



KOMPSAT-3/3A 광학 위성영상으로부터 산업단지 및 채석장 분류를 위한 GeoAI 데이터셋

박채원^{1,2} , 정형섭^{3,4,5,*} , 이원진⁶ , 이광재⁷ , 오관영⁸ , 장재영⁸ , 이명진⁹ , 한근혁¹⁰ , 최일훈¹¹

¹석사과정생, 서울시립대학교 공간정보공학과, 서울특별시 동대문구 서울시립대로 163, 02504, 대한민국

²석사과정생, 서울시립대학교 스마트시티학과, 서울특별시 동대문구 서울시립대로 163, 02504, 대한민국

³교수, 서울시립대학교 공간정보공학과, 서울특별시 동대문구 서울시립대로 163, 02504, 대한민국

⁴교수, 서울시립대학교 스마트시티학과, 서울특별시 동대문구 서울시립대로 163, 02504, 대한민국

⁵방문연구원, 서던 메소디스트 대학교 지구과학과, 6425 Boaz Lane, Dallas, 75205 TX, 미국

⁶환경연구원, 국립환경과학원 환경위성센터, 인천광역시 서구 환경로 42, 22689, 대한민국

⁷책임연구원, 한국항공우주연구원 위성활용부, 대전광역시 유성구 과학로 169-84, 34133, 대한민국

⁸선임연구원, 한국항공우주연구원 위성활용부, 대전광역시 유성구 과학로 169-84, 34133, 대한민국

⁹연구위원, 한국환경연구원 환경계획연구실, 세종특별자치시 시청대로 370, 30147, 대한민국

¹⁰이사, 네이버시스템(주), 서울특별시 송파구 중대로 135, 05717, 대한민국

¹¹상무, 네이버시스템(주), 서울특별시 송파구 중대로 135, 05717, 대한민국

GeoAI Dataset for Industrial Park and Quarry Classification from KOMPSAT-3/3A Optical Satellite Imagery

Che-Won Park^{1,2}, Hyung-Sup Jung^{3,4,5,*}, Won-Jin Lee⁶, Kwang-Jae Lee⁷, Kwan-Young Oh⁸, Jae-Young Chang⁸, Moun-jin Lee⁹, Geun-Hyouk Han¹⁰, and Il-Hoon Choi¹¹

¹Master Student, Department of Geoinformatics, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, 02504 Seoul, South Korea

²Master Student, Department of Smart Cities, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, 02504 Seoul, South Korea

³Professor, Department of Geoinformatics, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, 02504 Seoul, South Korea

⁴Professor, Department of Smart Cities, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, 02504 Seoul, South Korea

⁵Hamilton Visiting Fellow, Department of Earth Sciences, Southern Methodist University, 6425 Boaz Lane, Dallas, 75205 TX, USA

⁶Senior Researcher, Environmental Satellite Center, National Institute of Environmental Research, Hwangeong-ro 42, Seo-gu, 22689 Incheon, South Korea

⁷Principal Researcher, Satellite Application Center, Korea Aerospace Research Institute, 169-84 Gwahak-ro, Yuseong-gu, 34133 Daejeon, South Korea

⁸Senior Researcher, Satellite Application Center, Korea Aerospace Research Institute, 169-84 Gwahak-ro, Yuseong-gu, 34133 Daejeon, South Korea

⁹Research Fellow, Division for Environmental Planning, Korea Environment Institute, 370 Sicheong-daero, 30147 Sejong, South Korea

¹⁰Director, Neighbor System, 135 Jungdae-ro, Songpa-gu, 05717 Seoul, South Korea

¹¹Managing Director, Neighbor System, 135 Jungdae-ro, Songpa-gu, 05717 Seoul, South Korea

Received November 28, 2023

Accepted December 18, 2023

Corresponding Author

Hyung-Sup Jung

Tel: +82-2-6490-2892

E-mail: hsjung@uos.ac.kr

Air pollution is a serious problem in the world, and it is necessary to monitor air pollution emission sources in other neighboring countries to respond to the problem of air pollution spreading across borders. In this study, we utilized domestic and international optical images from KOMPSAT-3/3A satellites to build an AI training dataset for classifying industrial parks and quarries, which are representative sources of air pollution emissions. The data can be used to identify the distribution of air pollution emission sources located at home and abroad along with various state-of-the-art models in the image segmentation field, and is expected to contribute to the preservation of Korea's air environment as a basis for establishing air-related policies.

