

화재 탐지 영역의 이미지와 동영상 인식 사이 인공지능 모델 성능 비교 연구

A Comparative Study on Artificial in Intelligence Model Performance between Image and Video Recognition in the Fire Detection Area

이정록¹ · 이대웅² · 정서현³ · 정상^{4*}Jeong Rok Lee¹, Dae Woong Lee², Sae Hyun Jeong³, Sang Jeong^{4*}¹Research Director, Research Laboratory, AI Reader Co., Ltd, Seoul, Republic of Korea²Assistant manager, Research Laboratory, AI Reader Co., Ltd, Seoul, Republic of Korea³Researcher, Research Laboratory, AI Reader Co., Ltd, Seoul, Republic of Korea⁴Research Professor, Department of Parliament ICT convergence safety, University of Chung-Ang, Seoul, Republic of Korea

*Corresponding author: Jeong Sang, safepeople@cau.ac.kr

ABSTRACT

Purpose: We would like to confirm that the false positive rate of flames/smoke is high when detecting fires. Propose a method and dataset to recognize and classify fire situations to reduce the false detection rate. **Method:** Using the video as learning data, the characteristics of the fire situation were extracted and applied to the classification model. For evaluation, the model performance of Yolov8 and Slowfast were compared and analyzed using the fire dataset conducted by the National Information Society Agency (NIA). **Result:** YOLO's detection performance varies sensitively depending on the influence of the background, and it was unable to properly detect fires even when the fire scale was too large or too small. Since SlowFast learns the time axis of the video, we confirmed that detects fire excellently even in situations where the shape of an atypical object cannot be clearly inferred because the surrounding area is blurry or bright. **Conclusion:** It was confirmed that the fire detection rate was more appropriate when using a video-based artificial intelligence detection model rather than using image data.

Keywords: Fire Detection, Fire Situation Recognition, Fire Training Dataset, An Artificial Intelligence Model, False-detection rate

요약

연구목적: 화재 탐지시 불꽃/연기의 오탐지율이 높은 것을 확인하고 오탐지율을 낮추기 위해 화재 상황을 인식하여 분류하는 방법과 데이터셋을 제안하고자 한다. **연구방법:** 동영상을 학습데이터로 활용하여 화재 상황의 특징을 추출하여 분류모델에 적용하고, 평가는 한국정보화진흥원(NIA)에서 진행되는 화재 데이터셋을 이용하여 Yolov8, Slowfast의 모델 성능을 비교 및 분석하였다. **연구결과:** YOLO는 배경의 영향에 따라 탐지 성능이 민감하게 변화하며, 화재의 규모가 너무 크거나 작을 때에도 화재를 제대로 감지하지 못했다. SlowFast는 동영상의 시간 축을 같이 학습하기 때문에 비정형 객체에 대해 주변이 흐리거나 밝아 형상을 명확하게 유추할 수 없는 상황에서도 우수하게 화재를 탐지하는 것을 확인했다. **결론:** 화재 탐지율은 이미지 데이터 방식보다는 동영상 기반의 인공지능 인식(Detection) 모델을 활용했을 때 더 적절했음을 확인했다.

핵심용어: 화재감지, 화재상황인식, 화재훈련 데이터 집합, 인공지능 모델, 오탐지율

Received | 7 December, 2023

Revised | 20 December, 2023

Accepted | 21 December, 2023

OPEN ACCESS



This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted noncommercial use, distribution, and reproduction in anymedium, provided the original work is properly cited.

© Society of Disaster Information All rights reserved.

