10 March 2025

English

United Nations Group of Experts on Geographical Names
2025 session
New York, 28 April – 2 May 2025
Item 5 (b) of the provisional agenda *
Technical expertise: Geographical names data management

Utilization of deep learning and computer vision technology in geographical names management

Submitted by Indonesia **

Summary:

The Ministry of Public Works (MPW) actively contributes to implementing the Government Regulation No. 2 of 2021 on Geographical Names Standardization. This includes collecting, reviewing, and validating geographical names within MPW's jurisdiction. The validated data are submitted to the *Sistem Informasi Nama Rupabumi* (SINAR), managed by the Indonesian Geospatial Information Agency (*Badan Informasi Geospasial* - BIG) through an Application Programming Interface (API). The submission process ensures that MPW has thoroughly reviewed and validated all geographical names.

The workflow for geographical names management is governed by Standard Operating Procedures (SOP) of MPW No. 147/2024. It begins with collecting secondary data from data producers via APIs, followed by a detailed review and validation process, and culminates in field verification surveys regulated by MPW SOP No. 182/2024. The finalized draft of geographical names is formally signed by the Secretary General of MPW, establishing its legal status. This process is critical for ensuring the legitimacy and governance of geographical names under MPW's jurisdiction.

Given the large volume and variety of infrastructure data under the MPWs' jurisdiction, including 49 infrastructure categories and 72,603 data entities across Indonesia, using manual review would be time-consuming, inefficient as well as prone to errors. Development of the PNR.AI tool, leveraging deep learning and computer vision technologies, is an innovative solution to this challenge.

Given the extensive scope of MPW's infrastructure data—spanning 49 categories and comprising 72,603 data entities across Indonesia—manual review methods are inefficient, time-intensive, and prone to errors. To address this challenge, MPW has developed the PNR.AI tool, which integrates deep learning and computer vision technologies to streamline and enhance the process.

PNR.AI offers significant advantages in speed, accuracy, and efficiency. Capable of processing up to 10 times more data simultaneously than manual methods, it drastically reduces the

* GEGN.2/2025/1

** Prepared by Komang Sri Hartini, Amalia Siti Rohmah, and Khoirunnisaa Ronaa F, Indonesia, The Ministry of Public Work of Indonesia.

time required for infrastructure name validation. Moreover, its automated systems minimize human error, ensuring compliance with the legal requirements established by Government Regulation.

Utilization of deep learning and computer vision technology in geographical names management

Introduction

To accelerate the process of managing geographic names, especially in collecting and reviewing geographical names under the jurisdiction of The Ministry of Public Works (MPW), the Center for Data and Information Technology (*Pusat data dan Informasi*-PUSDATIN) developed the SRI-PU tool using deep learning technology that utilizes spatial data from Google Maps through the Google Cloud Platform (GCP) API. It is used to detect the shape and category of infrastructure. This infrastructure includes reservoirs, reservoirs, roads, bridges, landfills, etc. As of January 2025, the infrastructure of MPW has reached 72,603.

The computer vision technology uses the You Only Look at Once (YOLO) model. It is an advanced neural network used to detect objects in real time. This technology identifies the geographical names of infrastructure from images. The combination of these technologies offers a cutting-edge approach to process and validate infrastructure names while complying with the legal requirements outlined in the Head of Geospatial Agency Regulation No. 06 Year 2023. SRI-PU offers great benefits in terms of speed, accuracy, and efficiency, using several artificial intelligences such as computer vision, API and cloud connections, and artificial neural networks.

Geographical Names for MPW Infrastructure API Integration

API is a kind of connection that is connectional and secure. API conns secure. API connection right now is the best fit and an API is a collection of definitions and protocols that allow two software applications to communicate with each other. APIs provide a way for developers to access features or data from other applications, services, or hardware without having to understand the technical details behind how those features are implemented.

Based on the draft Decree of the Secretary-General number 616/KPTS/SJ/2023 on the Team for the Implementation of Earthly Names in the Field of Public Works and Public Housing, the review of earthly names of the Ministry of Public Works is carried out at the Center for Data and Information Technology (PUSDATIN), Secretariat General. The activity of reviewing geographical names has been carried out since 2022 and is still ongoing until 2024, with the following methods and work steps.

(i) The activity of reviewing the name of the geography of infrastructure starts from collecting infrastructure data from each data warehouse of each directorate general. Data collection is done using API so that updating the number of infrastructures can be done automatically. (ii) The next step is to review the name of the geographical name infrastructure for each infrastructure. Collectively the amount of data is more than 50 thousand data. (iii) The next stage is to confirm the data from the review of each work unit and field visit. (iv) Updating the data and legalizing the infrastructure name by making a decree.

