INEGI

UN Big Data Regional Hub for Africa

International webinar on Earth Observations for Agricultural Statistics (EO-STAT):

EO Data cubes, smart classification algorithms, impacts of disasters on crops, and a live demo

Use of EO data cube in Mexico to calibrate an algorithm to obtain the Agriculture Frontier

M. Sc. Eduardo Vázquez Andrade

Rwanda

December 18, 2023

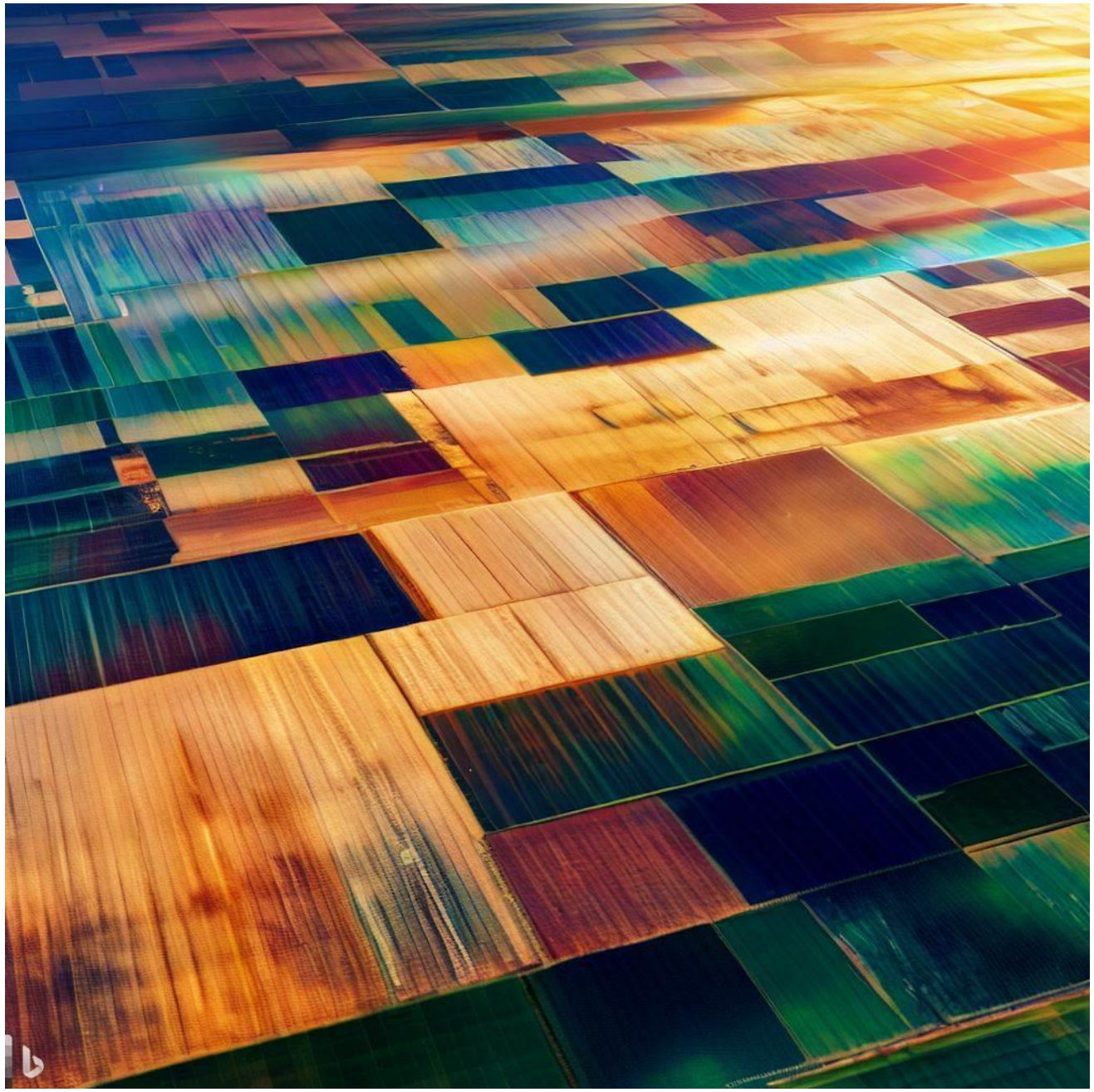
Background

Background in the Use of

Earth Observations

In 2009 began the first exercises to use Earth Observations to obtain agricultural statistic information. Some projects were developed between 2012 y 2018.

However, the high cost of images and software limited their use



Data from

Agricultural Census

2007 Agriculture Census. One of the main results was a digital archive of all censused lands, with the primary activity (Agricultural, livestock, or forestry) included as an attribute. Subsequent update projects began with the Update of the Agricultural Census Framework 2016 (AMCA):

- 2016 AMCA, at land level
- 2017 ENA, only simple selected
- 2019 ENA, only simple selected
- 2018-2019 Review of AMCA with satellite imagery
- 2019-2020 Comparison of AMCA with other sources of agricultural frontiers

Agricultural Land Use Identification

Agricultural frontier concept, Territorial distribution of areas in Mexico with agricultural activity, and lands cultivated in the last 5 years.

* According to the concepts of SIAP and INEGI

COD_ACT	DESCRIPCION	KM²	
A	Completely agricultural	200,257.75	AGRICULTURAL
C	At least 30% agricultural	109,783.30	
M	Mixed	3,728.26	
F	Formerly agricultural	5,007.09	NOT AGRICULTURAL
N	No agricultural activity	117,662.82	
U	Urban	4,342.64	
V	Verified (no agricultural activity)	1,471,232.10	
W	Body of water	8,098.02	
B	Roads	16.49	
I	Flood zones	6.33	

INEGI's Data Cube

In 2019 Geosciences Australia advised a group of experts of INEGI

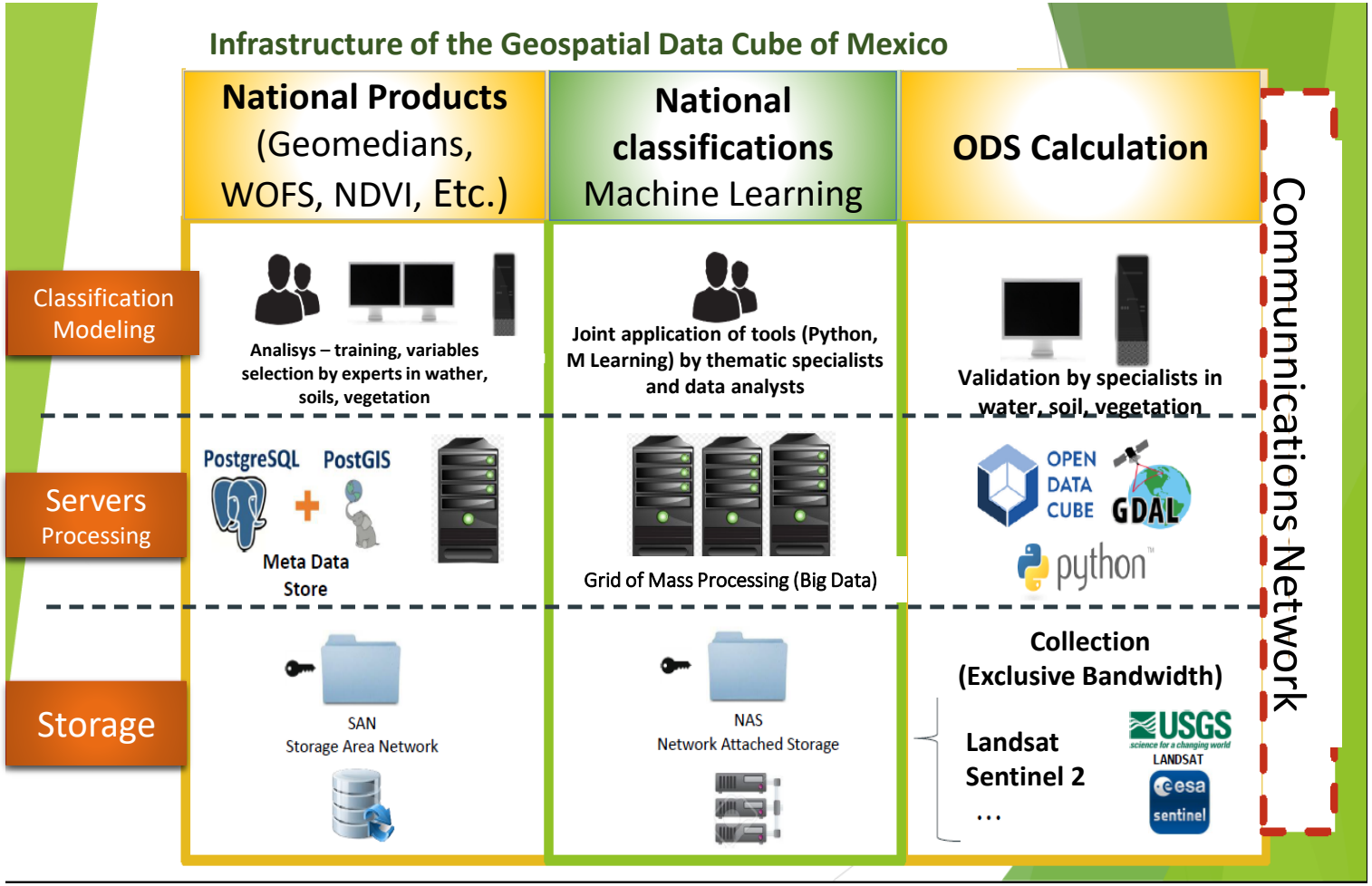
Implementation of the Data Cube.

- Landsat images

-Sentinel 2 images

- Integration of Geomedian processes.

Exercises of application.



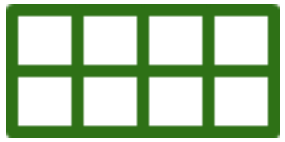
Problem Statement

Statement

of the Problem

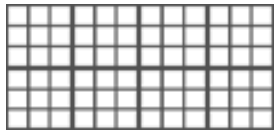
The problem to address is how to produce timely, cost-effective, and reliable estimates of the national agricultural frontier using Earth Observations combined with artificial intelligence algorithms.

Landsat (~38 years) & Sentinel-2 (~7 years)



30m. Spatial resolution

Landsat 4,5,7,8,9



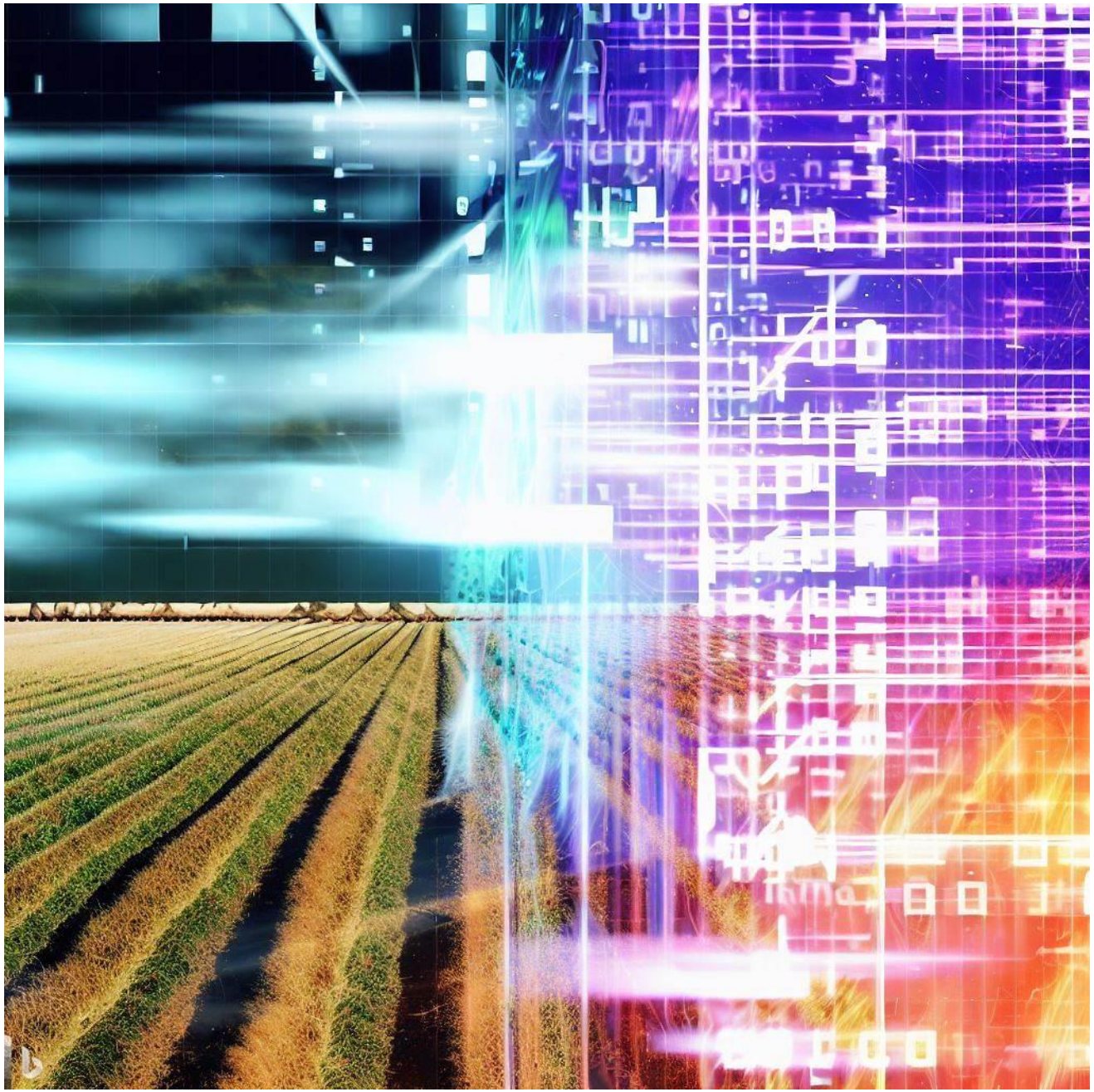
10m. Spatial resolution

Sentinel 2A, 2B

Objective

Objective

Calibrate an algorithm using Artificial Intelligence and SENTINEL 2 satellite imagery to estimate the National Agricultural Frontier.



