



WV Building Footprint Extraction

West Virginia University
WV GIS Technical Center
Compiled by Shannon Maynard
March 2024

Contents

Overview	2
Building Footprint Totals, Comparisons, and Statistics	2
Table 1. Building Footprint Products for West Virginia and Match Rates	2
Table 2. WVGISTC Building Footprint Information by County	
Methods	4
Convert MrSID files to TIFF.....	4
Split county-level TIFF into tiles.....	4
Deep learning.....	4
Post-processing.....	5
Calculate Statistics.....	5
Findings and Observations	6
Conclusions	8
Reference Links	9

Project Highlights

- Generated 2.1 million building footprints from best available leaf-off imagery for all 55 counties using Esri's software deep-learning package.
- The WVGISTC generated building footprint layer had a statewide match rate of 85% compared to the Microsoft and FEMA USA Structure building footprint datasets that had match rates of 82% and 70%, respectively.
- The WVGISTCT building footprints using the Esri algorithm extracted row houses and residential structures better but was not as accurate for large commercial structures. The imagery resolution and imagery quality (e.g., shadows, color balance) are also factors in the building footprint accuracy.
- Building footprints reference layer published to the WV Flood Tool
- Building footprints have multiple uses including:
 - 3D and 2D flood visualizations.
 - Thematic maps specialized to visualize a particular subject or theme.
 - Training data input for land cover classification
 - Depicting destroyed and news buildings before and post disaster.
 - Building counts of structures in floodplain or other geographies

Overview

Using ESRI's USA Building Footprints deep-learning package, over two million building footprints for the entire state of West Virginia were extracted from high-resolution leaf-off, aerial imagery (2018-2023) acquired at the county level. Some of the county imagery was acquired via a cost share from a FEMA HMGP grant of previous years.

The [WV Best Leaf-Off photography](#) statewide coverage is the highest temporal and resolution imagery in the State. West Virginia ranks as the third most forested state in the nation and often has a dense forested canopy that makes identify structures remotely more difficult. Leaf-off imagery provides a reliable source for identifying building footprints.

The resolution of the imagery to create the statewide building mosaic from 55 counties ranged between 3" and 6" resolution, with a majority 4" resolution. Compared to Microsoft's 2018 building footprint dataset, the ESRI output or WVGISTC building footprint version provides a higher average accuracy across many counties across the state, as well as improved footprint attributes, including orientation and building-edge tracing accuracy.

Findings reveal that the ESRI deep learning building extraction algorithm excels at extracting primary residential structures, but does not delineate large commercial buildings (e.g., warehouses, shopping centers) as well as the Microsoft algorithm. However, the ESRI algorithm delineates individual row houses or commercial buildings better than Microsoft, which tends to coalesce highly concentrated, tightly spaced structures into a single footprint polygon. Pixel resolution and imagery quality (e.g., shadows, color balance, sharpness) may be factors in the effectiveness of automated building extractions from high-resolution aerial photography.

Quality verified against primary structures collected from the WV floodplain building inventory dataset, a detailed, visually-checked structure layer. The comparative analysis shows that footprints extracted by the WVGISTC intersected more structures (85% match rate), and thus had higher overall statewide accuracy match, than Microsoft (82%) and FEMA's USA Structures / Oak Ridge Research Lab (70%).

Building Footprint Totals, Comparisons, and Statistics

For West Virginia's 55 counties, 2,121,130 building footprints were extracted using county-level, leaf-off aerial imagery, collected between the years 2018 and 2023, using ESRI's Building Footprint Extraction deep learning model, which resulted in a 108% increase from the 1,020,048 footprints identified by Microsoft in 2018 and a 95% increase from FEMA USA's 1,085,876 total footprints.

Table 1. Building Footprint Products for West Virginia

Building Footprint Product	Dates	Building Footprints	Statewide Match Rate
WVGISTC Building Footprints using ESRI Algorithm	2018-2023	2.1 million	85%
Microsoft Building Footprints	Before 2016	1.0 million	82%
FEMA's USA Structures		1.1 million	70%

The WV floodplain building inventory dataset was used as a quality control check to compare building extraction accuracies. Averaged across all 55 counties, FEMA USA/Oak Ridge National Laboratory returned an accuracy of 70.1%, Microsoft 82.1%, and the WVGISTC 85%. The WVGISTC building footprints had a higher accuracy match rate of 72% (40 of the 55 counties) when compared to Microsoft's 2018 building footprints, and a higher match rate of 95% (52 of 55) when compared to the FEMA USA Structures data.