Methodology

The activities of organizing geographical names at the MPW are regulated in SOP 147/SOP/BDI, hereinafter referred to as geographic naming activities, which have the following flow, the method to review geographical names of infrastructures carried out by internal CDIT with several category classes, then confirmation is carried out to each organizational unit. If confirmation is needed regarding further information, a survey is carried out based on SOP number 182/SOP/BDI. Updating and surveying further location data can be carried out independently by each organizational unit with a form that has been

prepared by CDIT. and the preparation of NR Decree legalized by the secretary general, as explained in the diagram below.

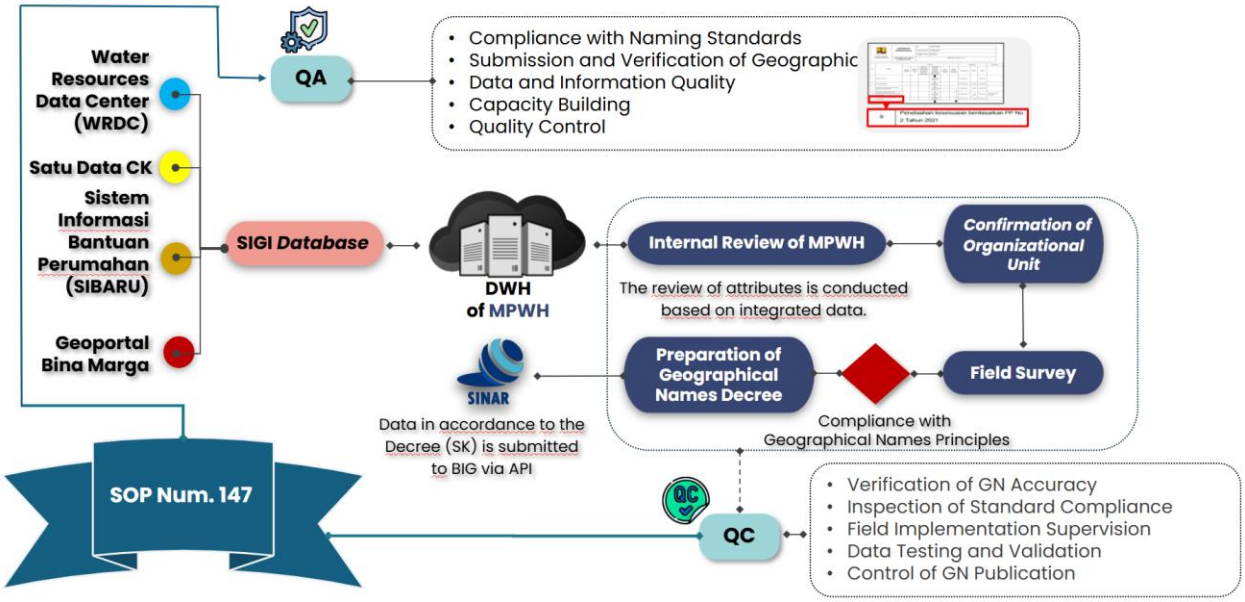


Figure 2. Flow to Review of Geographical Names of Infrastructures Carried Out by Internal CDIT

The diagram above is the flow of thinking of the internal review method by CDE. The concept is a standard reference in compiling the framework of the research process and the preparation of this SRI-PU. The following is the working method for the preparation of this SRI-PU.

