and detached garages at the spatial level. The YOLO outputs were subsequently used as a secondary validation layer to support or reject lower-confidence ResNet classifications.

At the conclusion of training, the ResNet18 dataset contained 11,093 manually verified samples:

- 7,494 “House + Garage” instances
- 3,599 “House Only” instances

After 12 training epochs, the ResNet18 classifier achieved 89.1% validation accuracy. Although later epochs demonstrated signs of overfitting through divergence between training and validation loss, performance remained highly reliable within elevated confidence bands.

Manual verification of predictions within the 0.90–0.99 confidence interval yielded approximately 99% precision (798 correct identifications out of 800 reviewed samples). Confidence bands between 0.80–0.89 produced an observed accuracy rate of approximately 87%, while lower confidence intervals between 0.50–0.70 demonstrated substantially reduced reliability.

To improve detection consistency within ambiguous confidence ranges, the YOLO detection framework was iteratively refined using 6,705 manually labeled bounding-box annotations for houses and detached garages.

Final YOLO model performance metrics included:

- mAP50: 0.934
- Precision: 0.893
- Recall: 0.891

The convergence of training and validation loss curves indicated strong generalization performance without significant spatial overfitting. When YOLO detections independently confirmed detached garage presence within lower-confidence ResNet prediction bands, observed identification accuracy increased meaningfully relative to the standalone classifier.

Prior to formal model development and following the 2023 permit reconciliation review, a manual spatial audit of Cook County parcels was initiated using QGIS and historical parcel shapefiles from 2024.

Residential improvements were isolated by filtering the “AssessorBL” characteristic field within the parcel dataset. Parcel geometries were then overlaid against high-resolution aerial imagery and longitudinal satellite imagery to visually identify parcels carrying active improvement records despite the apparent absence of extant structures.

To ensure findings reflected the most current administrative record, a Python-based reconciliation workflow was developed to cross-reference identified PINs against multiple real-time open-data APIs. If the api returned the parcel was still carrying a building improvement it would write the parcel to a CSV. This process prevented parcels recently reclassified as vacant from being incorporated into the final inventory.

This initial manual audit identified 370 parcels carrying active improvement value for structures no longer physically present on-site.

Spatial distribution analysis demonstrated that the phenomenon was heavily concentrated within:

- The Southern Triad of Cook County
- Southern townships within the City Triad

Comparative reviews of commercial classifications, northern townships, and northwest suburban regions yielded minimal or no occurrences of this specific discrepancy pattern, suggesting the phenomenon was geographically concentrated rather than system-wide.

To scale the audit beyond manual GIS observation, I initiated a series of Freedom of Information Act (FOIA) requests with municipalities in Cook County’s southern triad. The objective was to secure all electronic demolition permit records from the earliest digital entries to the present to establish where to look.

Municipal Data Retrieval Summary

The FOIA campaign highlighted the following administrative categories regarding demolition data:

- **Absence of Electronic Records:** Several municipalities were unable to provide digital permit data. This included the **City of Hometown, Village of Posen, Glenwood, Robbins, City of Calumet City, Markham, and Harvey**. Notably, **Crestwood** responded to the electronic request with a physical paper note.
- **Successful Electronic Integration:** Digital demolition records were successfully retrieved from **Lansing, Bridgeview, Orland Park, Village of Chicago Ridge, Oak Forest, Evergreen Park, Palos, and Homewood**.

- **Administrative Friction (Withdrawals):** Requests for **Burnham, Blue Island, and Dixmoor** were met with extensions and were subsequently withdrawn due to the administrative hurdles presented by those offices.
- **High-Volume Challenges:** The **Village of Tinley Park** identified the request as "voluminous,".

Due to the extreme fragmentation of municipal demolition records and the inconsistent availability of electronic permit infrastructure across jurisdictions, I determined that municipality-by-municipality reconciliation would be insufficient for conducting a comprehensive review of the assessment roll at county scale.

