

구급차 회피 기동을 위한 딥러닝 및 컴퓨터 비전 기반 센서 퓨전 알고리즘

김 동 욱¹⁾

한양대학교 미래자동차공학과¹⁾⁾

Emergency Vehicle Evasion Maneuvers: Deep Learning and Computer Vision-based Sensor Fusion Algorithm

Donguk Kim¹⁾

¹⁾⁾Department of Automotive Engineering, Hanyang University, Seoul 04763, Korea

Abstract : Urban traffic systems continually evolve, posing challenges that demand advanced solutions. Emergency vehicle deployment, crucial for saving lives, contributes to congestion and accident risks. This study focuses on developing technology for emergency vehicle evasion. A sensor fusion algorithm, blending deep learning and computer vision, is designed for safe interactions during emergencies. This approach aims to prevent collisions, enhancing road user safety. We developed a sensor fusion algorithm integrating camera, audio, and ultrasonic data, conducting experiments with ResNet50-trained vision models on proprietary datasets. This distinct sensor fusion showcases technical prowess in enabling emergency vehicle evasion maneuvers. Our research contributes to an advanced system, ensuring effective collision avoidance with emergency vehicles, bolstering road user safety.

Key words : Emergency Vehicle(구급차), Siren(사이렌) Sensor Fusion(센서 퓨전), Computer Vision(컴퓨터 비전), ResNet50(ResNet50), Deep Learning(딥러닝), Self-Driving(자율주행), Fine-Tuning(파인튜닝)

1. 서 론

도로 교통은 현대 도시에서 지속적으로 발전하고 있는 동시에 빠르게 변화하고 있는 환경 중 하나이다. 특히, 긴급 상황 차량의 출동은 삶을 구할 수 있는 중요한 활동이지만, 이로 인한 교통 혼잡과 사고 위험은 급증하고 있다. 이러한 도전에 대응하기 위해 첨단 기술의 개발이 필요하며, 본 연구는 이에 중점을 두고 있다.

자율주행자동차는 주변 환경을 인식하여 스스로 주행 결정을 내릴 수 있는 첨단 기술을 기반으로 한다. 이를 위해 차량 내외부에 탑재된 라이다, 카메라, 레이더 등 다양한 센서를 활용하여 도로상의 다양한 요소들을 식별한다.⁴⁾ 도로 사용자, 교통 신호등, 노면의 상태 등을 정확하게 감지하고 GPS 신호와 융합하여 차량의 위치를 결정하여 다양한 자율주행 인지, 판단, 제어 알고리즘을 적용시킨다.

본 연구에서는 특히 구급차 회피 기동을 위한 기술의 개발에 초점을 맞추고 있습니다. 딥러닝과 컴퓨터 비전을 기반으로 한 센서 퓨전 알고리즘을 도입하여 긴급 상황에서 안전하게 교통 차량과 상호작용하는 시스템을 설계하고자 한다. 이를 통해 긴급 차량과의 충돌을 효과적으로 회피하고 도로 사용자의 안전을 보장하는 첨단 시스템의 구축을 목표로 한다. 본 연구는 다양한 센서 데이터를 융합하는 센서 퓨전 알고리즘을 활용하여 구급차의 회피 기동을 가능하게 하는 기술적인 차별성을 제시할 것으로 기대된다.

2. 관련 연구

2.1 Object Detection

Object Detection은 컴퓨터 비전의 중요한 작업 중 하나로, 이미지나 동영상 내의 객체를 식별

하고 위치를 찾는 데 중점을 둡니다. 이 도전적인 작업은 객체의 위치와 경계를 정확하게 파악하고 특정 범주에 할당하는 것을 포함하며, 이미지 분류 및 검색과 함께 자율주행 기술에서 주변 환경 인식과 물체를 탐지하는 인지 알고리즘에서 큰 비중을 차지하는 기술이다. 현재까지 많은 연구가 진행 중에 있으며 일반적인 카메라 2D 이미지 뿐 아니라 라이다, 레이더 센서를 사용하여 3D 이미지를 연구하는 3D Object Detection 연구도 활발히 진행 중에 있다.