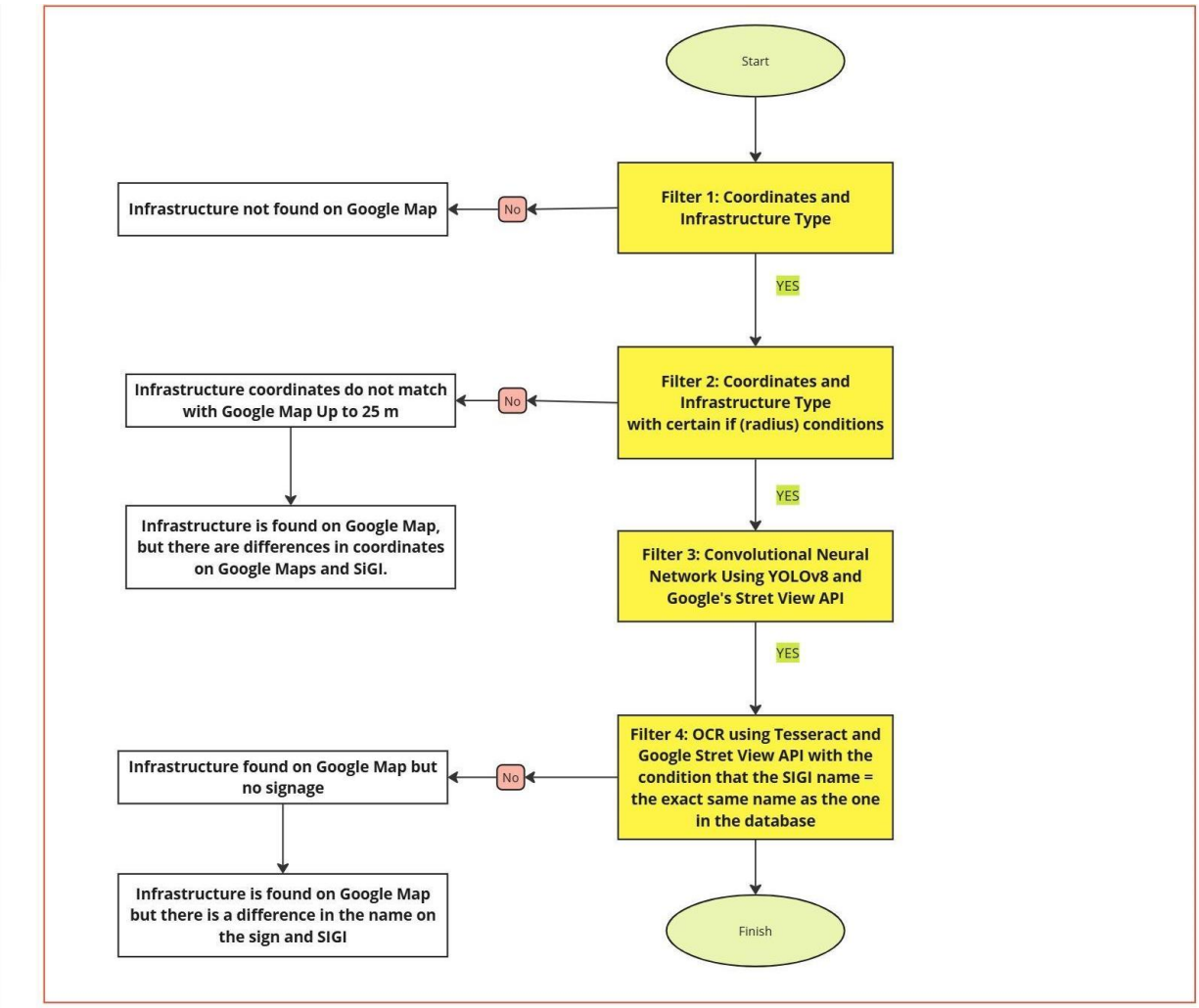


Figure 3. Mind Map of SRI-PU

The steps, methods, and workings will be explained in the description below.

1. Geolocation Connection API (Show endpoint from SIGI and the diagram of SIGI connection. To retrieve the infrastructure data from Warehouse data of MPWH. Autocorrection of the location between SIGI and Google Streetview using Geolocation and API cons. The result of differences in distance between the coordinates of latitude and Longitude will be fixed based on Google coordinates. The concept of this geo-correction is to connect the tagging name from Google Maps and match it with one of attributes from SIGI data, which is the “name of Infrastructure”. If the initial name of infrastructure does not match there are differences of more than 200 meters so that the data could be eliminated.
2. Stage of next elimination is about the correct coordinate being left and the left amount. The next step is to identify the location of infrastructures with technology of Deep Learning. The implementation of CNN is applied here. One of the architectures from CNN that is created especially to detect objects precisely and in real-time is YOLOv8.

The statistical tests for trust in the SRI-PU prototype are the Word Error Rate (WER) test (using Jitsi Word Error Recognition library) and the T-test which are used to test the work of the tesseract and artificial intelligence used. WER is a metric that measures the number of errors in OCR-generated text compared to text, with the Formula:

Formula WER:

$$WER = \frac{S + D + I}{N}$$

- S = Number of misrecognized words (Substitution).
- D = Number of missing words (Deletion).
- I = Number of incorrectly added words (Insertion).
- N = Number of words in the original text.

So, the conclusion is that Word Error Rate (WER) is a metric that measures the error in the OCR result text compared to the original text (ground truth). The lower the WER value, the more accurate the OC result.

Result

This Prototype generates a multiple-step process of the result. There are 54 types of infrastructure in the MPW and the data warehouse owned by the Directorates General of Water Resources, Human Settlements, and Highways. In this study, the infrastructure from the Directorate General of Water Resources, namely Reservoir, is used. In adjusting and implementing the standardization of SOPs that are poured into automation, the filter system used is as follows.

1. **Geopy: Distance Difference Elimination**

API connection between the Open API of the SIGI Geoportal and the Google Map API, to correct the coordinate data from the coordinates in SIGI to the Google coordinates. This coordinate correction has a constraint by calculating the difference if the difference is more than 25 meters (as a reference on a 1 scale map: 50.000) then it is considered compliant and if more than 25 does not meet. With the help of a Python library, namely geodesic. The Geodesic library in Python is part of *Geopy*, which is used to calculate the distance between two points on the earth's surface based on the ellipsoid model. This is more accurate than calculations using the Euclidean distance or Haversine formula because it considers the imperfect shape of the earth (ellipsoid, not a perfect sphere).