Methodology

Data

Sources

Agricultural frontier

2019

Sentinel-2 Geomedian (12)

2019

Spectral indexes (20)

2019

Texture filters (48)

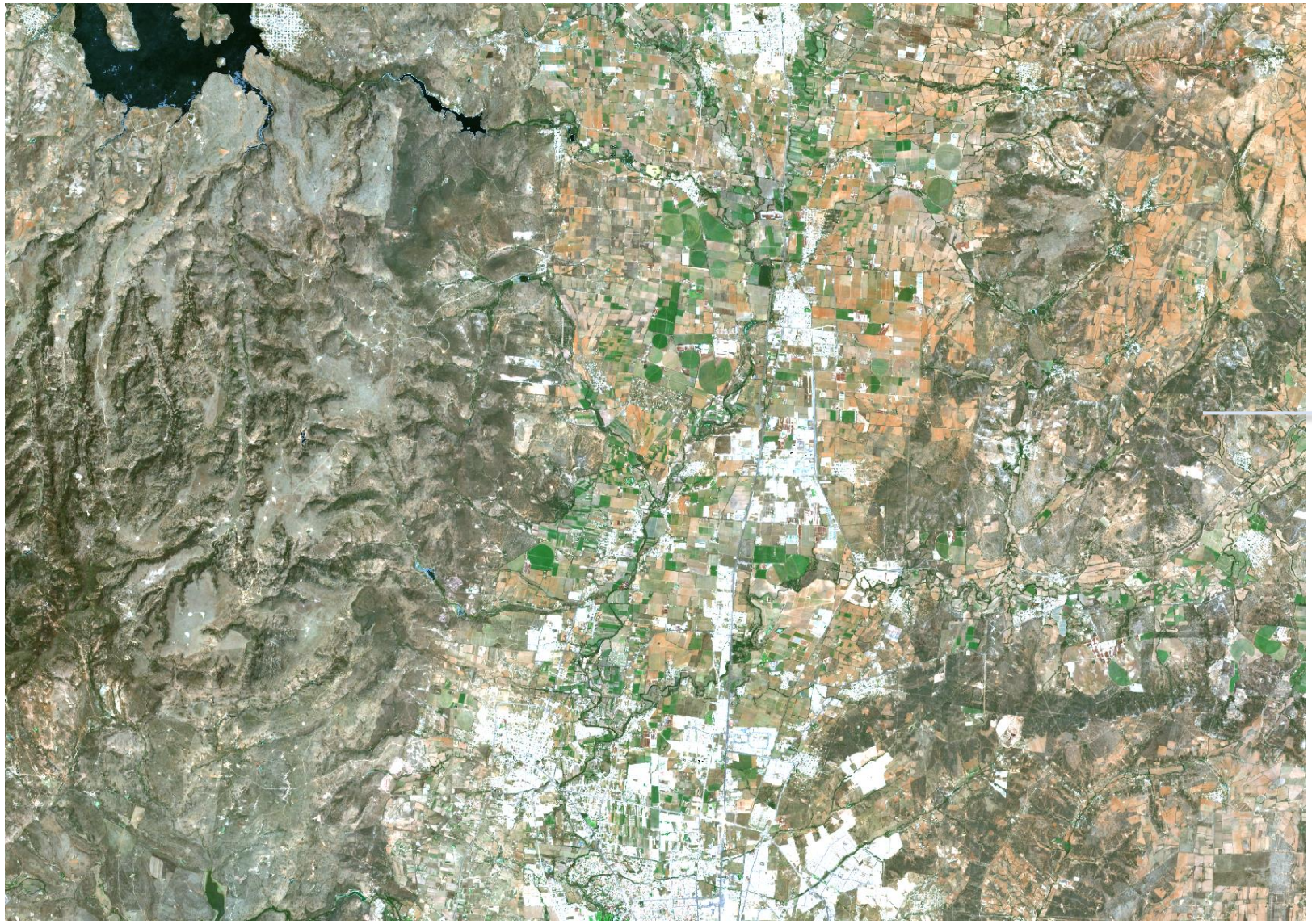
2019

Reference

Labels

COD_ACT	DESCRIPCION	KM²	
A	Completely agricultural	200,257.75	AGRICULTURAL
C	At least 30% agricultural	109,783.30	
M	Mixed	3,728.26	
F	Formerly agricultural	5,007.09	NOT AGRICULTURAL
N	No agricultural activity	117,662.82	
U	Urban	4,342.64	
V	Verified (no agricultural activity)	1,471,232.10	
W	Body of water	8,098.02	
B	Roads	16.49	
I	Flood zones	6.33	

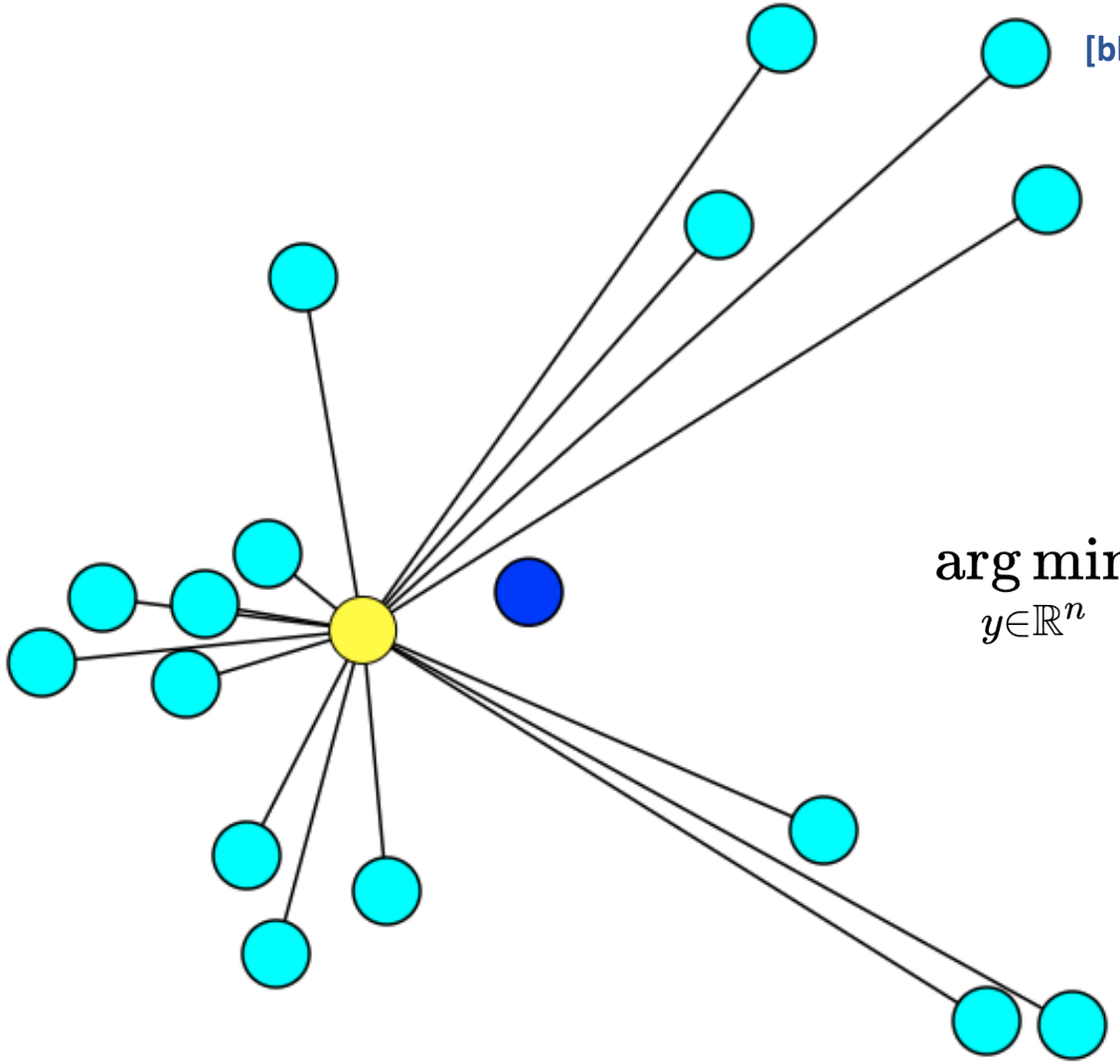
Geomedian



1. Coastal Aerosol
2. Blue
3. Green
4. Red
5. Vegetation 5
6. Vegetation 6
7. Vegetation 7
8. Near-Infrared
9. Vegetation 8
10. Water Vapour
11. Short Wave Infrared 1
12. Short Wave Infrared 2

Geomedian = Geometric Median

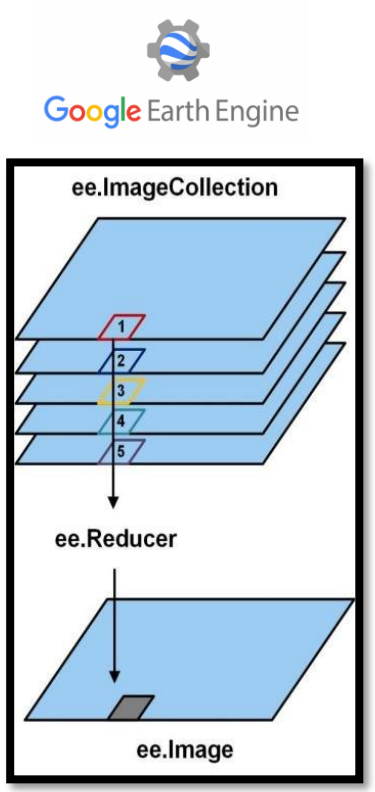
[blue, green, red, nir, swir1, swir2, ...]

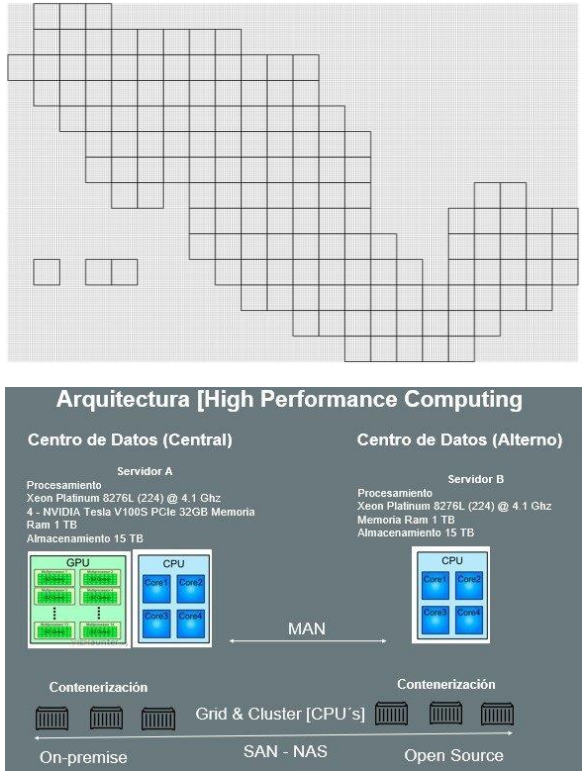


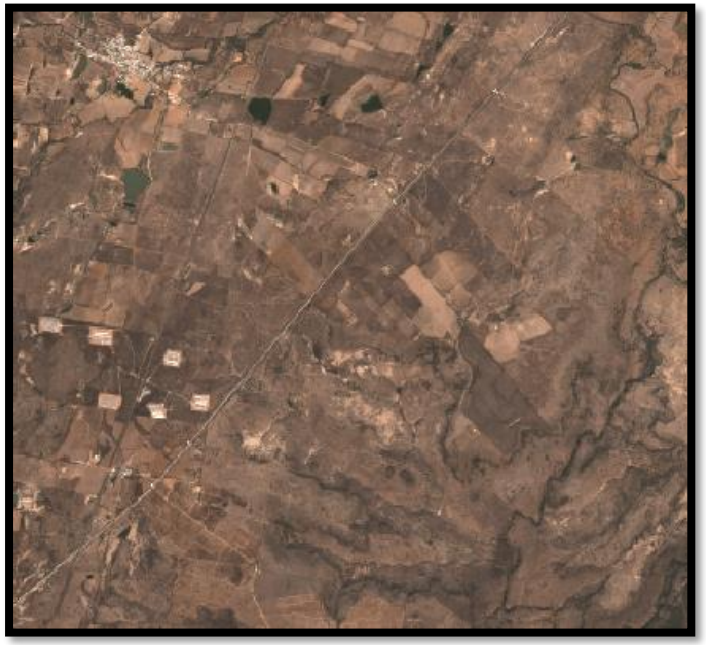
$$\arg\min_{y \in \mathbb{R}^n} \sum_{i=1}^{m} \|x_i - y\|_2$$

[blue, green, red, nir, swir1, swir2, ...]