서론

소개

최근 기후재난 이슈와 더불어 캐나다 산불, 하와이 산불과 같은 화재로 인해 엄청난 재산과 인명피해가 발생하고 있다. 화재 예방하기 위한 노력과 함께 피해를 최대한 줄일 수 있게 화재를 조기에 탐지하기 위한 연구가 진행되고 있다. 현재 소방 분야에서는 화재를 탐지하기 위한 불꽃, 연기, 가스 등을 감지하는 센서를 설치하여 관리하고 있다(Khan, 2022). 센서를 활용한 방식은 가격이 저렴하고 작동이 편리하여 건물 및 공장 등에 많이 설치되어 있다. 하지만 정확도가 낮아 오탐이 많다. 이를 보완하기 위해 딥러닝 인공지능 기술인 CNN, Transformer 등의 신경망을 이용한 화재 탐지 모델에 대한 연구가 활발하다(Na et al., 2020; Jin et al., 2023). 불꽃 및 연기 객체를 탐지하는 화재 탐지 모델은 입력 영상에서 클래스 정의한 객체의 특징을 추출하여 인식하는 기술로 실시간성을 확보하기 위해 모델을 경량화하는 쪽으로 연구가 계속되고 있다. 여러 CCTV 영상을 동시에 분석할 수 있게 경량화된 모델은 CCTV기반 화재 탐지 시스템에 적용되고 적은 비용으로 많은 공간을 넓게 모니터링할 수 있다(Feichtengofe et al., 2019). 그러나 경량화로 인해 오탐지율이 높아 이를 개선하기 위한 연구가 활발하다. 본 연구에서는 오탐지율을 낮추기 위해 화재 상황 분류 모델과 학습데이터셋을 제안한다. 기존 탐지 모델과 학습방법에 대해 실험을 통해 비교, 평가해 제안방법의 우수성을 증명한다.

관련 연구

딥러닝 기술을 기반으로 한 객체 탐지 기술은, 연산 속도를 비약적으로 단축한 YOLO 모델이 2015년 처음 발표되면서부터 실시간 탐지 프로그램에 적용되기 시작하였다. 초창기에는 연산 시간을 단축하는 것에 치중한 나머지 예측 성능이 떨어진다는 평가를 받기도 하였지만, 2023년을 기준으로 8번째 버전까지 고도화를 거치면서 정확도와 실용성을 모두 갖춘 기술로 평가받고 있다. 하지만 불꽃과 연기를 탐지하도록 훈련된 모델들은 실제 화재를 감지할 때 사용하기 어려운 성능 수준을 보인다. 이는 객체 탐지 모델은 탐지하고자 하는 객체의 형태가 명확할수록 성능이 올라가지만, 화재에서 발생하는 불꽃과 연기는 시각적인 경계가 모호하고, 객체의 크기와 형태 또한 다양한 양상을 보이기 때문이다(Kim et al., 2021). 이러한 특징들은 학습데이터를 구축할 때도 일관된 레이블링 작업을 어렵게 만들기 때문에, 모델의 성능이 더욱 떨어지는 원인이 된다(Kim et al., 2021). 예를 들면 길쭉하게 뻗어 퍼지고 있는 연기의 경우, 직사각형 모양으로 해당 객체를 표시했을 때, 연기를 포함하는 영역보다 포함하지 않는 영역이 더욱 커질 수 있어, 모델의 성능을 저하하는 원인이 된다.

이러한 객체 탐지 모델의 한계점을 극복하기 위해 기존 YOLO 모델에 optical flow 같은 광학적 특성들을 전처리 과정을 통해 모델에 입력하거나(Hu et al., 2018) 다양한 크기의 불꽃을 학습하기 위해 모델 구조에 변화를 주는 방식의 시도들이 있었으나(Chen et al., 2023; Yan et al., 2023), 복잡한 전처리가 필요하거나 연산 속도가 느려지는 등의 한계가 존재한다. 따라서 본 논문에서는 기존 이미지 데이터를 사용한 방식보다는 여러 프레임을 한번에 고려하는 동영상 기반의 딥러닝 모델이 정확한 화재탐지에 더 적절하다는 것을 실험으로 증명하려 한다.

본론

모델 사용

이미지 기반 모델은 실시간 탐지 프로그램에 적용할 수 있으며, 2023년을 기준으로 8번째 버전까지 고도화를 거쳐 정확도와 실용성을 모두 갖춘 YOLOv8 (Jocher, 2023) 모델을 선정했으며, COCO(Microsoft Common Objects in Context) (Lin et al, 2014) 데이터를 사전 훈련한 모델에 대해 NIA 화재 데이터셋을 추가 학습했다. Fig. 1은 YOLO 모델 구조를 보여준다.

동영상 기반 모델은 공간과 시간 축의 프레임 수를 달리하여 빠르게 변화하는 모션을 인지하는 네트워크와 느리게 변화하는 모션을 인지하는 네트워크로 구분하는 SlowFast(Feichtengofer, 2019) 모델을 선정했으며, Kinetics-400 (Kay et al., 2017) 데이터를 사전 훈련한 모델에 대해 NIA 화재 데이터셋을 추가 학습했다. Fig. 2는 SlowFast 모델 구조를 보여준다.

