Generating Useful Data with Computer Vision Tools:
2 Use Cases (PDFs & Imagery)

Erik Neemann
10 May 2023

gis.utah.gov/presentations

gis.utah.gov

Overview

- Two Computer Vision (CV) Projects
  - UDOT Parcel Detection Project
  - DHHS Cooling Towers Project
- Motivation
- Process
- Details/Tools
- Results

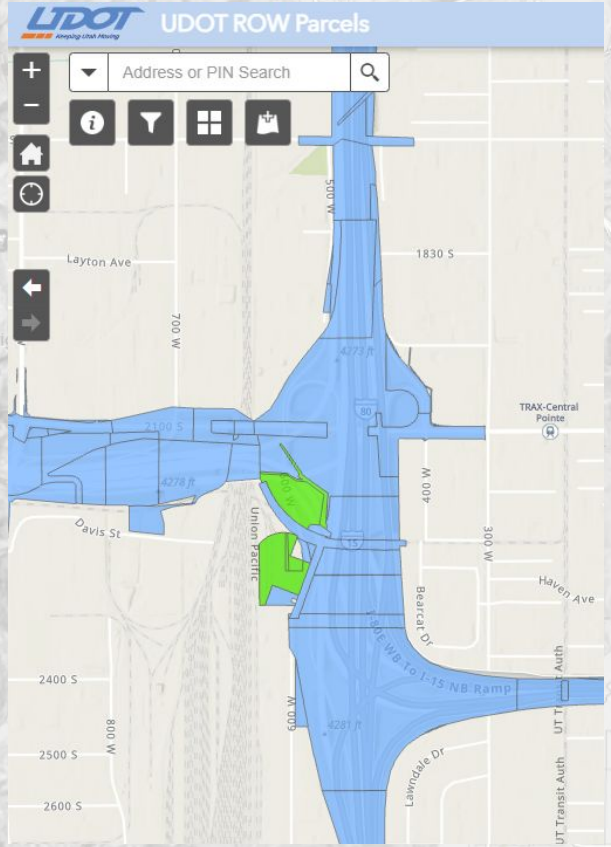
Utah Department of Health & Human Services

Utah Geospatial Resource Center

gis.utah.gov

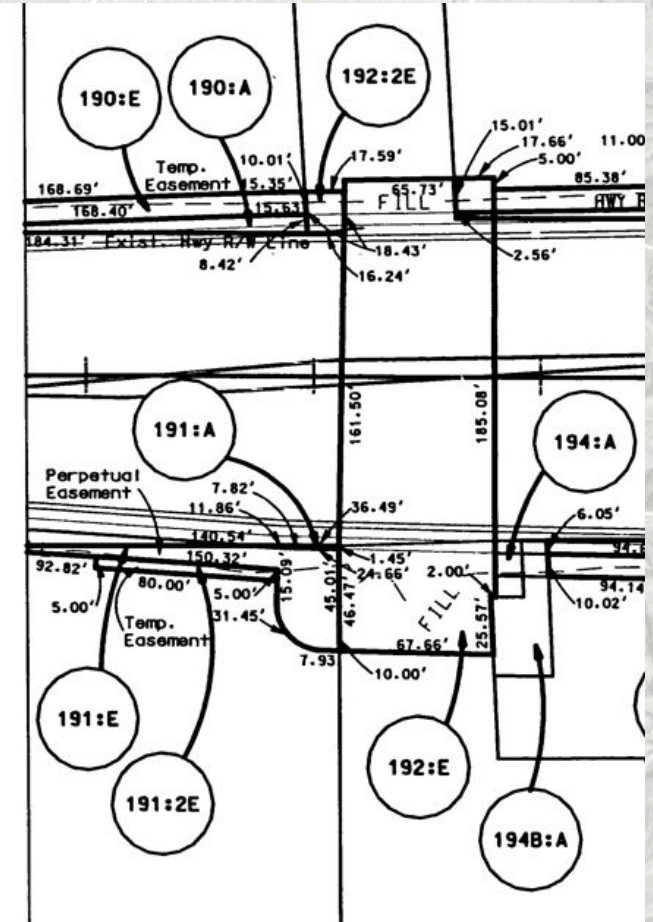
Motivation - UDOT Parcel Detection

- UDOT has acquired a LOT of property over 100+ years
  - ...but they didn't keep track of where that property was located
- UDOT has a tons (90K) of project plan documents
  - …but it's difficult to find a specific parcel within those documents
- How can they untangle where everything is?
  - Manually sift through 90,000 documents?
    - NO!
  - Let the machines do the work!



UDOT ROW Parcels

Motivation - UDOT Parcel Detection

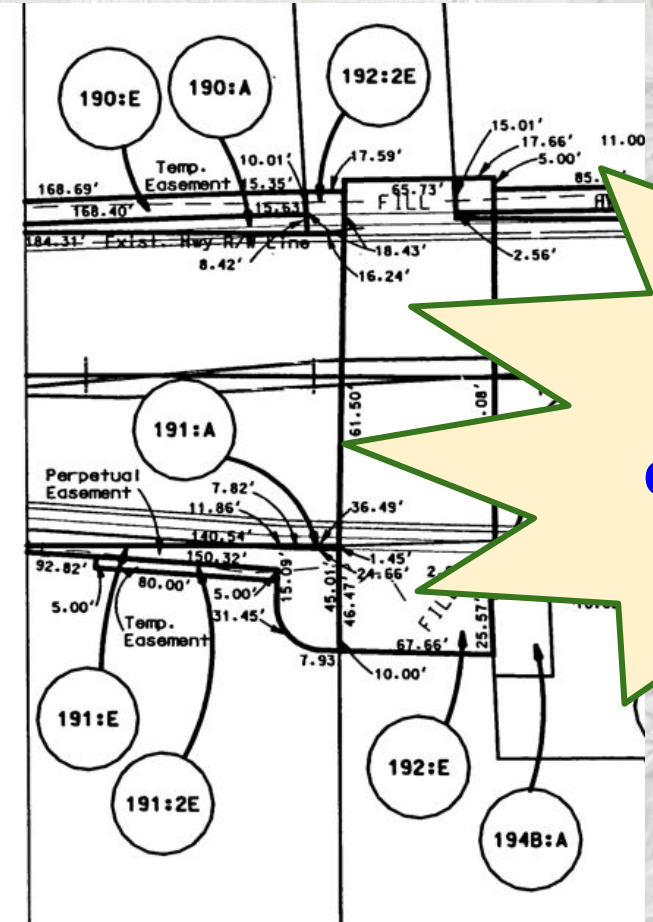


- Parcel information is annotated in a very specific way - circles around the text
- Plan documents fairly consistent in format
- Circles take up roughly the same size in any given document
- Parcel text follows patterns, w/ defined rules
  - numbers/letters, colon, numbers/letters
  - 193B:2A
  - 191:E
  - 41BNT:2E

UDOT Keeping Utah Moving

gis.utah.gov

Motivation - UDOT Parcel Detection

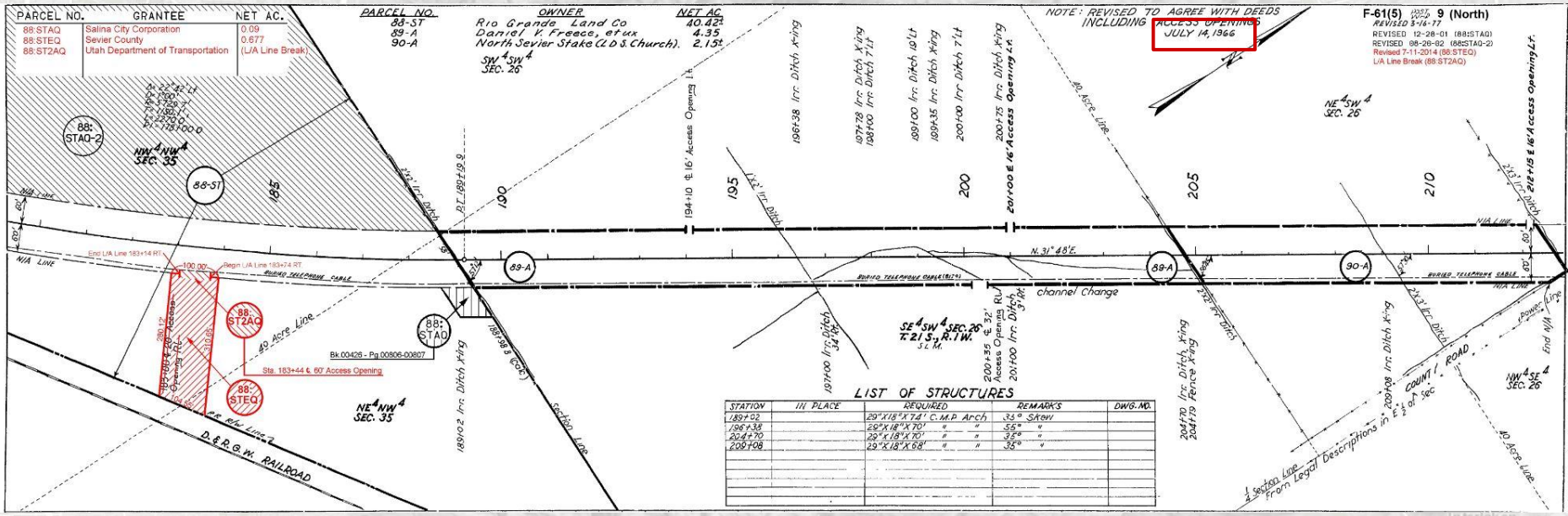


- Parcel information is annotated in a very specific way - [illegible]cles around the text
- Plan[illegible] fairly consistent in format
  - [illegible] size in any
  - [illegible]terns, w/ defined rules
  - [illegible]on, numbers/letters
  - [illegible]A
  - [illegible]91:E
  - 41BNT:2E

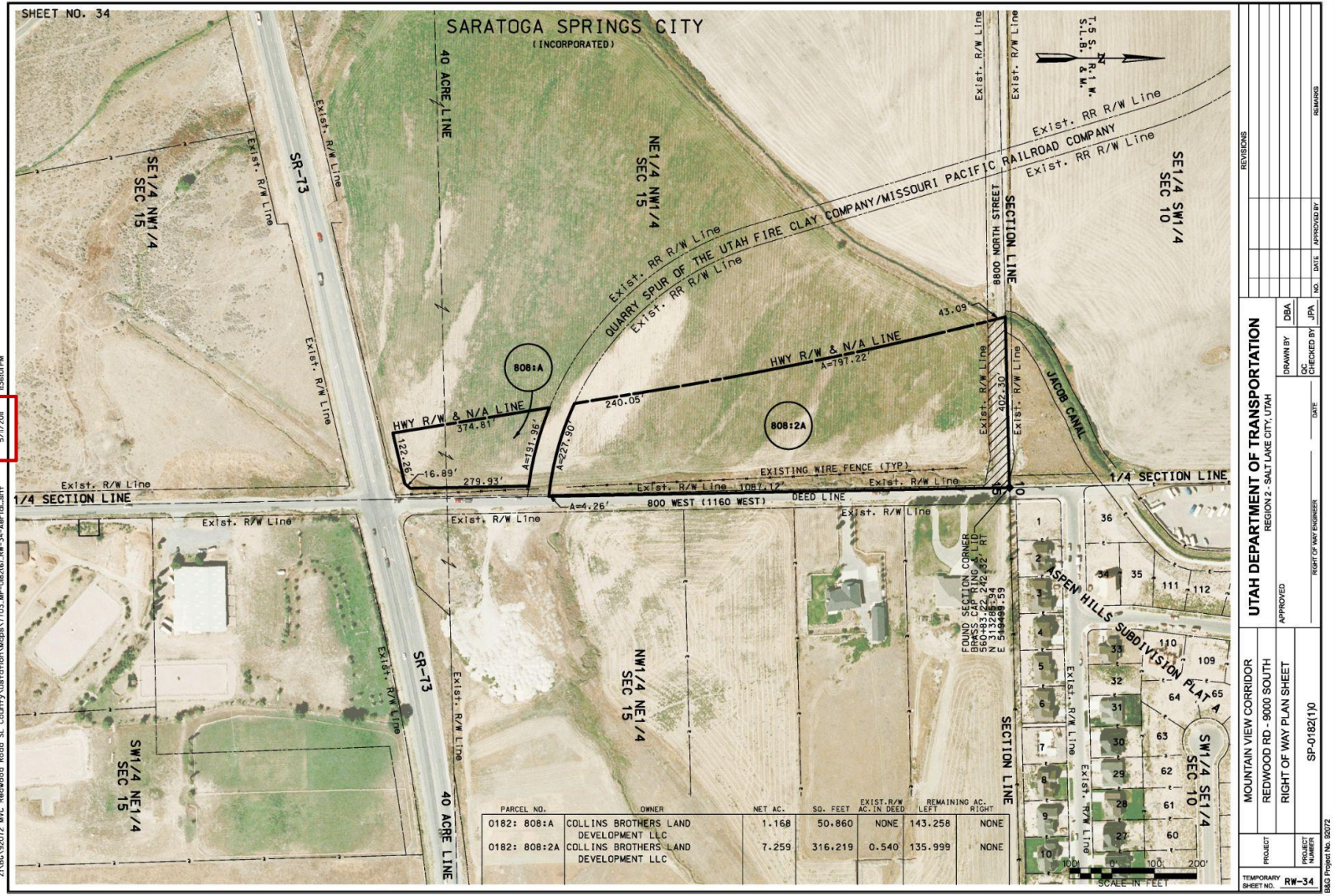
This problem is solvable with computer vision!