Keywords: Satellite imagery; Deep learning; Semantic segmentation; Industrial park; Quarry

1. 서론

대기오염은 그 영향이 국경을 넘어 확산되는 월경성 특성을 가지고 있어, 국내뿐만 아니라 인접 국가들에 대한 철저한 모니터링이 요구된다. 특히, 대한민국의 경우 대기오염물질 농도에서 국외 요인의 영향이 약 40-70%로 높게 추정되고 있어 (Han et al., 2018; Kumar et al., 2021) 국제적 차원에서의 대기오염 문제 해결이 필요하다. 이러한 상황에서 해외 대기오염물질 배출원의 정확한 위치 정보는 오염물질의 이동과 분포를 이해하고, 원인을 추적하는 데 핵심적인 자료로 활용될 수 있다. 그러나 해외 대기오염 배출원에 대한 정보 획득은 물리적인 거리 및 접근성 제한과 같은 요소로 인해 상당한 어려움이 존재한다.

위성 원격탐사는 접근이 어렵거나 불가능한 대상에 대해 원격으로 정보를 획득하는 기술이다. 이는 접근성 제한, 물리적 거리, 또는 다른 제약 조건으로 인해 직접 탐사가 어려운 지역에 대한 정보 수집에 있어 매우 유용하다. 이와 같은 특성은 위성 원격탐사가 대기오염 배출원 데이터 수집 및 관리 분야에서 핵심적인 역할을 수행할 수 있음을 시사한다. 한편 위성 원격탐사 자료는 최근 인공지능 기법과의 접목을 통해 다양한 연구 분야에서 향상된 결과를 도출하고 있다(Kim et al., 2022). 위성영상으로부터 인공지능 기술을 적용하여 대기오염물질 배출원을 탐지한 사례는 거의 드물지만, 토지 피복이나 도시기능 지역 분할 연구에서 배출원과 유사한 객체를 효과적으로 구분한 연구 사례는 다수 존재한다(Baek et al., 2022; Zhang et al., 2018). 이러한 선행 연구를 바탕으로, 최신 인공지능 기술과 위성자료를 함께 활용한다면, 대기오염 배출원에 해당하는 객체를 보다 효과적으로 식별 가능할 것으로 예상된다.

이에 따라 본 연구에서는 KOMPSAT-3/3A 광학위성으로부터 촬영된 국내의 위성영상을 활용하여, 대표적인 대기오염 배출원에 해당하는 산업단지 및 채석장을 분류하기 위한 AI 학습데이터셋을 구축하였다. 산업단지는 굴뚝을 통해 대량의 오염물질을 배출하며, 이는 대기오염 영향 분석에 있어 매우 중요한 오염원으로 간주된다. 또한 채석장은 채석 작업에서 발생하는 분진을 여과 시설 없이 대기 중으로 직접 배출하기 때문에, 탐지 필요성이 높은 오염원으로 인식된다. 구축된

산업단지 및 채석장 분류 데이터셋을 영상 분할 분야에서 뛰어난 성능을 보인 U-Net 네트워크에 적용하여 결과를 평가하였다.

2. 연구 자료

본 연구에서는 산업단지 및 채석장 분류를 위한 GeoAI 데이터셋 제작을 목표로 KOMPSAT-3/3A 위성영상을 수집하였다. 정확한 위치 정보를 가진 대한민국과 상대적으로 위치 정보가 부족한 주변 국가인 태국을 대상 지역으로 선정하여, 2021년부터 2022년 사이에 촬영된 영상 중 운량이 10% 이하인 영상을 선별적으로 수집하였다.