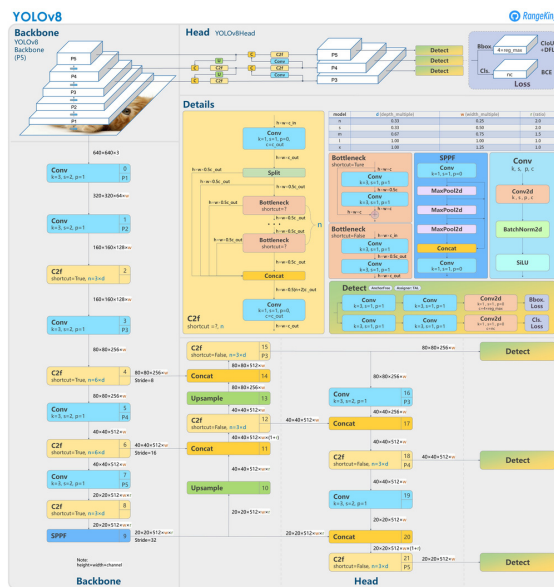


Fig. 1. YOLOv8 frame(locher, 2023)

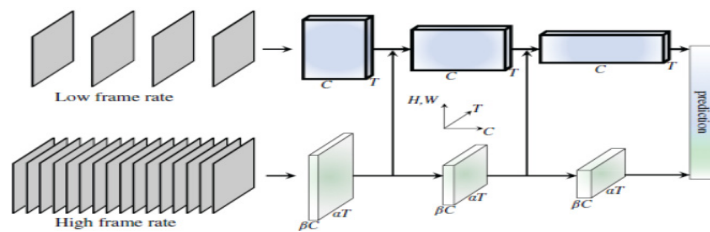


Fig. 2. SlowFast frame(Feichtehofer, 2019)

데이터셋 구성

데이터는 한국정보화진흥원(NIA)에서 제공하는 데이터셋을 사용했으며, 데이터는 1920x1080과 1280x720 크기로 12초

길이의 30FPS 클립으로 구성되어 있다. 클래스는 “정상(NONE)”, “연기(SMOKE)”, “불꽃(FLAME)” 3개로 구분하였으며, 영상에는 불꽃과 연기 등이 다양하게 분포되어 있다. 이미지 기반 모델 동영상 데이터를 이미지로 변환하여 학습을 진행한다. 이 때 정상은 학습할 수 없기 때문에 연기와 불꽃 탐지만 할 수 있도록 학습을 진행하며, 이후 후처리를 통해 정상 클래스를 분류한다. 각 불꽃, 연기, 정상은 300개 클립으로 Table 1과 같이 구성하였고 Table 2는 학습, 검증, 테스트가 8:1:1 비율로 구성됨을 보여준다.

Table 1. Data for each class, 1:1:1 composition of proportions



No	Class	Size	clip length	Frames	Number	Example Image
1	Fire	1920*1080	12	360	300	
		1280*720				
2	Smoke	1920*1080	12	360	300	
		1280*720				
3	Normal	1920*1080	12	360	300	
		1280*720				

Table 2. Learning by class, verification, Test data, 8:1:1 composition of proportions

No	Class	Training	Validation	Test
1	Fire	240	30	30
2	Smoke	240	30	30
3	Normal	240	30	30

성능평가

평가는 동영상 데이터에 대해 정상, 연기, 불꽃 클래스를 올바르게 분류했는지를 P(Precision), R(Recall), F(F1-Score)로 평가한다. 이미지와 동영상 기반 모델을 동일한 조건으로 평가하기 위해 YOLO는 후처리를 통해 분류 형식에 맞도록 반환 값을 조정한다. 1개의 클립은 360장의 이미지로 이뤄진 데이터를 사용하고 있기 때문에 YOLO는 1개의 클립에 대해 총 360장 이미지를 읽어 각 이미지 별로 클래스를 탐지한다. 탐지된 우선순위는 불꽃 > 연기 > 정상으로 판단한다. 불꽃이 발생하는 경우엔 필연적으로 연기가 같이 발생하기 때문에 하나의 이미지에서 연기와 불꽃을 모두 탐지한 경우는 불꽃에 더 높은