gis.utah.gov

Examples



gis.utah.gov



PARCEL NO.	OWNER	NET AC.	SQ. FEET	EXIST.R/W AC. IN DEED	REMAINING AC. LEFT	REMAINING AC. RIGHT
0182: 808:A	COLLINS BROTHERS LAND DEVELOPMENT LLC	1.168	50.860	NONE	143.258	NONE
0182: 808:2A	COLLINS BROTHERS LAND DEVELOPMENT LLC	7.259	316.219	0.540	135.999	NONE

UTAH DEPARTMENT OF TRANSPORTATION
REGION 2 - SALT LAKE CITY, UTAH

MOUNTAIN VIEW CORRIDOR
REDWOOD RD - 9000 SOUTH
RIGHT OF WAY PLAN SHEET
SP-0182(10

APPROVED
RIGHT OF WAY ENGINEER
DATE

DRAWN BY: DBA
CHECKED BY: JPA
QC

REVISIONS
NO. DATE APPROVED BY REMARKS

PROJECT
PROJECT NUMBER

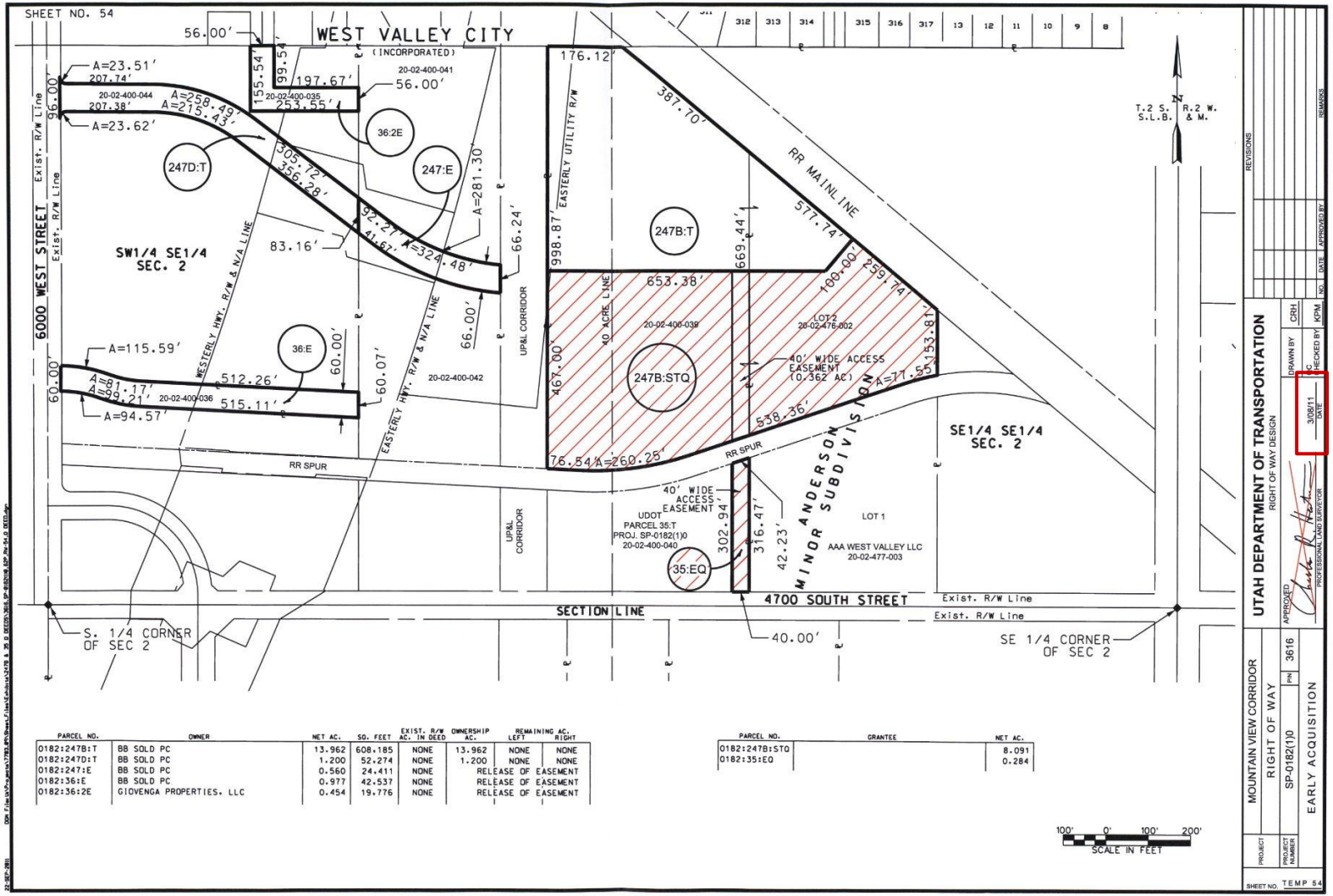
TEMPORARY SHEET NO. RW-34

B&G Project No. 92072

SCALE IN FEET

5/1/2018 8:36:01 PM

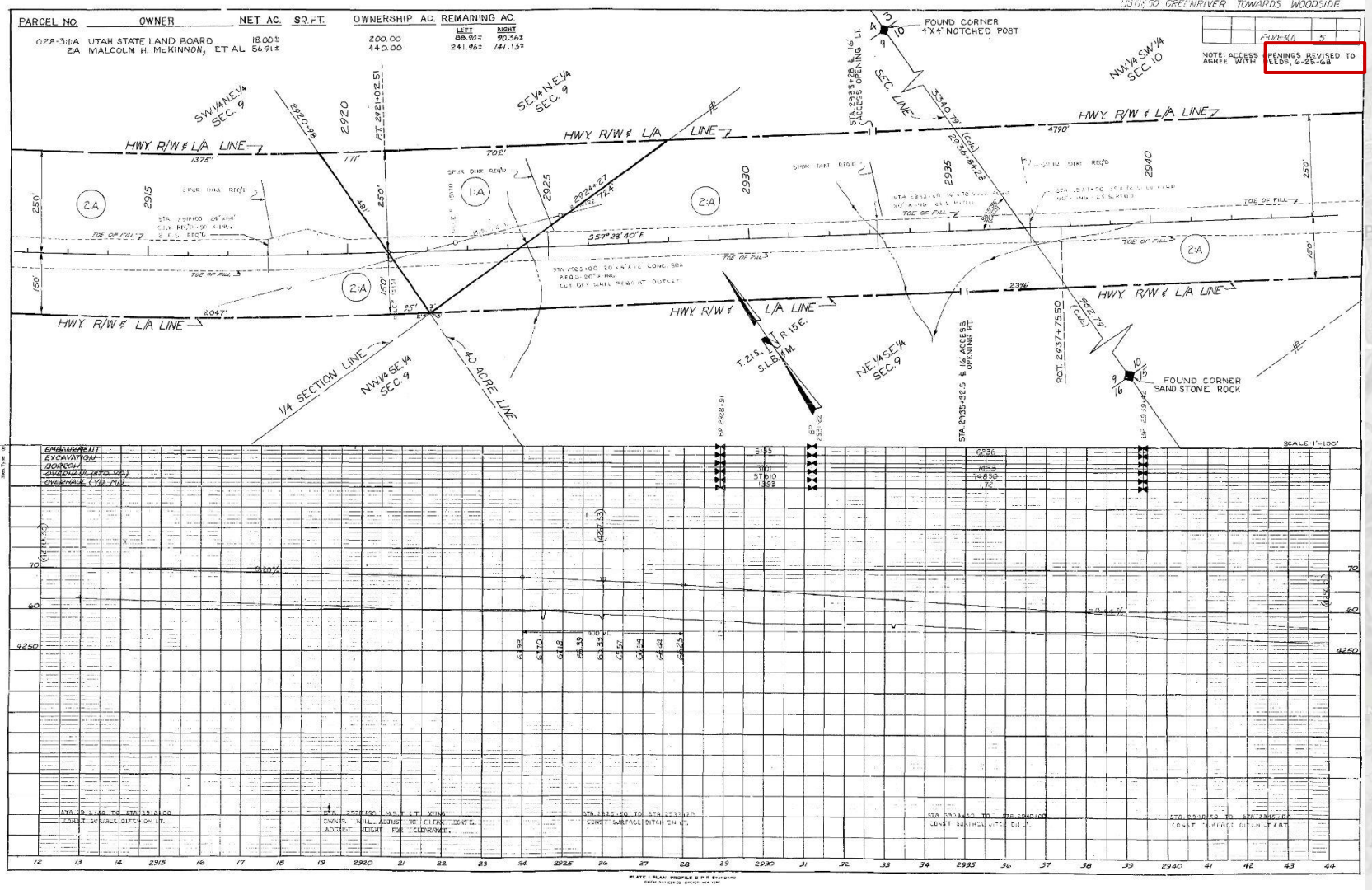
Z:\B&G\92072 MVC Redwood Road SL County\Utah\Maps\1703_MP-0182(6)_RW-34-Aerial.dgn



PARCEL NO.	OWNER	NET AC.	SQ. FEET	EXIST. R/W AC. IN DEED	OWNERSHIP AC.	REMAINING AC. LEFT	REMAINING AC. RIGHT
0182:247B:T	BB SOLD PC	13.962	608.185	NONE	13.962	NONE	NONE
0182:247D:T	BB SOLD PC	1.200	52.274	NONE	1.200	NONE	NONE
0182:247:E	BB SOLD PC	0.560	24.411	NONE	RELEASE OF EASEMENT		
0182:36:E	BB SOLD PC	0.977	42.537	NONE	RELEASE OF EASEMENT		
0182:36:2E	GIOVENGA PROPERTIES, LLC	0.454	19.776	NONE	RELEASE OF EASEMENT		

PARCEL NO.	GRANTEE	NET AC.
0182:247B:STQ		8.091
0182:35:EQ		0.284

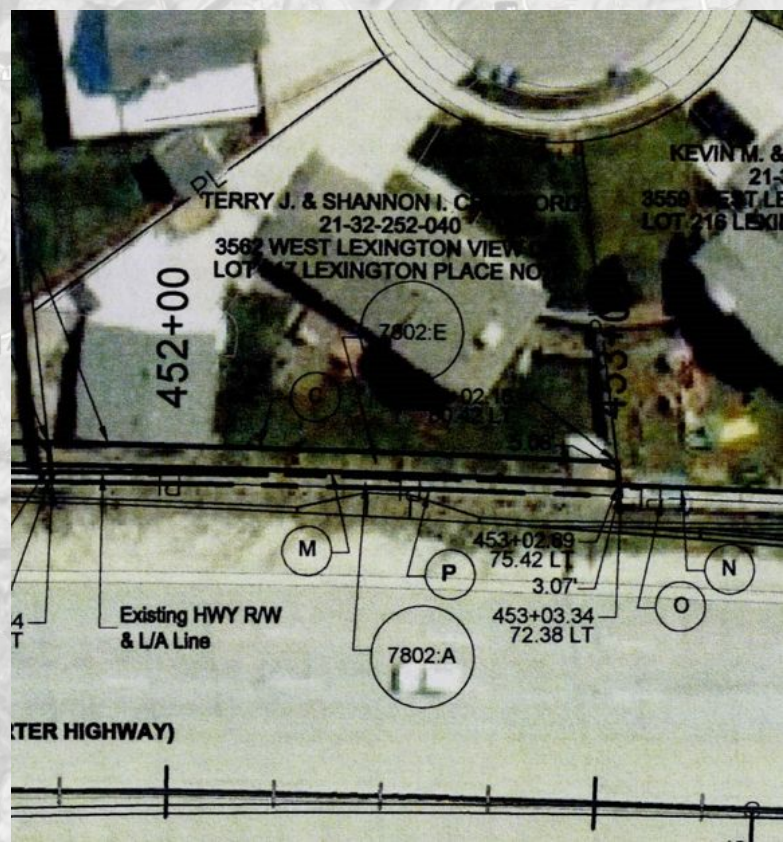
gis.utah.gov


PARCEL NO. OWNER NET AC. SQ.FT. OWNERSHIP AC. REMAINING AC. LEFT RIGHT
028-3:1A UTAH STATE LAND BOARD 18.00± 200.00 88.90± 90.36±
2A MALCOLM H. McKINNON, ET AL 56.91± 440.00 241.96± 141.13±
FOUND CORNER 4"X4" NOTCHED POST
NOTE: ACCESS OPENINGS REVISED TO AGREE WITH FIELDS, 6-25-68
HWY R/W & L/A LINE
SW1/4NE1/4 SEC. 9
SE1/4NE1/4 SEC. 9
NW1/4SW1/4 SEC. 10
SEC. LINE
1/4 SECTION LINE
NW1/4SE1/4 SEC. 9
40 ACRE LINE
NE1/4SE1/4 SEC. 9
T. 2 S., R. 1 E. S.L.B.&M.
FOUND CORNER SAND STONE ROCK
SPUR DIKE REQ'D
TOE OF FILL
SCALE 1"=100'
EMBANKMENT
EXCAVATION
BORROW
OVERHAUL (STA-YD)
OVERHAUL (YD-MI)
PLATE I PLAN-PROFILE & P R Standard