As a result, the research transitioned from localized administrative record reconciliation toward the development of a generalized parcel-scale detection framework. A dual CNN architecture (Swin and ResNet) was developed to conduct a uniform scan of the southern triad independent of municipal reporting quality or record availability.

The models were trained on a binary classification task: "Is Vacant." For purposes of this study, "Vacant" was defined as a parcel absent of any major structure. Parcels improved solely with gravel, concrete, or asphalt surfaces were classified as vacant.

Through iterative labeling, recursive dataset auditing, and active-learning refinement, the supervised training inventory expanded to more than 23,000 labeled images. During this process, approximately 1,200 ambiguous or low-quality samples were removed following manual review to improve dataset consistency and reduce false-positive classifications.

These refinements substantially improved model performance within low-structure-signal environments, ultimately increasing accuracy within the lowest structure-confidence bands to approximately 96%. The refined review process isolated approximately 700 additional confirmed ghost improvements, increasing the identified inventory to roughly 1,550 parcels.

Additional retraining efforts focused specifically on difficult vacant-parcel scenes that visually retained residual structural signatures despite the absence of extant improvements. Incorporating these edge-case environments into the vacant training class further improved model discrimination and reduced false-positive detections associated with long-demolished parcels.

Following these refinements, the analysis transitioned from 2024 ortho imagery to 2025 imagery. This final review identified an additional 184 parcels, including approximately 50 parcels previously obscured by earlier classification bias. The remaining parcels reflected recent demolitions occurring during 2024 and 2025 rather than long-standing inventory discrepancies.

At the conclusion of training, the supervised dataset had expanded to 69,122 labeled samples, including 32,519 machine-generated negative labels generated through agreement between the Swin, ResNet, and YOLO architectures.

Technical Observations

The Swin-T model exhibited higher stability in validation loss throughout the cycle, with a final loss of **0.1000** compared to ResNet18's **0.1157**. While ResNet18 reached a slightly higher peak validation accuracy earlier in the training, the Swin-T's lower final validation loss suggests a more consistent generalization across the dataset.

YOLO Integration and Geographic Validation

Following the CNN-driven identification, the YOLO model was deployed to the southern triad to scan for structural absence. This secondary layer of detection was utilized to capture any anomalies that may have been filtered out by the ResNet or Swin-T confidence thresholds. This phase identified an additional **10 confirmed cases** of ghost improvements.

To test the geographic universality of these administrative gaps, a random control scan was conducted in the North and Northwest townships. The results of this scan are as follows:

- **Total North/Northwest Samples:** 80,000 parcels.
- **Confirmed Ghost Improvements:** 12.

This near-zero finding in the northern jurisdictions suggests that the "Ghost Improvement" phenomenon is not a byproduct of a global system-wide glitch or a failure in the Assessor's reconciliation methods. Instead, it shows a localized drift concentrated within the Southern Triad and Southern City townships.

Conclusion of Ghost improvement identifications:

1,610 parcels or 1,544 improvements when accounting for prorations, \$12,500,000 2024 BOR certified assessed building values for cases where a building was demolished before 2023. 607 of these parcels or 605 improvements have demolition years which could not be resolved looking as far back as 1999.

22 parcels or 22 improvements where the improvement is recorded on a parcel that intersects the boundaries of public roadways. BOR certified assessed building values equal \$581,106

135 parcels or 131 improvements when accounting for prorations, \$1,206,183 2024 BOR certified assessed building values for cases where a building was demolished after 2024.

39 parcels and 39 improvements that represent false positives identified during review. This includes cases where the house is recorded on the identified vacant parcel, but the house is recorded as vacant and the two parcels are under common ownership. As well as cases where the prorations are incorrect but represent a true existing structure.

Of the 1,610 parcels analyzed, only 2.1% (35 parcels) initiated any form of appeal following demolition. This exceptionally low engagement rate suggests that many of these properties exist in a state of effective administrative abandonment, with no active steward participating in the standard correction process.

- Board of Review: 29 appeals were filed between 1 and 22 years post-demolition.
- Assessor's Office: 6 appeals were identified during the same period.