KOMPSAT-3/3A는 한국항공우주연구원에서 개발하고 운용 중인 국내 광학위성에 해당한다. KOMPSAT-3와 KOMPSAT-3A는 각각 0.7 m와 0.55 m에 해당하는 고해상도 광학영상을 제공한다. 본 연구에서는 해당 위성들의 특정 파장 대역을 활용하였으며, 이와 관련된 세부 사항은 Table 1에 나타나 있다.

3. 데이터 제작 방법

3.1 데이터 정제

KOMPSAT-3/3A 위성의 원시 데이터에 팬 샤프닝(pan-sharpening) 기술을 적용하여 고해상도 컬러 영상을 제작하였다. 이후 공간 해상도가 다른 두 종류의 센서 영상들을 동일한 딥러닝 모델에 적용하기 위해, 모든 영상을 1 m 해상도로 리샘플링하였다. 리샘플링 과정은 영상이 일관된 품질을 유지하도록 보장하며, 이는 모델 학습의 정확도 및 신뢰

Table 1. Spectral bands used in the study

Sensor	KOMPSAT-3	KOMPSAT-3A
Multispectral resolution (m)	2.8	2.2
Panchromatic resolution (m)	0.7	0.55
Wavelength (nm)		
Panchromatic	450-900	450-900
Blue	450-520	450-520
Green	520-600	520-600
Red	630-690	630-690

성을 향상하는 데 중요한 역할을 하여 필수적이다. 리샘플링된 영상에서 필요한 밴드(B, G, R)를 추출하여 이를 16비트의 Geo TIFF 형식으로 병합하는 작업을 수행하였다. 또한, 산업단지 및 채석장 위치와 크기를 기준으로 영상을 2,048×2,048 px² 또는 9,216×9,216 px² 크기로 잘라내어 원천 데이터를 정의하였다.

3.2 데이터 가공

본 연구에서 라벨 데이터는 course annotation 방식을 사용하여 산업단지와 채석장의 경계를 대략적으로 정의하였다. Fig. 1은 본 연구에서 제작한 산업단지와 채석장 라벨 데이터 예시를 나타낸다. 산업단지는 그 경계를 기준으로 내부 영역을 구획하였다. 이 과정에서 4차선 이상의 도로, 외곽 가로수, 초지는 제외하였으나 주차장은 포함하여 구획하였다. 채석장은 구조물을 기준으로 나지와 구분하여 구획하였으며, 근처에 위치한 시멘트 공장 등 사업장에 해당하는 객체는 산업단지로 분류하였다. 구획된 산업단지, 채석장 그리고 비대상지 영역에 순서대로 분류 코드 1, 2, 0을 할당하고 TIFF 형식으로 변환하였다.

3.3 AI 데이터셋 정의

데이터 정제 및 가공 단계를 거쳐 제작된 원천 데이터와 라

벨 데이터를 딥러닝 학습에 적합한 512×512 px² 크기의 패치 데이터로 변환하였다. 패치 크기로 절단된 이미지 데이터는 99% 최대 최소 정규화 수식(Eq. 1) 적용하여, 각 이미지 패치마다 0에서 1 사이의 동일한 값의 분포를 갖도록 변환하였다.

$$x' = \frac{x - \min_{0.99}(x)}{\max_{0.99}(x) - \min_{0.99}(x)} \quad [\text{Eq. 1}]$$