SHEET NO. 11
WOLF POINT CIR.
670 EAST
SANDY CITY
INCORPORATED
SOUTH VALLEY
ESTATES PHASE 1
700 EAST
N 00°29'02" E
SECTION LINE
Monument Line
N 00°18'39" E 2639.92 (Measured)
Exist. ℄ Calc. From Record
S 0°27'39" W 2063.86'
DUSTY
CREEK AVE.
11250 SOUTH
STORM MOUNTAIN
TERRACE NO. 10
STORM MOUNTAIN
TERRACE NO. 5
STORM
MOUNTAIN
TERRACE NO. 5
ALTARA HEIGHTS
SUBD. NO. 1
SANDY DUNES DR.
720 EAST
40 ACRE LINE
SEC. 19
SEC. 24
118+00
119+00
120+00
121+00
122+00
123+00
TEMP. CONST. EASE.
FILL
HWY R/W LINE
Exist. R/W Line
LOT 1
LOT 2
LOT 3
LOT 23
LOT 524
LOT 523
NOTE: STATION AND OFFSET VALUES SHOWN FOR NEW RIGHT OF WAY AND EASEMENT LINES ARE MEASURED FROM PROJECT CONTROL LINES. OFFSET VALUES FOR EXISTING RIGHT OF WAY LINES ARE MEASURED FROM EXISTING MONUMENT LINES OR EXISTING STREET CENTERLINES AS NOTED.
SCALE IN FEET
REVISIONS
UTAH DEPARTMENT OF TRANSPORTATION
REGION 2
SALT LAKE CITY
APPROVED
PROFESSIONAL LAND SURVEYOR
DRAWN BY JDB
CHECKED BY KPM
DATE 9-18-09
700 EAST; 11400 S. TO CARNATION
RIGHT OF WAY
F-0071(23)9
PIN 5865
PLAN SHEET
SHEET NO. RW-11

UDOT Parcels - Challenges

- UDOT completed a bulk export of files from their project software into Google Cloud Storage
  - …but many files are not relevant!
- Various file types (documents, images, other files)
- Images vary in format and file type
- Images vary in size, resolution, DPI
- Images vary in font style (typed, handwritten, different typed fonts)
- Images vary in orientation
- Images vary in quality, noise, consistency



Examples

LEGACY

MILESTONE DESIGN TRANSMITTAL

Legacy Parkway Design Build Project

PLAN PACKAGE (if applicable)

Segment: **1-15** Type: Sequence:

Original DocC Number: 23DCN87

Title: **Design Change Notice**

SUBJECTS **(Check box for each category that defines transmittal contents)** X

Roadway		Drainage		Geotechnical		Pavement		Structures	
MOT		Signals		Lighting	X	Striping		Signs	
ATMS		Utility		Landscape		Aesthetics			
Other:									

REVIEW ITEMS

Use additional sheet to list items

1	Revision 1 to Phase 1 Lighting Plans
2	Check prints
3	DCN Form — NO COMMENTS
4	
5	
6	

IQF FOR AUDIT

1st DocC Extension #: 23DCN87-1

Transmitted By: Dr. K. N. Gunalan, FAK-LLC Design Manager, 360 N. 700 W. Suite F, North Salt Lake, UT 84054

Transmitted To: Mr. John Bale, IQF Manager, 360 N. 700 W. Suite F, North Salt Lake, UT 84054

K. N. Gunalan — Print FAK Designer — Signature — Date: 10/16/03

Non-Audit — Print IDQM — 1022023D-1-NA — Signature — Date: 10/17/03

IDQM sign and return (with comments) to FAK Design Manager

UDOT FOR REVIEW

2nd DocC Extension #: 23DCN87-2

Transmitted By: Dr. K. N. Gunalan, FAK-LLC Design Manager, 360 N. 700 W. Suite F, North Salt Lake, UT 84054

Transmitted To: Mr. Todd Jensen, UDOT Design Manager, 360 N. 700 W. Suite F, North Salt Lake, UT 84054

K. N. Gunalan — Print FAK Designer — Signature — Date: 10-22-03

Bethany Stonington — Print UDOT Oversight — Signature — Date: 11/4/03

UDOT sign and return (with comments) to FAK Design Manager within five (5) days of receipt. UDOT signature demonstrates concurrence with the subject of this transmittal when applied to the specifically stated subject. Signature is not a Release for Construction.

F:\Admin\Forms\Milestone Transmittal.doc

Fluor Ames Kraemer, LLC

360 North 700 West, Ste F
North Salt Lake City, UT 84054

Phone: 801-951-1900
Fax: 801-951-1840

TRANSMITTAL
No. 01892

PROJECT: Legacy Parkway Design-Build

DATE: 5/13/2004

TO: Utah Department of Transportation
360 North 700 West, Ste F
North Salt Lake , UT 84054

REF: Plan Set, Park and Main, final revision, from CRS AFC

ATTN: Rick Campagna

WE ARE SENDING:	SUBMITTED FOR:	ACTION TAKEN:
☐ Shop Drawings	☐ Approval	☐ Approved as Submitted
☐ Letter	☑ Your Use	☐ Approved as Noted
☐ Prints	☑ As Requested	☐ Returned After Loan
☐ Change Order	☐ Review and Comment	☐ Resubmit
☐ Plans		☐ Submit
☐ Samples	**SENT VIA:**	☐ Returned
☐ Specifications	☑ Attached	☐ Returned for Corrections
☐ Other:	☐ Separate Cover Via: hand	☐ Due Date:

ITEM NO.	COPIES	DATE	ITEM NUMBER	REV. NO.	DESCRIPTION	STATUS
001	6	5/13/2004			Plan Set Park & Main, Final Revision from CRS	AAP

Remarks: 58M. 1 lg.

CC: file 220, John Bale, Alan Beane, Doc Cntrl North, Cliff Barber, Lewis Young, Jereme Frank, Bryan Keck, Dan Openshaw

Signed: Deena Farmer

Expedition®

gis.utah.gov

UDOT – STORM WATER POLLUTION PREVENTION PLAN

1. SITE DESCRIPTION

PROJECT LIMITS:
I-15 RP. 282.3, 287.0

PROJECT DESCRIPTION:
LANDSCAPE TREATMENTS AT I-15 / SR-73 AND SR-92 INTERCHANGES.

MAJOR SOIL DISTURBING ACTIVITIES:
(CHECK THE FOLLOWING AS THEY APPLY)
- X CLEAR AND GRUBBING
- __ EXCAVATION
- X GRADING
- __ PLACEMENT OF FILL
- __ CUT AND FILL
- __ OTHERS

TOTAL PROJECT AREA:
2.2 AC.

TOTAL AREA TO BE DISTURBED:
2.2 AC.

WEIGHTED RUNOFF COEFFICIENT (AFTER CONSTRUCTION):
0.40

EXISTING CONDITION OF SOIL & VEGETATIVE COVER AND % OF VEGETATION COVER:
SOIL, NATIVE GRASSES, APPROX 80% VEGETATED

NAME OF RECEIVING WATERS:
UTAH LAKE

2. CONTROLS:

2a. EROSION AND SEDIMENT

- **STABILIZATION PRACTICES:**

	STD. SPEC. NO.
__ TEMPORARY OR PERMANENT SEEDING	02922
__ TURF SODDING	02922
__ PLANTING	02931, 02932
__ MULCHING	02911
__ EROSION CONTROL BLANKETS	02376
__ VEGETATIVE BUFFER STRIPS	
__ PRESERVATION OF TREES	
__ ROUGHENED SURFACE GRADING PRACTICES	02912

- **STRUCTURAL PRACTICES:**

	STD. DWG. NO.	STD. SPEC. NO.	DET. SHT. NO.
__ SILT FENCES	EN 2	01571	
__ STRAW OR HAY BALE BARRIERS	EN 1	01571	
__ TEMPORARY BERMS	EN 3	01571	
__ TEMPORARY SLOPE DRAINS	EN 3	01571	
__ BRUSH BARRIERS			
__ DRAINAGE SWALES/DITCHES		02318	
__ CHANNEL LINERS		02376	
__ RIPRAP		02372	
__ CHECK DAMS	EN 1	01571	
__ SEDIMENT TRAPS	EN 5	01571	
__ TEMP. OR PERMANENT SEDIMENT BASINS		01571	
X PIPE INLET PROTECTION			
X PIPE OUTLET PROTECTION		02373	
__ DROP-INLET BARRIERS	EN 4	01571	
__ CURB INLET BARRIERS	EN 5	01571	
__ CURBS AND GUTTERS			
__ OFF-SITE VEHICLE TRACKING CONTROL			

- WILL CONSTRUCTION OR PLACEMENT OF EROSION CONTROL STRUCTURES IMPACT REGULATED WETLANDS ?
  YES___ NO X

THE PLACEMENT OF EROSION CONTROL STRUCTURES ARE INCLUDED IN THE TOTAL PROJECT WETLAND IMPACTS AND HAVE BEEN PERMITTED THROUGH THE ARMY CORPS OF ENGINEERS.

2b. STORM WATER MANAGEMENT:

STORM WATER RUNOFF DURING AND AFTER CONSTRUCTION BY DITCHES, CHANNELS, AND DROP INLETS INTO CROSS CULVERTS OR STORM DRAIN.

2c. OTHER CONTROLS:

- **WASTE MATERIALS AND DISPOSAL:**
  COMPLY WITH SECTION 00725 ARTICLE "FINAL CLEANUP" SECTION 01455 ARTICLE "FINISHING LOCAL MATERIAL SOURCE SITES", AND SECTION 00820
- **OFFSITE VEHICLE TRACKING AND DUST CONTROL:**
  COMPLY WITH SECTION 01572
- **HAZARDOUS WASTE (INCLUDING SPILL REPORTING):**
  COMPLY WITH SECTION 01355 AND SECTION 00820
- **SANITARY WASTE:**
  COMPLY WITH SECTION 00820

2d. APPROVED STATE OR LOCAL PLANS:

THIS STORM WATER POLLUTION PREVENTION PLAN HAS BEEN DEVELOPED IN ACCORDANCE WITH THE PROVISIONS OF THE MEMORANDUM OF UNDERSTANDING (MOU) BETWEEN THE UDOT AND THE UTAH DEPARTMENT OF ENVIRONMENTAL QUALITY AND THE BEST MANAGEMENT PRACTICE (BMP) PLANS AND HAS BEEN APPROVED BY THE UTAH DIVISION OF WATER QUALITY.

3. MAINTENANCE:

COMPLY WITH SECTION 00725 ARTICLE "FINAL CLEANUP" SECTION 00727 ARTICLE "MAINTAINING THE WORK DURING CONSTRUCTION" AND SECTION 01571 ARTICLE "INSTALLATION"
ACCOMPLISH NECESSARY MAINTENANCE OF CONTROL MEASURES TO PROVIDE CONTINUED EFFECTIVENESS. SCHEDULE AND ACCOMPLISH MAINTENANCE EFFORTS AFTER EACH STORM EVENT AND PRIOR TO ANTICIPATED STORM EVENTS.

4. INSPECTION:

COMPLY WITH SECTION 00727 ARTICLE "INSPECTION OF WORK". INSPECT BY QUALIFIED PERSONNEL, DISTURBED AREAS, STORAGE AREAS, EROSION/SEDIMENT CONTROL MEASURES AND CONSTRUCTION ACCESS LOCATIONS:
- AT LEAST ONCE EVERY FOURTEEN DAYS FOR SITES NOT FINALLY STABILIZED.
- PRIOR TO ANTICIPATED STORM EVENTS THAT COULD RESULT IN SUBSTANTIAL RUNOFF.
- WITHIN 24 HOURS AFTER A STORM THAT RESULTS IN 0.5 INCH OF RUNOFF OR GREATER
- AT LEAST ONCE A MONTH, WHEN RUNOFF IS UNLIKELY OR WHERE SITES HAVE BEEN STABILIZED.

PREPARE A REPORT FOR EACH INSPECTION. INCLUDE THE SCOPE OF THE INSPECTION, DATE, NAMES OF INSPECTORS, MAJOR OBSERVATIONS, FAILED CONTROLS AND ACTIONS TAKEN.

RETAIN COPIES OF INSPECTION REPORTS, THE SWPPP, THE NOI AND OTHER RECORDS FOR 3 YEARS FOLLOWING.

5. NON-STORM WATER DISCHARGES:

COMPLY WITH APPROPRIATE SECTIONS.

6. KEEPING PLANS CURRENT:

AMEND THIS STORM WATER POLLUTION PREVENTION PLAN WHENEVER:
1. THERE IS A CHANGE IN DESIGN, CONSTRUCTION, OPERATION OR MAINTENANCE THAT HAS A SUBSTANTIAL EFFECT ON THE DISCHARGE OF POLLUTANTS TO THE WATERS OF THE STATE.
2. INSPECTIONS OR INVESTIGATIONS BY OFFICIALS INDICATE THAT THE PLAN IS INEFFECTIVE IN ELIMINATING, MINIMIZING OR CONTROLLING THE DISCHARGE OF POLLUTANTS ASSOCIATED WITH CONSTRUCTION ACTIVITY.
3. THERE IS A NEW CONTRACTOR OR SUBCONTRACTOR THAT IMPLEMENTS A MEASURE OF THE STORM WATER POLLUTION PREVENTION PLAN.

GENERAL NOTES:

1. IMPLEMENT STORM WATER POLLUTION PREVENTION PLAN IN ACCORDANCE WITH UDOT'S CURRENT STANDARD SPECIFICATIONS FOR ROAD AND BRIDGE CONSTRUCTION, WITH SPECIAL EMPHASIS ON SECTIONS 01355, 01571, 01574 AND STANDARD DRAWING SHEETS EN 1 - EN 5.
2. COMPLY WITH THE REQUIREMENTS OF THIS STORM WATER POLLUTION PREVENTION PLAN FOR ANY ADDITIONAL CONSTRUCTION ACTIVITIES PERFORMED WITHIN THE PROJECT LIMITS NOT COVERED UNDER THE CONTRACT.