In several cases, repeated filings did not result in corresponding characteristic reconciliation within the active roll. This suggests that traditional owner-driven correction mechanisms may be insufficient in environments characterized by chronic delinquency, transient ownership, or absent property stewardship.

Notification Inertia and System Rollovers

Programmatic queries of the Assessor's API identified **107 demolition permits** related to the subject parcels. **77 of these permits** were tied to the mentioned 2023 system rollover. While the notifications exist within the broader data environment, they remain unresolved in the characteristic master roll.

Chronic Delinquency and Transient Ownership

Reconciliation with tax sale data reveals that **75% of the identified parcels** were certified for annual or scavenger sales with erroneous assessment data.

Scavenger Sale Persistence: 126 parcels appeared in the scavenger sale, many appearing multiple times over several decades. In most cases, physical demolition occurred *during* the property's long-term delinquency. This suggests a state of **transient ownership**; the properties are trapped in a cycle of failed tax sales where the lack of an active, interested owner ensures the "Improved" status remains unchallenged and unchanged.

The clustering of demolitions in specific corridors, often characterized by mass "**Emergency Wreck and Removal**" permits, indicates that local municipalities are the primary agents of structural removal, filling the void left by absent owners.

Ultimately, these "Ghost Improvements" represent a persistent disconnect between administrative records and physical property conditions. Without a built-in mechanism to flag properties in a persisted state of abandonment, the system continues to generate debt based on non-existent assets. This cycle is not a result of administrative intent, but rather a byproduct of orphaned properties falling outside the reach of traditional, owner-driven or reactive system correction channels.

The persistence of demolished improvements within the assessment roll introduces multiple downstream challenges for valuation accuracy, redevelopment activity, and long-term inventory integrity.

Within mass-appraisal systems, parcel characteristics frequently function as inputs for comparable selection, regression modeling, and broader valuation analysis. When a physically vacant parcel remains classified as improved, the parcel continues to exist within the administrative inventory as a structurally misclassified observation. Over time, these records introduce noise into comparable selection and reduce alignment between the physical environment and the assessment baseline.

The effects of long-term inventory drift extend beyond valuation modeling. Parcels carrying improvement characteristics inconsistent with physical conditions often accumulate tax liabilities disproportionate to their actual utility as vacant land. In chronically delinquent environments, this can contribute to prolonged tax-sale cycling, unstable ownership patterns, and reduced redevelopment activity.

The issue also creates future reconciliation risk when redevelopment eventually occurs. If a new structure is constructed on a parcel already carrying stale “improved” characteristics, the distinction between historical and newly constructed improvements becomes administratively obscured. In these cases, the assessment roll may continue reflecting outdated structural characteristics despite substantial physical change on-site.

Unlike a traditional “vacant-to-improved” transition, where new construction creates a clear binary signal, redevelopment occurring on parcels with preexisting stale improvement records can remain partially hidden within the existing administrative baseline. This increases the difficulty of maintaining long-term inventory accuracy within jurisdictions operating at significant parcel scale.

Collectively, these findings illustrate how prolonged inventory drift can create persistent divergence between assessment records and physical property conditions, particularly within environments characterized by chronic delinquency, fragmented reporting systems, and long-term abandonment.

Proposed Reconciliation Framework

The Kaegi administration has implemented significant modernization efforts aimed at improving data reconciliation and inventory integrity, including partnerships with vendors such as Pushpin

and expanded use of sales and permit analysis workflows. These systems are highly effective at identifying current physical changes, including recent demolitions and new construction activity, thereby reducing the future accumulation of inventory discrepancies.

Supporting these systems is a formal reclassification framework requiring documentation such as demolition permits, affidavits of unimproved real estate, and photographic evidence before a parcel is transitioned to a Class 1-00 designation.

However, the inventory identified in this study exists largely outside the reach of these reactive trigger systems. Many of the identified parcels are characterized by chronic delinquency, absent ownership participation, limited market activity, and prolonged abandonment.