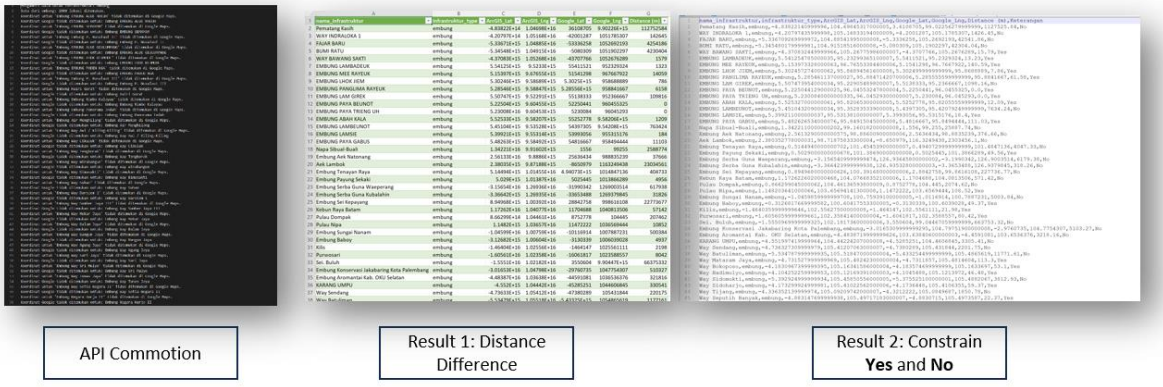


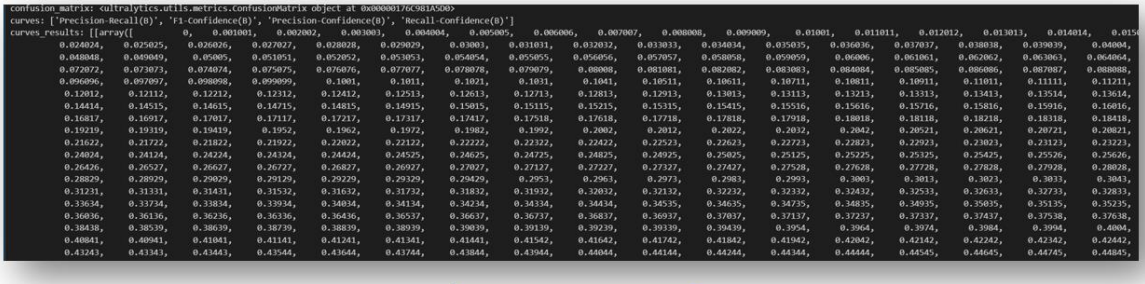
Figure 4. Result of *Geopy* library and API Connection

The resulting output is a .CSV file. The controlled variable is “nama_innfrastruktur” which is one of the attribute data in the database and is matched with location tagging and place names on Google. Because the simple concept applied is to calculate the coordinates of the destination with an ellipsoid earth model and WGS 84 datum, the library used is *geopy.distance.GeodesicDistance*. There are 3005 Reservoir data contained in SIGI, but until the first result only 1533 data can be found on Google Maps and only 821 data have precise coordinates.

2. **Deep Learning Concept: CNN and Computer Vision**

To find out that what is in the list of attributes “nama_infrastruktur” is an infrastructure in the form of a Reservoir, the implementation of Deep learning using computer vision technology is applied. In its application, the author needs to create a dataset commonly called Common Object in Context or COCO. The COCO dataset is created or collected to be trained to recognize and detect objects called reservoirs.

Reservoir COCO data is then trained with the concept of neural learning with a 50 -100 times epoch iteration concept. COCO data is usually in .yaml format. To be used as a YOLOv8 data model, it is necessary to train the COCO model with the composition of the Random Forest statistical concept. Using a composition of 75% training data and 25% test data and evaluating the model using a confusion matrix with accuracy and precision values respectively, the model is quite good at detecting Reservoir, with a precision of 79% and recall of 75%. More errors come from false negatives (undetected embankment) than false positives (wrong detection). The mAP50 (0.76) is quite good, but the mAP50-95 (0.45) can still be improved, suggesting the model struggles when IoU requirements are stricter. Inference speed is around 30ms, indicating the model is fast enough for detection in real-time scenarios, as shown in Figure 4 below.