Geomedian = Geometric Median







`ee.Reducer.geometricMedian`

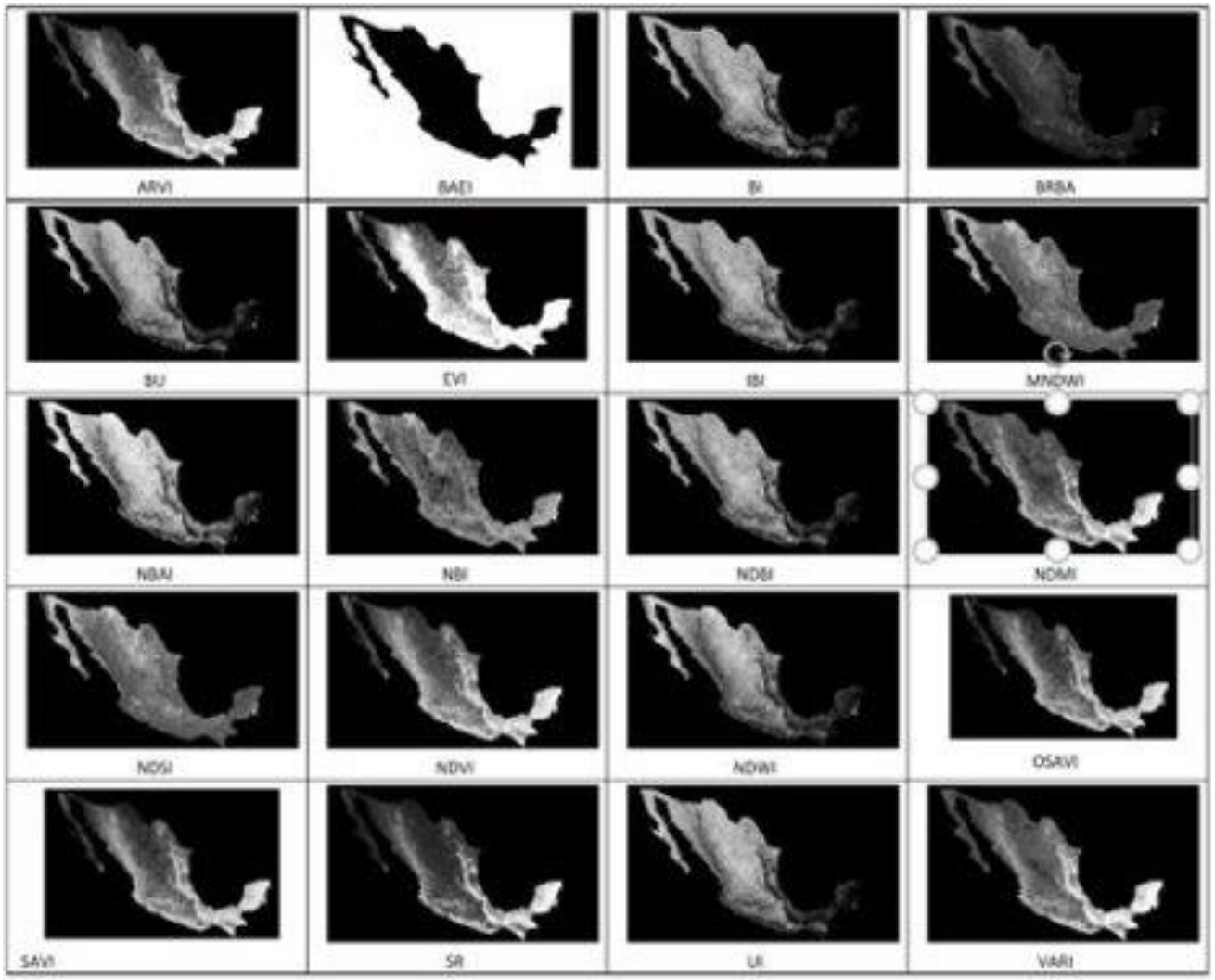
Imange Reprojection & Alignment

GeoTIFF Images
12 Bands

Sources:

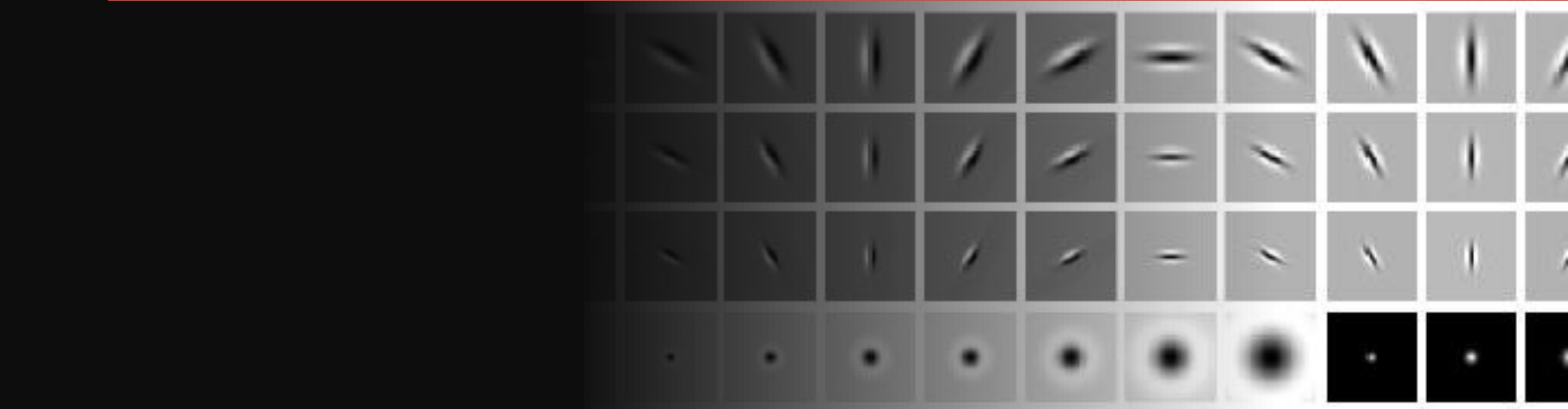
https://developers.google.com/earth-engine/apidocs/ee-reducer-geometricmedian
https://www.researchgate.net/figure/The-reducer-operation-provided-by-Google-Earth-Engine-GEE-17_fig3_349430332
https://en.wikipedia.org/wiki/Geometric_median

20 Spectral Indexes

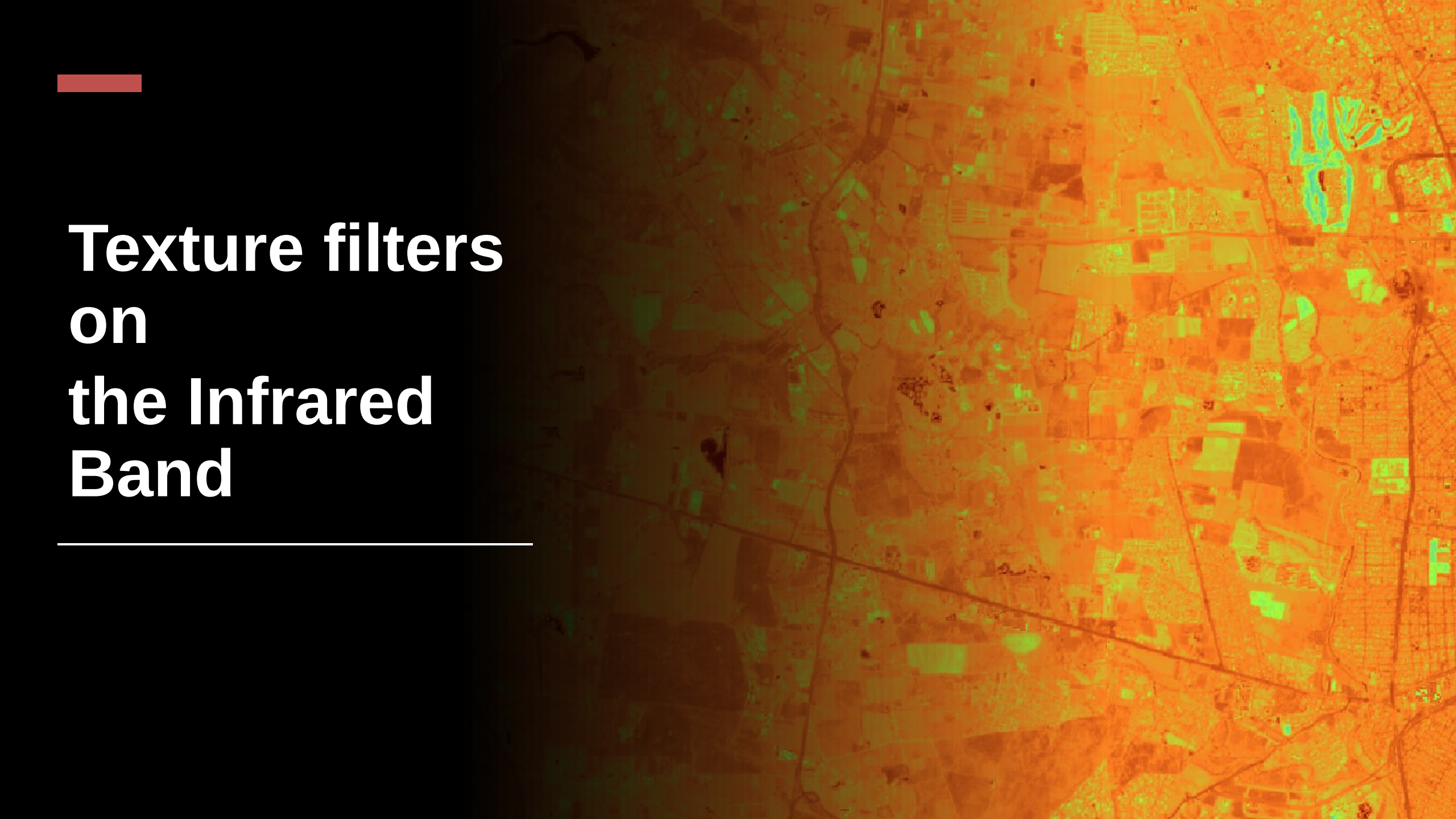


Dimensiones de 110,000 pixeles x 70,000 pixeles

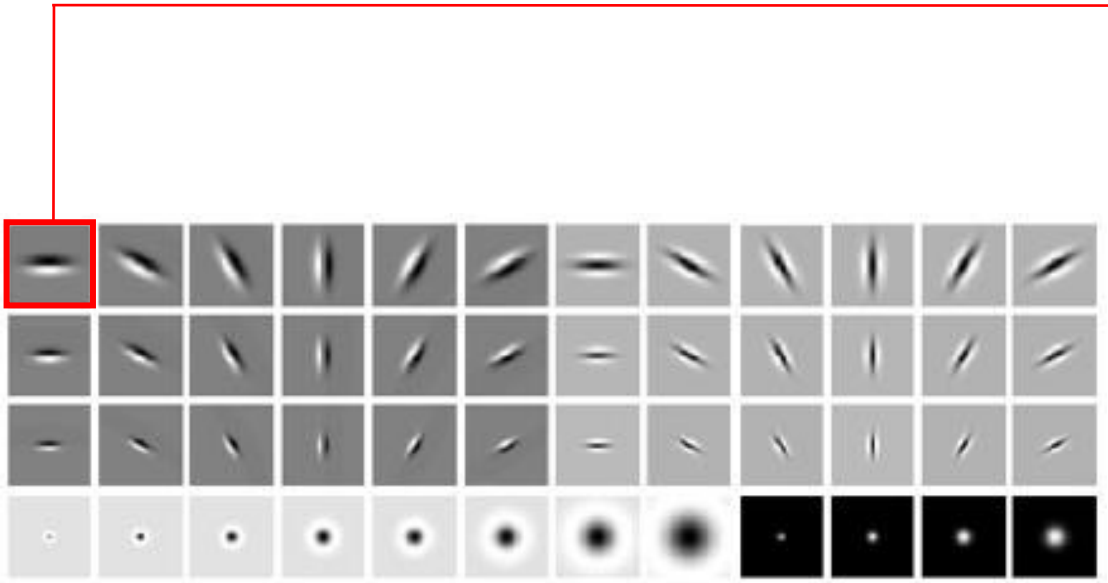
48 Texture Filters



Texture filters on the Infrared Band

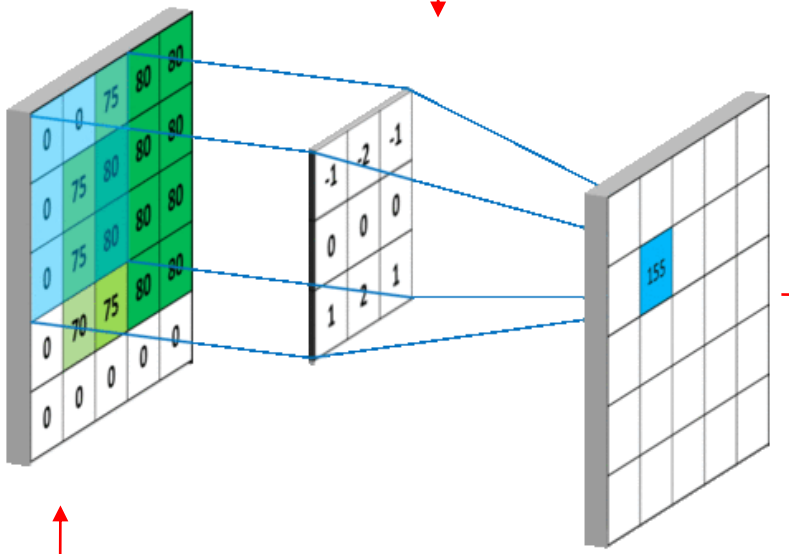


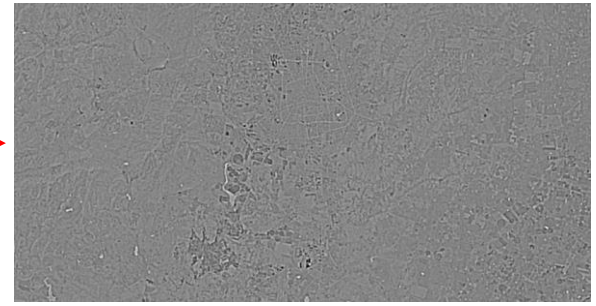
48 Texture Filters



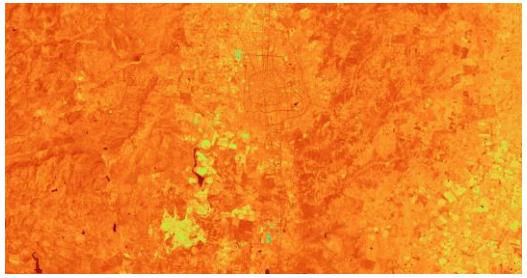
Leung-Malik Texture Filter Bank

Leung, Thomas, and Jitendra Malik. "Representing and recognizing the visual appearance of materials using three-dimensional textons." International journal of computer vision 43, no. 1 (2001): 29-44.