변환된 학습 데이터를 훈련 및 테스트 목적으로 활용하기 위해 7 대 3의 비율로 분할하였다. 이때 훈련 데이터는 촬영 국가와 클래스에 따라 수량적인 불균형을 포함한다. 이는 데이터가 대한민국과 태국의 두 지역에서 수집되었으며, 각 지역마다 산업단지와 채석장이 차지하는 비율이 원래부터 서로 다르기 때문이다. 본 연구에서는 촬영 국가와 클래스별 데이터 불균형 문제를 해결하고자, 오버 샘플링과 데이터 증강 기법을 사용하였다. 이를 위해, 모든 데이터를 산업단지, 채석장, 산업단지-채석장, 배경 클래스로 구분하였으며, 훈련 데이터에서 각 클래스가 차지하는 비율을 Table 2와 같이 설정하여 오버 샘플링하였다. 이때 산업단지와 채석장이 결합된 영역을 포함하는 원본 데이터의 수량이 매우 부족하였기 때문에, 이를 해결하기 위해 컷믹스(CutMix) 증강 기법을 적용하여 네

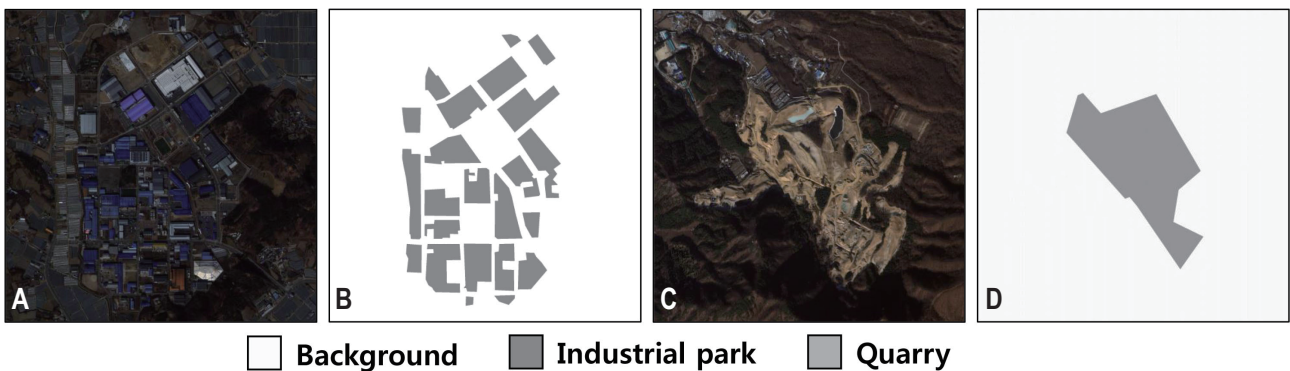


Fig. 1. Labeling data samples for industrial park and quarry classification. (A) Industrial park image data. (B) Industrial park label data. (C) Quarry image data. (D) Quarry label data.

Table 2. Training data composition ratio

	Industrial park	Quarry	Industrial park and quarry	Background
South Korea	1,000	1,000	1,000	1,000
Thailand	1,000	1,000		1,000

개의 데이터를 부분적으로 결합한 데이터를 사용하였다. 이러한 방법은 데이터의 불균형 문제를 해결하여 AI 모델의 일반성을 향상시키는 역할을 한다.

4. 결과

본 연구에서는 효율적인 산업단지 및 채석장 분류를 위한 GeoAI 데이터 8,350셋을 구축하였으며, 이 데이터셋은 훈련 데이터 7,000셋과 테스트 데이터 1,350셋으로 구성되어 있다. 구축된 데이터셋의 성능을 평가하기 위해 의미론적 분할에 뛰어난 성능을 보인 U-Net 모델을 적용하였다 (Ronneberger et al., 2015). 그 결과, 모델 Accuracy는 0.85, F1-score는 0.78을 달성하였고, 산업단지과 채석장 클래스에 대한 F1-score는 각각 0.74와 0.82로 높은 성능을

나타냈다.

