Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.33, NO.6, Dec. 2023 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2023.33.6.1099

카메라-라이다 융합 모델의 오류 유발을 위한 스케일링 공격 방법*

임 이 지,[1†] 최 대 선[2‡]
[1,2]숭실대학교 (대학원생, 교수)

Scaling Attack Method for Misalignment Error of Camera-LiDAR Calibration Model*

Yi-ji Im,[1†] Dae-seon Choi[2‡]
[1,2]Soongsil University (Graduate student, Professor)

요 약

자율주행 및 robot navigation의 인식 시스템은 성능 향상을 위해 다중 센서를 융합(Multi-Sensor Fusion)을 한 후, 객체 인식 및 추적, 차선 감지 등의 비전 작업을 한다. 현재 카메라와 라이다 센서의 융합을 기반으로 한 딥러닝 모델에 대한 연구가 활발히 이루어지고 있다. 그러나 딥러닝 모델은 입력 데이터의 변조를 통한 적대적 공격에 취약하다. 기존의 다중 센서 기반 자율주행 인식 시스템에 대한 공격은 객체 인식 모델의 신뢰 점수를 낮춰 장애물 오검출을 유도하는 데에 초점이 맞춰져 있다. 그러나 타겟 모델에만 공격이 가능하다는 한계가 있다. 센서 융합 단계에 대한 공격의 경우 융합 이후의 비전 작업에 대한 오류를 연쇄적으로 유발할 수 있으며, 이러한 위험성에 대한 고려가 필요하다. 또한 시각적으로 판단하기 어려운 라이다의 포인트 클라우드 데이터에 대한 공격을 진행하여 공격 여부를 판단하기 어렵도록 한다. 본 연구에서는 이미지 스케일링 기반 카메라-라이다 융합 모델(camera-LiDAR calibration model)인 LCCNet 의 정확도를 저하시키는 공격 방법을 제안한다. 제안 방법은 입력 라이다의 포인트에 스케일링 공격을 하고자 한다. 스케일링 알고리즘과 크기별 공격 성능 실험을 진행한 결과 평균 77% 이상의 융합 오류를 유발하였다.

ABSTRACT

The recognition system of autonomous driving and robot navigation performs vision work such as object recognition, tracking, and lane detection after multi-sensor fusion to improve performance. Currently, research on a deep learning model based on the fusion of a camera and a lidar sensor is being actively conducted. However, deep learning models are vulnerable to adversarial attacks through modulation of input data. Attacks on the existing multi-sensor-based autonomous driving recognition system are focused on inducing obstacle detection by lowering the confidence score of the object recognition model.However, there is a limitation that an attack is possible only in the target model. In the case of attacks on the sensor fusion stage, errors in vision work after fusion can be cascaded, and this risk needs to be considered. In addition, an attack on LIDAR's point cloud data, which is difficult to judge visually, makes it difficult to determine whether it is an attack. In this study, image scaling-based camera-lidar We propose an attack method that reduces the accuracy of

Received(10. 24. 2023), Modified(11. 24. 2023), Accepted(11. 24. 2023)
* 이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원 (No. 2021-0-00511, 엣지 AI 보안을 위한 Robust AI 및 분산 공격탐지기술 개발)과 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1A2C1014813)
† 주저자, ezim@soongsil.ac.kr
‡ 교신저자, sunchoi@ssu.ac.kr(Corresponding author)

LCCNet, a fusion model (camera-LiDAR calibration model). The proposed method is to perform a scaling attack on the point of the input lidar. As a result of conducting an attack performance experiment by size with a scaling algorithm, an average of more than 77% of fusion errors were caused.

Keywords: Multi-Sensor Fusion, Camera-LiDAR Calibration model, Downscaling, Autonomous Driving

I. 서 론

자율주행 기술은 지난 몇 년간 주목받고 있으며 급속도로 발전해왔다. 하지만 레벨 4[1] 이상의 높은 수준의 자율주행[2]기술을 제공하기 위해선 정확한 상황 인지와 안전한 운전을 보장해야 한다. 이를 위해 자율주행에서는 카메라, 라이다, 레이더 등의 센서를 사용하고 있으며 센서로부터 얻은 데이터는 주변 장애물들을 실시간으로 감지하고 충돌 회피와 같은 안전에 중요한 결정에 직접적인 영향을 미친다[3-11]. 자율주행 차량은 멀티 모달 센싱 시스템으로 각 감지 센서의 상호 보완적인 특성을 융합하여 정확도와 신뢰성을 향상한다.

센서들로 정보를 융합하는 것은 서로 다른 센서 좌표계 간의 상대적인 6-DoF 변환을 추정하는 과정이다. 이는 자율주행과 같은 분야에서 객체 검출을 위해 많이 사용된다. 이러한 센서를 사용하여 수집된 의미 있는 데이터들은 보정 매개 변수에 매우 민감하므로 정확하게 보정하여 센서 간 융합을 해야 한다. 센서 간 융합이 제대로 되지 않으면 오차로 인해 융합 데이터를 사용하는 비전 작업에 대한 연쇄적인 오류를 유발하므로 안전이 중요한 자율주행 기술에서는 더욱 치명적인 결과를 초래할 수 있다. 다중 센서 기반 모델의 공격에 관한 선행 연구에서 적대적 물체로 객체 인식을 공격[12]하거나 센서를 모두 공격하여 객체 인식을 방해[13]하는 인식 시스템에 대해 다양한 공격을 했다. 그러나 오분류, 객체 인식 등 목표만 수행 가능하며 타겟 모델에서만 공격할 수 있다는 한계가 있다. 센서 간 융합과정에서의 오류는 추후의 비전 작업 전체에 대한 공격으로 이어질 수 있다. 센서 간 융합 모델에 대한 공격은 없었으며 본 논문에서 최초로 제안하고자 한다.

센서 간 융합 모델에 대한 스케일링 공격은 딥러닝 모델에 데이터가 입력될 때 일반적으로 고정된 입력 차원을 요구하기 때문에(예: LCCNet [256,512]) 데이터 전처리 작업에서 스케일링이 널리 사용된다는 것에 착안하여 시도되었다. 특히, 스케일링 공격은 사용하는 보간 알고리즘(interpolation algorithm)에 따라 스케일링 된 이미지를 생성하기 때문에 특정 픽셀만 조작된다. 따라서, 라이다 데이터에 스케일링 공격을 했을 때, 라이다 데이터에 대한 영향은 최소화하여 공격을 인지하기 어려우나 보정된 결과에 심각한 영향을 미친다. 또한 시각적으로 익숙한 정보인 이미지가 아닌 라이다 포인트 클라우드 데이터는 사람이 공격을 판단하기 어렵다. 스케일링 공격은 적대적 예제와 달리 특정 모델이나 Dataset에 의존하지 않으며, 모델에 사용된 스케일링 알고리즘과 스케일링 크기를 추론[20]하여 스케일링 공격에 사용하므로 블랙박스 공격이 가능하다.