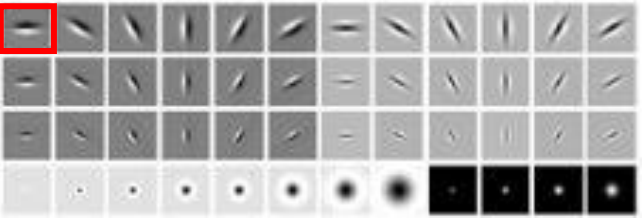




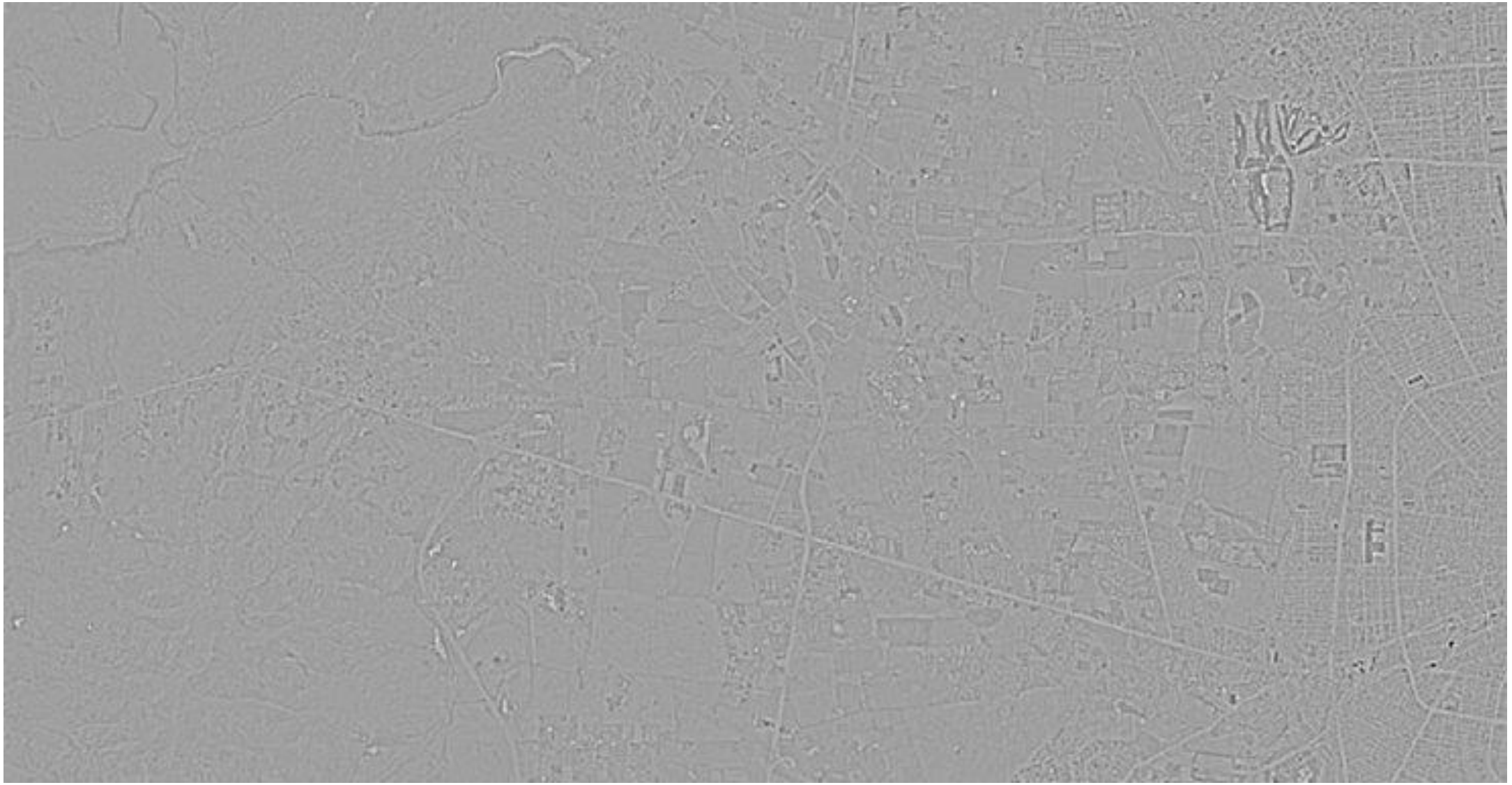
New Filtered Image

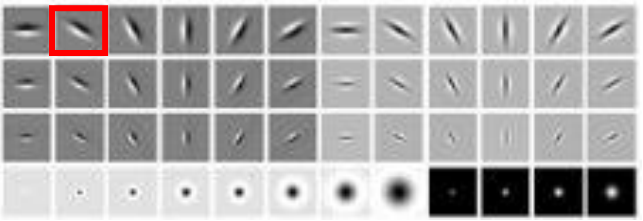


NIR

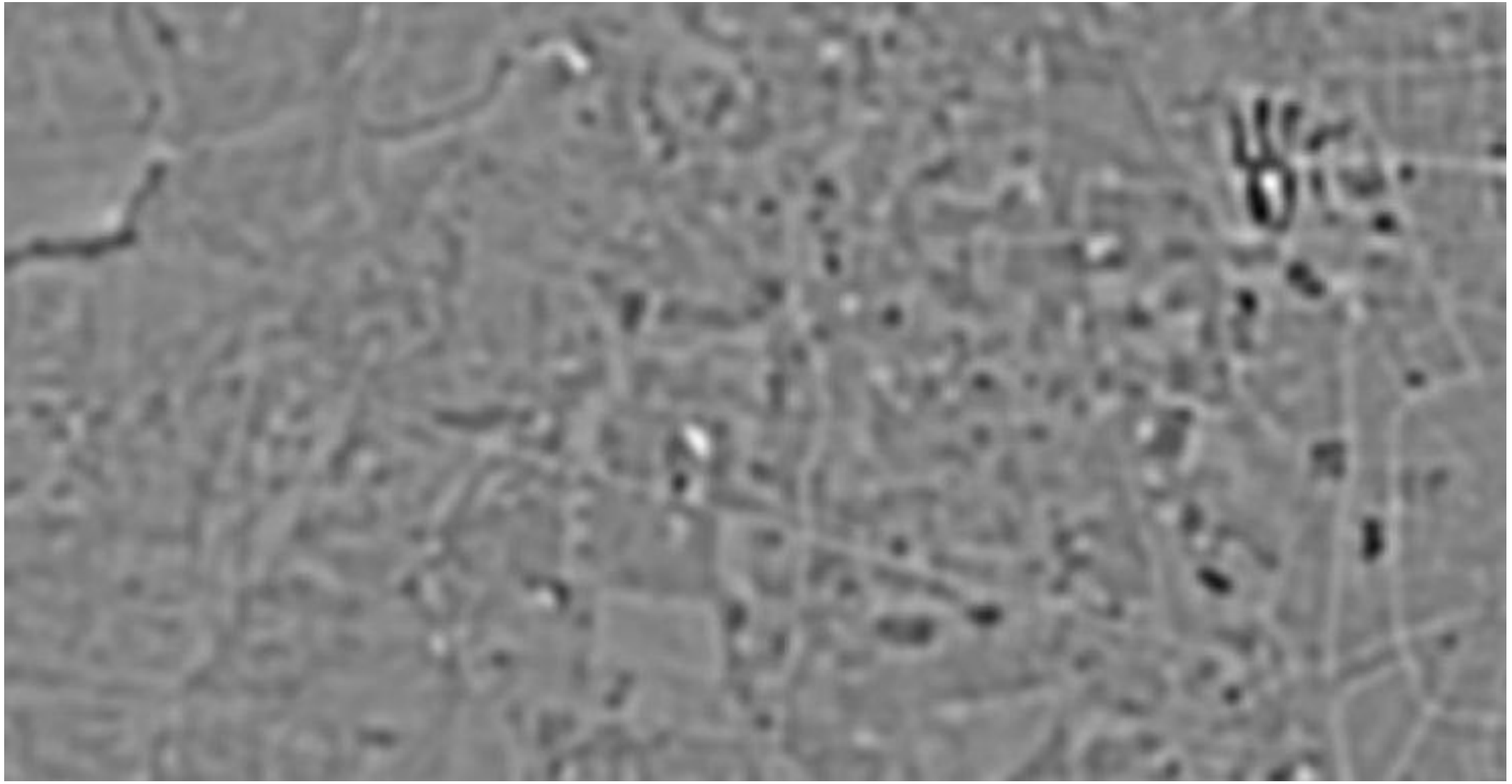


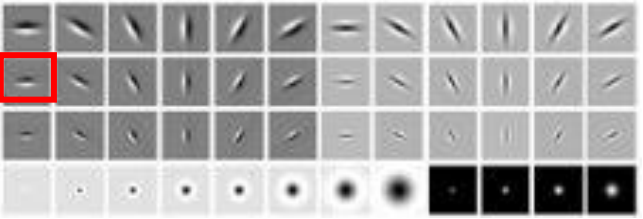
48 Texture Filters



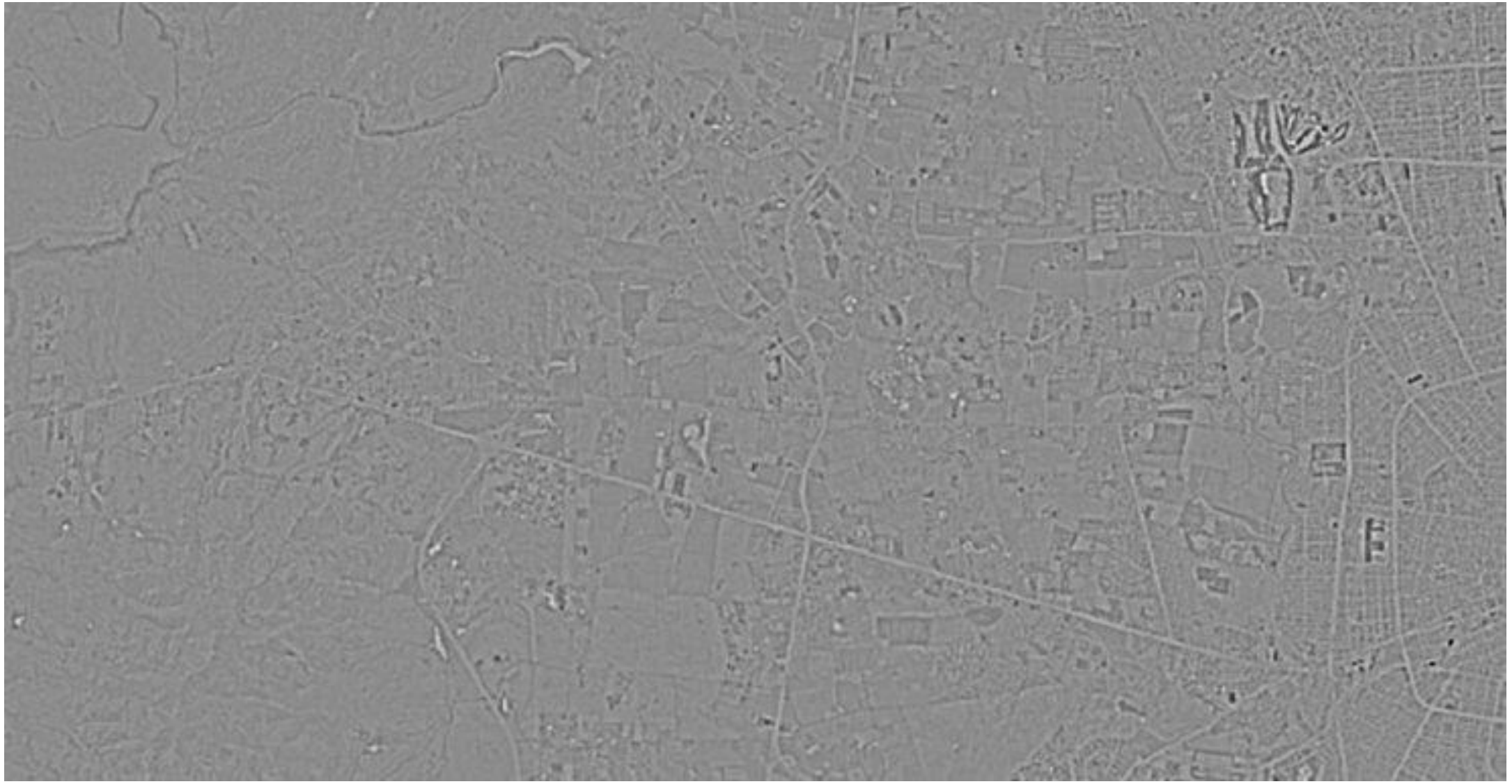


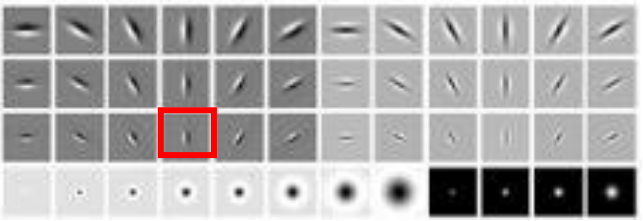
48 Texture Filters



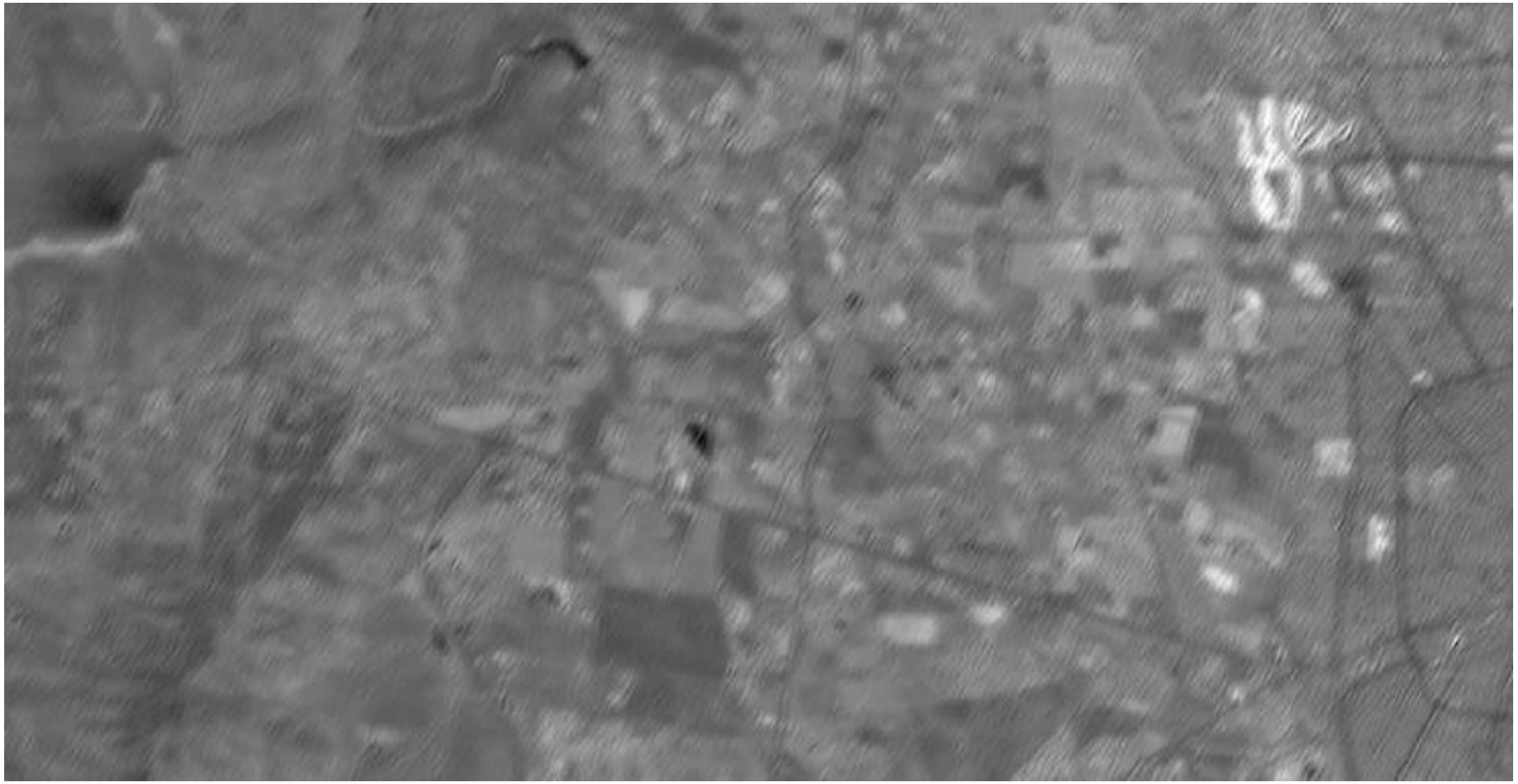


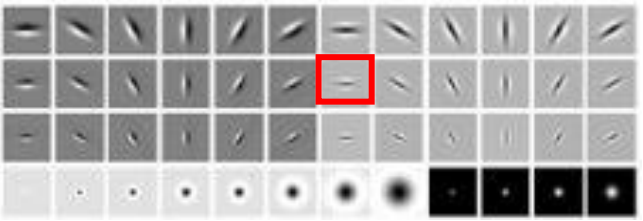
48 Texture Filters



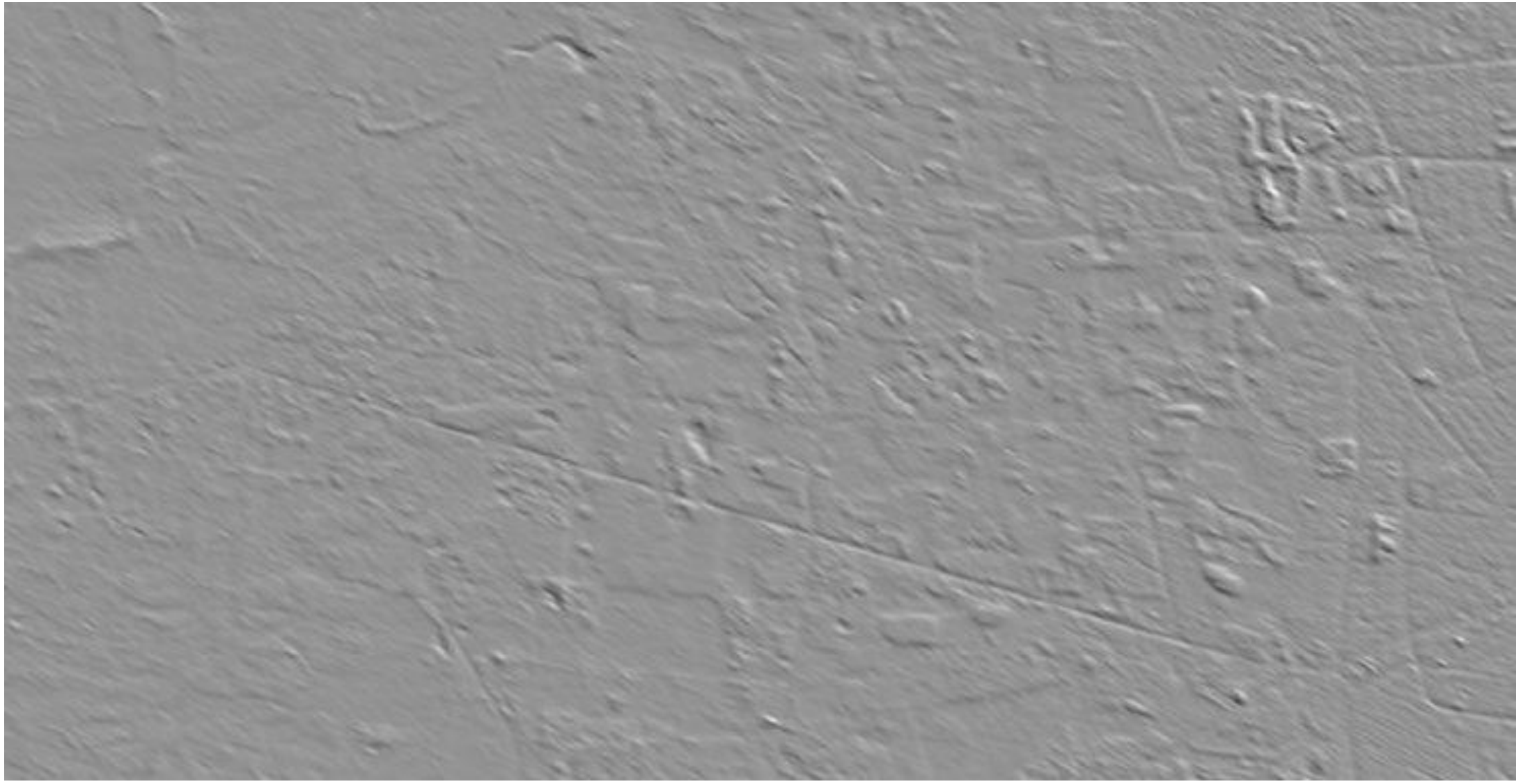


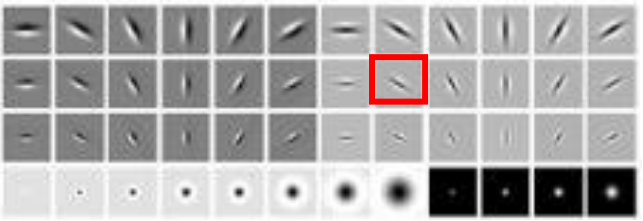
48 Texture Filters



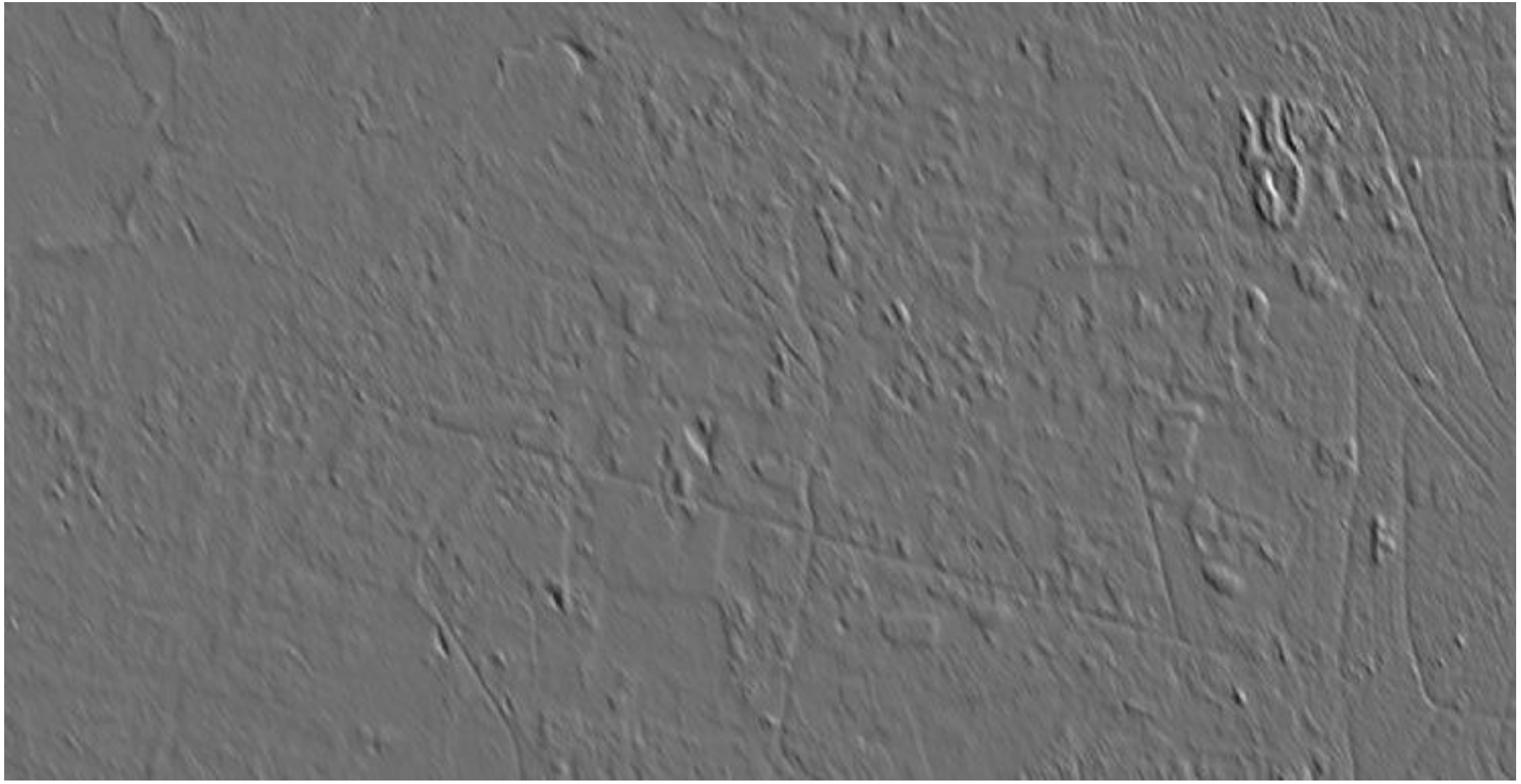


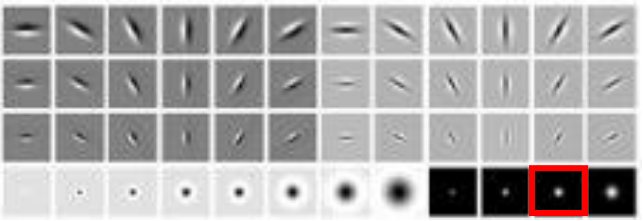
48 Texture Filters



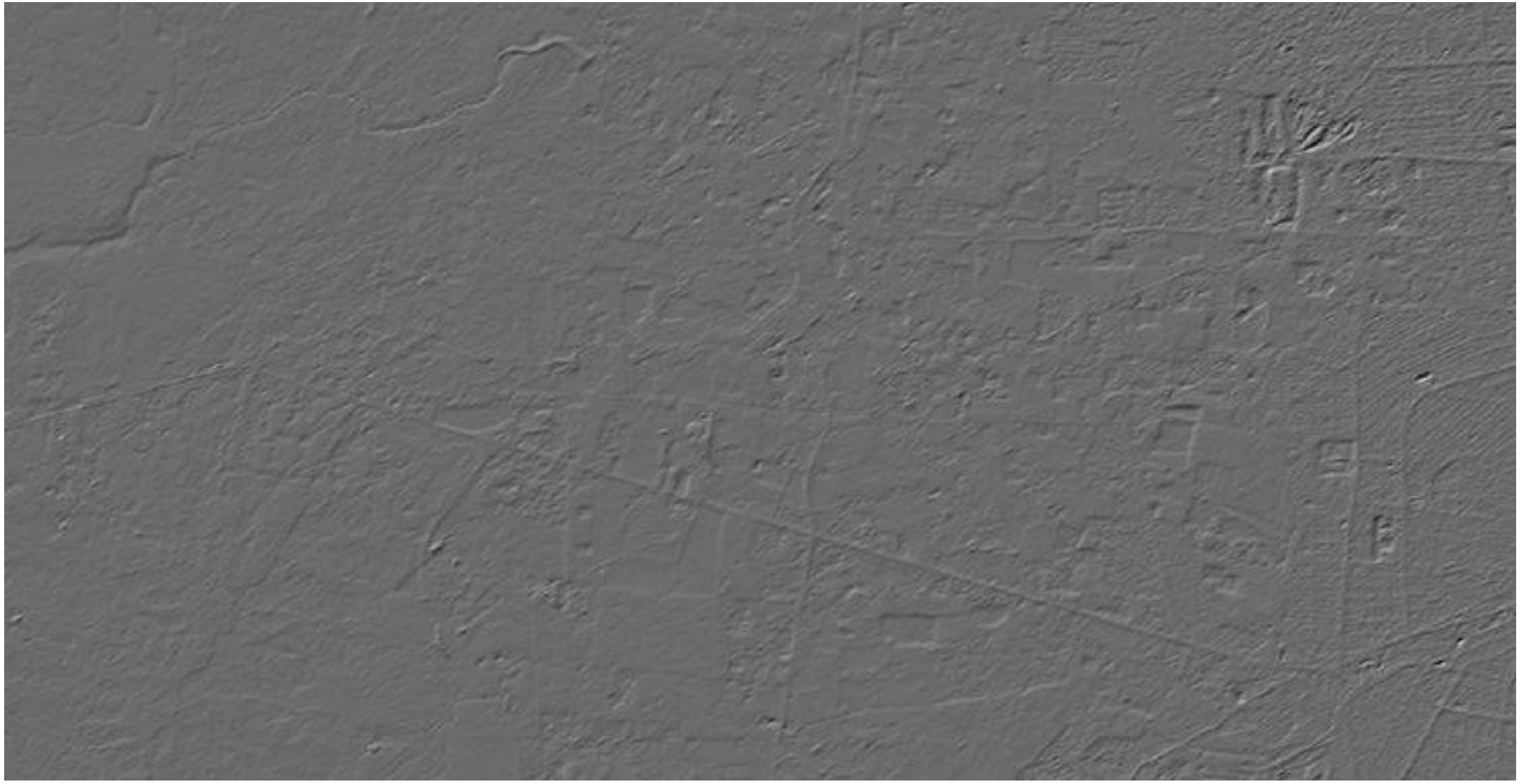


48 Texture Filters



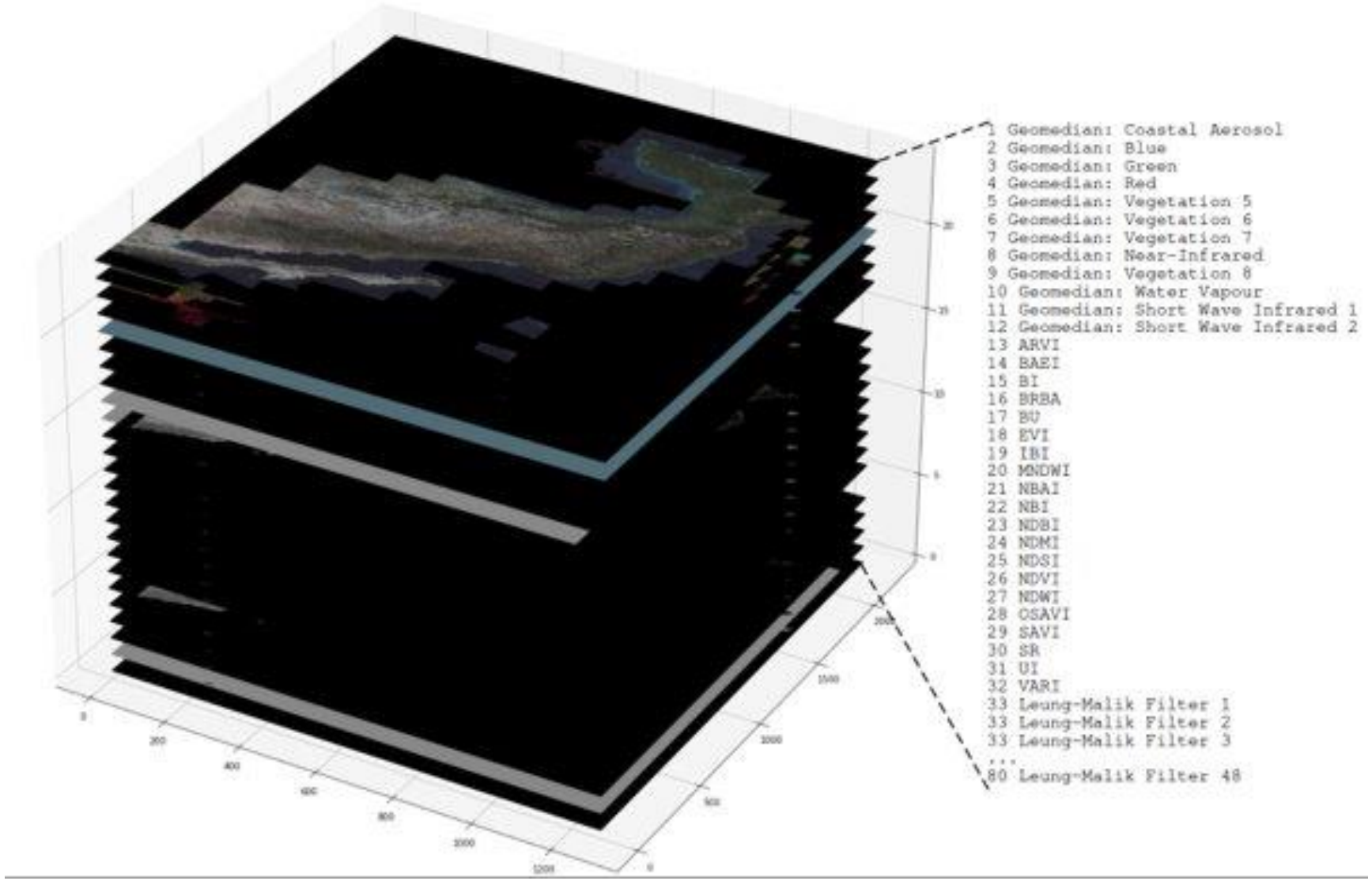


48 Texture Filters



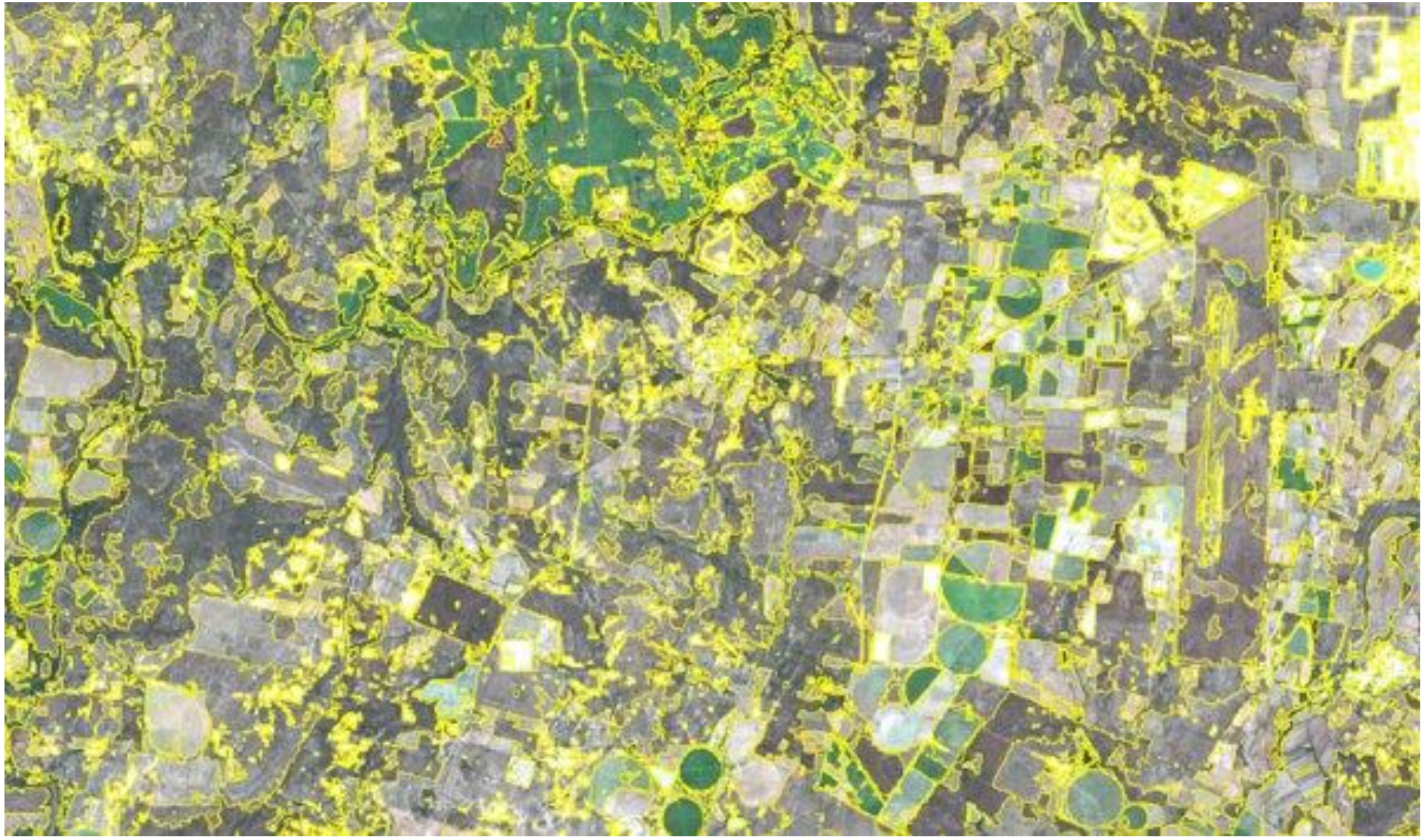
80 Raster Layers

17.2 TB