Fig. 2는 테스트 이미지 영상(A, D), 라벨 영상(B, E) 및 이에 해당하는 모델 추론 결과 영상(C, F)을 나타낸다. 산업단지에 대한 라벨 데이터 제작 시 산업단지 내부의 지형지물을 최대한 제외하였으나, 딥러닝 모델 추론 결과는 이러한 부분까지 산업단지로 인식하여 경계가 라벨과 다소 다른 경향을 보였다. 채석장의 경우, 예측 결과가 라벨 데이터에 비해 더 세밀하고 면밀하게 나타났다. 이러한 라벨 데이터와 예측 결과 간의 경계 표현 차이에도 불구하고, 실제 이미지 영상과의 비교 분석 결과 이 두 경향성 모두 유의미하게 나타났다. 이는 모델의 실제 성능이 연구에서 도출된 값보다 더 높을 수 있음을 암시하며, 본 연구 결과가 대기오염 배출원의 정확한 탐지에 유용할 것으로 기대된다.

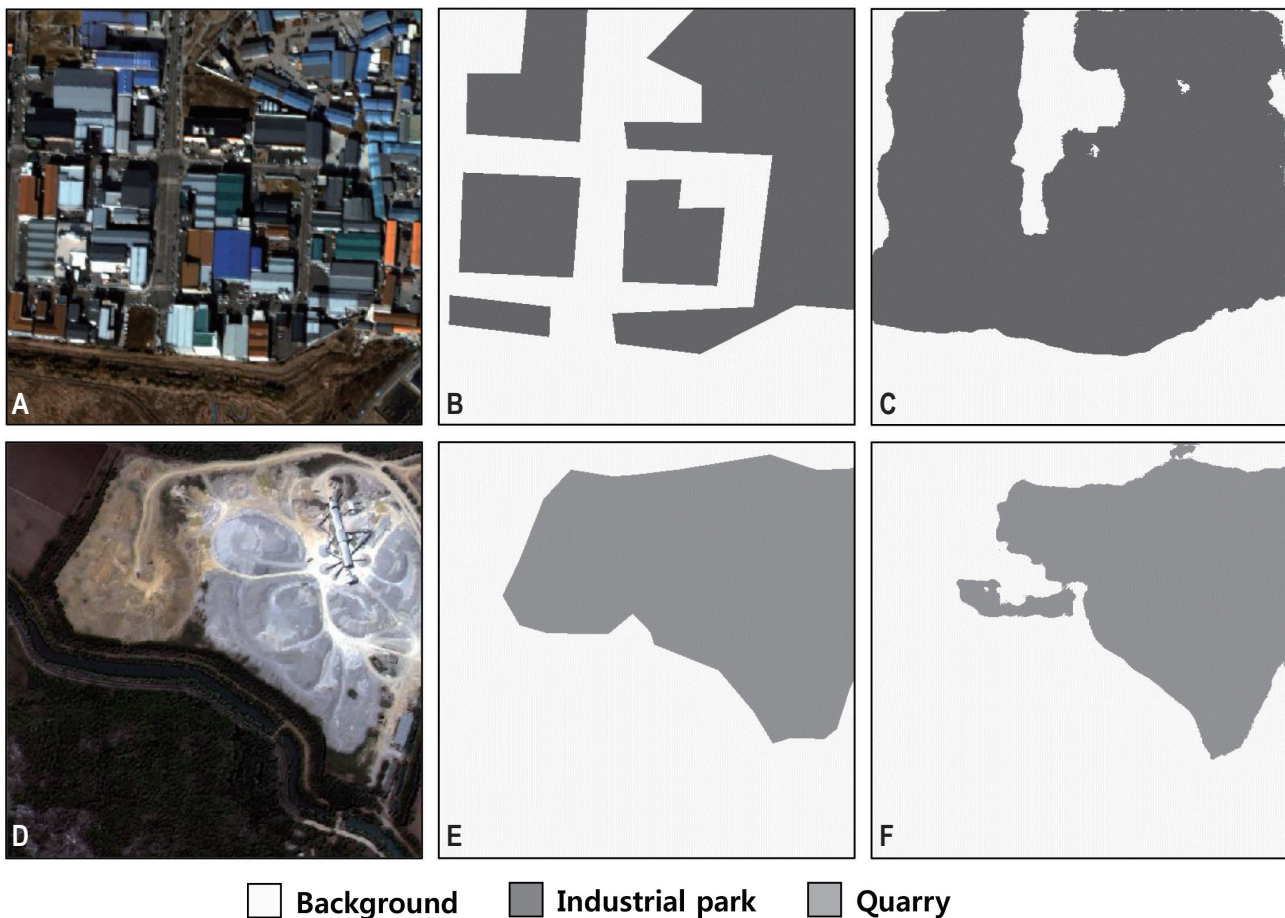


Fig. 2. Predicted industrial park and quarry classification results. (A) Industrial park image data. (B) Industrial park label data. (C) Industrial park prediction result. (D) Quarry image data. (E) Quarry label data. (F) Quarry prediction result.

5. 결론 및 토의

본 연구는 국내외 대표적인 대기오염물질 배출원에 해당하는 산업단지와 채석장을 효과적으로 분류하기 위해 대한민국과 태국 지역의 KOMPSAT-3/3A 영상 자료를 이용하여 AI 학습용 데이터셋을 구축하였다. 구축된 데이터셋을 U-Net 모델에 적용한 결과, 산업단지와 채석장 클래스에 대한 F1-score는 0.74와 0.82로 높게 도출되었다. 해당 데이터셋은 향후 대한민국과 태국에서 대기오염물질의 현황을 파악하고, 배출원을 탐지 및 모니터링하는 데 도움을 줄 것으로 예상된다. 특히, 탐지된 산업단지와 채석장의 위치 데이터는 정지궤도 환경 위성 GEMS를 통한 대기오염물질의 관측 결과와 결합하여 대기 관련 정책 수립 및 이행을 위한 중요도 높은 참고자료로 활용될 수 있다. 또한 태국 지역에 대해 입증된 분류 정확도는 대한민국의 대기오염물질 농도에 영향을 미치는 타국에 대한 향후 연구를 위한 기초자료로서 유용하게 활용될 것으로 기대된다.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Funding Information

This work is financially supported by Korea Ministry of Land, Infrastructure and Transport (MOLIT) as 「Innovative Talent Education Program for Smart City」.

Data Availability Statement

The data that support the findings of this study are openly available in DataON at <https://doi.org/10.22711/idr/1005>.