Object Detection는 주어진 이미지에서 특정 클래스에 속하는 물체의 존재 여부를 확인하는 Classification 문제와 해당 물체의 정확한 위치를 박스 형태로 지정하는 Localization 문제로 구성되어 있다. 먼저, Classifier 구조에서는 대체로 사용하는 일반적인 Classification Model을 사용하여 각 Class 별로 분류하여 Class 확률에 대한 출력을 Output으로 내보낸다. 주로 CNN 기반 딥러닝 기법을 사용한 모델들이 있고 ResNet, VGG 16, MobileNet 등이 있고 최근에 자연어처리 분야에서 떠오르는 Transforemer를 이미지 분류 모델에 적용시켜 사용하였는데 뛰어난 성능을 보여주어 현재 많은 연구가 이루어지고 있고 ViT(Vision Transformer)가 대표적이다. 다음으로 물체에 위치를 파악하는 Localization에서는 중간에 위치한 신경망에서 이미지의 특성을 파악하는 Feature Map에서 Anchor Box를 통해 물체가 포함되는 영역을 확률로 판단하여 예측하는 구조로 되어있다. 이 과정을 통해서 최종적으로 Boundary Box라는 물체의 위치를 알려주는 박스 모양 플롯을 출력하여 물체의 정확한 위치를 파악할 수 있다.

Classification과 Localization 과정이 동시에 일어나게 하는 One-Stage Detector와 각각 따로 신경망 구조로 되어 학습하는 Two-Stage Detector로 Detection 모델이 나뉘게 된다. 두 분류의 모델들은 각자의 장단점을 가지고 있는데 One-Stage Detector의 경우 빠른 학습과 처리 속도가 우수하다는 것이 장점이지만 이미지를 분류하거나 위치 추정에 있어 정확도가 떨어진다는 점이 단점이다. 반대로 Two-Stage Detector는 신경망 구조가 따로 구성되어 있어 정확성이 뛰어나다는 점이 있지만 처리 속도가 느려 실시간

Object Detection 모델에는 성능이 떨어진다는 단점을 가지고 있다. One-Stage Detector 모델로써는 YOLO(Fig 1.), SSD, RetinaNet 등이 있고 Two-Stage Detector 모델로는 대체적으로 R-CNN기반의 fast R-CNN, faster R-CNN, Mask R-CNN 등이 있다.

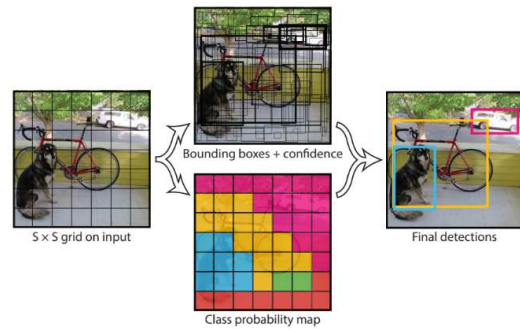


Fig. 1 YOLOv1 structure, the main model of One-Stage Detector for object detection¹⁾

3. 모델링

3.1 Vision Model

3.1.1 모델 설정

카메라를 통해 불러오는 영상들을 실시간으로 분석하여 구급차를 판단하는 모델을 구축하기 위해서 Fig 2.와 같이 순서로 알고리즘을 구축하였다.

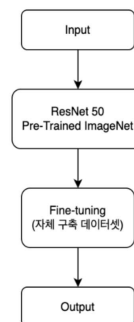


Fig. 2. Algorithm flowchart of vision model

먼저, 구급차 인지 알고리즘의 기초가 되고 기본적인 성능을 보여주기 위해 기존의 선행 학습(Pre-Train)된 모델을 활용하였다. Vision Model에서 Base가 되는 Backbone 모델로는 ResNet50⁶⁾, ResNet101, VGG-16⁷⁾, VGG-19와 같이 CNN기반의 이미지 분류 모델을 사용하였다. 각각의 모델들을 1000개의 Class, 1400만 개의 일반 이미지로 구성되어있는 ImageNet 데

이터 셋을 통해 선행 학습을 진행하였다. 이 4가지의 분류 모델을 초기의 설정한 이유는 모델이 비교적 가벼워 저화질 이미지를 빠른 속도로 예측할 수 있다는 점과 빠른 처리 속도에도 불구하고 높은 정확도를 보여주기 때문이다. 또한, ImageNet 데이터 셋에는 수많은 구급차의 데이터를 포함하고 있고 407번 Class가 구급차를 나타내기에 구급차 회피 기동 알고리즘 학습에 적절한 데이터 셋이라고 판단하였다.