본 논문은 라이다의 포인트 클라우드에 대한 스케일링 공격을 통해 카메라-라이다 융합 모델인 LCCNet의 성능을 저해하는 방법을 제안하며, LCCNet의 모델 구성은 Fig. 1.과 같다.

본 논문이 기여한 바는 다음과 같다.

- 카메라-라이다 융합 모델의 오류를 유도하는 공격을 처음으로 제안한다.
- 라이다로부터 수집한 포인트 클라우드의 특징을 추출하여 이미지 스케일링 공격[14]을 적용하여 라이다 데이터에는 영향을 최소화하여 공격 여부를 시각적으로 판단하기 어렵다.
- 카메라와 라이다 센서의 융합 모델 LCCNet에 공격을 수행하여 성능을 저해한다.

본 논문의 구성은 다음과 같다. 2장에서 카메라-라이다 센서의 융합과 LCCNet, 스케일링 공격 및 스케일링 공격 방어 방법을 소개한다. 3장에서 LCCNet의 오류를 유도하는 스케일링 공격을 소개한다. 4장에서는 본 논문에서 제안한 방법을 평가하기 위해 KITTI 주행 거리 측정 Dataset[15]와 NuScenes Dataset[16]에 대한 스케일링 알고리즘과 스케일링 크기별 실험 결과를 도출하고 5장에서는 연구에 대한 고찰, 6장에서는 결론으로 논문을 맺는다.

II. 관련 연구

2.1 카메라-라이다 센서의 융합

카메라-라이다 센서의 융합은 이미지의 픽셀에 대

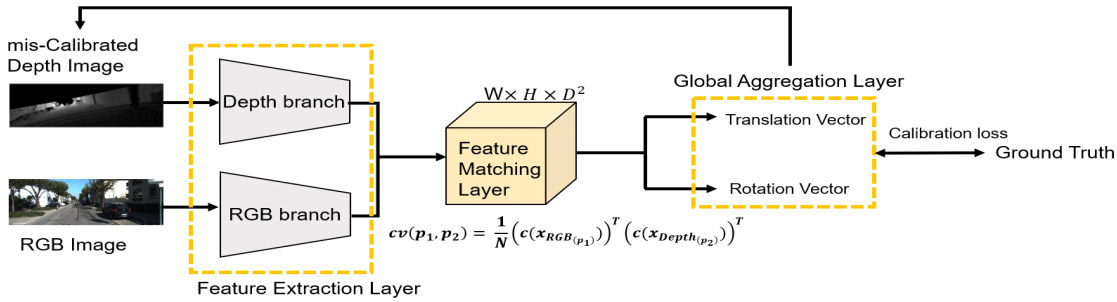


Fig. 1. The workflow of LCCNet

한 깊이 정보를 얻기 위해 라이다와 카메라의 데이터를 동일한 좌표계로 변환하는 것이다. 2D인 카메라의 이미지 데이터와 3D인 라이다의 포인트 간의 대응관계를 설정하여 출력을 융합한다. 카메라가 색감, 질감, 모양 정보를 캡처하는 동안 라이다는 이미지의 3D 구조적 정보를 캡처한다. 두 센서의 데이터를 융합하여 장면을 정확하게 식별한다. 이는 자율주행, 네비게이션, 로봇공학 등에서 3D 이미지 재구성과 객체 검출을 위해 널리 사용된다.

2.2 LCCNet

본 연구에서 사용된 카메라-라이다 융합 모델은 LCCNet[17]으로 end-to-end 심층 신경망을 훈련하여 외부 파라미터를 직접 예측하기 위해 제안된 네트워크이다. LCCNet은 기존의 외부 파라미터와 카메라의 내부 파라미터를 통해 라이다의 포인트를 2D로 투영하여 depth image를 생성한다. 생성된 Depth image와 RGB image는 Depth branch와 RGB branch에서 각 피처를 추출한다. 피처 추출 네트워크의 입력 크기는 256x512이다. feature matching layer에서 PWC-Net의 correlation layer를 사용해 Depth 피처맵 x_{Depth}과 RGB 피처맵 x_{RGB}의 픽셀 정보를 매칭하여 cost volume $cv(p_1,p_2)$을 계산한다.

$$cv(p_1,p_2)=\frac{1}{N}(c(x_{RGB(p_1)}))^T(c(x_{Depth(p_2)})) \quad (1)$$

여기서 $c(x)$는 피처맵 x의 flatten vector이며, N은 피처맵 x의 채널 크기($=c(x)$의 크기)이다. cost volume의 dimension은 $W\times H\times D^2$ 이며, $|p_1-p_2|_\infty \le D$ 이다. W, H는 피처맵 크기이다. cost volume feature는 feature global aggregation 과정을 통해 6-DoF 강체 변환(rigid-body transformation)을 추론한다. 네트워크의 출력은 1x3 translation vector와 1x4 quaternion이다.

LCCNet의 손실함수는 Regression loss L_T 와 Point cloud distance loss L_P 를 사용한다.

$$L=\lambda_T L_T+\lambda_P L_P \quad (2)$$

2.2.1 Regression Loss

회전에 대한 사원수(quaternion)는 방향 정보이므로 Euclidean distance는 두 개의 사원수에 대해 설명할 수 없다, 그러므로 Angular distance를 이용해 두 사원수의 차이를 설명한다.

$$L_R=D_a(q_{gt},q_{pred}) \quad (3)$$

D_a는 Angular distance이다. 전체 Regression Loss는 다음과 같다.

$$L_T=\lambda_t L_t+\lambda_q L_R \quad (4)$$

L_t는 $smooth\,L_1$ loss를 적용한 translation vector이고, λ_t와 λ_q는 각각의 loss weight이다.