Rather than replacing current systems, these mechanisms would function as an additional reconciliation layer focused specifically on long-standing legacy discrepancies.

Several potential approaches emerged during the course of this research:

1. Predictive Structural-Life Review

Certain structures may exhibit statistically elevated probabilities of long-term inventory drift based on age, size, construction type, or geographic conditions. Periodic review of historically vulnerable improvement classes could assist in identifying parcels requiring additional verification.

2. Integration of Existing Environmental and Demolition Records

Demolition activity already generates multiple administrative records across county and municipal systems, including:

- 10-day demolition notices
- Asbestos abatement permits
- Demolition debris diversion filings
- Emergency demolition and collapse-response documentation

While these records exist for regulatory and environmental compliance purposes, they also represent potential reconciliation signals capable of assisting inventory verification workflows.

Leveraging these existing administrative datasets alongside geospatial review systems may allow jurisdictions to improve long-term alignment between assessment records and physical property conditions without relying exclusively on owner-initiated appeals or market-triggered corrections.

While the statutory responsibility for reporting structural demolition ultimately remains with the property owner, the findings of this study suggest that reactive, owner-driven reconciliation systems alone are insufficient for maintaining long-term inventory integrity within large legacy jurisdictions.

Through manual verification and parcel-scale spatial analysis, this research identified more than 1,600 parcels carrying active improvement characteristics despite the absence of extant primary structures. The geographic concentration of these discrepancies within areas characterized by chronic delinquency, abandonment, and fragmented municipal reporting suggests that long-term inventory drift is not random, but structurally associated with regions where conventional administrative correction mechanisms are least likely to activate.

The study further demonstrates that demolition intelligence already exists within the broader administrative environment through environmental filings, municipal records, permit systems, and geospatial imagery. However, these signals often remain fragmented across disconnected workflows and are not consistently reconciled into the active characteristic roll.

Rather than indicating institutional failure, these findings illustrate the operational difficulty of continuously reconciling physical property conditions across approximately 1.8 million parcels within a predominantly reactive administrative framework.

The integration of proactive spatial verification workflows, environmental reconciliation signals, and targeted inventory-review systems offers a scalable path toward improving long-term alignment between assessment records and physical reality while complementing existing modernization efforts already underway within the County.

Technical appendix

1. Administrative & Master Roll Data

Assessor Parcel Universe (Current Year Only) - [Assessor - Parcel Universe \(Current Year Only\) | Cook County Open Data](#) The definitive source for current parcel classifications

Assessor - Single and Multi-Family Improvement Characteristics - [Assessor - Single and Multi-Family Improvement Characteristics | Cook County Open Data](#). The primary baseline used to identify "Improved" status and "0-car" garage codes

Assessor – Permits - [Assessor - Permits | Cook County Open Data](#)
Used to find permits with various keyword associated with parcels numbers.

Appeals 2 sources

[Assessor - Appeals | Cook County Open Data](#)
[Board of Review Appeal Decision History | Cook County Open Data](#)

Used to identify historical appeals following demolitions.

Parcel shapes

[-Parcels - Historical - 2024 | Cook Central](#)

Used to extract parcel geometry to query image servers and filter parcels on QGIS.

2. Physical Verification & Geospatial Imagery

Cook County 6-inch RGB Ortho-imagery (2024, 2025): The source for training and running all models.

[Cook County Aerial Imagery 2024 | Cook Central](#)

[Cook County Aerial Imagery 2025 | Cook Central](#)

Vexcel: The manual verification of all findings.

Google Earth Pro: Utilized for longitudinal spatial verification (1999–present) to establish structural removal date

3. Tax sale Datasets

County Clerk delinquent dataset https://d16ff5nedhorcz.cloudfront.net/20YrDelinqFile_2025-05-01.zip

Treasurer - Annual Tax Sale - datacatalog.cookcountyil.gov/resource/55ju-2fs9.json

Treasurer – Scavenger Sale - <https://datacatalog.cookcountyil.gov/resource/ydgz-vkrp.json>