Result 3: Confusion Matrix

	A	B	C	D	E	F	G	H	I
1	nama_infrastruktur	infrastruktur_type	ArcGIS_Lat	ArcGIS_Lng	Google_Lat	Google_Lng	Distance (m)	Keterangan	Verifikasi YOLO
2	WAY BAWANG SAKTI	embung	-4.37083E+15	1.05268E+16	-43707766	1052676289	1579	Yes	Benar Embung
3	EMBUNG LAMBADEUK	embung	5.54125E+15	9.5233E+15	55411521	952329324	1323	Yes	Benar Embung
4	EMBUNG MEE RAYEUK	embung	5.15397E+15	9.67655E+15	51541298	967667922	14059	Yes	Benar Embung
5	EMBUNG LHOK JIEM	embung	5.30246E+15	9.58689E+15	5.3025E+15	958688889	786	Yes	Benar Embung
6	EMBUNG PANGLIMA RAYEUK	embung	5.28546E+15	9.58847E+15	5.28556E+15	958841667	6158	Yes	Benar Embung
7	EMBUNG PAYA BEUNOT	embung	5.22504E+15	9.60455E+15	52250441	960455325	0	Yes	Benar Embung
8	EMBUNG PAYA TRIENG UH	embung	5.23008E+16	9.60453E+15	5230084	96045293	0	Yes	Benar Embung
9	EMBUNG ABAH KALA	embung	5.52533E+15	9.58207E+15	55252778	9.58206E+15	1209	Yes	Benar Embung
10	EMBUNG LAMSIE	embung	5.39921E+15	9.55314E+15	53993056	955315176	184	Yes	Benar Embung
11	EMBUNG PAYA GABUS	embung	5.48263E+15	9.58492E+15	54816667	958494444	11103	Yes	Benar Embung
12	Embung Payung Sekaki	embung	5.029E+15	1.01387E+16	5025445	1013866289	4956	Yes	Benar Embung
13	Pulau Nipa	embung	1.1482E+15	1.03657E+16	11472222	1036569444	10852	Yes	Benar Embung
14	Embung Baboy	embung	-3.12682E+15	1.00604E+16	-3130339	1006039028	4937	Yes	Benar Embung
15	Kilis	embung	-1.46404E+16	1.02556E+16	-1464147	1025561111	2198	Yes	Benar Embung
16	Purwosari	embung	-1.60561E+16	1.02358E+16	-16061817	1023588557	8042	Yes	Benar Embung
17	Way Mataram Jaya	embung	-4.73153E+15	1.05482E+16	-47311857	1054814604	1133	Yes	Benar Embung
18	Way Bokoposo	embung	-4.1831E+16	1.05163E+16	-4.18357E+15	1051633697	531	Yes	Benar Embung
19	Way Hadimulyo	embung	-4.10425E+15	1.05122E+16	-41045488	1051213972	4648	Yes	Benar Embung
20	Way Sidoharjo	embung	-4.1733E+15	1.0541E+16	-41736448	1054106355	5937	Yes	Benar Embung
21	Way Seputih Banyak	embung	-4.88315E+15	1.05497E+16	-48830715	1054973587	2237	Yes	Benar Embung
22	Way Lahat 1	embung	-4.35283E+15	1.05174E+16	-4352684	10517453	3425	Yes	Benar Embung
23	Way Pagar Jaya	embung	-4.37327E+15	1.05157E+16	-43727931	1051569158	5277	Yes	Benar Embung
24	Kota Baru	embung	-5.30006E+15	1.05415E+16	-5.29863E+15	1054145154	18418	Yes	Benar Embung
25	Way Kramat	embung	-4.80209E+15	1.0493E+16	-48022145	1049295556	2891	Yes	Benar Embung
26	Way Sinar Karya	embung	-5.43366E+15	1.05487E+16	-5.43325E+15	1054865619	4482	Yes	Benar Embung
27	Solo Valley Tapelan	embung	-7.227E+14	1.11555E+16	-7.22746E+15	1115528027	21367	Yes	Benar Embung
28	Petirsari	embung	-8.12541E+15	1.10823E+16	-81277219	11082223	2577	Yes	Benar Embung
29	Embung Titang Krajan	embung	-7.6808E+15	1.11316E+16	-76811483	1113162299	4246	Yes	Benar Embung
30	Karangtengah	embung	-7.27639E+15	1.10506E+16	-7.277E+15	1105065997	7198	Yes	Benar Embung
31	Embung Digal	embung	-8.12596E+15	1.10806E+16	-8.12698E+15	1108056905	11596	Yes	Benar Embung
32	Taman Arum	embung	-7.72765E+15	1.11344E+16	-7.7281E+15	111344092	6032	Yes	Benar Embung
33	Embung Sumber Suro	embung	-8.2009E+15	1.10984E+16	-82010489	1109835742	3257	Yes	Benar Embung
34	Sambeng Lamongan	embung	-7.32183E+15	1.12229E+16	-7.3214E+15	1122290094	483	Yes	Benar Embung
35	Musuk 1	embung	-7.53475E+15	1.10547E+16	-75347109	11054704	433	Yes	Benar Embung
36	Selotinatah	embung	-7.68667E+15	1.11319E+16	-7.68621E+15	1113168203	23369	Yes	Benar Embung
37	Embung Brangkal	embung	-7.18977E+15	1.12091E+16	-71896314	112091399	1677	Yes	Benar Embung
38	Embung Temboro	embung	-7.58716E+15	1.11392E+16	-758707	1113917344	1468	Yes	Benar Embung

Result 4: Result of each infrastructure Dam Identification

Figure 5.