우선순위를 부여한다. 360장 이미지에 대해 각 클래스를 탐지한 후 후처리를 위해 S_TH(Score_Threshold)를 0.1, 0.2, ..., 1.0으로 0.1 STEP씩 S_TH를 증가하며, 360장 이미지에서 탐지된 불꽃, 연기 비율을 카운팅한다. 이때 카운팅 수량이 S_TH를 넘어서면 해당 클래스로 분류했다고 판단하고 S_TH를 넘기지 못하면 정상 클래스로 분류했다고 판단한다. Fig. 3은 탐지 모델인 YOLO의 출력값을 분류로 후처리하는 방법을 보이며, Table 3은 실험에 사용한 PC의 사양을 보여준다.

```

Algorithm 1: YOLO 예측값 후처리 방법
Input:
  classes: inferred classes for n frames
  S_TH: fixed floating number ∈ (0.0, 1.0]
Result: result ∈ {FLAME, SMOKE, NONE}
i ← 0;
flame_cnt ← 0;
smoke_cnt ← 0;
while i < n do
  if classes[i] = FLAME then
    | flame_cnt ← flame_cnt + 1
  end
  if classes[i] = SMOKE then
    | smoke_cnt ← smoke_cnt + 1
  end
  i ← i + 1
end
if flame_cnt / n ≥ S_TH then
  | return FLAME
end
if smoke_cnt / n ≥ S_TH then
  | return SMOKE
else
  | return NONE
end
    
```

Fig. 3. YOLO Post-processing methods

Table 3. Server specs

Module	Specificaion
CPU	Intel(R) Xeon(R) Gold 6330N CPU @ 2.20GHz
RAM	128GB
GPU	RTX A4000, 1EA
SSD	1TB
OS	Ubuntu 22.04.2 LTS

성능비교

테스트셋은 Table 2에 기재된 불꽃, 연기, 정상 클래스별 각 30개의 클립을 사용하며, YOLO 후처리 별 성능은 Table 4와 같고 두 모델에 대한 성능은 Table 5, 클래스 별 성능은 Table 6과 같다(YOLO, S_TH 0.6기준). 제한된 테스트셋 기준에선 전반적으로 이미지 기반의 탐지모델인 YOLO를 사용한 경우가 동영상 기반의 SlowFast 모델보다 정확도가 높게 나타났다. 하지만 YOLO는 Fig. 4와 같이 화재가 발생하였는데 불꽃이 작아 연기 위주인 경우에는 불꽃을 제대로 탐지하지 못했고,

Fig. 5와 같이 불꽃의 밝기로 그 주변까지 밝아져 불꽃 형상이 명확하게 보이지 않는 경우에도 1개의 불꽃을 여러 개로 탐지하는 등의 오탐이 존재했다. SlowFast는 YOLO와 다르게 Fig. 4 경우처럼 불꽃이 작고 연기 위주인 경우에도 불꽃을 잘 탐지했으며, Fig. 5처럼 불꽃의 빛 번짐으로 불꽃 형상이 명확하지 않더라도 올바르게 탐지함을 확인하였다. 다만, Table 6에 정상 recall과 연기 precision을 보면 두 클래스 사이 오탐이 성능을 저하시킴을 볼 수 있고 이는 정상에 대한 특징이 명확하지 않아 정상 영상에 존재하는 희미한 구름 등이 연기로 오탐을 일으키는 것으로 유추되어 정상 데이터를 불꽃, 연기 보다 배로 확보하여 정상, 연기, 불꽃 별 성능 추이를 확인해볼 필요가 있다. 마지막으로 두 모델 모두 화재를 올바르게 탐지하지만 현재 테스트 케이스는 명확한 정상, 연기, 불꽃으로 나누어졌기에 YOLO가 보다 우수한 성능을 가짐을 확인할 수 있다. 하지만 노이즈가 많은 데이터를 입력으로 넣은 경우에는 Fig. 4,5와 같이 YOLO의 오탐지가 더 높음을 확인할 수 있다. 이는 너무 소량의 테스트 데이터를 사용하였기에 현재 성능을 그대로 받아들이기 어려운 점을 시사한다.