UTAH DEPARTMENT OF TRANSPORTATION
REGION 3, OREM, UTAH
ROADWAY DESIGN

I-15; LEHI INTERCHANGE, LANDSCAPING PHASE II

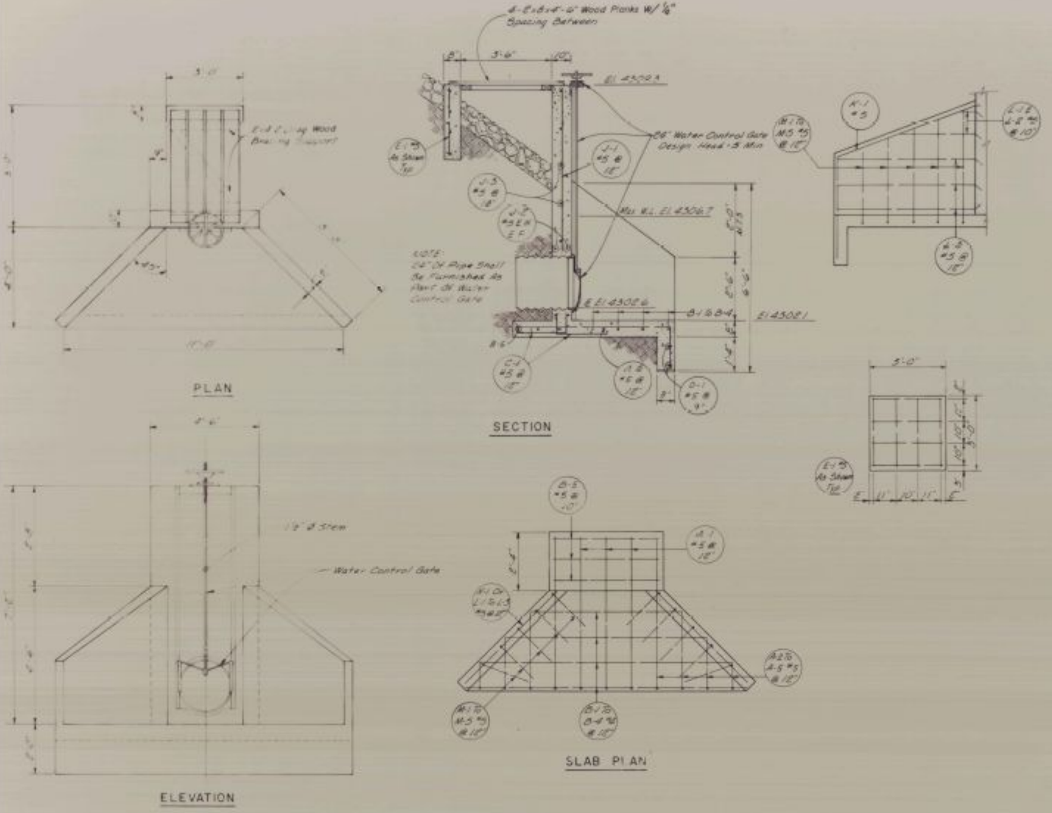
STORM WATER POLLUTION PREVENTION PLAN

PROJECT NUMBER ISTP-15-6(143)280

UTAH COUNTY

SHEET NO. 1E

8/20/2003



INLET STRUCTURE, STATION 765+75 RT.

LOCATION	MARK	SIZE	NO. BARS	LENGTH	TOTAL LENGTH	SKETCH
SLAB	A-1	5	3	7'-6"	59'-6"	a = 5'-10"
	A-2	5	2	7'-6"		a = 5'-10"
	A-3	5	2	4'-8"		a = 3'-0"
	A-4	5	2	3'-8"		a = 2'-0"
	A-5	5	2	2'-8"		a = 1'-0"
SLAB	B-1	5	1	11'-2"	46'-9"	
	B-2	5	1	8'-9"		
	B-3	5	1	7'-8"		
	B-4	5	1	5'-8"		
	B-5	5	3	4'-2"		
SLAB	C-1	5	8	2'-6"	20'-0"	
CUT OFF WALL	D-1	5	2	11'-2"	22'-4"	
Platform Footing	E-1	5	8	2'-8"	21'-4"	
HEAD WALL	J-1	5	6	7'-6"	CUT IN FIELD AS NEEDED	
	J-2	5	6	4'-0"		
MAS WALL	J-3	5	5	4'-2"	90'-4"	
WING WALLS	K-1	5	2	11'-5"	22'-10"	
WING WALLS	L-1	5	2	3'-0"	56'-4"	b = 2'-0"
	L-2	3	2	5'-2"		b = 4'-2"
	L-3	5	6	6'-8"		b = 5'-8"
WING WALLS	M-1	5	2	4'-11"	56'-2"	c = 3'-1"
	M-2	5	2	5'-3"		c = 3'-5"
	M-3	5	2	5'-7"		c = 3'-9"
	M-4	5	2	6'-0"		c = 4'-2"
	M-5	5	2	6'-4"		c = 4'-6"
			TOTAL		395'-8" @ 1.043#/ft = 413#	

ESTIMATED QUANTITIES

Concrete "Class AA" (AE)	3.533	Cu Yds
Reinforcing Steel	413	Lbs
24" Water Control Gate, Type I	1	Each
Timber Untreated	0.030	1,000 Bd Ft

NOTES

All Structural Steel Shall Be Painted In Accordance With The Utah Department Of Highway Standard Specifications, Dated 1970 And Revisions Thereto

NO	BY	DATE	TYPE	REMARKS

REVISIONS

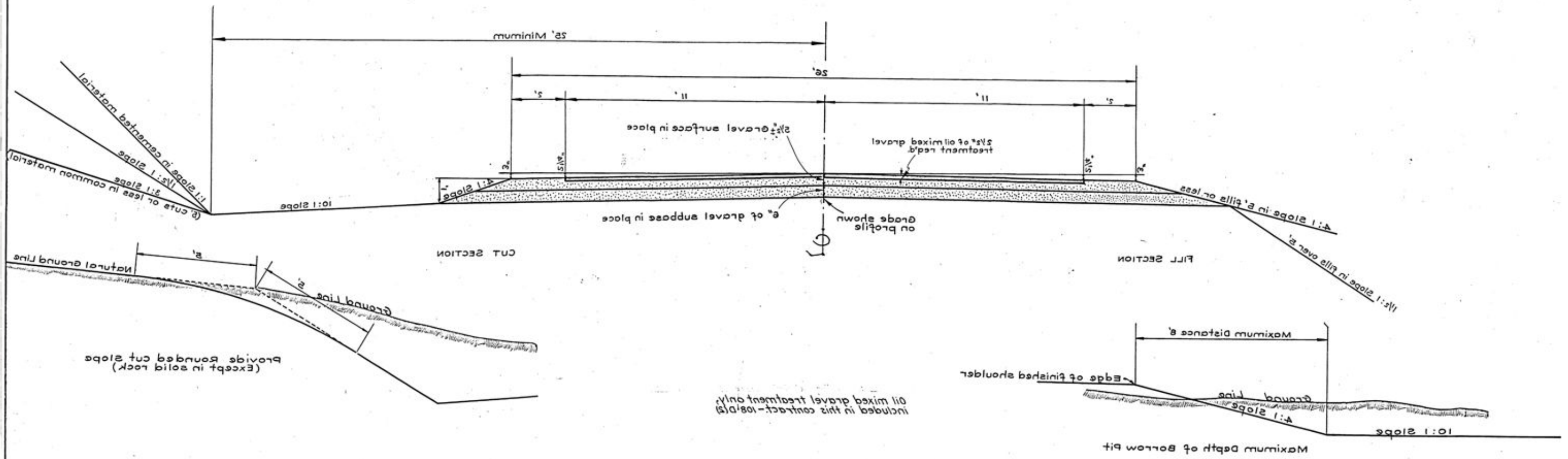
UTAH DEPARTMENT OF TRANSPORTATION
SALT LAKE CITY, UTAH
ROADWAY DESIGN

LAGOON TO LAYTON
IRRIGATION DETAILS

DESIGN C.H.S. 76	CHECK C.S. 76	REVIEW
DRAWN R.E.L. 76	CHECK C.H.S. 76	DESIGN R.L.R. 10-76
QUANT C.H.S. 76	CHECK D.S. 76	R/W
APPROVAL RECOMM 10-1-76 [signature]		DAVIS COUNTY
APPROVED 10/1/76 [signature]		
PROJECT NUMBER I-15-7(21)320		SHEET NO 46

FED. ROAD DIST. NO.	STATE	F.A.P. PROJ. NO.	FISCAL YEAR	SHEET NO.	TOTAL SHEETS
12	UTAH	108-D(1)	1938	2	
"	"	108-D(2)	1940	2	3

TYPICAL CROSS SECTION



Work started on F.A.P. 108-D(2) 6-24-39
" Completed " " " 8-2-39

Revised 6-17-39

UTAH STATE ROAD COMMISSION
SALT LAKE CITY - UTAH
K.C. WRIGHT - CHIEF ENGINEER

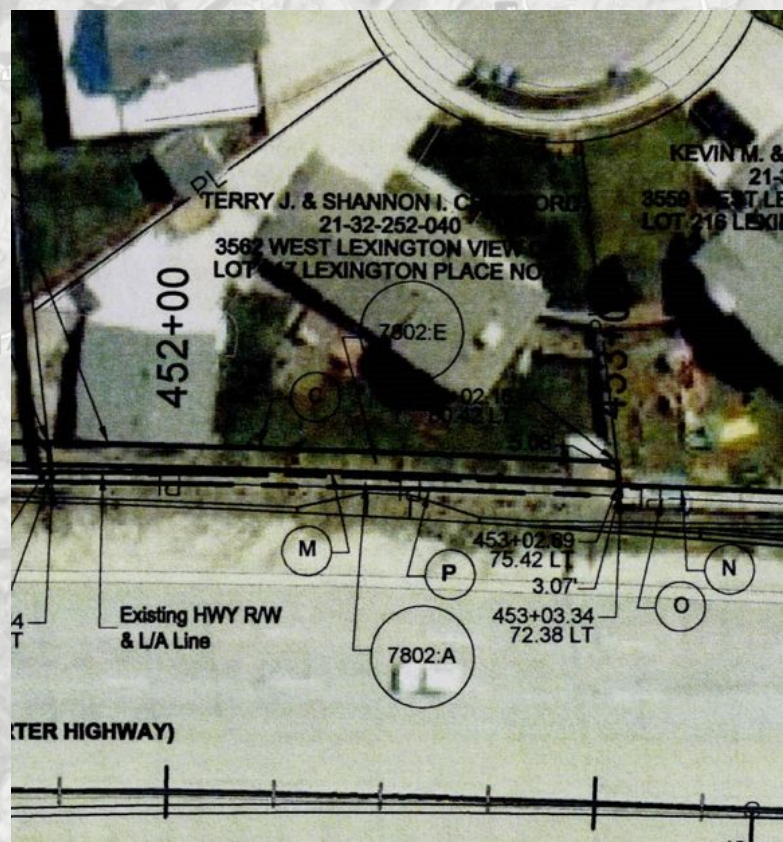
— TYPE —

26' GRAVEL SURFACE ROAD
2½"x 22' OIL MIXED GRAVEL

Widen and superelevate curves according to J-343-R5

UDOT Parcels - Challenges

- UDOT completed a bulk export of files from their project software into Google Cloud Storage
  - …but many files are not relevant!
- Various file types (documents, images, other files)
- Images vary in format and file type
- Images vary in size, resolution, DPI
- Images vary in font style (typed, handwritten, different typed fonts)
- Images vary in orientation
- Images vary in quality, noise, consistency



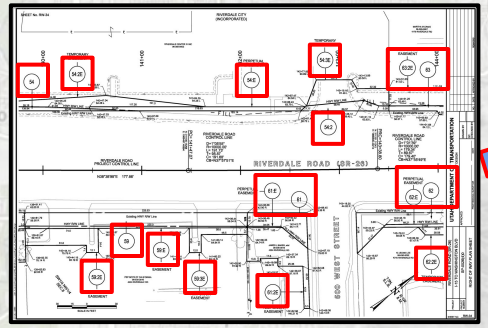
UDOT Parcels - Process Overview

- 1) Initial file processing
- 2) Detect circles and build mosaic images
- 3) Detect text
- 4) Combine results
- 5) Post-process results

JSON

```
  ],
  "text": "59:2E\n62:2E\n54:E\n61\n54:3E\n61:2E\n63\n54\n59\n54:2\n59:E\n61:E\n54:2E\n63:2E\n62\n59:3E\n62:E",
  "url": ""
}
```

3



2

59:2E	62:2E
54:E	63
61	54:2
54:3E	61:E
61:2E	

	54	63:2E
	59	62
	59:E	59:3E
	54:2E	62:E

4/5

	text
/archive1/SP-0026(4)0_RW-33_Plan_Rev-2.TIF	58:E
/archive1/SP-0026(4)0_RW-33_Plan_Rev-2.TIF	58:2E
/archive1/SP-0026(4)0_RW-33_Plan_Rev-2.TIF	54:2E
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	59:2E
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	62:2E
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	54:E
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	61
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	54:3E
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	61:2E
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	63
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	54
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	59
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	54:2
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	59:E
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	61:E
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	54:2E
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	63:2E
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	62
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	59:3E
/archive1/SP-0026(4)0_RW-34_Plan_Rev-1.TIF	62:E
/archive1/SP-0026(4)0_RW-36_Plan_Rev-1.TIF	65:2E
/archive1/SP-0026(4)0_RW-36_Plan_Rev-1.TIF	70:E