2.2.2 Point Cloud Distance Loss

포인트 클라우드의 거리 제약을 위한 손실함수이다. 사원수 q_{pred}를 rotation matirx R_{pred}로 변환한 후 homogeneous matrix T_{pred}를 추출한다.

$$T_{pred} = \begin{bmatrix} R_{pred} & t_{pred} \\ 0 & 1 \end{bmatrix} \quad (5)$$

$P = \{P_1, P_2, ..., P_N\}, P_i \in R^3$ 는 라이다의 포인트 클라우드이고 Point Cloud distance loss L_P는 다음과 같다.

$$L_P = \frac{1}{N}\sum_{i=1}^{N} \| T_{LC}^{-1} \cdot T_{\mathrm{Pred}}^{-1} \cdot T_{\mathrm{init}} \cdot \mathrm{P_i} - \mathrm{P_i} \|_2 \quad (6)$$

N은 포인트 클라우드 수, $\| \cdot \|_2$는 L_2 정규화이다. T_{LC}는 카메라-라이다 extrinsic matrix이다. 보정되지 않은 라이다와 카메라 간의 extrinsic matrix $\hat{T}_{init}$는 융합 모델의 예측 결과 T_{pred}와 초기 보정 파라미터 T_{init}을 조합하여 얻는다.

이동 〔1.5m, 1.0m, 0.5m, 0.2m, 0.1m〕, 회전 〔20°,10°,5°,2°,1°〕의 다양한 범위를 선택하여 network를 학습시킨다. 가장 큰 범위인 (1.5m, 20°〕로 이동·회전 범위로 입력하여 예측한 T_{pred}를 T_0로 간주하고 $T_{-1}^0 \cdot T_{init}$로 라이다 포인트 클라우드를 재투영하여 더 많은 Depth image를 생성한다. 새로운 Depth image와 동일한 RGB image는 두 번째 범위 (1.0m,10°)으로 이동·회전 범위를 설정하여 T_1을 예측한다. 앞서 언급한 프로세스를 5회 반복하여 최종 extrinsic matrix를 구한다.

$$\widehat{T_{LC}} = (T_0 \cdot T_1 \cdots T_5)^{-1} \cdot T_{\mathrm{init}} \quad (7)$$

2.3 이미지 스케일링 알고리즘

이미지 스케일링은 컴퓨터 비전의 표준 절차이며 기계학습의 일반적인 전처리 단계이다. 이미지의 크기나 해상도를 변경할 때 픽셀 간의 값을 채우는 과정을 보간법(interpolation)이다. 보간법에는 nearest interpolation, bilinear interpolation, bicubic interpolation, lanczos interpolation 등이 있다. 본 논문에서는 각 보간법을 적용한 스케일링 과정을 스케일링 알고리즘이라고 표현한다. nearest 알고리즘은 주변의 픽셀을 그대로 복사하여 픽셀 사이에 넣는 방식이다. 픽셀이 눈에 띄기 때문에 해상도가 낮은 이미지에 주로 이용한다. bilinear 알고리즘은 주변의 4개 픽셀을 사용하는 방식이다. 간단한 방식이기 때문에 속도가 빠르다는 장점이 있다. bicubic 알고리즘은 주변의 16(4x4)개 픽셀을 사용하는 방식이다. bilinear 알고리즘보다 많은 픽셀을 고려하기 때문에 계산량이 많고 더 깔끔한 이미지를 얻을 수 있다. 인접한 픽셀 간의 색조값의 차이로 인한 불연속이 있는 경우 해상도가 낮다. lanczos 알고리즘은 각 픽셀의 사인함수의 합을 사용하는 방식이며 가장 해상도가 높은 이미지를 얻을 수 있는 알고리즘이다.

이미지를 이산 및 2차원 신호로 고려했을 때, 주파수 영역에 표현할 수 있다. 이미지의 크기를 작게 조정하면 높은 주파수가 손실된다. 이 과정은 고주파 신호가 저주파 신호로 변환되는 신호 처리에서의 다운샘플링 과정과 관련이 있다. 다운 샘플링의 주요 문제점은 주어진 해상도가 모든 이미지의 주파수를 설명할 수 없다는 것이다. 나이키스트-샤넌 정리〔18〕에 따르면, 샘플링 속도 f_T가 신호의 가장 높은 주파수 $f_{\max}$보다 적어도 두 배 높은 경우, 이산 샘플링 포인트에서 신호 $s(t)$를 재구성할 수 있다

$$f_T \geq 2 \cdot f_{\max} \quad (8)$$

주파수 f_T가 임계값보다 낮은 경우 샘플링 된 점은 원래 신호와 구별하기 힘들다. Fig. 2.는 샘플링된 점들에 의해 설명되는 두 신호 $s(t)$와 $\hat{s}(t)$중 하나를 결정하는 것이 불가능한 이 현상의 예를 보여준다. 재구성된 신호는 앨리어싱 현상으로 인해 원해 신호와 다를 수 있다. 이미지 스케일링 공격은 앨리어싱 현상을 기반으로 신호를 교묘하게 조작하여 다운 샘플링된 버전이 새로운 신호가 되도록 한다.

본 논문이 의미하는 스케일링 알고리즘은 단순히 이미지의 빈도를 줄이는 것이 아니라 앨리어싱 현상을 완화하기 위해 이미지를 축소하기 전에 소스 이미지의 픽셀을 보간한다. OpenCV, Pillow 등의 이미징 라이브러리의 스케일링 알고리즘은 먼저 이미지의 크기

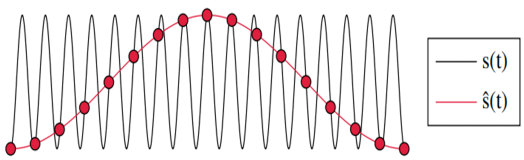


Fig. 2. An example of an undersampled signal $s(t)$〔14〕

를 수평으로 조정한 다음 수직으로 조정하여 이미지를 축소한다.

$$ScalingFunc(X_{m*n}) = CV_{m'*m} * (X_{m*n}) * CH_{n*n'} \quad (9)$$

$ScalingFunc(X_{m*n})$은 $m*n$ 크기의 이미지 X_{m*n}를 $m'*n'$크기인 이미지 $SA_{m'*n'}$로 축소하는 스케일링 알고리즘이다. $CV_{m'*m}$와 $CH_{n*n'}$은 수직 스케일링 $(m*n' \rightarrow m'n')$, 수평 스케일링 $(m*n \rightarrow m*n')$을 하는 계수 행렬이다. $ScalingFunc()$의 구현 세부 정보를 오픈 소스 패키지에서는 알 수 있으므로 $CV_{m'*m}$와 $CH_{n*n'}$을 정확하게 계산할 수 있다.