Table 4. YOLO post-processing, performance by S_TH(0-100)

Score	S_TH									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
F1 Score	94.11%	95.29%	95.29%	95.29%	95.29%	95.29%	92.94%	89.41%	85.88%	74.11%

Table 5. SlowFast, YOLO model performance(0-100)

No	Model	Precision	Recall	F1-Score
1	SlowFast	74.44%	74.44%	74.44%
2	YOLOv8(0.6)	95.29%	95.29%	95.29%

Table 6. Performance of SlowFast, YOLO model by class(0-100)

No	Model	class	Precision	Recall	F1-Score
1	SlowFast	NONE	72%	60%	65.45%
2		SM	64.7%	73.33%	68.75%
3		FL	87.09%	90%	88.52%
4	YOLOv8 (0.6)	NONE	89.28%	100%	94.33%
5		SM	96.66%	96.66%	96.66%
6		FL	100%	90%	94.73%



Fig. 4. If flames are seen in far away places (YOLO / SlowFast)

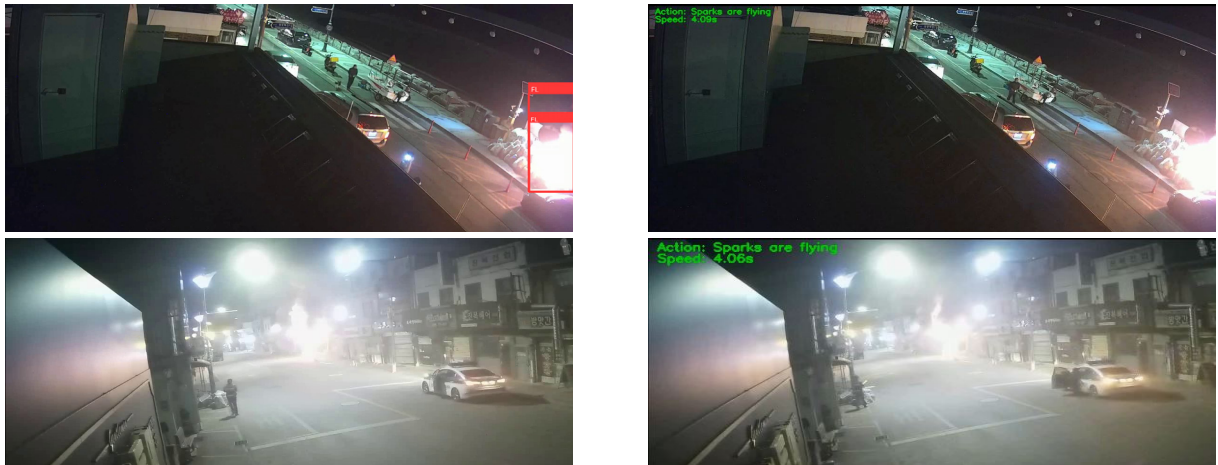


Fig. 5. If the shape of the flame is not clearly visible due to the strong flame (YOLO / SlowFast)

결론

화재탐지 분야에서 이미지 기반의 모델 YOLO와 동영상 기반의 모델 SlowFast 성능을 비교하여, 둘 중 어느 접근법이 화재탐지 분야에 더 유효한지 실험하였다. YOLO는 배경의 영향을 많이 받아 주변이 너무 밝거나 흐리면 탐지 성능이 급감하며, 화재의 규모가 너무 크거나 작을 때에도 화재를 제대로 감지하지 못했다. 이는 객체 탐지를 위해 레이블링 할 때 주변이 밝거나 흐린 경우 불꽃이나 연기의 형상을 일관되게 잡을 수 없기 때문으로 보이며, 이는 모델 학습 시 성능 저하가 더 가중된 것으로 판단된다. 동영상 기반 모델 SlowFast는 FastPath와 SlowPath를 이용하여 빠르게 변화하는 모션과 느리게 변화하는 모션을 인지하는 네트워크를 구분하여 학습하며, 동영상의 시간 축을 같이 학습하기 때문에 비정형 객체에 대해 주변이 흐리거나 밝아 형상을 명확하게 유추할 수 없는 상황에서도 우수하게 화재를 탐지하는 것을 확인했다. 다만, SlowFast는 불꽃을 잘 예측하지만 정상과 연기 사이 오탐하는 경우가 빈번하다. 이를 개선하기 위해 정상 데이터 분포를 연기, 불꽃 대비 2-5배 가량 증가시키며 연기, 불꽃 성능은 유지하되 정상인 경우의 오탐을 줄이는 등 모델 성능 고도화를 위한 연구가 필요하다.