Spread sheet

Original File

Mosaic

Utah Geospatial Resource Center

gis.utah.gov

Step 1 - Initial File Processing

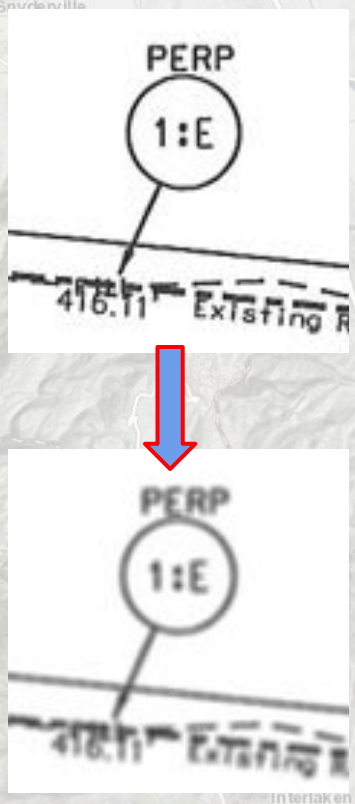
- About 90K objects in cloud storage bucket
- Several files were irrelevant and discarded (.xlsx, .doc, others)
- Many files types to deal with
  - ~46K PDF Documents (.pdf)
  - ~42K images (.tif, .jpg)
- Convert PDFs to images
  - Multipage PDFs to multiple images

Utah Geospatial Resource Center

gis.utah.gov

Step 2 - Detect Circles

- Image preprocessing to improve circle detection
  - Convert to grayscale
  - Add slight blur
- Detect Circles
  - Assume circle radius ~2.5% of image width
    - Iterate through up to 6 values (smaller, bigger, smaller, etc.)
    - Stop when 1-100 circles are found
- Build Mosaics
  - Crop out square around detected circle
  - Mask out area outside circle
  - Inset the mask to remove circle outline
  - Stitch together all cropped squares into a mosaic
- Upload to Google Cloud Storage
- 10,000(!) worker tasks running in parallel in Google Cloud Run
  - 2.5 run of 10K or 25K, each took about 3 minutes = 75,000 minutes of processing
  - 52 days or 7.4 weeks of processing time completed in less than 4 hours!!!!!



Grayscale & Blur

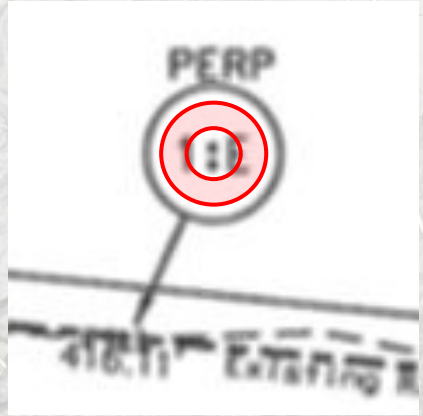
Utah Geospatial Resource Center

gis.utah.gov

Step 2 - Detect Circles

- Image preprocessing to improve circle detection
  - Convert to grayscale
  - Add slight blur
- Detect Circles
  - Assume circle radius ~2.5% of image width
    - Iterate through up to 6 values (smaller, bigger, smaller, etc.)
    - Stop when 1-100 circles are found
- Build Mosaics
  - Crop out square around detected circle
  - Mask out area outside circle
  - Inset the mask to remove circle outline
  - Stitch together all cropped squares into a mosaic
- Upload to Google Cloud Storage
- 10,000(!) worker tasks running in parallel in Google Cloud Run
  - 2.5 run of 10K or 25K, each took about 3 minutes = 75,000 minutes of processing
  - 52 days or 7.4 weeks of processing time completed in less than 4 hours!!!!!

radius too small



Iterate Through Radius Sizes

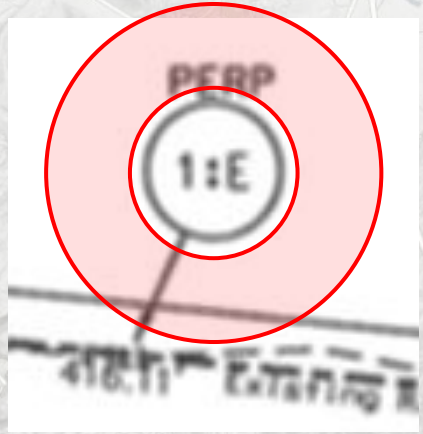
Utah Geospatial Resource Center

gis.utah.gov

Step 2 - Detect Circles

- Image preprocessing to improve circle detection
  - Convert to grayscale
  - Add slight blur
- Detect Circles
  - Assume circle radius ~2.5% of image width
    - Iterate through up to 6 values (smaller, bigger, smaller, etc.)
    - Stop when 1-100 circles are found
- Build Mosaics
  - Crop out square around detected circle
  - Mask out area outside circle
  - Inset the mask to remove circle outline
  - Stitch together all cropped squares into a mosaic
- Upload to Google Cloud Storage
- 10,000(!) worker tasks running in parallel in Google Cloud Run
  - 2.5 run of 10K or 25K, each took about 3 minutes = 75,000 minutes of processing
  - 52 days or 7.4 weeks of processing time completed in less than 4 hours!!!!!

radius too big



Iterate Through Radius Sizes

Utah Geospatial Resource Center

gis.utah.gov

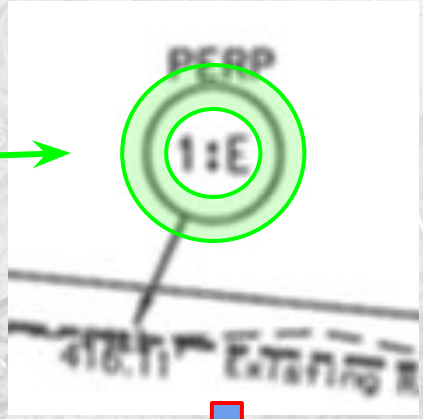
Step 2 - Detect Circles

- Image preprocessing to improve circle detection
  - Convert to g
  - Add slight b
- Detect Circles
  - Assume cir
    - Iterate
    - Stop w
- Build Mosaics
  - Crop out sq
  - Mask out ar
  - Inset the ma
  - Stitch toget
- Upload to Google Cloud Storage
- 10,000(!) worker tasks running in parallel in Google Cloud Run
  - 2.5 run of 10K or 25K, each took about 3 minutes = 75,000 minutes of processing
  - 52 days or 7.4 weeks of processing time completed in less than 4 hours!!!!!

Goldilocks Zone

r, etc.)

radius just right



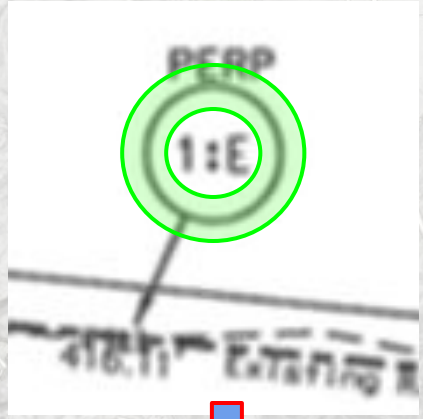
Crop & Mask
(original image)

Utah Geospatial Resource Center

gis.utah.gov

Step 2 - Detect Circles

- Image preprocessing to improve circle detection
  - Convert to grayscale
  - Add slight blur
- Detect Circles
  - Assume circle radius ~2.5% of image width
    - Iterate through up to 6 values (smaller, bigger, smaller, etc.)
    - Stop when 1-100 circles are found
- Build Mosaics
  - Crop out square around detected circle
  - Mask out area outside circle
  - Inset the mask to remove circle outline
  - Stitch together all cropped squares into a mosaic
- Upload to Google Cloud Storage
- 10,000(!) worker tasks running in parallel in Google Cloud Run
  - 2.5 run of 10K or 25K, each took about 3 minutes = 75,000 minutes of processing
  - 52 days or 7.4 weeks of processing time completed in less than 4 hours!!!!!

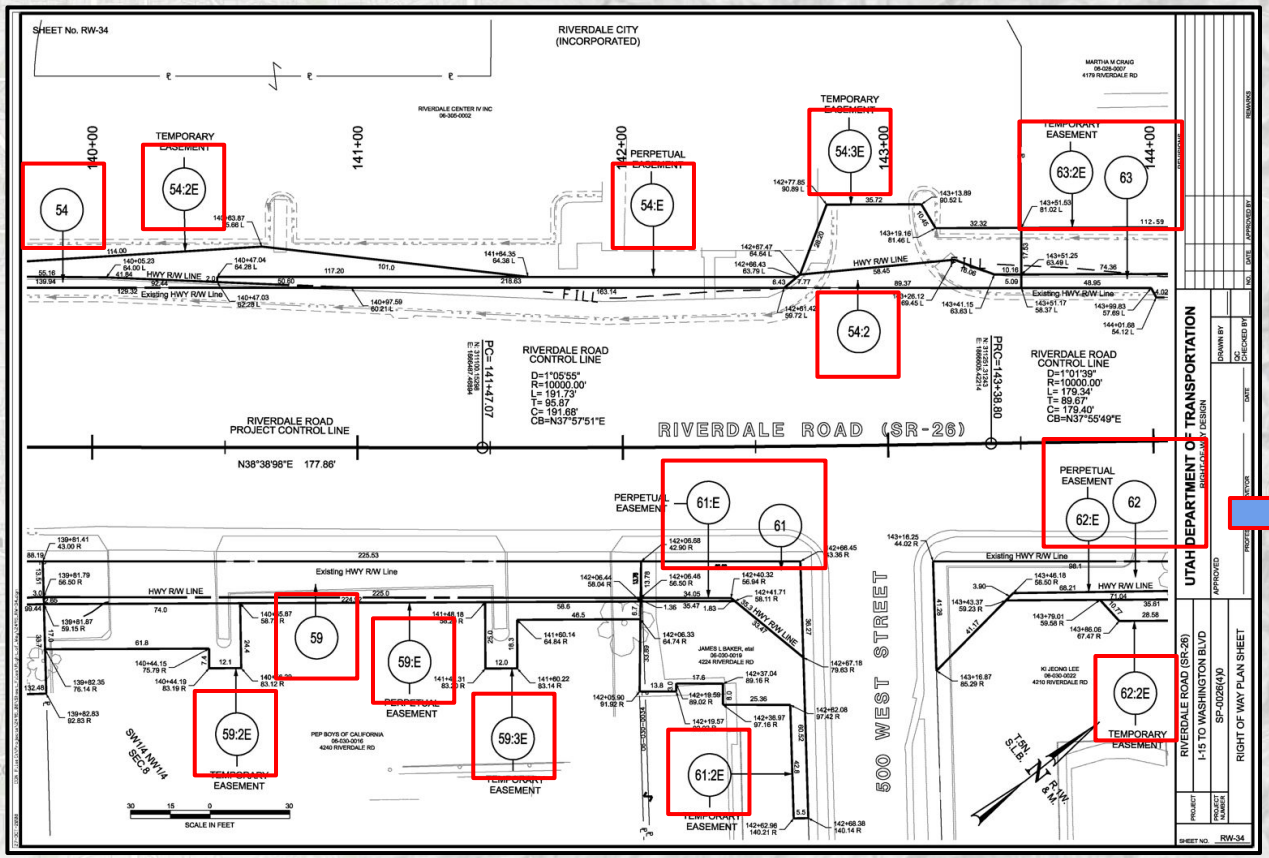


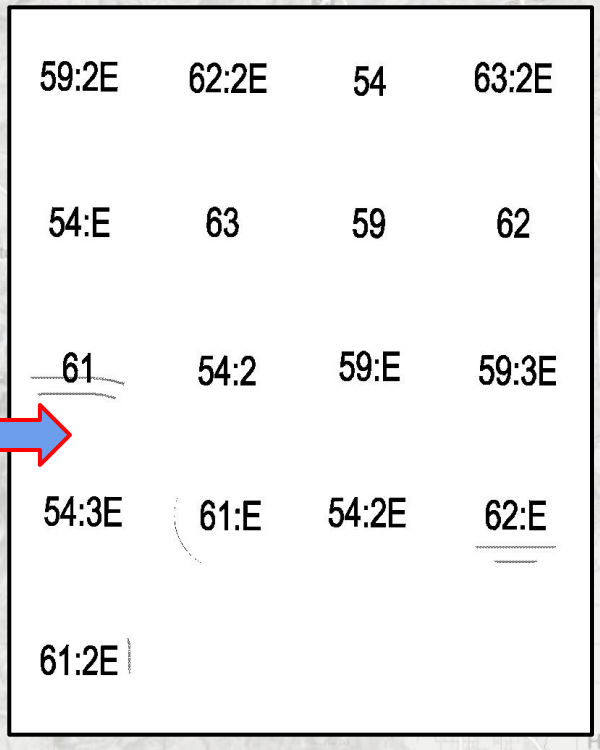
Crop & Mask
(original image)