2.4 스케일링 공격

Xiao[14]은 이미지 스케일링 알고리즘에 대한 공격을 제안했다. Fig. 3.과 같이 공격자가 소스 이미지를 교란하여 새로운 이미지를 생성하며 생성된 공격 이미지가 특정 치수로 스케일링 되었을 시 타겟 이미지가 출력되는 기존의 공격 방법이다. 제안된 연구는 스케일링 알고리즘들이 스케일링 된 이미지를 계산하기 위해 소스 이미지의 모든 픽셀을 동등하게 고려하지 않기 때문에 스케일링 조정 공격이 가능하다는 점을 이용한다. Fig. 4.는 LCCNet에 스케일링 공격된 이미지를 입력했을 경우 결과 이미지이다. 스케일링 공격을 수행하는 방법은 다음과 같다.

$$\begin{aligned} &CV_{m'*m} * (S_{m*n} + \Delta_1) * CH_{n*n'} = SA_{m'*n'} \\ &SA_{m'*n'} = T_{m'*n'} + \Delta_2 \\ &\|\Delta_2\|_\infty \le \epsilon * IN_{\max} \end{aligned} \quad (10)$$

$m*n$ 크기의 소스 이미지 S_{m*n}에 섭동 Δ_1을 추

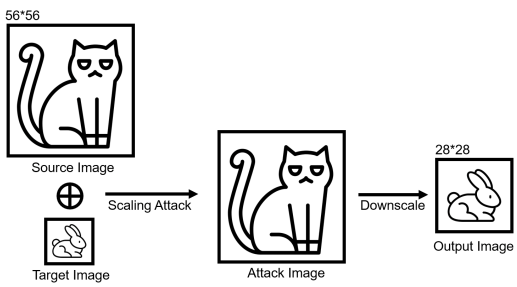


Fig. 3. An example of Image Scaling Attack

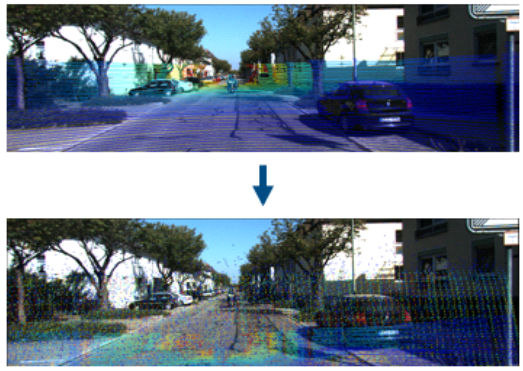


Fig. 4. Normal output and output with attack image in LCCNet

가한 후 $m'*n'$크기인 이미지 $SA_{m'*n'}$로 축소한다. $SA_{m'*n'}$은 타겟 이미지 $T_{m'*n'}$에 섭동 Δ_2를 추가한 것과 같아야 한다.

$CV_{m'*m}$와 $CH_{n*n'}$을 정확하게 계산하는 것은 가능하지만 계수 행렬이 커지고 보간 방법이 복잡해질수록 절차가 어려워진다. 따라서 계수 행렬을 쉽게 추론하는 방법을 사용한다. 다음과 같은 식을 설정 할 경우:

$$\begin{aligned} &CV_{m'*m} * (I_{m*m} * IN_{\max}) = CV_{m'*m} * IN_{\max} \\ &(I_{n*n} * IN_{\max}) * CH_{n*n'} = CH_{n*n'} * IN_{\max} \end{aligned} \quad (11)$$

$S = I_{m*m} * IN_{\max}$ 로 가정하면 다음을 얻을 수 있다.

$$\begin{aligned} &Scaling * Func(S) = \\ &unsigned\,int(CV_{m'*m} * IN_{\max}) \\ &\rightarrow \widehat{CV_{m*m}} \approx \frac{SA}{IN_{\max}} \end{aligned} \quad (12)$$

I_{m*m}과 I_{n*n} 은 단위 행렬이다. 이론적으로 계수 행렬의 각 행의 요소 합은 1이므로 식(12) 와 같이 각 행에 대한 정규화를 수행하면 $CV_{m'*m}$의 근사치를 얻을 수 있다. $CH_{n*n'}$도 같은 방법으로 진행하여 구할 수 있다.

$$\begin{aligned} &\widehat{CV_{m*m}}[i,:] = \frac{\widehat{CV_{m*m}}[i,:]}{\sum_{j=0}^{m-1} (\widehat{CV_{m*m}}[i,j])} \\ &(i = 0,1,\ldots,m'-1) \end{aligned} \quad (13)$$

섭동 행렬 Δ_1은 m*n 차원인 행렬이다. 계산복잡

도가 $O(n^2)$이상이므로 이미지 크기가 커질수록 최적화 비용이 많이 든다. 따라서 약간의 정밀도 손실을 감안하고 이미지 생성 과정을 두 개로 분리하여 사용한다. 계수 행렬 계산을 마쳤다고 가정하고 섭동 행렬을 여러 개로 분리하여 계산할 수 있다. $A_{m^*n} = S_{m^*n} + \Delta_1$일 때, 수직 방향으로 스케일링할 경우 이미지 변환은 식(13)과 같이 나타낼 수 있으며, 이러한 방식으로 컬럼별 하위 최적화 문제로 단순화 가능하다.

$$CV_{m'^*m}{}^*A = [CV^*A[:,0]_{(m^*1)} \cdots CV^*A[:,n-1]_{(m^*1)}] \tag{14}$$

$$\begin{aligned} &obj : \min/\max(\|\Delta_1[:,j]\|^2) \\ &s.t.\ CV^*A[:,j]_{(m^*1)} = T[:,j]_{(m'^*1)} + \Delta_2 \\ &\quad \|\Delta_2\|_\infty \le \epsilon^* IN_{\max} \\ &\quad (j = 0,1,\ldots,n-1) \end{aligned} \tag{15}$$

식(14)을 quadratic problem[21]으로 공식화한다. 공격 이미지의 각 요소는 $[0, IN_{\max}]$ 내에 존재한다. 제약조건은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} &O \le A[:,j]_{m^*1} \le IN_{\max} \\ &\|CV^*A[:,j]_{m^*1} - T[:,j]_{m'^*1}\|_\infty \le \epsilon^* IN_{\max} \end{aligned} \tag{16}$$

목적함수는 다음과 같다.

$$\begin{aligned} &\min/\max(\Delta_1[:,j]^T I_{m'^*m'} \Delta_1[:,j]) \\ &(j = 0,1,\cdots,n-1) \end{aligned} \tag{17}$$

목적함수를 식(16)을 식(15) 와 결합하여 최종적으로 m‘ 차원의 quadratic problem을 얻는다. 최적화 문제를 해결하기 위해 DCCP toolbox[19]를 이용한다.