Table 2. WVGISTC Building Footprint Product Information by County: imagery year, primary structure count, percent intersection with primary structures, and percent intersect comparisons between WV footprints and Microsoft/Oak Ridge.

County	Imagery Year	# Primary Structures	% MS Intersect	% FEMA Intersect	% WV Intersect	WV/MS	WV/FEMA
Barbour	2018	819	68.3%	58.4%	87.5%	19.3%	29.2%
Berkeley	2022	705	75.0%	58.6%	90.9%	15.9%	32.3%
Boone	2020	3,928	77.8%	73.3%	91.9%	14.1%	18.5%
Braxton	2019	824	73.1%	74.2%	90.9%	17.8%	16.7%
Brooke	2022	1,102	95.7%	85.5%	80.5%	-15.2%	-5.0%
Cabell	2021	3,390	86.8%	69.8%	87.9%	1.2%	18.1%
Calhoun	2019	621	69.4%	59.1%	93.1%	23.7%	34.0%
Clay	2019	1,036	53.6%	46.6%	90.0%	36.4%	43.3%
Doddridge	2021	770	72.6%	53.1%	86.9%	14.3%	33.8%
Fayette	2020	1,778	93.5%	61.7%	88.2%	-5.3%	26.5%
Gilmer	2019	516	70.3%	69.0%	86.2%	15.9%	17.2%
Grant	2019	320	83.8%	66.9%	87.8%	4.1%	20.9%
Greenbrier	2020	1,899	82.5%	77.3%	87.8%	5.3%	10.5%
Hampshire	2023	1,157	64.4%	60.2%	77.6%	13.2%	17.4%
Hancock	2019	548	96.4%	56.6%	90.9%	-5.5%	34.3%
Hardy	2018	599	81.6%	58.9%	90.3%	8.7%	31.4%
Harrison	2022	2,118	82.8%	75.6%	88.1%	5.3%	12.5%
Jackson	2019	1,063	83.3%	76.7%	92.4%	9.0%	15.7%
Jefferson	2020	734	96.5%	55.3%	88.1%	-8.3%	32.8%
Kanawha	2022	14,833	87.7%	72.1%	82.9%	-4.8%	10.8%
Lewis	2022	1,071	79.7%	71.8%	90.7%	10.9%	18.9%
Lincoln	2018	2,660	75.1%	66.0%	90.3%	15.3%	24.4%
Logan	2018	5,483	82.2%	75.9%	80.8%	-1.4%	4.9%
Marion	2020	1,708	74.9%	69.5%	89.5%	14.6%	20.0%
Marshall	2022	1,648	84.6%	63.1%	89.6%	5.0%	26.5%
Mason	2018	1,572	76.4%	57.6%	94.0%	17.6%	36.4%
McDowell	2018	3,654	88.9%	72.7%	57.8%	-31.0%	-14.8%
Mercer	2018	2,553	92.7%	75.2%	81.0%	-11.8%	5.7%
Mineral	2019	1,168	94.3%	79.6%	87.3%	-7.0%	7.7%
Mingo	2020	3,660	71.6%	69.4%	89.3%	17.8%	19.9%
Monongalia	2021	1,257	74.7%	59.6%	86.1%	11.4%	26.5%
Monroe	2018	534	80.9%	73.0%	90.6%	9.7%	17.6%
Morgan	2022	645	68.5%	57.8%	75.3%	6.8%	17.5%
Nicholas	2018	1,000	80.2%	66.3%	63.9%	-16.3%	-2.4%
Ohio	2019	3,463	91.2%	78.9%	90.3%	-1.0%	11.3%
Pendleton	2021	472	77.5%	53.4%	89.4%	11.9%	36.0%