Utah Geospatial Resource Center

gis.utah.gov

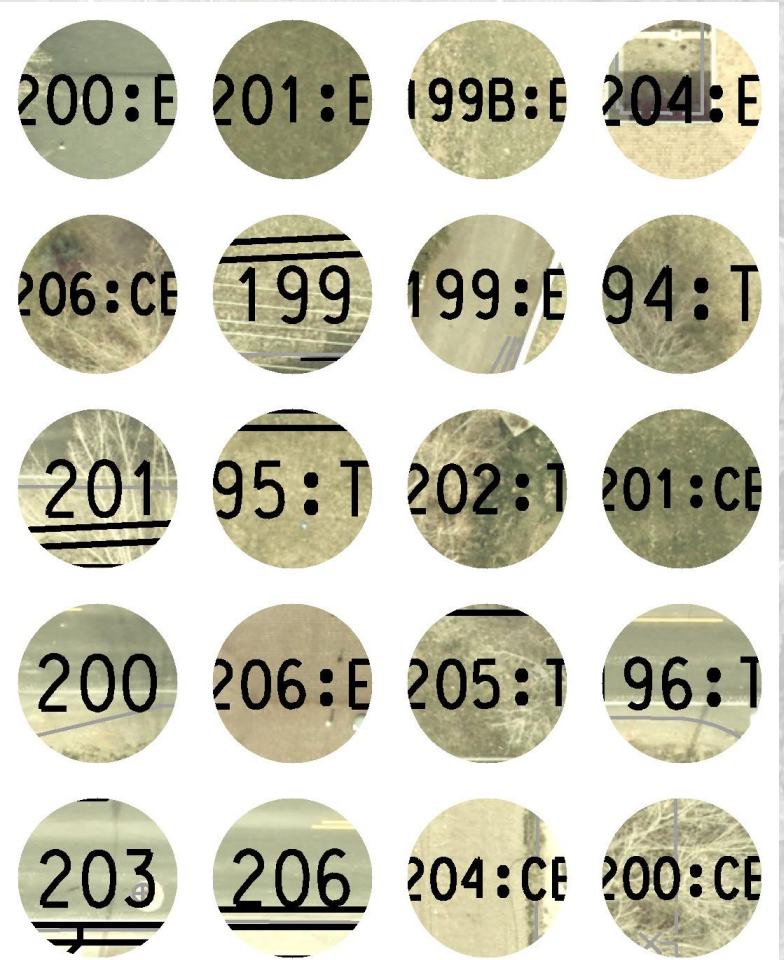
Step 2 - Detect Circles





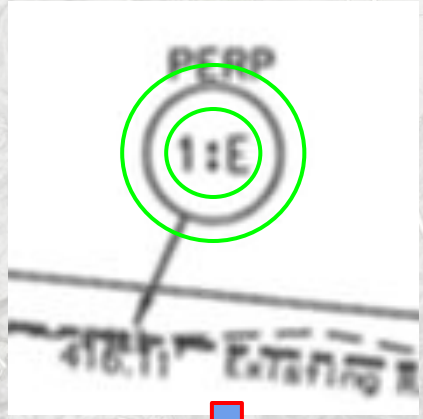
Build Mosaic

Build Mosaic - Extreme Example



Step 2 - Detect Circles

- Image preprocessing to improve circle detection
  - Convert to grayscale
  - Add slight blur
- Detect Circles
  - Assume circle radius ~2.5% of image width
    - Iterate through up to 6 values (smaller, bigger, smaller, etc.)
    - Stop when 1-100 circles are found
- Build Mosaics
  - Crop out square around detected circle
  - Mask out area outside circle
  - Inset the mask to remove circle outline
  - Stitch together all cropped squares into a mosaic
- Upload to Google Cloud Storage
- 10,000(!) worker tasks running in parallel in Google Cloud Run
  - 2.5 run of 10K or 25K, each took about 3 minutes = 75,000 minutes of processing
  - 52 days or 7.4 weeks of processing time completed in less than 4 hours!!!!!



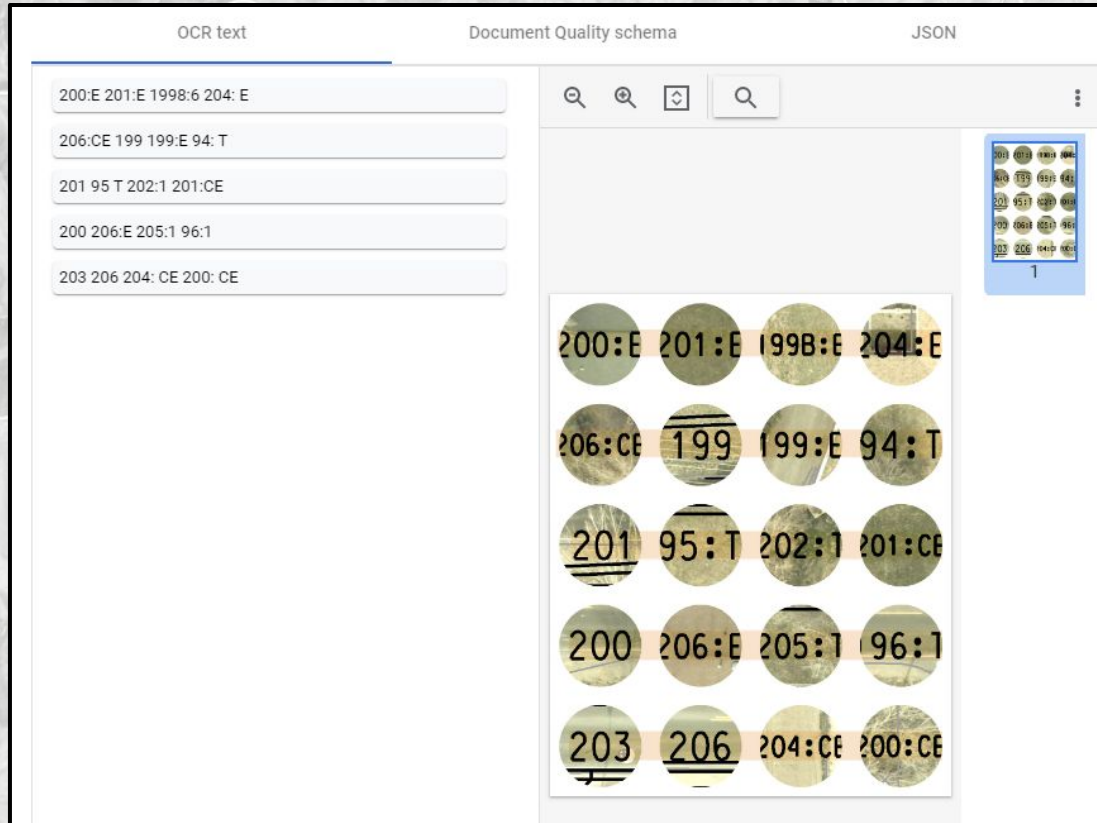
Crop & Mask
(original image)

Utah Geospatial Resource Center

gis.utah.gov

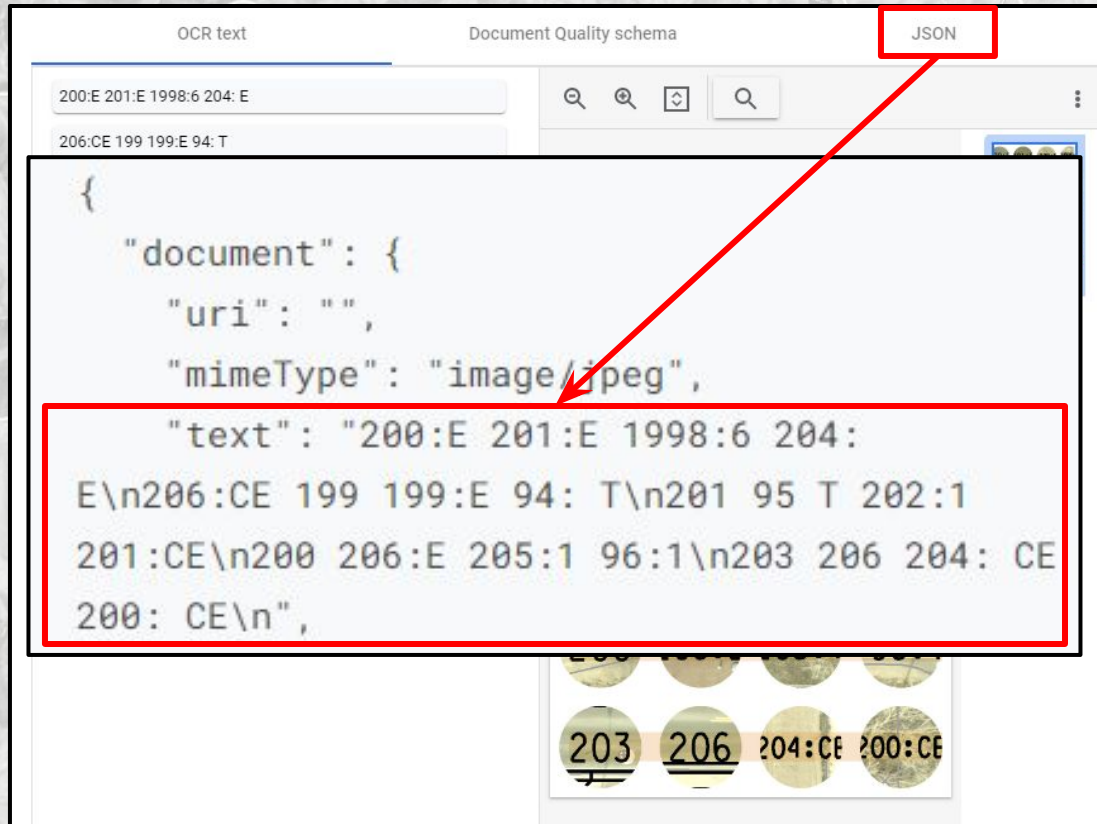
Step 3 - Detect Text

- Google Document AI tool used to perform Optical Character Recognition (OCR) on each mosaic file
  - API extracts text and other metrics, returns a JSON file with results
- Basic cleanup on JSON text string
  - Remove newlines, whitespace, empty text results
- Insert results into a dataframe, with filename
- Save dataframe as a CSV



Step 3 - Detect Text

- Google Document AI tool used to perform Optical Character Recognition (OCR) on each mosaic file
  - API extracts text and other metrics, returns a JSON file with results
- Basic cleanup on JSON text string
  - Remove newlines, whitespace, empty text results
- Insert results into a dataframe, with filename
- Save dataframe as a CSV



Step 4 - Combine Results

- Combine all result dataframes into a single dataframe (concatenate)
- Join additional fields from UDOT spreadsheets on filename field
  - project_number
  - project_name
  - guid from project management system (ProjectWise)
- Build URLs to ProjectWise, Cloud Storage files
- Explode text into multiple rows

	F	G	H	I	J	K	L	M
1	udot_file_name	project_number	project_name	guid	projectwise_url	udot_url	mosaic_url	text
2	archive1/SP-0089(88)313_RW-15_Plan_Rev-1.TIF	SP-0089(88)313	SR-89, State Street,	{c88d59d5-f1a6-442e-a6b3-0bdebbdf3c88}	https://connect-projec	https://storage.	https://storag	21:4
3	archive1/SP-0089(88)313_RW-15_Plan_Rev-1.TIF	SP-0089(88)313	SR-89, State Street,	{c88d59d5-f1a6-442e-a6b3-0bdebbdf3c88}	https://connect-pro	https://storage.	https://storag	24:E
4	archive1/SP-0089(88)313_RW-15_Plan_Rev-1.TIF	SP-0089(88)313	SR-89, State Street,	{c88d59d5-f1a6-442e-a6b3-0bdebbdf3c88}	https://connect-pro	https://storage.	https://storag	25
5	archive1/SP-0089(88)313_RW-15_Plan_Rev-1.TIF	SP-0089(88)313	SR-89, State Street,	{c88d59d5-f1a6-442e-a6b3-0bdebbdf3c88}	https://connect-pro	https://storage.	https://storag	24
6	archive1/SP-0089(88)313_RW-15_Plan_Rev-1.TIF	SP-0089(88)313	SR-89, State Street,	{c88d59d5-f1a6-442e-a6b3-0bdebbdf3c88}	https://connect-pro	https://storage.	https://storag	25:3E
7	archive1/SP-0089(88)313_RW-15_Plan_Rev-1.TIF	SP-0089(88)313	SR-89, State Street,	{c88d59d5-f1a6-442e-a6b3-0bdebbdf3c88}	https://connect-pro	https://storage.	https://storag	25:E
8	archive1/SP-0089(88)313_RW-15_Plan_Rev-1.TIF	SP-0089(88)313	SR-89, State Street,	{c88d59d5-f1a6-442e-a6b3-0bdebbdf3c88}	https://connect-pro	https://storage.	https://storag	24:3E
9	archive1/SP-0089(88)313_RW-15_Plan_Rev-2.TIF	SP-0089(88)313	SR-89, State Street,	{7f8fab8f-947e-4d06-aa4a-71f1505c43dd}	https://connect-proj	https://storage.	https://storag	25:E
10	archive1/SP-0089(88)313_RW-15_Plan_Rev-2.TIF	SP-0089(88)313	SR-89, State Street,	{7f8fab8f-947e-4d06-aa4a-71f1505c43dd}	https://connect-proj	https://storage.	https://storag	24:3E
11	archive1/SP-0089(88)313_RW-15_Plan_Rev-2.TIF	SP-0089(88)313	SR-89, State Street,	{7f8fab8f-947e-4d06-aa4a-71f1505c43dd}	https://connect-proj	https://storage.	https://storag	24:E
12	archive1/SP-0089(88)313_RW-15_Plan_Rev-2.TIF	SP-0089(88)313	SR-89, State Street,	{7f8fab8f-947e-4d06-aa4a-71f1505c43dd}	https://connect-proj	https://storage.	https://storag	25
13	archive1/SP-0089(88)313_RW-15_Plan_Rev-2.TIF	SP-0089(88)313	SR-89, State Street,	{7f8fab8f-947e-4d06-aa4a-71f1505c43dd}	https://connect-proj	https://storage.	https://storag	25:3E
14	archive1/SP-0089(88)313_RW-15_Plan_Rev-2.TIF	SP-0089(88)313	SR-89, State Street,	{7f8fab8f-947e-4d06-aa4a-71f1505c43dd}	https://connect-proj	https://storage.	https://storag	24
15	archive1/SP-0089(88)313_RW-15_Plan_Rev-2.TIF	SP-0089(88)313	SR-89, State Street,	{7f8fab8f-947e-4d06-aa4a-71f1505c43dd}	https://connect-proj	https://storage.	https://storag	21:4
16	archive1/SP-0089(88)313_RW-15_Plan_Rev-3.TIF	SP-0089(88)313	SR-89, State Street,	{eb9fc3b1-1d03-432f-800c-a39e0411b310}	https://connect-pro	https://storage.	https://storag	25