2.5 이미지 스케일링 공격 방어 방법

Xiao[14]은 이미지 스케일링 공격 알고리즘에 대한 공격 방어 방법을 제안한다.

첫 번째 방법은 스케일링 공격에 강한 스케일링 알고리즘이을 사용하는 것이다. OpenCV에서 제공하는 스케일링 알고리즘의 소스 코드를 조사한 결과, 스케일링 알고리즘이 nearest는 1, bilinear은 2, bicubic은 4, lanczos4는 8로 고정된 크기의 컨볼루션 커널과 함께 구현된다는 것을 관찰한다. 이러한 알고리즘은 스케일링 비율이 커널 폭을 초과하면 취약해지고 스케일링 중에 소스 이미지의 픽셀이 생략된다. 그러나 Area 스케일링 알고리즘은 동적 커널 폭으로 구현되어 스케일링 공격으로부터 보호할 수 있다. 그러나 Area 스케일링 알고리즘은 소스 이미지에 조작된 픽셀의 잔재가 크게 남아 눈으로 식별할 수 있어 사용이 제한적이다.

두 번째 방어 방법은 스케일링 알고리즘으로 처리된 픽셀을 식별한 후 나머지 픽셀을 사용하여 이미지를 재구성한다. 재구성 방법에는 몇가지 방법이 있으나 투명한 보안 특성을 가진 선택적 중앙값 필터와 선택적 무작위 필터를 사용한 재구성 방법을 제안한다. 선택적 중앙값 필터는 스케일링 알고리즘에 의해 고려되는 픽셀 주변의 모든 값의 중앙값을 취하여 이미지를 보정하는 것이다. 선택적 무작위 필터는 중앙값이 아닌 임의의 점을 취하는 필터이다. 이러한 방어는 기존 스케일링 알고리즘에 쉽게 사용할 수 있다.

III. 제안 방법

본 논문에서는 카메라-라이다 센서 융합 모델의 오류 유발을 위한 스케일링 공격 방법을 제안하며 공격 과정은 Fig. 5.와 같다.

카메라 이미지 평면에 투영된 라이다의 포인트 클라우드를 투영하여 2차원 라이다 이미지를 생성한다. 라이다의 포인트 클라우드는 360도 전방향에 대한 데이터가 기록되어 있다. 따라서 카메라 이미지와 같은 시점으로 매핑하는 전처리가 필요하다. 먼저, 라이다 좌표를 기준 좌표계(reference coordinate)로 변환한 후, reflection 과정을 통해 월드 좌표(world coordinate)를 구한다. 이 월드 좌표를 각 카메라의 projection matrix를 이용해 투영하여 이미지 좌표(image coordinate)를 구한다. 투영된 라이다 포인트를 라이다 이미지라고 한다.

라이다 이미지를 식(10)과 같은 방식으로 공격을 수행한다. 계수 행렬을 추론한 후 섭동 행렬을 추론한다. 라이다에 영향을 최소화하기 위해 quadratic problem으로 공격 라이다 이미지와 정상 라이다 이미지의 차이를 거의 없게 한다. 식(10)과 다른 점은 ϵ을 고정값으로 사용하지 않는다. quadratic problem solver가 문제를 해결 못 할 경우 1에서 100까지 점차 값을 증가시킨다. ϵ이 증가할수록 공격

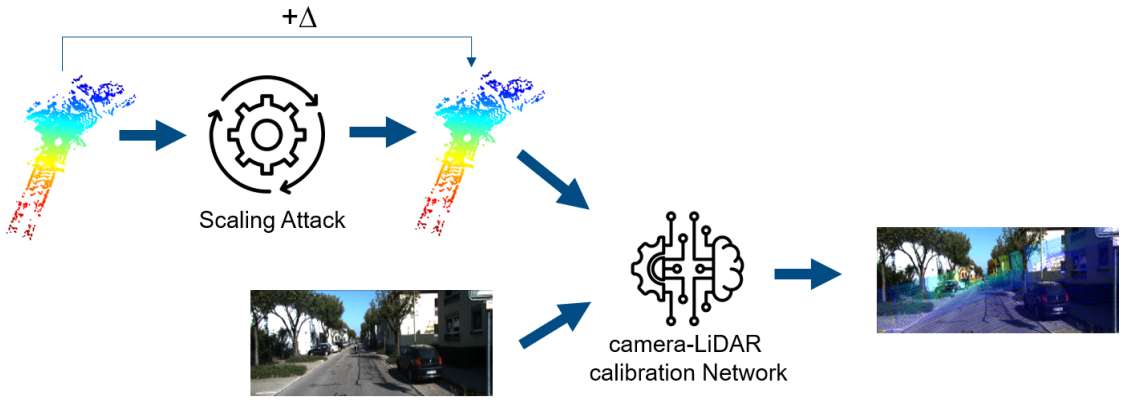


Fig. 5. The workflow of our proposed scaling attack method for misalignment error of camera-LiDAR calibration model.

성공률이 올라간다. 결과로 공격 라이다 이미지를 출력한다. 공격 라이다 이미지를 타겟 이미지 크기로 스케일링할 경우 타겟 이미지와 유사한 이미지가 된다.

융합 모델은 공격 라이다 이미지와 정상 RGB 이미지를 입력으로 하여 잘못 교정된 출력을 내놓는다.

IV. 실험 및 실험 결과

4.1 실험 환경 설정

본 논문에서 제안한 카메라-라이다 융합 모델 공격을 평가하기 위해 오픈 Dataset인 KITTI 주행 거리 측정 Dataset와 nuScenes Dataset를 사용했다. Dataset는 각 센서 사이의 정합 매개 변수를 제공하며 그 중 카메라와 라이다 간의 융합 매개 변수를 ground-truth로 사용했다.