The figure above also shows the result of Deep learning from the implementation of YOLOv8. There is an additional column on the right. The algorithm used is that if the YOLOv8 model finds that the coordinate point is in the form of infrastructure in the form of a reservoir then it will identify whether it is a reservoir or not. Then the output is the right column which is a statement that it is a Reservoir or not.

3. Optical Character Recognition (OCR) Implementation: Tesseract OCR

The final result of the review of the MPW Infrastructure Name is to identify whether the name in the “nama_infrastructure” attribute is the same as the name sign or information board of the place. As examples are shown in the image below.

Figure 6. Various Name Signs

OCR technology is used to recognize and extract Text from documents/images. In this research, the author uses OCR in character recognition, namely Tesseract OCR. The existing name signs have various shapes like the example above, so thresholding is not enough. This character recognition Tesseract technology, implemented on the identified infrastructure in the form of ponds (in result 4), is linked to the Google Streat View API. Tesseract OCR uses the Long Short-Term Memory (LSTM) model to recognize characters, which have previously been trained by machines to recognize many characters.[3]

The overall result of the process called SRI-PU is shown in the figure below. For one infrastructure, it will produce a Text/CSV file with information on whether the infrastructure name has a name sign under the name of the infrastructure in the SIGI data.

	A	B	C	D	E	F	G	H	I	J	K	L
1	nama_infrastruktur	infrastruktur_type	ArcGIS_Lat	ArcGIS_Lng	Google_Lat	Google_Lng	Distance (m)	Keterangan	Verifikasi YOLO	Verifikasi Nama	Nama yang ditemukan sekitar	WER
2	WAY BAWANG SAKTI	embung	-4.37083E+15	1.05268E+16	-43707766	1052676289	1579	Yes	Benar Embung	Tidak Cocok	Embung Bawang Tirto Mulyo	1.33333E+16
3	EMBUNG LAMBADEUK	embung	5.54125E+15	9.5233E+15	55411521	952329324	1323	Yes	Benar Embung	Tidak Cocok	Embung Lambadeuk	10
4	EMBUNG MEE RAYEUK	embung	5.15397E+15	9.67655E+15	51541298	967667922	14059	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
5	EMBUNG LHOK JIEM	embung	5.30246E+15	9.58689E+15	5.3025E+15	958688889	786	Yes	Benar Embung	Tidak Cocok	Embung Lhok Jiem	10
6	EMBUNG PANGLIMA RAYEUK	embung	5.28546E+15	9.58847E+15	5.28556E+15	958841667	6158	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
7	EMBUNG PAYA BEUNOT	embung	5.22504E+15	9.60455E+15	52250441	960455325	0	Yes	Benar Embung	Tidak Cocok	Embung Paya Beunot	10
8	EMBUNG PAYA TRIENG UH	embung	5.23008E+16	9.60453E+15	5230084	96045293	0	Yes	Benar Embung	Tidak Cocok	Embung Paya Trieng Uh, Kantor keuchik Mns ujong leubat	225
9	EMBUNG ABAH KALA	embung	5.52533E+15	9.58207E+15	55252778	958205556	1209	Yes	Benar Embung	Tidak Cocok	Embung Abah Kala	10
10	EMBUNG LAMSIE	embung	5.39921E+15	9.55314E+15	53993056	955315176	184	Yes	Benar Embung	Tidak Cocok	Embung Lamsie	10
11	EMBUNG PAYA GABUS	embung	5.48263E+15	9.58492E+15	54816667	958494444	11103	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
12	Embung Payung Sekaki	embung	5.029E+15	1.01387E+16	5025445	1013866289	4556	Yes	Benar Embung	Tidak Cocok	EMBUNG PAYUNG SEKAKI	10
13	Pulau Nipa	embung	1.1482E+15	1.03657E+16	11472222	1036569444	10852	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
14	Embung Baboy	embung	-3.12682E+15	1.00604E+16	-3130339	1006039028	4937	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
15	Kilis	embung	-1.46404E+16	1.02556E+16	-1464147	1025561111	2198	Yes	Benar Embung	Tidak Cocok	Tidak Ada	20
16	Purwosari	embung	-1.60561E+15	1.02358E+16	-16061817	1023588557	8042	Yes	Benar Embung	Tidak Cocok	GRAND AQUATIC PARK(wisata SUHA), Green Aquatic park(tempat wisat	80
17	Way Mataram Jaya	embung	-4.71153E+15	1.05482E+16	-47111857	1054814604	1133	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
18	Way Bokoposo	embung	-4.1831E+14	1.05163E+16	-4.18357E+15	1051633697	531	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
19	Way Hadimulyo	embung	-4.10425E+15	1.05122E+16	-41045488	1051213972	4648	Yes	Benar Embung	Tidak Cocok	Embung Way Hadimulyo	5
20	Way Sidoharjo	embung	-4.1733E+15	1.0541E+16	-41736448	1054106355	5937	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
21	Way Seputih Banyak	embung	-4.88315E+15	1.05497E+16	-48830715	1054973587	2237	Yes	Benar Embung	Tidak Cocok	Embung Way Seputih Banyak	3.33333E+15