Step 5 - Post-process Results

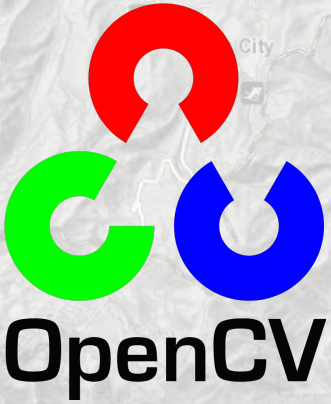
- Perform additional data cleanup
  - Upper-case all letters
  - Remove punctuation (except for colons)
  - Remove character accents
- Apply filtering rules to remove "invalid parcels", flag results:
  - with special characters
  - starting with a letter or non-digit
  - with ':P' pattern
  - with a colon, if a number is not present before the colon
  - longer than 13 characters
  - with 4 or more letters, if no colon is present
  - with 5 or more numbers in a row

- Remove duplicate rows
- Export final results into "good", "bad", and "all" spreadsheets

UDOT Project Tools

- OpenCV
  - Image manipulation
  - Circle detection
  - Cropping
  - Mosaicking
- Optical Character Recognition (OCR)
  - PyTesseract (original choice)
  - Google DocumentAI (better results)
- Pandas
  - Tabular data manipulation
  - Joins
  - String cleanup
  - Filtering rules

Google Cloud



pandas

Utah Geospatial Resource Center

github.com/agrc/udot-parcel-ml/

gis.utah.gov

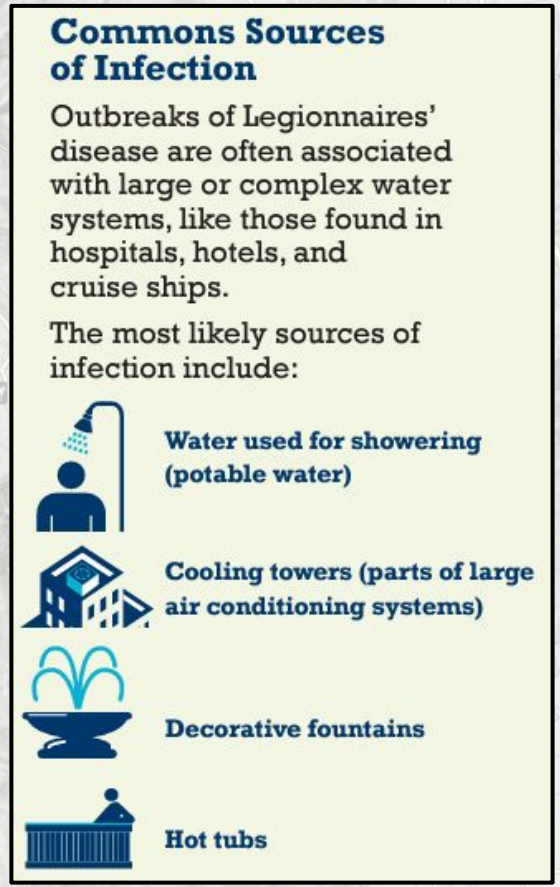
UDOT Results

- Over 240,000 "good" parcels were extracted from the documents
- Data Accuracy
  - Reviewed 50 documents to compare CV results to human results
  - Two "outlier" PDFs with 44 and 48 pages - might skew results
  - Circle detection
    - Including outliers: CV detected 1039/1244 circles (83.52%)
    - Excluding outliers: CV detected 662/698 circles (94.84%)
  - Text comparison on 897 valid parcels
    - Average edit distance: 0.229
    - Average correct letter percentage (truth_len - edit_dist)/truth_len: 95.15%
    - Number of results that were perfect: 878 (88.96%)
    - Number of results with edit distance <= 1: 921 (93.31%)

Motivation - DHHS Cooling Towers

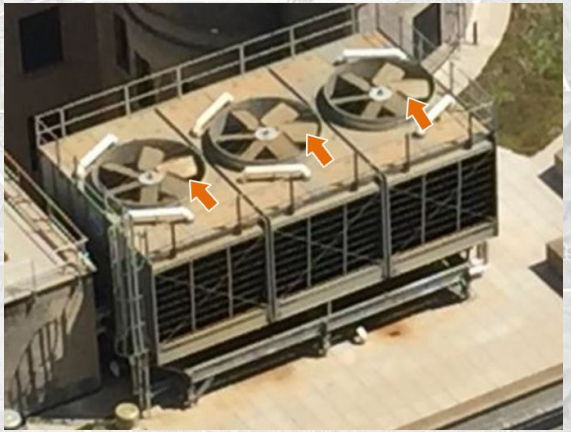
- Legionella bacteria can cause a serious type of pneumonia called Legionnaires' disease
- Legionella can grow/spread in large building water systems
  - Water tanks, HVAC, large/complex plumbing systems, cooling towers
- Cooling towers are concerning because they can release aerosolized water into the atmosphere
  - If Legionella is present, the aerosolized water can spread the bacteria over miles*

*CDC - Controlling Legionella in Cooling Towers



CDC - Legionnaires' Fact Sheet

Motivation - DHHS Cooling Towers



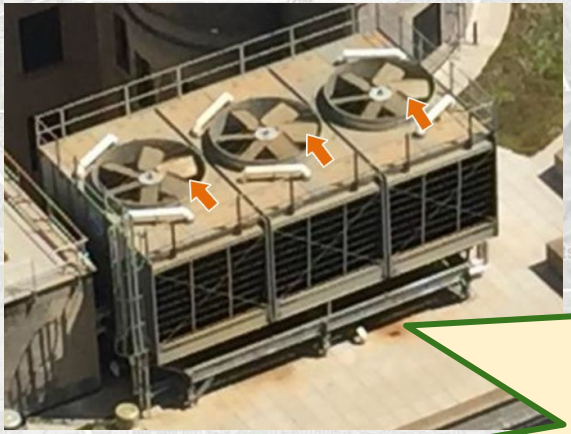
CDC - Photos of Cooling Towers

- Cooling towers can cause outbreaks of Legionnaires' disease when they are not adequately maintained
- They are often investigated & located using aerial imagery during Legionnaires' outbreaks
- Cooling towers have distinctive features that make them identifiable
- Researchers and the CDC have used object-detection models to identify potential cooling towers in aerial imagery (TowerScout)

Utah Department of Health & Human Services

CDC - Procedures for Identifying Cooling Towers

Motivation - DHHS Cooling Towers



- Cooling towers can cause outbreaks of Legionnaires' disease when they are not adequately maintained
- [illegible]ted & located using [illegible]gionnaires' outbreaks [illegible]nctive features that [illegible]
- [illegible]and the CDC have used [illegible]bject-detection models to identify potential cooling towers in aerial imagery

This problem is solvable with computer vision!

CDC - Photos of Cooling Towers

Utah Department of Health & Human Services

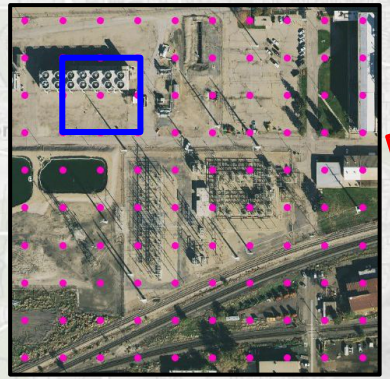
CDC - Procedures for Identifying Cooling Towers

DHHS Cooling Towers - Process Overview

- 1) Build Index & Footprint
- 2) Download Images
- 3) Detect & Locate Towers
- 4) Post-process results
- 5) Build Web Map

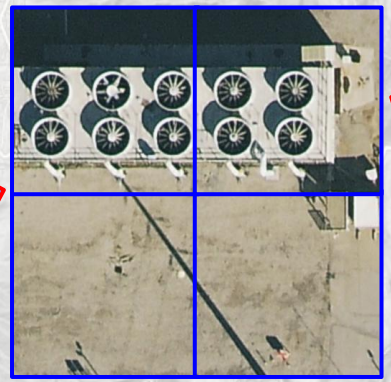
Utah Department of Health & Human Services

1



Build Index

2



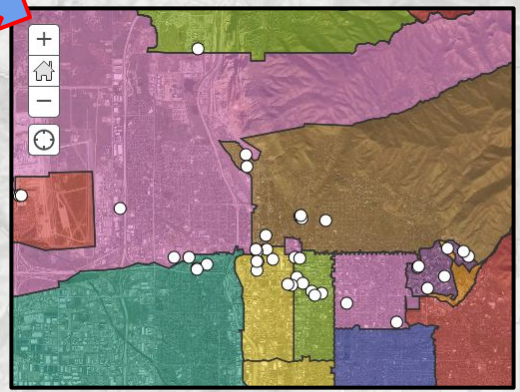
Download Images

3

Detect & Locate Towers

4/5

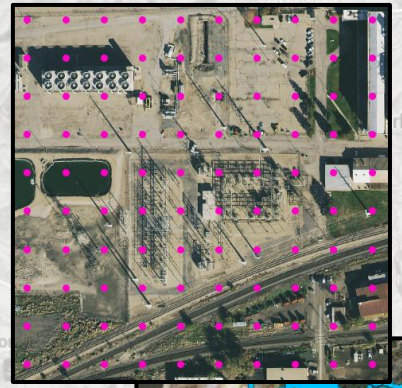
Build Web Map



gis.utah.gov

Step 1 - Build Imagery Index & Footprint

- Discover Web Map Tile Services (WMTS) used for statewide imagery
- Must build imagery index at highest zoom level (20)
- Create point for upper left corner of each WMTS tile (row, col)
  - Calculate latitude/longitude based on row/col values
  - 275,000,000 tiles cover the state of Utah!
- Create processing footprint to select a subset of all image tiles
  - Buffer census places by 800m
  - Buffer large buildings (>5k sq ft) by 800m
- Select WMTS indices of tiles in processing footprint
  - Load tile index and footprint into Google BigQuery
  - Run Spatial SQL query to select tiles to process (~6%)
  - (data is waaaayyy too big for desktop GIS)

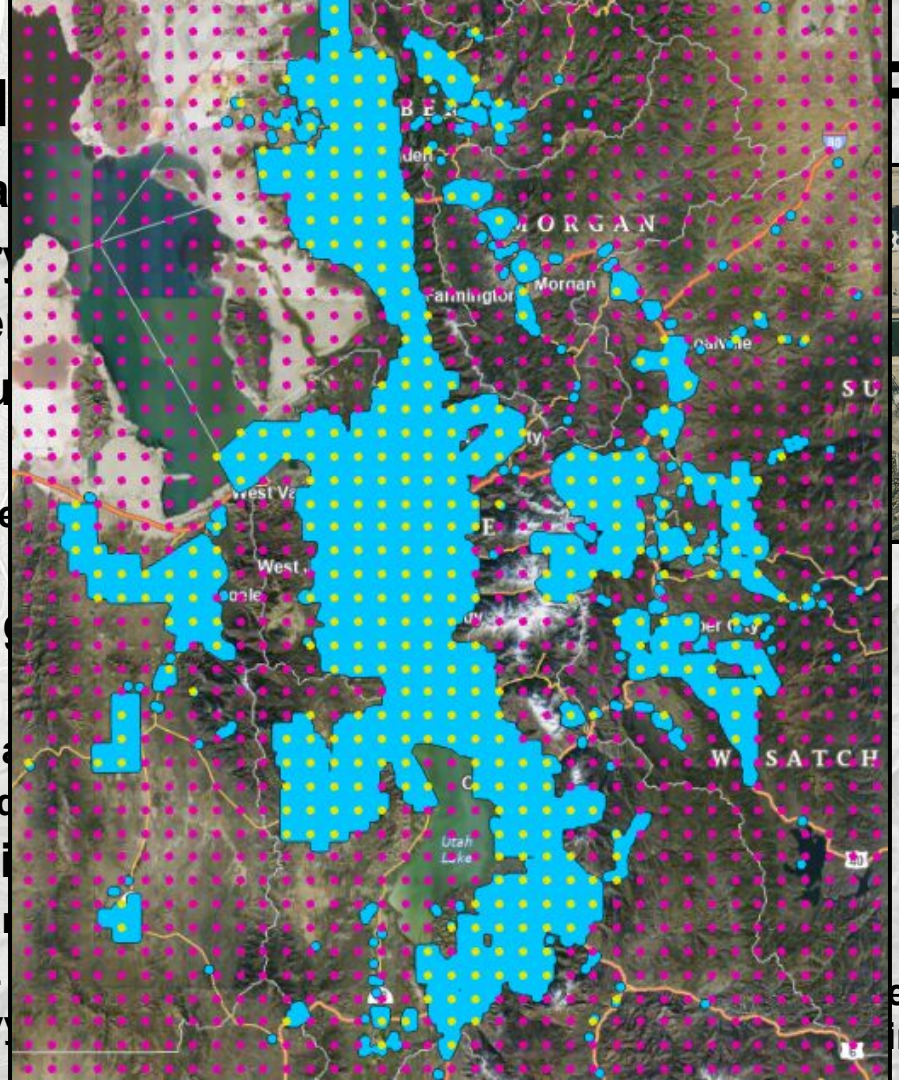


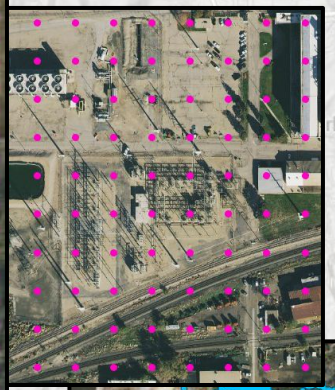
Build Index

Create Footprint

Step 1 - Bu Footprint

- Discover Web Ma
  statewide imagery
- Must build image
- Create point for u
  (row, col)
  - Calculate latitude
  - 275,000,000 tiles
- Create processing
  image tiles
  - Buffer census pla
  - Buffer large build
- Select WMTS indi
  - Load tile index ar
  - Run Spatial SQL
  - (data is waaaayyy