본 논문에서는 라이다와 카메라의 왼쪽 컬러 카메라 사이의 정합만 고려한다. KITTI 주행 거리 측정 Dataset의 20개 시퀀스(34350 프레임)를 train과 validation에 사용했으며, 1개의 시퀀스(4541 프레임)을 test로 사용했다. nuScenes Dataset는 700개의 씬을 train과 validation에 사용했으며, 150개의 씬(6008 프레임)으로 test로 사용했다.

공격 라이다 이미지는 nearest, bicubic, bilinear, lanczos4, 이렇게 4가지 보간법을 사용한 스케일링 알고리즘과 10x20, 60x90, 100x150, 256x512 4가지 스케일링 크기를 사용한 이미지 스케일링 공격을 통해 총 16종으로 생성한다. 스케일링 크기란 타겟 이미지의 크기를 의미한다. 학습 데이터를 늘리기 위해 extrinsic matrix에 설정한 범위 내의 편차를 추가한다. 최대 오보정 회전각은 20° 이고 최대 오보정 이동 거리는 1.5m로 설정했다. 편차값을 랜덤으로 설정하면 많은 양의 데이터를 얻을 수 있다.

공격 모델 대상인 LCCNet에 사용된 스케일링 알고리즘과 스케일링 크기를 추론[20]했다고 가정한다. LCCNet은 OpenCV 이미징 라이브러리의 이미지 스케일링 알고리즘을 사용하고 사용된 스케일링 알고리즘은 Bilinear, 스케일링 크기는 256x512이다.

4.2 카메라-라이다 정합 공격 성능평가

공격 성능평가는 스케일링 알고리즘과 스케일링 크기별 공격 후 평균 제곱 이동오류 E_t, 평균 사원수 각도 오류 E_R를 측정한다. E_t는 Translation vector의 절대 오차를 계산하며 다음과 같다.

$$E_t = \|t_{pred} - t_{gt}\|_2 \tag{18}$$

E_R는 Rotation vector를 Euler angles로 변환하고 Roll, Pitch, Yaw의 각도 오차를 계산한다.

KITTI 주행 거리 측정 Dataset의 스케일링 알고리즘과 크기별 공격 성능은 Table 1., Table 2.. 와 같다. non-attack은 공격하지 않은 데이터의 예측 결과 오류이다. non-attack보다 공격한 데이터의 예측 결과 오류가 더 큰 경우, 공격이 성공했다고 판단한다. non-attack 결과는 평균 제곱 이동오류

Table 1. An Example scaling attacked KITTI odometry dataset image with LiDAR point cloud by scaling algorithm and size.

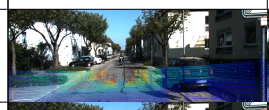
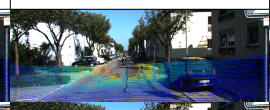
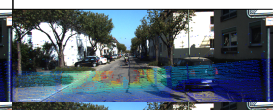
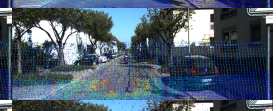
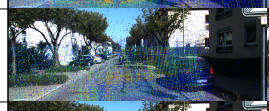
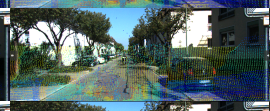
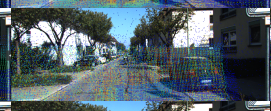
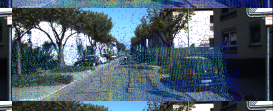
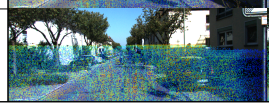
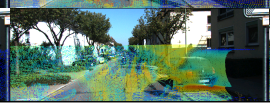
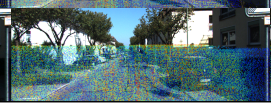
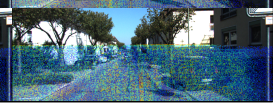
	nearest	bilinear	bicubic	lanczos4
10X20				
60X90				
100X150				
256X512				

Table 2. The result of the calibtration error for attacked KITTI Odometry Dataset

non-attack	E_t	0.191	E_R	0.006						
	nearest		bilinear		bicubic		lanczos4		mean	
Scale size	E_t	E_R	E_t	E_R	E_t	E_R	E_t	E_R	E_t	E_R
(10, 20)	0.141	0.102	0.294	0.148	0.164	0.128	0.247	0.131	0.212 (+0.021)	0.127 (+0.121)
(60, 90)	0.254	0.203	1.256	0.200	0.717	0.192	0.737	0.188	0.741 (+0.550)	0.196 (+0.190)
(100, 150)	2.418	0.228	3.786	0.323	3.337	0.404	3.302	0.325	3.211 (+3.020)	0.320 (+0.314)
(256, 512)	5.840	0.311	19.204	1.292	8.078	0.292	8.457	0.284	36.323 (+36.132)	2.179 (+2.173)
mean	2.163 (+1.972)	0.211 (+0.205)	6.135 (+5.944)	0.491 (+0.485)	3.074 (+2.883)	0.254 (+0.248)	3.186 (+2.995)	0.232 (+0.226)		

Table 3. The result of the calibtration error for attacked NuScenes Dataset

non-attack	E_t	1.611	E_R	0.365						
	nearest		bilinear		bicubic		lanczos4		mean	
Scale size	E_t	E_R	E_t	E_R	E_t	E_R	E_t	E_R	E_t	E_R
(10, 20)	1.808	0.369	1.722	0.371	1.860	0.369	1.641	0.368	1.758 (+1.147)	0.370 (+0.004)
(60, 90)	1.886	0.370	2.590	0.371	2.291	0.371	2.203	0.370	2.243 (+1.632)	0.371 (+0.005)
(100, 150)	3.129	0.371	3.808	0.371	3.622	0.370	3.716	0.370	3.569 (+2.958)	0.371 (+0.005)
(256, 512)	13.970	0.369	33.223	0.369	23.960	0.369	22.822	0.369	23.494 (+22.883)	0.369 (+0.004)
mean	5.198 (+3.587)	0.370 (+0.005)	10.336 (+8.725)	0.370 (+0.006)	7.933 (+6.322)	0.370 (+0.005)	7.596 (+5.985)	0.369 (+0.004)		

가 0.191, 평균 사원수 각도 오류가 0.006이다.

Nuscenes Dataset의 스케일링 알고리즘과 크기별 공격 성능은 Table 3.와 같다. non-attack 결과는 평균 제곱 이동 오류가 1.611, 평균 사원수 각도 오류가 0.365이다.