3.1.2 데이터 셋 구축

ImageNet 데이터셋으로 사전 훈련된 모델을 사용할 때, 해당 모델은 주로 고화질의 다양한 이미지에 대한 정보를 학습한 것이다. 그렇기 때문에 모델이 일반 이미지를 처리하는 데 뛰어나지만, 특정 저화질 이미지 클래스에 대한 예측 정확도가 떨어질 수 있습니다. 특히, 구급차 모형과 같은 특별한 저화질 이미지의 경우 모델의 성능이 더욱 미흡할 수 있다. 따라서 연구 목적에 맞게 모델을 Fine-Tuning하여 실제 환경에서 촬영된 데이터 셋에 적응시켜야 한다. 이렇게 함으로써 모델은 특정 저화질 클래스에 대해 높은 정확도를 보여줄 수 있게 되며, 연구에 더 적합한 결과를 얻을 수 있다. 웹캠을 이용하여 1280*720의 해상도 이미지로 데이터 셋을 생성했습니다. 이 중에서 구급차와 그렇지 않은 경우 각각 26000장의 이미지를 확보했으며, 이 중 18000장을 학습용 데이터로, 4000장을 검증용 데이터로 사용했다. 나머지 4000장은 테스트 데이터로 설정하여 전체적으로 균형 잡힌 데이터 셋을 구축하였다. (Fig 3.)

3.1.3 Fine-Tuning

기존의 Backbone 모델은 각 Class의 확률을 출력하기 위해 출력층 노드의 수를 Class 수와 동일하게 구성한다. 그러나 본 연구에서는 주로 구급차를 감지하는 것이 중요하므로 다중 분류가 아닌 이진 분류를 수행하고자 한다. 따라서 출력층의 노드 수를 2로 조정하여, 모델이 구급차의 존재 여부만을 간단하게 판단하도록 Fine-Tuning 기법을 사용하였다. 따라서, 전체 Feature Extraction 층을 학습시키지 않고, 그 대신 Feature Extraction 층의 가중치를 고정하고 출력층의 가중치만을 학습시켜서 기존 Backbone 모델을 Fine-Tuning 하였다. 출력층의 구조는 AveragePooling 2D를 사용하였고 활성화 함수로 구급차 확률을 0에서 1값 사이에 확률로 나타

내는 Sigmoid 함수로 설정하였다(Eq (1)). Average Pooling 2D 신경망을 사용한 이유는 모델의 파라미터를 줄이기 때문에 메모리 측면에서 처리 속도가 뛰어나 실시간 구급차 인지 알고리즘에 적합할 것이라고 판단하였다.

$$\text{sigmoid}(z) = \frac{1}{1+e^{-z}} \quad (1)$$

이렇게 Fine-Tuning 기법을 사용함으로써 모델이 구급차 확률을 나타내도록 하였고 검증 단계에서는 Threshold 값을 설정하여 모델의 정확도를 평가하고, 이를 기반으로 최종 모델을 선택하였다.



Fig. 3. Example of ambulance(up) data and not-ambulance (bottom) data

3.2 Audio Model

사이렌 탐지 및 음원 위치 추정 모델로써 사이렌 소리의 세기(dB)를 통해 사이렌 여부 및 상대적인 x, y 좌표의 3가지 라벨을 예측하는 모델이다. 음원의 원시 파형을 통해 시간-음원 강도 특성과 로그 멜 스펙트로그램, MFCC를 통해 시간-주파수 특성을 학습하였다. Conv2D, Dense, MaxPooling 등 12 layer와 약 220k parameter로 구성되었다. 출력층에서는 사이렌 여부에 대한 예측이 0에서 1 사이의 확률 값으로 나타나며, 사이렌 소리의 위치 예측은 x 및 y 값으로 출력된다. 학습에 사용된 데이터 셋은 마이크 센서 4개가 부착된 하드웨어를 통해서 자체적으로 구축하였다.