Geomedian Segmentation

Geomedian Segmentation



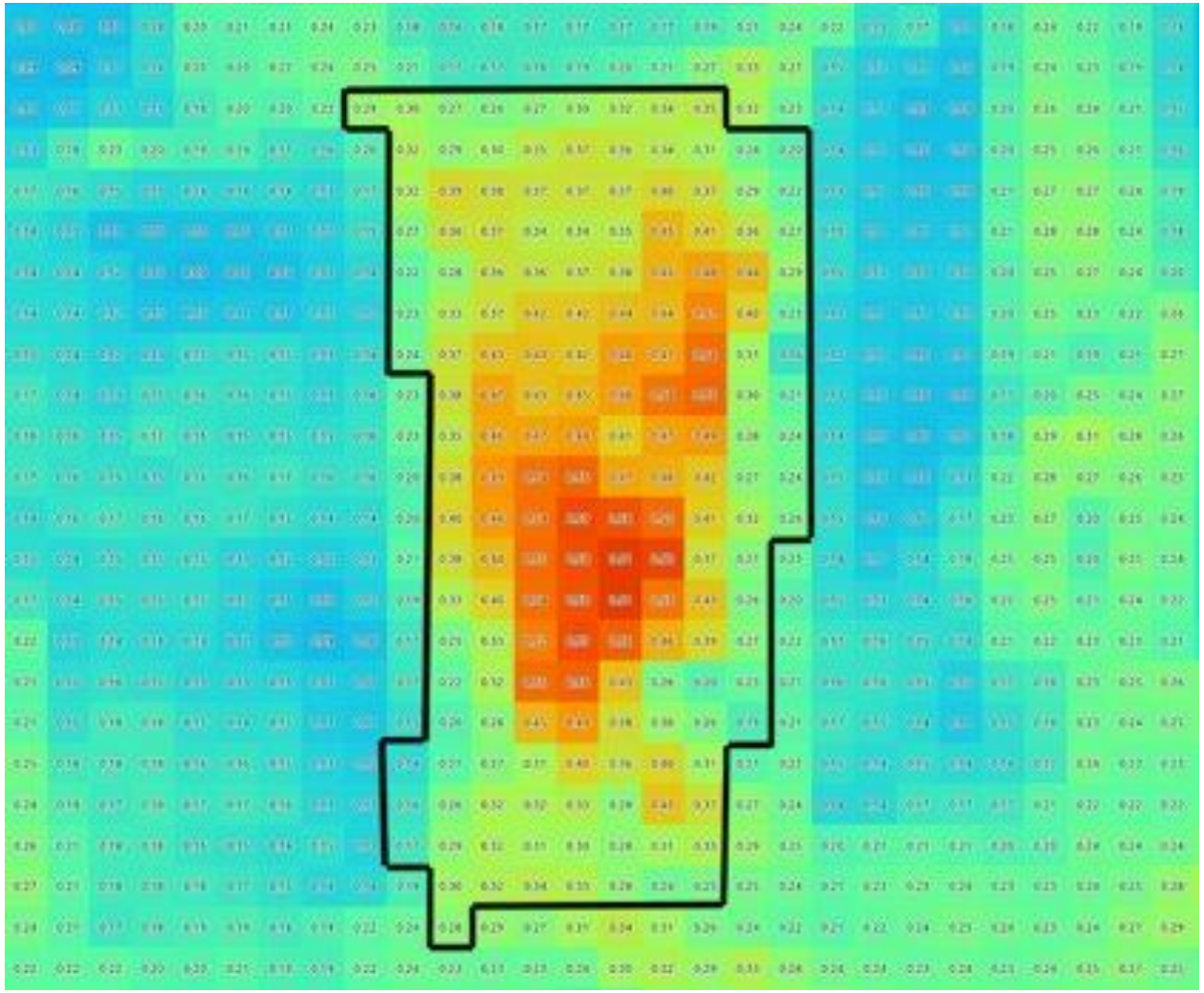
Segment characterization

ALL LAYERS

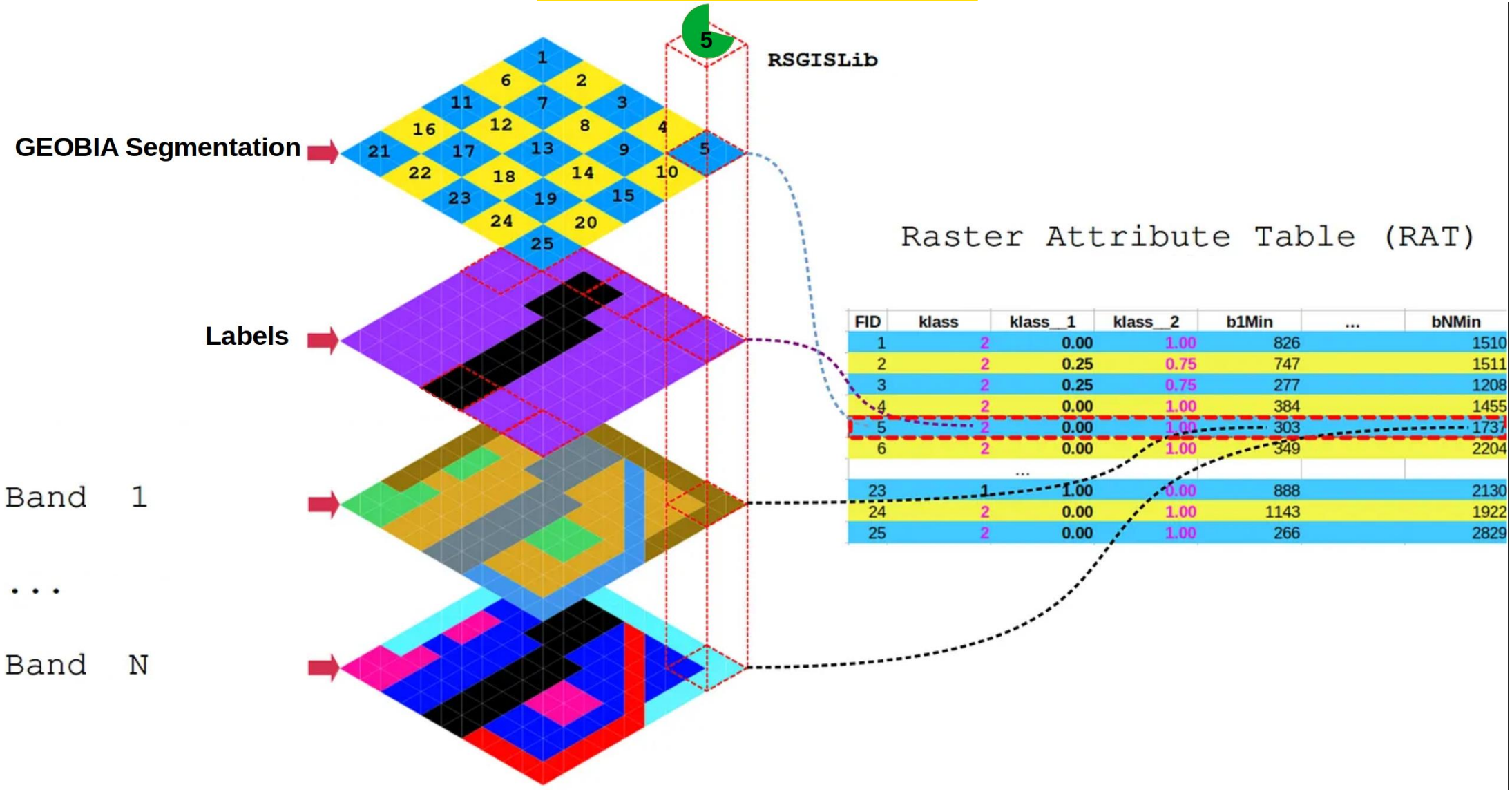
- Minimum
- Maximum
- Average
- Sum
- Standard Deviation

TEXTURE FILTERS

- Percentile 10 - 90



Segment characterization



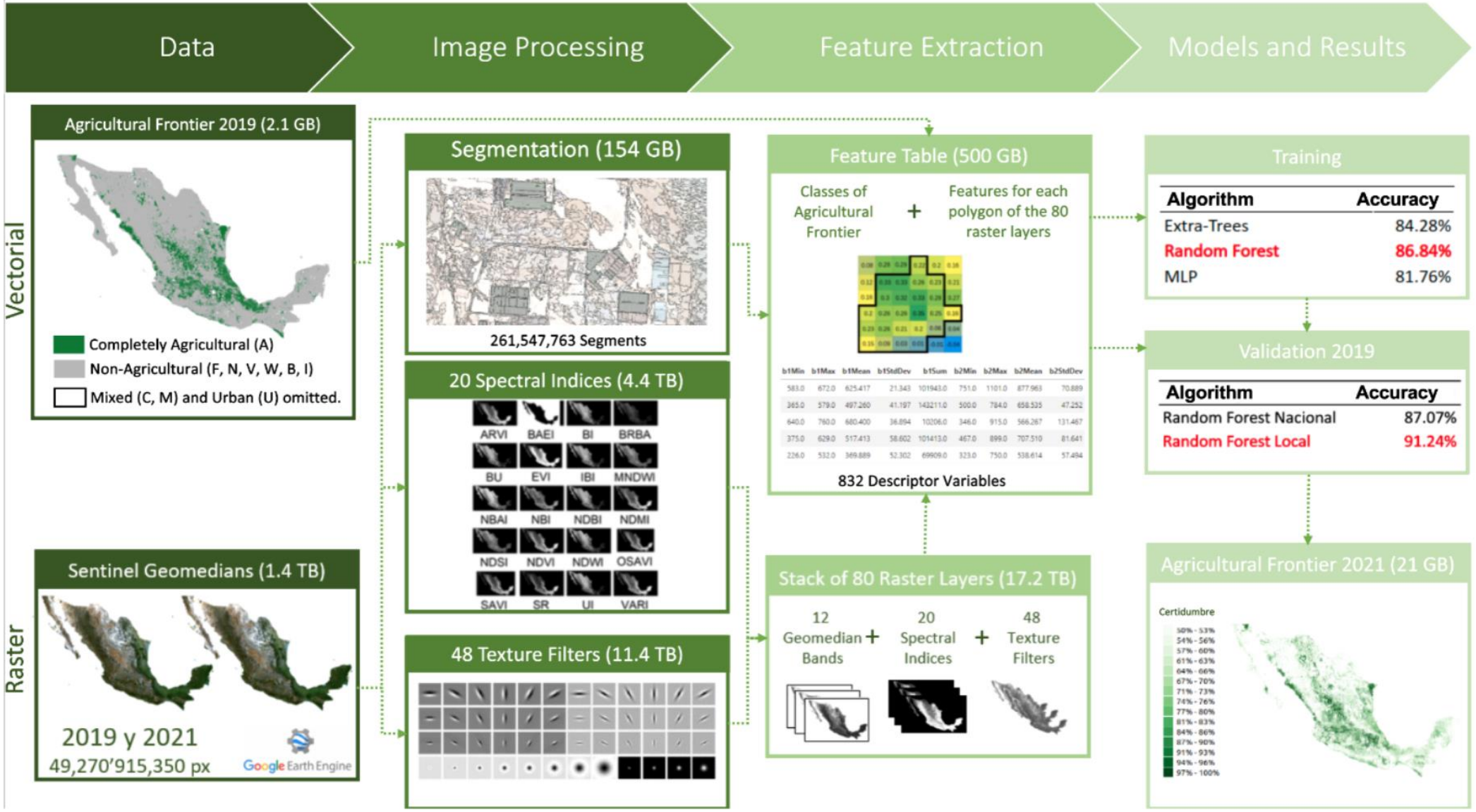
FID	klass	klass__1	klass__2	b1Min	...	bNMin
1	2	0.00	1.00	826		1510
2	2	0.25	0.75	747		1511
3	2	0.25	0.75	277		1208
4	2	0.00	1.00	384		1455
5	2	0.00	1.00	303		1737
6	2	0.00	1.00	349		2204
...						
23	1	1.00	0.00	888		2130
24	2	0.00	1.00	1143		1922
25	2	0.00	1.00	266		2829

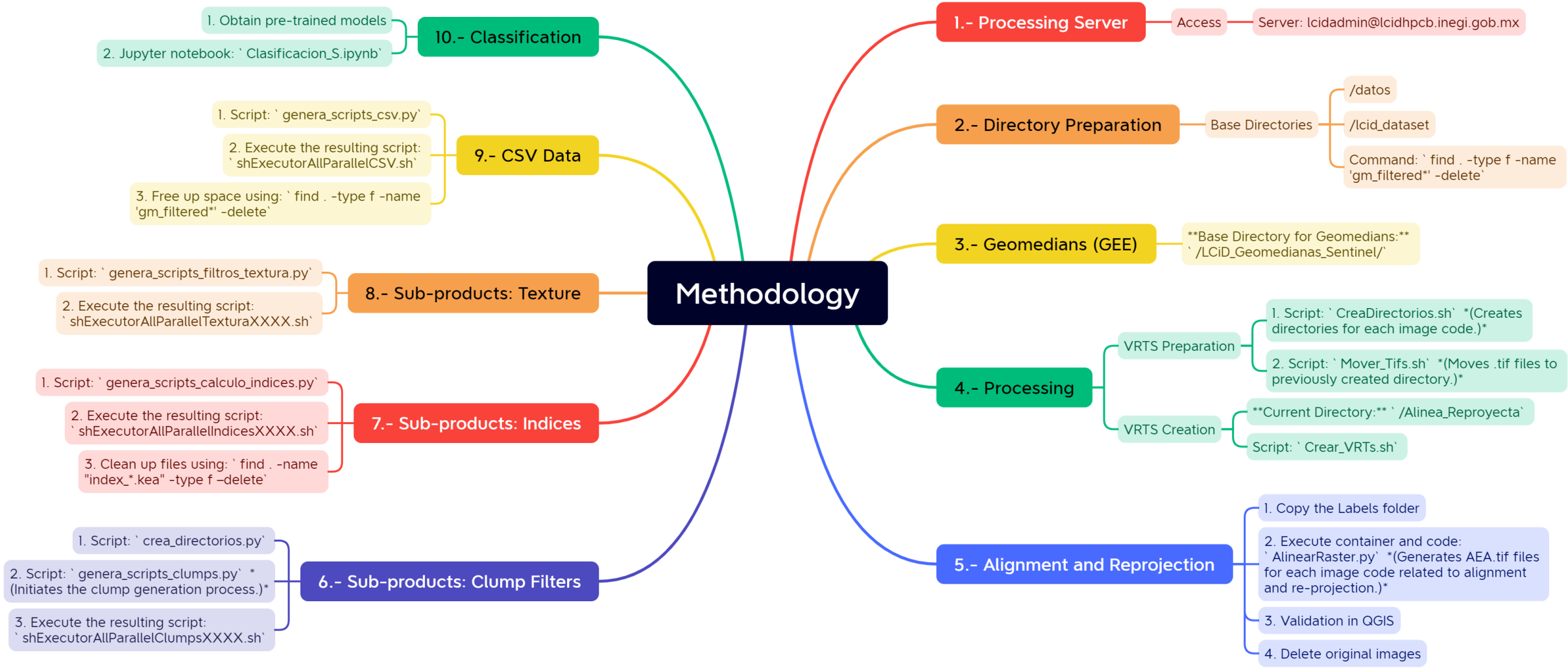
Segment characterization

Segment	Class	Geomedian-Blue-Minimum	Geomedian-Blue-Maximum	Geomedian-Blue-Mean	Geomedian-Blue-Sum	...	Filter 48 Percentile 80	Filter 48 Percentile 90
1	2	256	3235	1570	15023	...	0.26	0.074
...	...	...	...	...	...	...	...	...