4.2.1 KITTI 주행 거리 측정 Dataset의 공격 결과

KITTI Dataset의 스케일링 크기별 공격 성능을 평균 내었을 때,10x20 크기일 경우 평균 제곱 이동 오류가 +0.021, 평균 사원수 각도 오류가 +0.121, 60x90 크기일 경우 평균 제곱 이동오류가 +0.550, 평균 사원수 각도 오류가 +0.190, 100x150 크기일 경우, 평균 제곱 이동오류가 +3.020, 평균 사원수 각도 오류가 +0.314, 256x512 크기일 경우, 평균 제곱 이동 오류가 +36.132, 평균 사원수 각도 오류가 +2.173로 모든 경우에서 오류가 증가했다. 뿐만 아니라 LCCNet에서 사용하는 스케이링 사이즈인 256x512에서 훨씬 큰 오류가 발생했다.

KITTI 주행 거리 측정 Dataset를 스케일링 알고리즘별 공격 성능을 평균 내었을 때, LCCNet에서 사용하는 bilinear 알고리즘이 평균 제곱 이동 오류가 +5.944, 평균 사원수 각도 오류가 +0.465로 가장 오류가 증가하여 공격 성능이 높았고, nearest 알고리즘의 성능은 평균 제곱 이동 오류가 +1.972, 평균 사원수 각도 오류가 +0.205로 가장 낮았다. bilinear 알고리즘과 256x512 크기에서 평균 제곱 이동 오류는 3112%, 평균 사원수 각도는 7750% 오류가 증가했다. 뿐만 아니라 모든 경우에서 non-attack보다 오류가 커졌다.

4.2.2 NuScenes Dataset의 공격 결과

Nuscenes Dataset의 스케일링 크기 별 공격 성능을 비교했을 때, 10x20 크기일 경우 평균 제곱 이동 오류가 +1.147,평균 사원수 각도 오류가 +0.004, 60x90 크기일 경우 평균 제곱 이동오류가 +1.632, 평균 사원수 각도 오류가 +0.005, 100x150 크기일 경우, 평균 제곱 이동오류가 +2.958, 평균 사원수 각도 오류가 +0.005, 256x512 크기일 경우, 평균 제곱 이동오류가 +22.883, 평균 사원수 각도 오류가 +0.004이다. 평균 사원수 각도 오류가 ground-truth와 큰 차이가 없는 결과가 나왔다.

스케일링 알고리즘별 공격 성능을 평균 내었을 때, bilinear 알고리즘이 평균 제곱 이동 오류가 +8.725, 평균 사원수 각도 오류가 +0.006로 가장 오류가 증가하여 공격 성능이 높았고, Nearest 알고리즘의 성능은 평균 제곱 이동 오류가 +3.587 평균 사원수 각도 오류가 +0.005로 가장 낮았다. 평균 제곱 이동 오류의 최대 차이는 5.138이고, 알고리즘별로 비교했을 때, 평균 사원수 각도 오류의 최대 차이는 0.002이다.

스케일링 크기별 오류 최대 차이와 마찬가지로 평균 사원수 각도 오류의 최대 차이가 매우 작은 수치이다. 그러나 평균 제곱 이동 오류는 LCCNet에서 사용하는 알고리즘인 bilinear에서 확연하게 큰 오차 결과가 도출되었다.

V. 고 찰

본 논문은 카메라-라이다 융합 모델을 공격하는 방법으로 학습 기반의 정합 방식에서 필수적인 이미지 스케일링 단계를 활용하여 기존의 RGB image에서 행했던 이미지 스케일링 공격을 기반으로 라이다 포인트 클라우드의 Depth image에 적용하는 방법을 제안하였다. 스케일링 공격 과정에서 포인트 클라우드에 노이즈가 들어가게 된다. 뿐만 아니라 3D 포인트 클라우드를 학습에 활용하기 위해서 2차원의 평면에 투영해야 한다. 변환 과정에서 픽셀 누락이 있음을 염두해야 한다.

실험 결과, KITTI 주행 거리 Dataset와 NuScenes Dataset에 대부분의 공격에 성공했으며, 4가지 모든 알고리즘과 스케일링 크기에도 공격에 성공했다. LCCNet에서 사용된 스케일링 알고리즘과 크기뿐만 아니라 다른 경우에서도 공격이 성공했던 것은 이미지 투영 과정에서의 포인트 유실도 영향을 미쳤을 것이다.

LCCNet과 같은 다중 센서 융합 모델은 sparse한 입력에도 훌륭한 성능을 내기 위한 모델이다. 고해상도를 유지하면서 3D 데이터를 2D로 전환하는 코덱을 적용한다면 더 좋은 성능의 센서 융합 모델이 될 것으로 생각한다.

NuScenes Dataset에서 평균 사원수 각도 오류가 거의 일정하게 도출되었다. Rotation 부분에서 작용을 제대로 안 했다는 점에서 추후 연구가 필요하다. 기존의 이미지 스케일링 공격 논문에 제안된 방어 방법은 area 스케일링 알고리즘을 사용하는 것이다. area 알고리즘을 적용하여 실험한 결과 공격 성

공률이 매우 낮았다. 공격 결과 이미지에 노이즈가 확연히 드러나 공격 데이터로 사용하기에 부적합하다고 판단했다. 더 확실한 공격 방법과 센서 융합 기반의 모델 공격을 방어하기 위한 연구가 필요할 것으로 보인다.

본 연구는 모델의 스케일링 알고리즘과 크기를 추론하면 적용이 가능한 블랙박스 공격이다. 그러나 현실적으로 주행하는 차량의 입력에 공격 이미지를 주입하는 것은 어렵다. 그러나 자율주행 차량의 인식 시스템에서 대부분 사용되는 센서 융합 과정을 공격하는 것은 융합된 데이터를 사용하는 차선 감지, 객체 인식 등의 비전 작업등의 영향을 미칠 것이라는 접근은 바람직하다고 생각한다.