References

- Baek WK, Lee MJ, Jung HS (2022) The performance improvement of U-Net model for landcover semantic segmentation through data augmentation. *Korean J Remote Sens* 38(6):1663-1676
- Han S, Lee JY, Lee J, et al (2018) Estimation of the source contributions for carbonaceous aerosols at a background site in Korea. *Asian J Atmos Environ* 12(4):311-325
- Kim H, Kim M, Lee Y (2022) Research trend of the remote sensing image analysis using deep learning. *Korean J Remote Sens* 38(5): 819-834
- Kumar N, Park RJ, Jeong JI, et al (2021) Contributions of international sources to PM_{2.5} in South Korea. *Atmos Environ* 261:118542
- Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: *Proceedings of 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, 5-9 Oct 2015
- Zhang X, Du S, Wang Q, Zhou W (2018) Multiscale geoscene segmentation for extracting urban functional zones from VHR satellite images. *Remote Sens* 10(2):281

Metadata for Dataset

Sort	Field	Subcategory#1	Subcategory#2
Essential	*Title	Industrial park and quarry classification AI dataset	
	*DOI name	https://doi.org/10.22711/idr/1005	
	*Category	Geoscientificinformation	
	Abstract		
	*Temporal Coverage	From January 2021 to November 2022	
	*Spatial Coverage	Address WGS84 Coordinates	Republic of Korea, Kingdom of Thailand
	*Personnel	Name Affiliation E-mail	Hyung-Sup Jung University of Seoul hsjung@uos.ac.kr
	*CC License	CC BY-NC	
Optional	*Project	KARI	Construction of Satellite-Based Air Pollution Emission Source Data in East Asia
	*Instrument	KOMPSAT-3/3A	