County	Imagery Year	# Primary Structures	% MS Intersect	% FEMA Intersect	% WV Intersect	WV/MS	WV/FEMA
Pleasants	2020	446	83.0%	73.3%	85.9%	2.9%	12.6%
Pocahontas	2019	986	75.5%	75.4%	90.8%	15.3%	15.4%
Preston	2020	757	96.0%	69.6%	83.5%	-12.5%	13.9%
Putnam	2021	2,681	83.0%	66.8%	85.5%	2.5%	18.7%
Raleigh	2022	2,468	86.8%	80.3%	82.3%	-4.5%	1.9%
Randolph	2022	1,955	76.3%	71.4%	88.0%	11.7%	16.6%
Ritchie	2023	521	74.1%	67.4%	79.8%	5.8%	12.5%
Roane	2022	1,075	76.1%	52.0%	91.2%	15.1%	39.2%
Summers	2020	966	67.8%	58.3%	83.5%	15.7%	25.3%
Taylor	2020	427	77.8%	67.7%	86.4%	8.7%	18.7%
Tucker	2019	625	78.7%	74.9%	92.3%	13.6%	17.4%
Tyler	2021	838	78.2%	63.5%	90.8%	12.6%	27.3%
Upshur	2022	1,432	68.0%	69.5%	83.4%	15.4%	14.0%
Wayne	2021	2,876	78.9%	66.7%	84.4%	5.5%	17.7%
Webster	2020	1,125	68.8%	65.0%	88.0%	19.2%	23.0%
Wetzel	2021	2,130	85.3%	74.0%	92.8%	7.5%	18.7%
Wirt	2022	526	67.7%	65.8%	83.7%	16.0%	17.9%
Wood	2023	2,261	85.4%	73.2%	84.3%	-1.1%	11.0%
Wyoming	2018	2,726	89.9%	73.6%	79.1%	-10.7%	5.5%

Methods

Convert MrSID compressed files to TIFF

- Use the Export Raster function in ArcGIS Pro and set the Output Format as TIFF. Set the Pixel Type to 8 Bit Unsigned.

Split county-level TIFF into tiles

- To decrease runtime of the deep learning script, use the Split Raster function to split the county TIFF into 625 tiles. Place the tiles into a separate folder from the county imagery and ensure the Output Base Name ends with “_”.
 - “Split Method” = Number of Tiles.
 - “Output Format” as Geotiff (*.tif).
 - “Resampling Technique” as Nearest.
 - “Number of Output Rasters” as X = 25 & Y = 25.
 - Click the dropdown for “Other Options”
 - “Overlap” as 200.
 - “Units for Output...” as Feet

Deep learning

- On a computer equipped for deep learning, download ESRI’s Building Footprint Extraction – USA deep learning package (<https://www.arcgis.com/home/item.html?id=a6857359a1cd44839781a4f113cd5934>) and place

it into a new folder on your local machine titled DLFP. In the script below, replace the third block of highlighted text with the location of the deep learning package.

- In Jupyter Notebook, create a new script window using Python 3. Paste the following script, separating each section into different code blocks so you can see where any potential errors have occurred. Replace the highlighted text with relevant file paths:

```
>import arcpy

>arcpy.env.workspace = r'E:\Counties\Mercer\split2'

>rstLST = arcpy.ListRasters()
rstLST

>for i in rstLST:
    try:
        img = "E:/Counties/Mercer/split2/" + i
        buildOut = img + "_buildings"
        arcpy.ia.DetectObjectsUsingDeepLearning(img, buildOut,
        'C:/Users/smmaynard/Documents/DLFP/usa_building_footprints.dlpk',
        "padding 128;batch_size 4;threshold 0.9;return_bboxes False;tile_size 512", "NMS",
        "Confidence", "Class", 0, "PROCESS_AS_MOSAICKED_IMAGE")
    except AttributeError:
        pass
    print("Completed for " + i)
```

Post-processing

- Once footprints have been extracted for all 625 imagery tiles, use the Merge tool in Arc Pro to merge them into one file.
- If necessary, define the projection (“Define Projection”) to match the county’s original MrSID file.
- Regularize the footprints to remove “stair-step” edges using the Regularize Building Footprints tool. Set the Method to “Right Angles,” the Tolerance to 1, and the Precision to 0.25.
- Occasionally the deep learning algorithm generated footprints on the edge of the county where no imagery exists. For each county, scan of the perimeter of the imagery, selecting and deleting any footprints falling partially outside.
- Project the footprint layer to NAD83 UTM17N.
- For each county, remove overlapping footprints around county borders.

Calculate Statistics

- In Arc Pro, import Microsoft and FEMA USA Oak Ridge footprints as well as the floodplain building inventory layer for the county. Set a definition query to include only primary structures.
- Use Select by Location on the building inventory layer. Set the Selecting Features as the footprint layer you wish to obtain statistics for. Record the number of selected points divided by the total points. This is the intersection rate. Do this for Microsoft, Oak Ridge, and WVGISTC footprints for comparison.