22	Way Lahat 1	embung	-4.35283E+15	1.05174E+16	-4352684	10517453	3425	Yes	Benar Embung	Tidak Cocok	Embung Way Lahat	6.66667E+15
23	Way Pagar Jaya	embung	-4.37327E+15	1.05157E+16	-43727931	1051569158	5277	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
24	Kota Baru	embung	-5.30006E+15	1.05415E+16	-5.29863E+15	1054145154	18418	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
25	Way Kramat	embung	-4.80209E+15	1.0493E+16	-48022145	1049295556	2891	Yes	Benar Embung	Tidak Cocok	Embung Way Kramat	5
26	Way Sinar Karya	embung	-5.43366E+15	1.05487E+16	-5.43325E+15	1054865619	4482	Yes	Benar Embung	Tidak Cocok	Embung Way Sinar Karya	3.33333E+15
27	Solo Valley Tapelan	embung	-7.227E+14	1.11555E+16	-7.22746E+15	1115528027	21367	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
28	Petirsari	embung	-8.12541E+15	1.10823E+16	-81277219	11082223	2577	Yes	Benar Embung	Tidak Cocok	Tidak Ada	20
29	Embung Titang Krajan	embung	-7.6808E+15	1.11316E+16	-76811483	1113162299	4246	Yes	Benar Embung	Tidak Cocok	BENDUNGAN BALEASRI, EMBUNG TITANG KRAJAN	1.66667E+16
30	Karangtengah	embung	-7.27639E+15	1.10506E+16	-7.277E+15	1105065997	7198	Yes	Benar Embung	Tidak Cocok	Tidak Ada	20
31	Embung Digal	embung	-8.12596E+15	1.10806E+16	-8.12698E+15	1108056905	11596	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
32	Taman Arum	embung	-7.72765E+15	1.11344E+16	-7.7281E+15	111344092	6032	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
33	Embung Sumber Suro	embung	-8.2009E+15	1.10984E+16	-82010489	1109835742	3257	Yes	Benar Embung	Tidak Cocok	Gubuk asmoro kopian, Bendungan kali suro, Lapangan bola voli Tirto sur	50
34	Sambeng Lamongan	embung	-7.32183E+15	1.12229E+16	-7.3214E+15	1122290094	483	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
35	Musuk 1	embung	-7.53475E+15	1.10547E+16	-75347109	11054704	433	Yes	Benar Embung	Tidak Cocok	Embung musuk 1, Pemancingan Embung musuk, Warung Makan Embung	55
36	Selotinatah	embung	-7.68667E+15	1.11319E+16	-7.68621E+15	1113168203	23369	Yes	Benar Embung	Tidak Cocok	Tidak Ada	20
37	Embung Brangkal	embung	-7.18977E+15	1.12091E+16	-71896314	112091399	1677	Yes	Benar Embung	Tidak Cocok	Embung brangkal	5
38	Embung Temboro	embung	-7.58716E+15	1.11392E+16	-758707	1113917344	1468	Yes	Benar Embung	Tidak Cocok	Embung Temboro, Waduk trangkil	15
39	Ciroyom	embung	-7.5468E+15	1.0868E+16	-75467412	1086796637	1082	Yes	Benar Embung	Tidak Cocok	Tidak Ada	20
40	Bukit Japati	embung	-7.59913E+16	1.087E+16	-7.59914E+15	108699577	327	Yes	Benar Embung	Tidak Cocok	Embung bukit japati	15
41	Tunggilis	embung	-7.61501E+15	1.08711E+16	-7615882	1087098432	14143	Yes	Benar Embung	Tidak Cocok	Tidak Ada	20
42	Hanum	embung	-7.24089E+15	1.08621E+16	-7.2409E+15	1086215543	7447	Yes	Benar Embung	Tidak Cocok	Tidak Ada	20
43	Kasur	embung	-6.76488E+15	1.11678E+16	-6.76622E+15	1116796682	28567	Yes	Benar Embung	Tidak Cocok	Tidak Ada	20
44	Sumber	embung	-6.78975E+15	1.11262E+16	-6.78961E+15	1112620756	1565	Yes	Benar Embung	Tidak Cocok	Embung Sumber, Lapangan Volley Ball Bulak Sempu, BASVO Club	90
45	Tambak Agung	embung	-6.71637E+15	1.11243E+16	-6.71638E+15	11124259	2025	Yes	Benar Embung	Tidak Cocok	Embung Tambak Agung	5
46	Tlogomojo	embung	-6.72024E+15	1.11407E+16	-67210711	1114060466	14582	Yes	Benar Embung	Tidak Cocok	Tidak Ada	20
47	Kulur	embung	-6.90475E+15	1.1142E+16	-69047196	1114198582	403	Yes	Benar Embung	Tidak Cocok	Embung Kulur, Pos pantau embung Kulur	50
48	Sonokidul	embung	-7.09353E+15	1.11255E+16	-70932007	1112552412	3777	Yes	Benar Embung	Tidak Cocok	Embung Sonokidul	10
49	Cokrowati	embung	-6.94738E+15	1.11192E+16	-6947346	1111921147	1223	Yes	Benar Embung	Tidak Cocok	EMBUNG COKROWATI, Tukang listrik cokrowati, Rumah santai, PosAjal (	100
50	Kemiri (Kunduran)	embung	-7.08886E+15	1.11277E+16	-70905594	1112755598	27461	Yes	Benar Embung	Tidak Cocok	Tidak Ada	10
51	Kropak	embung	-7.1146E+15	1.11097E+16	-71146395	1110971987	2477	Yes	Benar Embung	Tidak Cocok	Toko sembako Bu Sarmi, Embung Kropak	50
52	Jurangjero (Bogorejo)	embung	-6.90836E+15	1.11508E+16	-6.90844E+15	1115083032	4425	Yes	Benar Embung	Tidak Cocok	Embung Jurangjero	10
53	Ketip	embung	-6.73265E+15	1.11169E+16	-67305807	1111697399	24633	Yes	Benar Embung	Cocok	Pkd Ketip, Jwn.secondhand, TK. Kartini 02 Ketip	60
54	Triguno	embung	-6.81795E+15	1.11133E+16	-68180396	1111334098	6794	Yes	Benar Embung	Tidak Cocok	Tidak Ada	20