261,547,763	1	129	2500	1120	12000	...	0.39	0.19

Data table with 834 columns = Segment + Class + 832 Variables

Methodology




Methodology
1.- Processing Server
Access
Server: lcidadmin@lcidhpcb.inegi.gob.mx
2.- Directory Preparation
Base Directories
/datos
/lcid_dataset
Command: ` find . -type f -name 'gm_filtered*' -delete`
3.- Geomedians (GEE)
Base Directory for Geomedians: ` /LCiD_Geomedianas_Sentinel/`
4.- Processing
VRTS Preparation
1. Script: ` CreaDirectorios.sh` *(Creates directories for each image code.)*
2. Script: ` Mover_Tifs.sh` *(Moves .tif files to previously created directory.)*
VRTS Creation
Current Directory: ` /Alinea_Reproyecta`
Script: ` Crear_VRTs.sh`
5.- Alignment and Reprojection
1. Copy the Labels folder
2. Execute container and code: ` AlinearRaster.py` *(Generates AEA.tif files for each image code related to alignment and re-projection.)*
3. Validation in QGIS
4. Delete original images
6.- Sub-products: Clump Filters
1. Script: ` crea_directorios.py`
2. Script: ` genera_scripts_clumps.py` * (Initiates the clump generation process.)*
3. Execute the resulting script: ` shExecutorAllParallelClumpsXXXX.sh`
7.- Sub-products: Indices
1. Script: ` genera_scripts_calculo_indices.py`
2. Execute the resulting script: ` shExecutorAllParallelIndicesXXXX.sh`
3. Clean up files using: ` find . -name "index_*.kea" -type f –delete`
8.- Sub-products: Texture
1. Script: ` genera_scripts_filtros_textura.py`
2. Execute the resulting script: ` shExecutorAllParallelTexturaXXXX.sh`
9.- CSV Data
1. Script: ` genera_scripts_csv.py`
2. Execute the resulting script: ` shExecutorAllParallelCSV.sh`
3. Free up space using: ` find . -type f -name 'gm_filtered*' -delete`
10.- Classification
1. Obtain pre-trained models
2. Jupyter notebook: ` Clasificacion_S.ipynb`