Findings and Observations

Overall, it was found that the use of ESRI's building footprint deep learning algorithm to create the statewide WVGISTC product produced more accurate footprints when compared to both Microsoft and FEMA USA, both in location and building outline delineation, although a slight shift between structures shown on aerial imagery and drawn building footprints is consistently present. This issue may be related to pixel shift that occurs during initial raster processing, and, to a lesser degree, shadows present in aerial imagery.

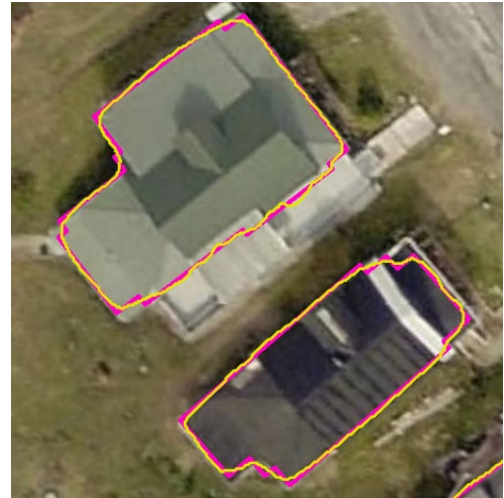


Figure 1. Unregularized (yellow) and regularized (pink) footprints in [Richwood, WV](#).

Processing time. Processing all 55 counties exceeded timeline expectations. Pre-processing for the first county began in March 2022 and post-processing for the last county was completed in January 2024, spanning nearly two years. Pre-processing and deep learning comprised most of the processing time; for each county, time spent preparing and splitting aerial imagery ranged from one to four days, while runtime for the deep learning script could exceed two weeks, in the case of Kanawha County.

Regularization. It was found that regularization of the raw output is necessary to square and reduce “stair-step” edges (Figure 1). In the case of some structures, a second round of regularization was necessary to further square building corners. Overregularization, however, can eliminate natural variation in the shape of the structure and reduce the footprint shape to a single rectangle, potentially eliminating or overcorrecting for portions of the building.

Issues defining large commercial structures. In general, analysis revealed that the ESRI deep learning algorithm did not excel at delineating large commercial structures, such as factories or shopping centers, as shown in Figure 2; commercial structures with dark-colored roofs were often missed entirely, likely mistaken for parking lots or shadows. Imagery quality and pixel size impact the correct delineation of both residential and commercial building footprints and may play a small role in assessing and misdrawing the boundaries of larger structures. Regularization did not seem to affect or influence amalgamations of overlapping “footprints” drawn over and around large commercial structures (see Figure 2).

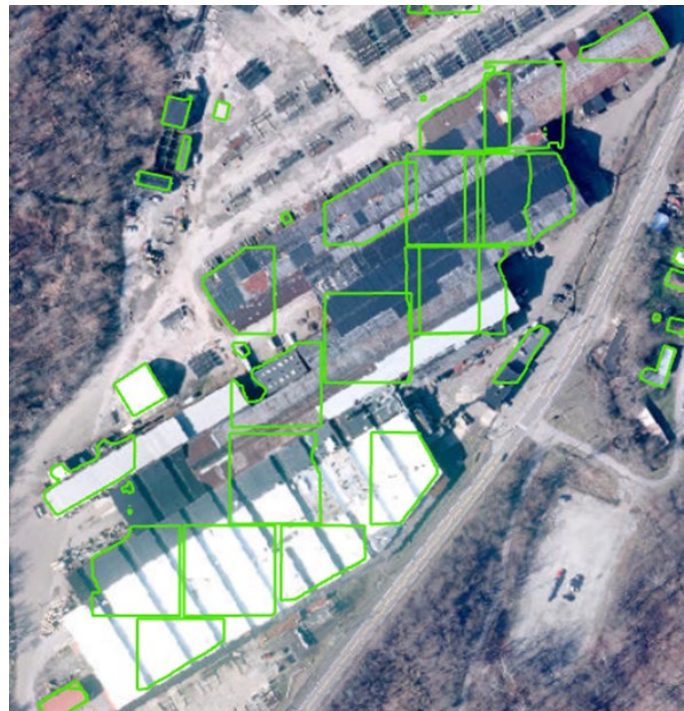


Figure 2. Aerial view of multiple commercial structures in [Brooke County, WV](#) overlain by building footprints created using ESRI's deep learning algorithm.



Figure 3. Footprints delineated by WVGISTC using ESRI's deep learning algorithm (green) and Microsoft (purple); [Wheeling, WV](#).