1cm 간격의 정밀측위 2D 보드를 제작하여 5~10cm 간격의 위치에서 데이터를 수집하였으며 약 20만 개의 데이터 셋을 구축하였다.^{5,8)}

4. 연구 방법

4.1 아두이노 하드웨어 제작

구급차 회피 기동 알고리즘을 구현하기 위해, 아두이노 우노 2개와 2개의 모터, 그리고 2WD 모형차에 모터 드라이버를 포함시킨 하드웨어를 개발하였다. 이 시스템은 Vision Model에 사용되는 카메라(웹캠), Audio Model에 사용되는 마이크 센서(Max9814) 4개, 그리고 초음파 센서를 포함한 다양한 센서들을 장착하였다. 하나의 아두이노 우노 보드에서는 마이크 센서와 초음파 센서의 감지 작업을 수행하고, 이러한 데이터를 처리하는 알고리즘을 실행하여 처리 결과를 다른 아두이노 우노 보드로 전송하는 역할을 한다. 반면에 다른 아두이노 우노 보드는 구급차 회피 기동 알고리즘에 중점을 두고, 알고리즘에서 전송된 값들을 받아들이며 모터를 적절하게 제어한다. 하드웨어 모형은 Fig 4.에서 확인할 수 있다. 구급차는 구급차 모형 RC카를 활용했으며, 구급차의 사이렌 소리는 스마트폰에서 다양한 사이렌 소리를 추출하여 구현하였다.

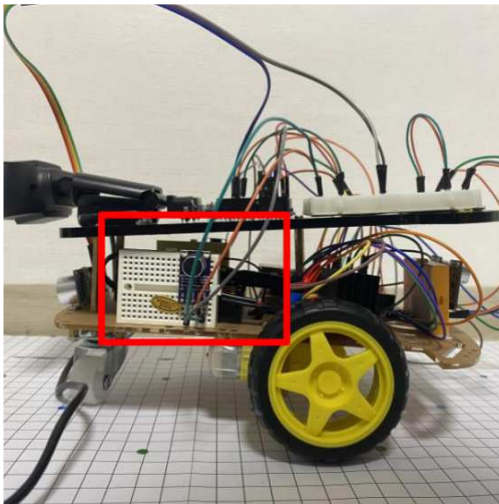


Fig. 4. Arduino hardware model car with webcam, mic sensor(Max9814) and ultrasonic sensor

4.2 시나리오 설정

구급차 회피 기동 알고리즘을 개발하기 위해, 모형차가 각 상황별로 스스로 구급차를 판단할 수 있는지 확인 위해서 다음과 같은 시나리오 1, 2, 3

을 설정하여 알고리즘의 성능을 평가하였다.

4.2.1 시나리오 1. 구급차 후방 접근

구급차가 후방에서 접근 시 상황이다. Vision Model을 통해서 구급차를 인지할 수 있으며, 초음파 센서로 구급차의 거리를 특정 지을 수 있고 Audio Model을 통해 사이렌 감지 및 위치를 탐지할 수 있는 시나리오 상황이다. 회피 기동 알고리즘을 통해서 시나리오 1이라는 상황을 모형차가 인식하면 차선변경 및 속도 제어를 통해 양보 주행이 이뤄진다.

4.2.2 시나리오 2. 구급차 교차로 접근

사각지대에서 구급차가 주행할 시 상황이다. 모형차는 사각지대에서 좌회전하는 구급차의 위치를 마이크 센서 만을 사용하여 인지하게 된다. 모형차가 카메라를 통해서 구급차를 확인할 수 없기 때문에 사이렌 소리만을 통해 구급차의 위치를 특정 지을 수 있게 된다. 시나리오 2 상황을 부여 받은 모형차는 속도를 제어하여 정지 상태에 이르고 구급차가 지나가고 난 후 좌회전 기동을 수행하게 된다.

4.2.3 시나리오 3. 구급차 전방 접근

구급차가 전방 반대편 차로로 접근 시 상황이다. 모형차와 반대편인 차로에서 구급차가 주행하는 상황이기에 이 시나리오에서도 마이크 센서 만을 사용하여 인지하게 된다. 하지만 시나리오 2 상황과 달리 구급차는 정상주행하고 있고 모형차도 어떠한 제어도 필요로 하지 않기에 모형차는 속도를 줄이지 않고 정상 기동을 하게 된다. 따라서 시나리오 3 상황을 인식하더라도 모형차는 같은 속도로 주행하게 된다.