Figure 7. Final Result

The picture above shows the final results of this study. From the file, it is explained that there are 2 conditions, namely “suitable” and “not suitable” which shows that if the name sign is the same as the

infrastructure, it will produce a suitable description and vice versa. Then this LSTM model can also provide information on the name signposts around the Reservoir within a 500 m radius. This can be a consideration for the Canter for Data and Information Technology of the MPW in deciding. The statistics accuracy used WER with a result of around 8,3%.

Discussion

The COCO training dataset used in the YOLO version 8 data model needs to be further developed in terms of epochs and the diversity of types of reservoirs in Indonesia. Indeed, infrastructure in Indonesia can be very diverse (there are 54 types of infrastructure but as a prototype test, the first infrastructure tested is a reservoir), so it is necessary to train a diverse dataset to more precisely recognize the selected infrastructure. In addition, the filters used in providing filter and edge detection as well as the RNN method used in LSTM for character recognition are very limited so it can cause ambiguity.

Statistical tests of the prototype results need to be added so that the WER value is <10%. For very diverse data and a short time, achieving the WER value limit below 10% requires training more and diverse OCR and LSTM tesseracts. There are several concerns in developing this prototype.

Availability of satellite image data. The weakness of remote sensing interpretation using Google images is the limitation of detection of objects that are covered by lush and high vegetation, the confidence value of successful object detection and name tagging is not necessarily a system error but can also be a Google location tagging error. The development direction of SRI-PU is expected to help review PU infrastructure data automatically for all types of infrastructure, namely Bina Marga, Cipta Karya, and Water Resources. This research and prototype can greatly assist the infrastructure name review work at the MPW.

Conclusion

1. Deep learning technology can be used to automatically review the names of MPW infrastructure, making it more efficient which was originally at a speed of 500 infrastructures in 1 month with SRI-PU can increase time efficiency by completing 500 infrastructures within 20 minutes or increasing by 250%, accurate, and fast in reviewing thousands of data. But besides that, the results still need to be trained and developed to be more accurate and precise.
2. The results of the review still require further scrutiny, namely in the form of additional filters in the character settings that will reduce the WER number because there are so many different nameplates used to name a location. In addition, the factors of code preparation and filter selection need to be continuously developed so that the statistical value is close to perfect
3. This prototype has a 79% confidence value for detecting images which means that the false positive value detected is only 20%. Geolocation confidence is around 97% which means that autocorrection is trusted. The WER confidence value is 8.3% which means that the model can be used for object detection and name tagging of MPW infrastructure automatically so that it can save time and budget more than 75%. The test of WER number from SRI-PU value is below 10% and the prototype is ready to develop.

Glosarium

MPW	= Ministry of Public Work
CDIT	= Cender of Data and Information Technology
SIGI	= Sistem Informasi Geospasial Infrastruktur
ML	= Machine Learning
DL	= Deep Learning
CNN	= Convolutional Neural Network
YOLO	= You Only Look at Once
COCO	= Common Objects in Context
WER	= Word of Error
OCR	= Optical Character Recognition
LSTM	= Long Short Term Memory
PNR.AI	= *Penelaahan Nama Rupabumi* (Geographical Names Verification) with Artificial Intelligence
SRI.PU	= Smart Review Infrastructure *Pekerjaan Umum* (The Ministry of Public Works-MPW)