Figure 4. Destroyed structures (purple) from 2016 flood and new structures (green) post flood. Location [Jordan Creek near Clendenin in Kanawha County, WV](#)

Improved delineation of individual structures. It was noted that ESRI's algorithm more accurately identified individual structures near one another, such as town houses or mobile homes, while Microsoft and FEMA Oak Ridge often consolidate these structures into one larger polygon (Figure 3). Differences in the algorithms and their priorities, advances in deep learning, and variations in aerial imagery vintage and resolution are likely contributors. The ESRI algorithm also identified a greater number of auxiliary structures, such as storage buildings, contributing to the much higher footprint count obtained by the WVGISTC. In addition, the area threshold limit of the other algorithms may have filtered more outbuildings and smaller sized polygons than the WVGISTC algorithm.

Improved locational accuracy. The ESRI algorithm produced more accurate footprint boundaries in comparison to Microsoft, as well as improved spatial accuracy (Figure 4), which majorly contributed to the improved intersection percentage between primary structure points and footprints. This is valuable not only for building identification, but also allows for more accurate structure depictions in the creation of 3D flood visualization models, which rely on building footprints to extrude the general shape of a structure. Building footprints are also useful for creating land cover datasets.

Identification of newly built or destroyed structures. The creation of the new footprint dataset also provides insight into structures built or demolished since the aerial imagery date used in the creation of Microsoft's and FEMA's data, which varies by county. This comparison is particularly valuable, for example, in identifying the many structures that were destroyed in West Virginia's 2016 flood, a disaster that majorly inundated several counties, and structures that were constructed post-flood (Figure 5).



Figure 5. Structures destroyed during or following WV's 2016 flood ([left](#)) present only in Microsoft's dataset, captured using pre-2016 imagery. Structures constructed post-flood, present only in WVGISTC's dataset, created using 2020 aerial imagery ([above](#)); White Sulphur Springs, WV

Conclusions

The use of ESRI's building footprint deep learning algorithm paired with West Virginia County aerial imagery produced a statewide footprint dataset with overall improved spatial accuracy and structure boundary delineation in comparison to Microsoft's and FEMA's datasets. Improvements in deep learning technology and aerial imagery resolution likely facilitated these improvements. Despite the refinement of residential structure footprints, the ESRI algorithm did not excel at the delineation of large commercial structure boundaries; the accuracy of commercial footprints is inconsistent and structures are often poorly or incompletely defined. Decreased accuracy was also noted in highly shadowed areas on aerial imagery and in counties with lower resolution imagery, as expected.

The WVGISTC dataset contains an abundance of auxiliary structures not identified by FEMA or Microsoft, leading to a near doubling of the total number of footprints. The overall accuracy of the footprints created using the ESRI algorithm exceeded Microsoft's and FEMA's statewide accuracy rates; on a county-by-county basis, the WVGISTC obtained a better accuracy rate for 40 WV counties compared to Microsoft and 52 counties compared to FEMA/Oak Ridge. Accuracy rates were determined using primary structures within the floodplain building inventory dataset, which only includes structures within the 1% annual chance floodplain and may not be representative of accuracy rates for the entire county.

In the future, the addition of attributes such as building occupancy type, as is being done by FEMA and Oak Ridge National Laboratory, would be beneficial for hazard mitigation and structural inventory

purposes, a process which would likely require a combination of reference datasets. Occupancy information is not yet available for West Virginia structures in the USA Structures database.

Reference Links

- **WV GIS Data Clearinghouse Data Download:**
 - WVGISTC Building Footprints (2018-23): <https://wvgis.wvu.edu/data/dataset.php?ID=509>
 - Microsoft Building Footprints (2017): <https://wvgis.wvu.edu/data/dataset.php?ID=476>

- **FEMA USA Structures Program:** <https://gis-fema.hub.arcgis.com/pages/usa-structures>

- **Methodologies**
 - **ESRI Building Footprint Extraction Deep Learning Model:**
<https://www.arcgis.com/home/item.html?id=a6857359a1cd44839781a4f113cd5934>
 - **Microsoft Building Footprints:** <https://hub.arcgis.com/datasets/esri::microsoft-building-footprints-features/about>

- **Building Footprint Report:** <https://data.wvgis.wvu.edu/pub/RA/resources/DataDev/Footprints/>

- **WV Flood Tool** (both Microsoft and WVGISTC building footprints can be found under the Reference tab): <https://www.mapwv.gov/flood>