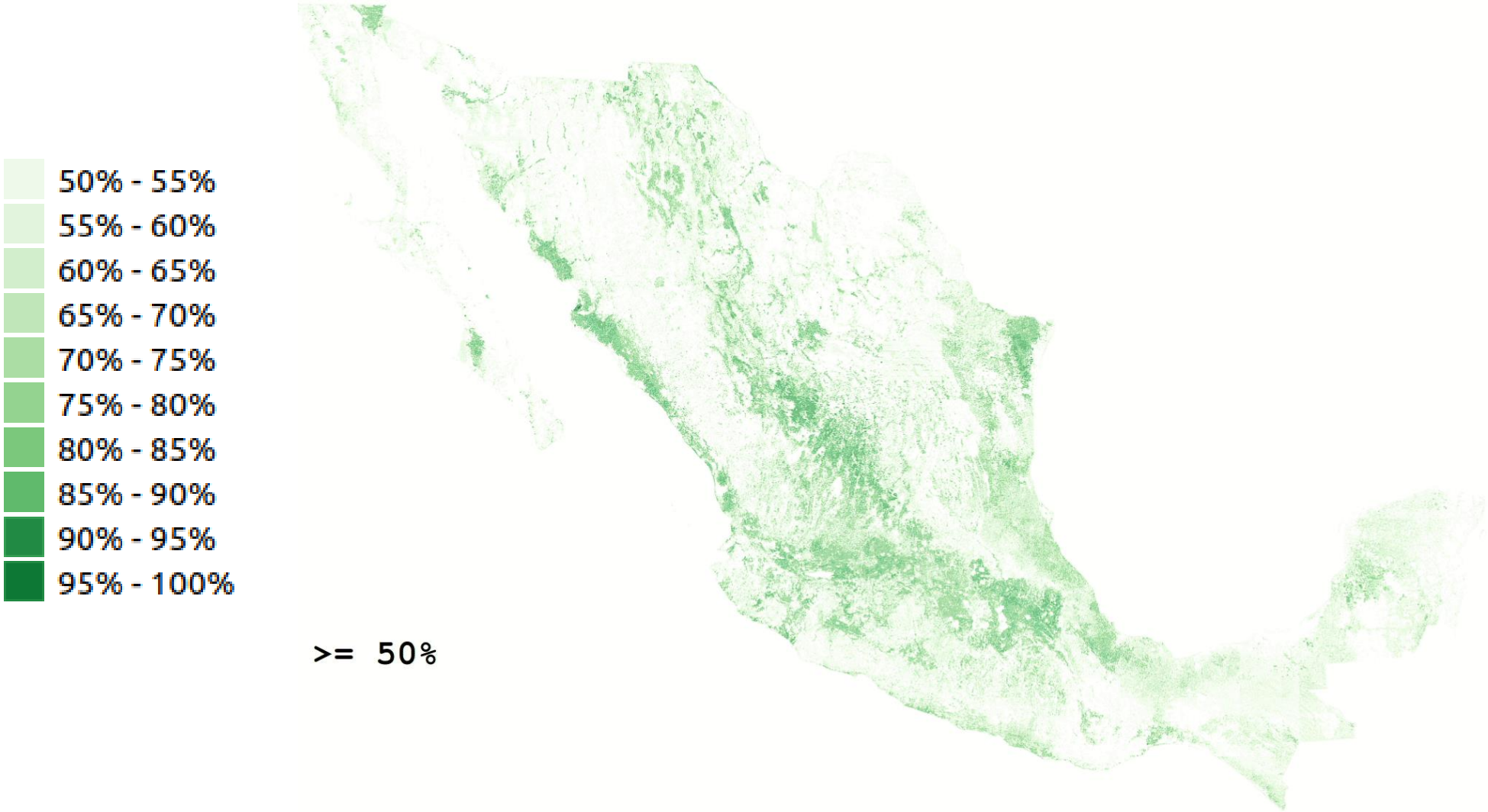
Results

Results

Landsat | Sentinel-2

AMCA 2016				AMCA-2016	First Iteration	Second Iteration	Third Iteration	Fourth Iteration	First Iteration Sentinel (2019)	First Iteration Sentinel (2021)
COD_ACT	DESCRIPTION	Hectares								
A	Completely agricultural	20,025,775.21								
C	At least 30% agricultural	10,978,329.61	AGRICULTURAL	31,376,931	55,797,681	43,915,831	40,326,623	38,806,173	32,457,571	33,241,915
M	Mixed	372,826.08								
F	Formerly agricultural	500,708.84								
N	No agricultural activity	11,766,281.74								
U	Urban	434,263.75								
V	Verified (no agricultural activity)	147,123,209.70	NOT AGRICULTURAL	160,636,548	139,432,164	151,314,013	154,903,222	156,423,671	162,772,274	161,987,929
W	Body of water	809,801.94								
B	Roads	1,649.32								
I	Flood zones	632.80								
	Accuracy				79%	80%	82%	83%	91%	*

Results



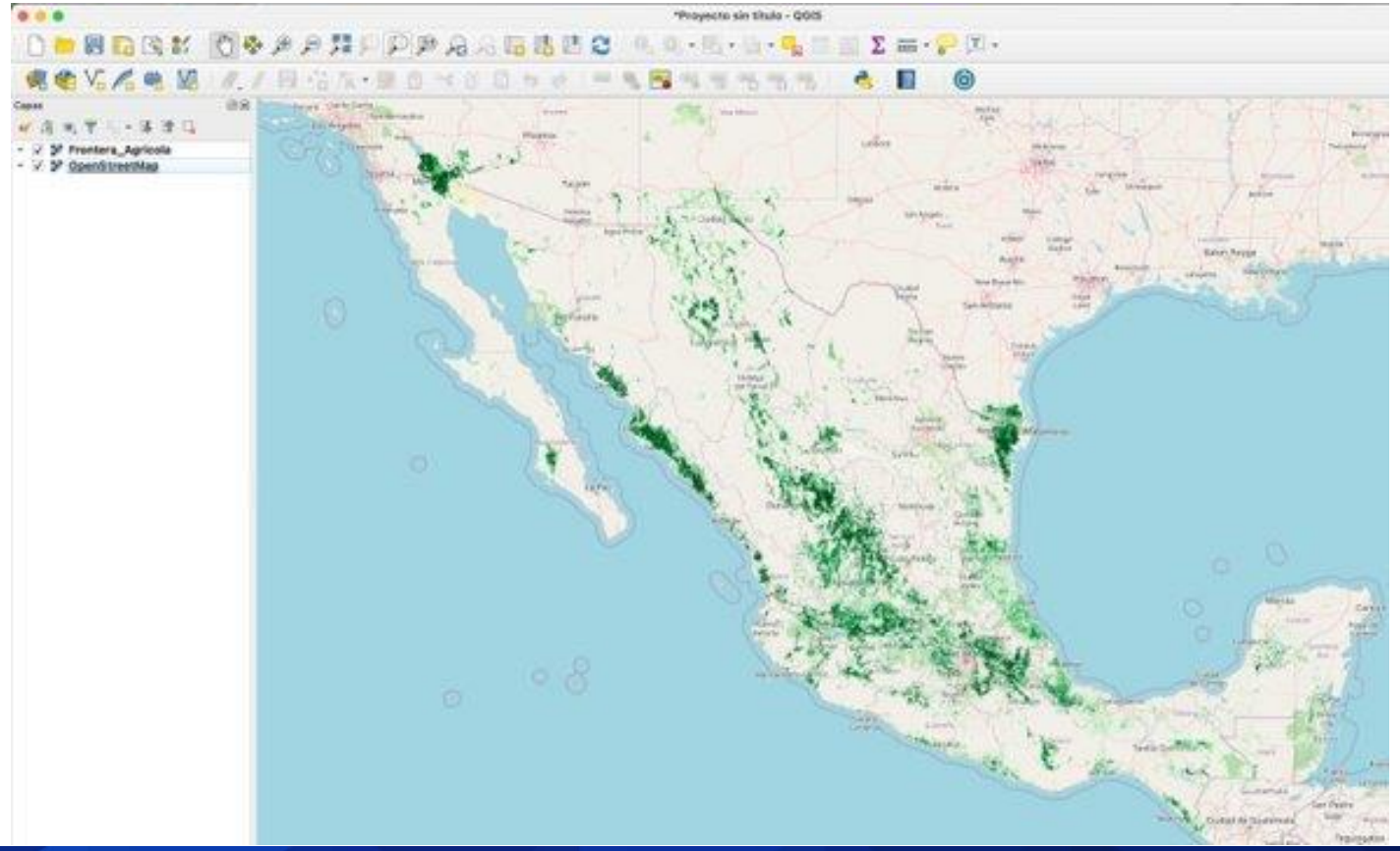
Classifier certainty measure	Hectares
0.50	39,174,530
0.55	33,241,915
0.60	28,185,354
0.65	23,832,196
0.70	20,105,247
0.75	16,745,200
0.80	13,752,280
0.85	10,979,343
0.90	8,229,267
0.95	5,243,554
1.00	1,181,944

Web App


INEGI Frontera Agricola
Find address or place
Layer List
Layers
Filtro Frontera Agricola Primavera 2021
Filtro Frontera Agricola Sin Localidad 2021 clasificador regional v2
Filtro Frontera Agricola 2021 clasificador regional v2
Frontera_Agricola_Primavera_2021
Frontera_Agricola_Sin Localidad_2021 clasificador_regional v2
Frontera_Agricola_2021_clasificador_regional_v 2
Frontera Agricola 2019
Frontera_Agricola_2021_clasificador_regional_v 1
Frontera_Agricola_2021_clasificador_general
-101.094 14.097 Degrees
Earthstar Geographics
esri

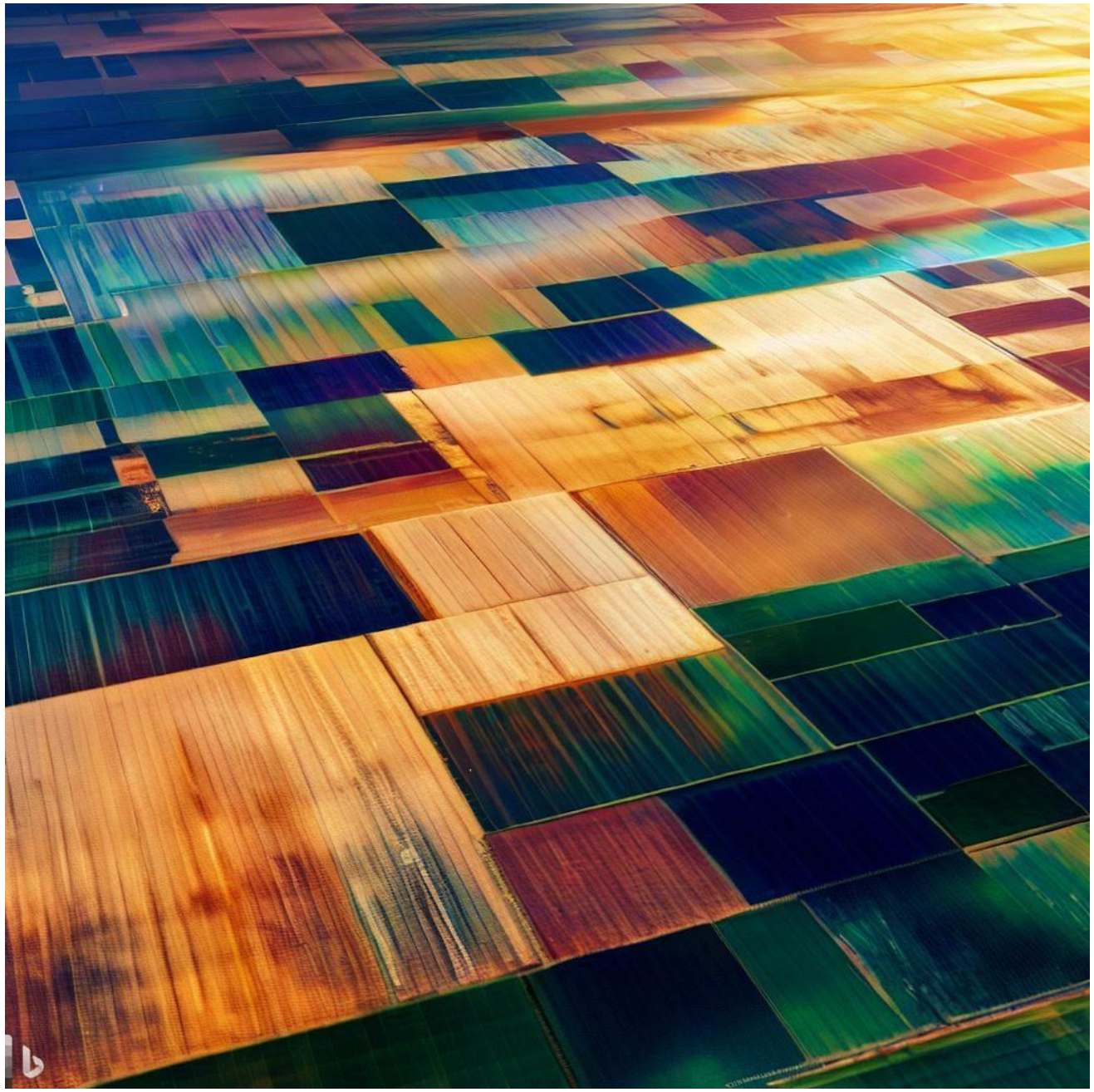
WMTS


Crear una nueva conexión WMS/WMTS
Detalles de la conexión
Nombre
Frontera Agrícola 2021 WMTS
URL
erver/rest/services/Frontera_Agricola/MapServer/WMTS?


*Proyecto sin título - QGIS
Capas
Frontera_Agricola
OpenStreetMap

Next Steps

- Process a new run with images corresponding to the census period, from October 2021 to September 2022.
- Compare the result of the new run with the results of the Census.
- Identify areas with good and poor algorithm performance
- Algorithm refinement from review results



Conclusions

Conclusions

Open Data

Big Data

Successful Collaboration

Replicable Methodology

Relevant Results

Scan me!

https://ecastats.uneca.org/regionalhub/

Thank you!

eduardo.vazquez@inegi.org.mx

Conociendo México

800 111 46 34
www.inegi.org.mx
atencion.usuarios@inegi.org.mx

INEGI Informa