References

- [1] Ang G. Aritejh Kr Goil, Henryk Chan, Jieyi Jeric Lew, Xin Chun Lee, Raihan Bin Ahmad Mustaffa, Timotius Jason, Ze Ting Woon, Bingquan Shen. (2023). "A novel application for real-time arrhythmia detection using YOLOv8." arXiv preprint arXiv:2305.16727.
- [2] Feichtengofer, C, Haoqi Fan Jitendra Malik Kaiming He. (2019). "SlowFast networks for video recognition." Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1-10.
- [3] Hu C., Tang, P., Jin, W., He, Z., Li, W. (2018). "Real-time fire detection based on deep convolutional long-recurrent networks and optical flow method." Proceedings of the 37th Chinese Control Conference. pp. 9061-9066.
- [4] Huo, Y., Zhang, Q., Zhang, Y., Zhu, J., Wang, J. (2022). "3DVSD: An end-to-end 3D convolutional object detection network for video smoke detection." Fire Safety Journal, Vol. 134, pp. 1-11.
- [5] Kim, K.-J., Jang, I.-S., Lim, K.-T. (2021a). "Construction of wild-fire smoke data-set and comparative analysis of detection method based on deep neural network." Electronics and Telecommunications Research Institute Winter

Conference, pp. 1172-1173.

- [6] Jeong, Y.-S., Kim, Y.-W., Yim, J.-I. (2023). "A study on the development of an automatic classification system for life safety prevention service reporting images through the development of AI learning model and AI model serving server." *Journal of the Society of Disaster Information*, Vol. 19, No. 2, pp. 432-438.
- [7] Jin, C., Wang, T., Alhusaini, N., Zhao, S., Liu, H., Xu, K., Zhang, J. (2023). "Video fire detection methods based on deep learning : dataset, methods, and future directions." *Fire*, Vol. 6, No. 8, pp. 3-15.
- [8] Jocher, G. (2023). YOLO by Ultralytics (Version 8.0.0). <https://github.com/ultralytics/ultralytics>.
- [9] Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., Viola, F., Green, T., Back, T., Natsev, A., Suleyman, M., Zisserman, A. (2017) "The kinetics human action video dataset." arXiv preprint.
- [10] Khan, F., Xu, Z., Sun, J., Khan, F.M., Ahmed, A., Zhao, Y. (2022). "Recent advances in sensors for fire detection." *Sensors*, Vol. 22, No. 9, pp. 3-10.
- [11] Kim, C.Y., Lee, H.-S., Lee, K.Y. (2022). "Implementation of a deep learning based realtime fire alarm system using a data augmentation" *Journal of IKEEE*, Vol. 26, No. 3, pp. 468-474.
- [12] Kim, J.-S., Park, S.-M., Hong, C.-H., Park, S.-H., Lee, J.-W. (2022). "Development of AI detection model based on CCTV image for underground utility tunnel." *Journal of the Society of Disaster Information*, Vol. 18, No. 2, pp. 364-373.
- [13] Kim, K.-J., Jang, I.-S., Lim, K.-T. (2021b). "Analysis of video-based fire detection learning model". *Electronics and Telecommunications Research Institute Summer Conference*, pp. 240-241.
- [14] Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P. (2014). "Microsoft COCO common objects in context." *European Conference on Computer Vision*. pp. 740-755.
- [15] Na, Y.-M., Hyun, D.-H., Park, D.-H., Hwang, S.-H., Lee, S.-H. (2020). "AI fire detection & notification system." *The Korea Society of Computer and Information*, Vol. 25, No. 12, pp. 63-71.
- [16] Nam, G.-T., Seo, K.-J., Choi, D.-C. (2022). "A study on the development of AI-based fire fighting facility design technology through image recognition." *Journal of the Society of Disaster Information*, Vol. 18, No. 4, pp. 883-890.
- [17] Park, J., Cho, Y.K., Kim, S. (2022). "Deep learning-based UAV image segmentation and inpainting for generating vehicle-free orthomosaic." *International-Journal-of-Applied-Earth-Observation-and-Geoinformation*, Vol. 115, 103111.
- [18] Yan, C., Wang, Q., Zhao, Y., Zhang, X. (2023). "YOLOv5-CSF: An improved deep convolutional neural network for flame detection." *Application of Soft Computing*, Vol. 27, pp. 19013-19023.