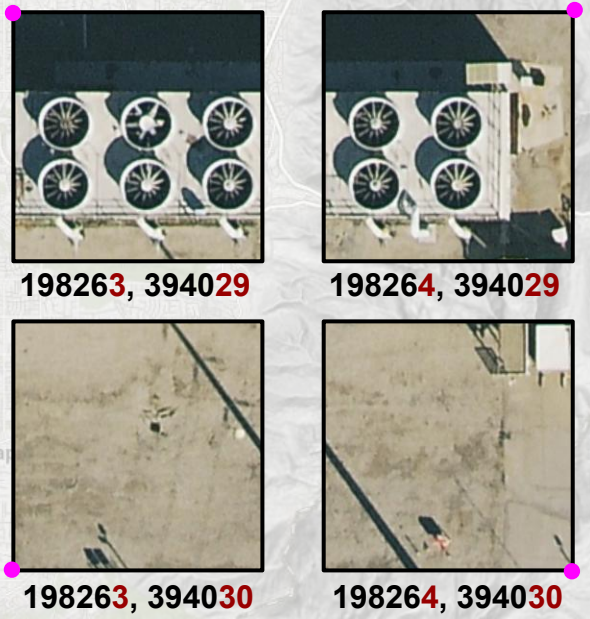


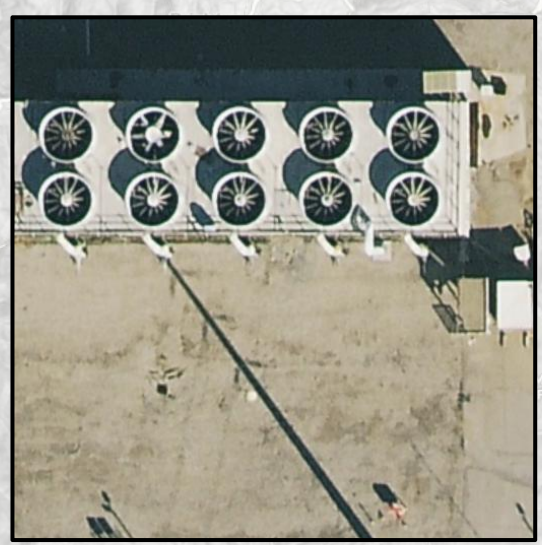
Build Index

Step 2 - Download Images

- Iterate through tile indices within footprint
  - Download primary tile and 3 neighboring tiles with HTTPS GET requests:
  - https://discover.agrc.utah.gov/login/path/{quad-word}/tiles/15cm_hexagon_utah/20/{col}/{row}
- Build mosaic image
  - Each WMTS tile is 256x256 pixels
  - Model input performs best on 512x512 images

{col}, {row}



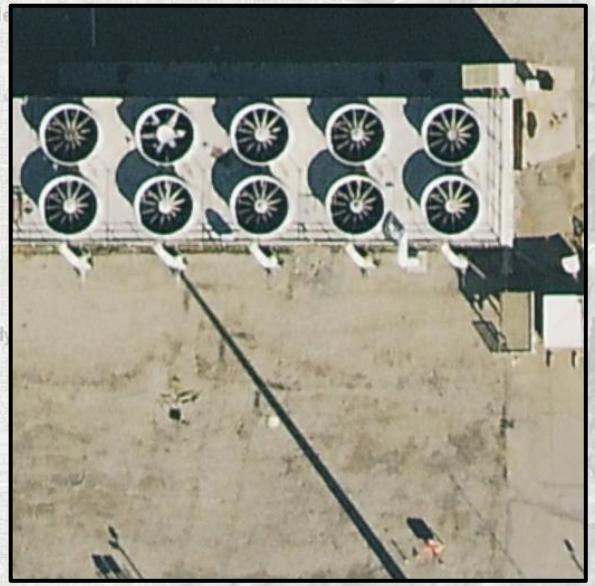


Utah Geospatial Resource Center

gis.utah.gov

Step 3 - Detect & Locate Towers

- **Run PyTorch model on each mosaic**
  - **Pre-trained "TowerScout" model, provided by CDC**
  - **YOLOv5 backbone**
- **Get results as a dataframe**



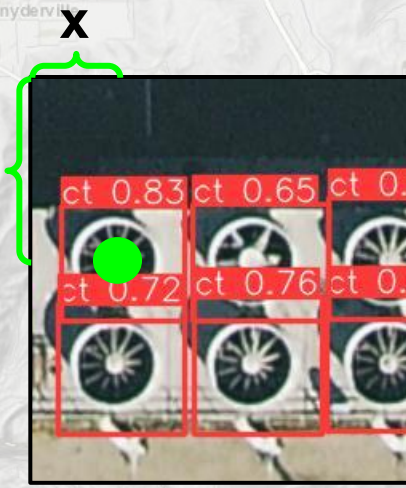
Detect Towers

Utah Geospatial Resource Center

*TowerScout created by Karen Wong, Jia Lu, Gunnar Mein, and Thaddeus Segura, licensed under CC-BY-NC-SA-4.0

Step 3 - Detect & Locate Towers

- Calculate X/Y coordinates of tower centroid
  - PyTorch model provides tower location bounding box in pixels from upper-left corner
  - Centroid is calculated from xmin, xmax, ymin, ymax values
  - Able to convert pixels into geographic space because we know:
    - 1) coordinates of the tile's upper-left corner
    - 2) pixel size at WMTS zoom level 20 (0.1492910708688 meters)



	xmin	ymin	xmax	ymax	confidence	x_centroid_px	y_centroid_px	x_centroid_3857	y_centroid_3857
9	265.932	130.687	347.821	217.394	0.633	306.877	174.041	-12460145.259	4978279.451
8	100.593	78.307	185.93	151.344	0.654	143.262	114.826	-12460169.686	4978288.291
7	184.386	135.914	268.444	220.131	0.676	226.415	178.023	-12460157.272	4978278.857

```
#: x, y = upper-left corner coordinates of tile in web mercator
#: calculate centroid x/y coords in web mercator
meters_per_pixel = 0.1492910708688
results_df["x_centroid_3857"] = x + results_df["x_centroid_px"] * meters_per_pixel
results_df["y_centroid_3857"] = y - results_df["y_centroid_px"] * meters_per_pixel
```

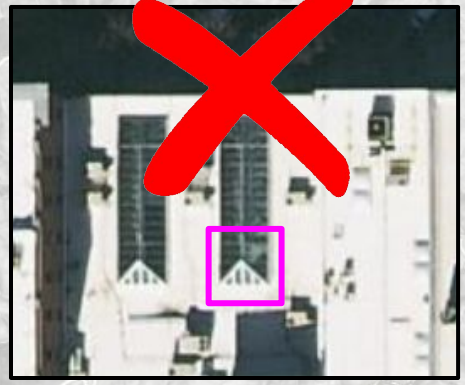
Calculate X/Y

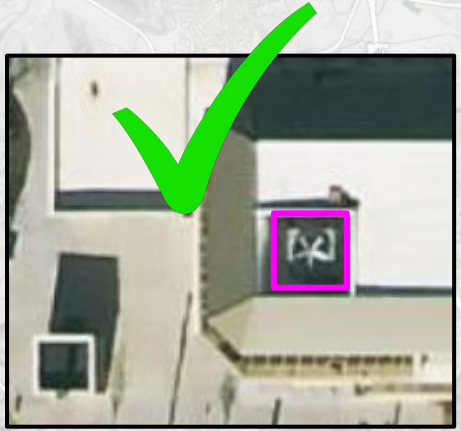
- Upload results to BigQuery

Utah Geospatial Resource Center

Step 4 - Post-process Results

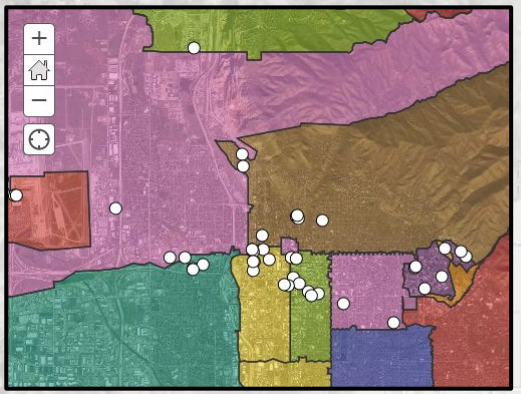
- Manually validate detected cooling towers
- Enrich data with attributes from other statewide datasets
  - Nearby address
  - County
  - City
  - Zip Code
  - Small Health Statistical Area





Step 5 - Build Web Map

- Create hosted feature layer of cooling tower locations
- Build a web map for DHHS users
- Add other relevant, health-related layers and tools

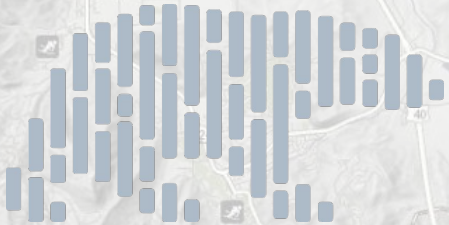


Utah Geospatial Resource Center

gis.utah.gov

DHHS Tools

- Mercantile - WMTS tile index to lat/lon
- Polars - large dataframe creation
- Requests - http requests and downloads
- Pyproj - coordinate conversion
- PyTorch - object detection model
- Google Cloud Platform
  - BigQuery - massive tabular data, querying, spatial SQL
  - Cloud Run - cloud computing



PYPROJ

Google Cloud

PyTorch

Requests
http for humans

Utah Geospatial Resource Center

github.com/agrc/dhhs-cooling-towers

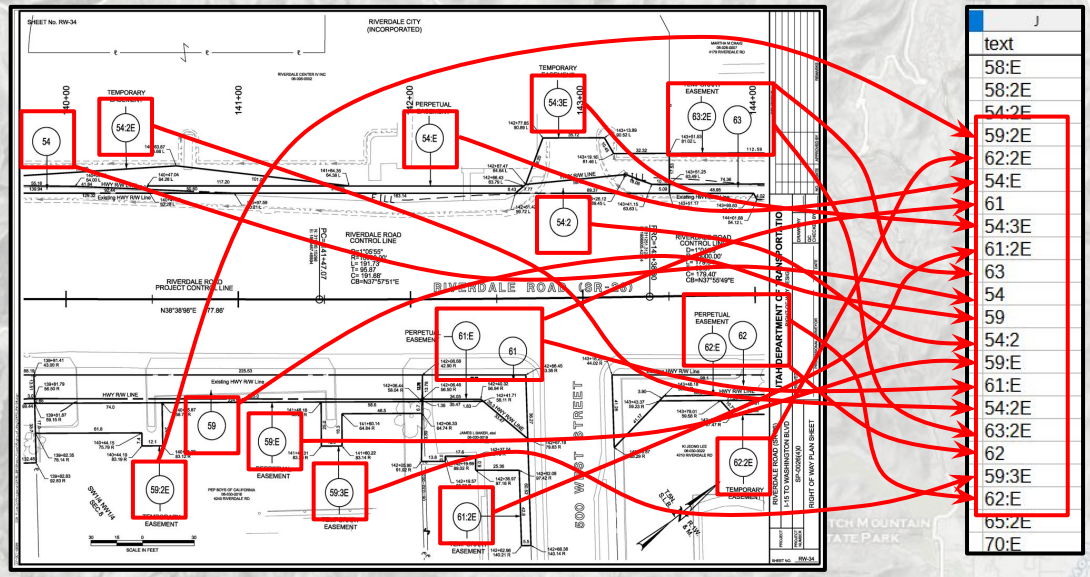
gis.utah.gov

DHHS Project Results

- To be determined…
- Most of the code is written, but the processing is in-work
- Check back later for an update! (UGRC blog post)

Final words

- A lot of useful data and information can be locked away in documents and imagery
- Computer Vision tools can help unlock that data
- Cloud computing can reduce processing time by orders of magnitude
- UDOT parcel detection and DHHS cooling towers projects highlight these possibilities



Utah Geospatial Resource Center

gis.utah.gov

Questions?

Erik Neemann

email: eneemann@utah.gov

twitter: @Erik_UGRC

gis.utah.gov/presentations

gis.utah.gov