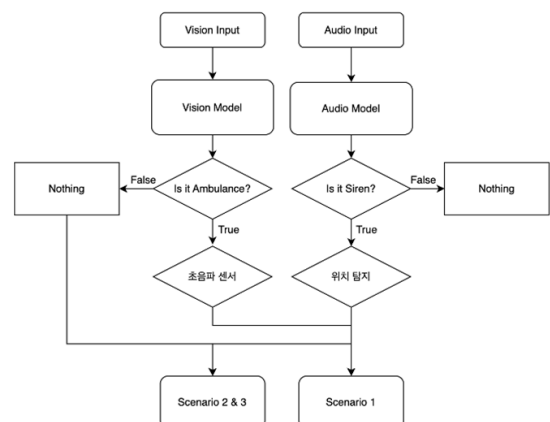


Fig. 5. Ambulance avoidance maneuver algorithm

4.3 구급차 회피 기동 알고리즘

전체적인 구급차 회피 기동 알고리즘의 구조는 Fig 5.와 같다. 카메라, 마이크, 초음파 센서의 퓨전을 통해 판단 제어 알고리즘을 제작하였다. 카메라를 통해 이미지를 실시간으로 전송하고, 아두이노 우노를 통해 센서 값들을 받아들이고 엡지 컴퓨터로 전송시키게 되면 Vision Model과 Audio Model이 동시에 수행된다. 두 모델에서 먼저 구급차 탐지를 시작하게 되면서 회피 기동 알고리즘이 작동하게 된다. And, Or 등의 논리 연산자를 활용하여 센서 퓨전 알고리즘을 구축하게 되고 각 시나리오 별 상황에 맞는 제어 값을 다른 아두이노 우노로 전송하여 모터를 제어할 수 있다.

시나리오 1에서는 Vision Model, Audio Model이 구급차를 동시에 인식하게 되고 Audio Model에서의 구급차 위치와 초음파 센서로 구급차의 위치를 파악한 결과를 통해서 판단한다. 초음파 센서, Audio Model이 구급차와의 거리를 20cm 이하로 판단하게 되면 시나리오 1이라는 값을 아두이노 보드로 전송하게 되고 모형차는 3초간 우회전, 5초간 직진, 3초간 좌회전 기동을 통해 구급차 회피 기동을 수행하게 된다.

시나리오 2에서는 Vision Model은 구급차를 인식할 수 없게 되고 Audio Model만이 구급차를 인지하게 된다. Audio 모델이 파악한 구급차의 위치를 토대로 구급차와 모형차 거리가 30cm 이하가 되면 시나리오 2라는 값을 아두이노 보드로 전송하게 된다. 전송에 성공하면 5초간 정지하게 되고 8초간 좌회전 기동을 통하여 구급차가 지나간 후 모형차가 교차로에서 벗어날 수 있게 된다.

마지막으로 시나리오 3에서는 시나리오 2와 마찬가지로 Audio Model 만이 구급차를 인지하게 되는데 Audio Model이 출력한 x값을 토대로 반대편 차선에 구급차의 위치를 판단하게 되고 시나리오 3 값을 아두이노 보드에 전송하게 된다. 시나리오 3 값을 전송받게 되면 기존에 직진 기동을 수행하게 된다.

5. 연구 결과

5.1 Vision Model 성능 평가

Vision Model에서 최종으로 선택한 사전훈련

Backbone 모델은 ResNet50이다. 그 이유는 ResNet50, ResNet101, Vgg-16, Vgg-19 모델들을 사용하여 Vision Model의 성능을 비교하였을 때 ResNet50 모델이 0.9871의 Test Accuracy 값으로 가장 좋았다. Table 1은 4개의 모델을 사용하였을 때의 Accuracy 값을 나타낸다.

Table 1. Comparison of performance values (accuracy) of ResNet50, ResNet101, Vgg-16, and Vgg-19 models

	ResNet50	ResNet101	Vgg-16	Vgg-19
Val_accuracy	0.9895	0.9877	0.9678	0.9589
Test_accuracy	0.9871	0.9703	0.9608	0.9584

최종 모델인 ResNet50은 ImageNet 데이터셋 Pre-Trained 과정에서 다음과 같은 파라미터로 학습을 진행시켰다. 224*224*3으로 input shape을 지정해 주었고 Optimizer는 Adam, Learning rate는 1e-4, Loss 함수는 이진 Crossentropy 함수를 사용하였다. Vision Model을 사용하여 웹캠을 통해 실시간으로 30fps 속도로 이미지를 분류하는 모델 결과는 Fig 6에서 확인할 수 있다.



Fig. 6. Vision model ambulance classification result (up: true, bottom: false)

5.2 시나리오 별 아두이노 테스트 결과

시나리오 1. 구급차 후방 접근

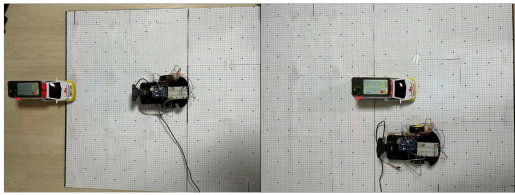


Fig. 7. Scenario 1. Ambulance rear approach

시나리오 2. 구급차 교차로 접근

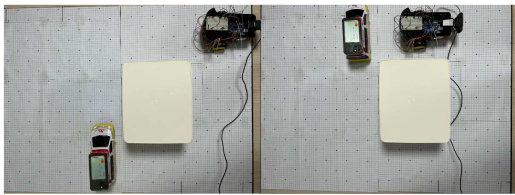


Fig. 8. Scenario 2. Ambulance approaching intersection

시나리오 3. 구급차 전방 접근

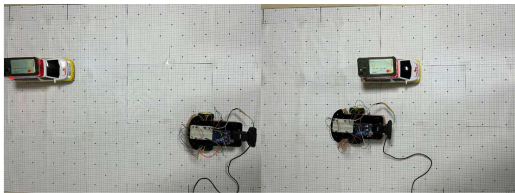


Fig. 9. Scenario 3. Ambulance front approach

6. 결 론

1) 본 연구는 미래 자율주행 시스템에서 필수적인 구급차 회피 기동 알고리즘을 제안한다. 대다수의 자율주행차는 카메라, 라이다, 레이더 센서를 활용하여 환경을 탐색하고 물체를 감지한다. 그러나 본 연구에서 제안하는 구급차 회피 기동 알고리즘은 마이크 센서를 활용해 소리를 기반으로 한 센서 퓨전 알고리즘이다. 이를 통해 비전 센서가 탐지하지 못하는 사각지대에서 빠르게 물체를 감지하고 인식함으로써 기존의 자율주행 기능을 향상시킬 것으로 기대된다.

2) 하드웨어 제작 및 ResNet50 기반 비전 모델 개발을 포함하고 있다. 이 종합적인 접근은 센서 퓨전 알고리즘의 현실적인 구현을 보장하여 현실적인 시나리오에서의 유용성을 입증한다. 또한, 자체적으로 구축한 이미지, 오디오 데이터 셋을 활용하여 기존의 비전 모델 및 오디오 모델의 성능을 향상시킬 수 있었다. 이를 통해 향후 자율주행 시스템의 긴급 차량에 대한 회피 기동 성능 개선 및 기타 상황에도

적용이 가능할 것으로 기대된다.

후 기

지난 4년간 한양대학교 미래자동차공학과에서의 학부 과정은 저에게 넓고 다양한 지식을 제공해주었습니다. 이 동안에 배운 기초적인 학문들은 내가 어떻게 성장해왔는지를 돌아보는 계기가 되었습니다. 그리고 나아가 본인이 진심으로 흥미를 느낀 컴퓨터 비전 분야에 대한 개인적인 탐구는 새로운 시야를 열어주었습니다. 4년 동안의 학습으로 쌓아온 지식과 스킬을 토대로, 나만의 아이디어와 열정을 담아 연구를 진행하였습니다. 함께 1년간 연구를 진행해 온 김찬주 학우에게 진심어린 박수를 보냅니다.

References

- 1) - J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", 2015.
- 2) - J. H. Kim, "Driving control algorithm based on sensor fusion and deep learning for autonomous driving", Ph. D. Electronics, Hanyang University, Seoul, 2021.
- 3) - X. Zou, "A Review of Object Detection Techniques", IEEE, 2019,
- 4) - J. Fayyad, M. A. Jaradat, D. Gruyer and H. Najjaran, "Deep Learning Sensor Fusion for Autonomous Vehicle Perception and Localization: A Review", MDPI, 2020,
- 5) - V. Tran, W. Tsai, "Audio-Vision Emergency Vehicle Detection", IEEE, 2021,
- 6) - K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", CVPR, 2015,
- 7) - K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", IEEE, 2015.
- 8) - H. Sun, X. Liu, K. Xu, J. Miao and Q. Luo, "Emergency Vehicles Audio Detection and Localization in Autonomous Driving", IEEE, 2021.