VI. 결 론

본 논문에서는 입력 라이다 이미지에 대한 스케일링 공격을 통해 카메라-라이다 융합 모델의 오류를 유발하는 방법을 최초로 제안한다. 선행 연구에 카메라와 라이다 센서를 융합을 기반으로 하는 객체 인식·분류 모델에서의 공격은 있었으나 센서 간 융합 모델에 대한 공격은 없었다. 공격자는 자율주행 모델의 인식 시스템에 사용되는 센서 융합 모델에 블랙박스 접근 권한을 가진 상황에서 공격을 수행한다. 카메라와 라이다 융합 모델은 입력 라이다에 스케일링 공격을 하여 융합 오차를 높임으로써 잘못 교정된 결과를 출력한다. 이후 잘못 교정된 결과 이미지를 사용한 비전 작업에서 오류를 유발하여 사고를 유도할 수 있다. 카메라와 라이다의 정합 모델에 해당 공격을 테스트 함으로써 자율주행 모델의 인식 시스템에 대한 안전성을 시험할 수 있으며, 해당 공격에 대한 방어 기술을 마련할 수 있다. 실험을 통해 라이다에 대한 스케일링 공격은 카메라-라이다 융합 모델 오류에 효과적임을 확인했으며, 스케일링 공격이 포인트 클라우드 데이터에 적용할 수 있음을 보였다. 스케일링 알고리즘과 크기별 공격 성능 실험을 통해 공격 전보다 정합 오류가 증가함을 보였다. 실험 결과, KITTI 주행 거리 측정 Dataset와 NuScenes Dataset에서 LCCNet이 사용하는 bilinear 알고리즘의 공격 성능이 각각 평균 제곱 이동 오류가 +5.944, +8.725이고 평균 사원수 각도 오류가 +0.0.465, +0.006으로 오류 증가율이 가장 높았다. 또한 LCCNet이 사용하는 256x512 크기에서 각각 평균 제곱 이동 오류가 +36.132, +22.866. 평균 사원수 각도 오류가 +2.173,+0.006으로 오류 증가율이 가장 높았다. 그뿐만 아니라 모든 경우에서도 공격이 성공적임을 확인했다. 본 연구를 통해 라이다 센서에 대한 공격으로 평균 77% 이상의 융합 오류를 유발하는 결과를 보였다.

References

[1] Shadrin, Sergej S., and Anastasiia A. Ivanova. "Analytical review of standard Sae J3016 "taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles" with latest updates. Avtomobil. Doroga. Infrastruktura, March. 2019.

[2] cbinsights, "40+ Corporations Working On Autonomous Vehicles", https://www.cbinsights.com/research/autonomous-driverless-vehicles-corporations-list. 2020-12-16

[3] D. Frossard and R. Urtasun, "End-to-end Learning of Multi-sensor 3D Tracking by Detection", ICRA. IEEE, pp. 635 - 642, May. 2018

[4] M. Liang, B. Yang, S. Wang, and R. Urtasun, "Deep Continuous Fusion for Multi-Sensor 3D Object Detection", ECCV, pp. 641-656, 2018.

[5] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-View 3D Object Detection Network for Autonomous Driving", CVPR, pp. 1907-1915, 2017.

[6] D. Xu, D. Anguelov, and A. Jain, "PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation,", CVPR, 2018, pp. 244 - 253. 2018

[7] M. Liang, B. Yang, Y. Chen, R. Hu, and R. Urtasun, "Multi-Task MultiSensor Fusion for 3D Object Detection,", CVPR, pp. 7345-7353, 2019.

[8] X. Du, M. H. Ang, and D. Rus, "Car Detection for Autonomous Vehicle: LIDAR and Vision Fusion Approach Through Deep Learning Framework,", IR

OS, pp. 749-754, September. 2017.
[9] J. Ku, M. Mozifian, J. Lee, A. Harakeh, and S. L. Waslander, "Joint 3D Proposal Generation and Object Detection from View Aggregation,", IROS, pp. 1 - 8. October. 2018,
[10] X. Ma, Z. Wang, H. Li, P. Zhang, W. Ouyang, and X. Fan, "Accurate Monocular 3D Object Detection via Color-Embedded 3D Reconstruction for Autonomous Driving," in CVPR, pp. 6851 - 6860. 2019,
[11] X. Du, M. H. Ang, S. Karaman, and D. Rus, "A General Pipeline for 3D Detection of Vehicles", ICRA IEEE, pp. 3194 - 3200. May. 2018,
[12] Cao, Y., Wang, N., Xiao, C., Yang, D., Fang, J., Yang, R., ... & Li, B."Invisible for both camera and lidar: Security of multi-sensor fusion based perception in autonomous driving under physical-world attacks", In IEEE Symposium on Security and Privacy (SP), pp. 176-194, May, 2021
[13] Abdelfattah, M., Yuan, K., Wang, Z. J., & Ward, R. "Adversarial attacks on camera-lidar models for 3d car detection", IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 2189-2194, September. 2021
[14] Quiring, E. Klein, D, Arp, D. Johns, M. & Rieck, K. "Adversarial preprocessing: Understanding and preventing image-scaling attacks in machine learning", USENIX Conference on Security Symposium, pp. 1363-1380. August. 2020,
[15] Geiger, A., Lenz, P., & Urtasun, R. " Are we ready for autonomous driving? the kitti vision benchmark suite." IEEE conference on computer vision and pattern recognition, pp. 3354-3361, June. 2012
[16] Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q. & Beijbom, O "nuscenes: A multimodal data set for autonomous driving." IEEE/CVF conference on computer vision and pattern recognition. pp. 11621-11631, 2020.
[17] Lv, X., Wang, B., Dou, Z., Ye, D., & Wang, S. "LCCNet: LiDAR and camera self-calibration using cost volume network", IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2894-2901, 2021
[18] A. V. Oppenheim, J. R. Buck, and R. W. Schafer. "Discrete-time Signal Processing"; 2nd ed. Prentice-Hall, 1999.
[19] github, "DCCP source code", https://github.com/cvxgrp/dccp, 2017-09-03
[20] Q. Xiao, Y. Chen, C. Shen, Y. Chen, and K. Li "Seeing is not believing: Camouflage attacks on image scaling algorithms." 28th USENIX Security Symposium, pp. 443-460. 2019.
[21] S. Boyd and L. Vandenberghe. "Convex Optimization", Cambridge university press, 2004.

〈저자소개〉

임 이 지 (Yi-ji Im) 학생회원
2021년 2월: 영남대학교 컴퓨터공학과 학사
2022년 9월~현재: 숭실대학교 소프트웨어학과 석사과정
〈관심분야〉 컴퓨터 비전, 전산 이미징, AI 보안

최 대 선 (Dae-seon Choi) 종신회원
1995년 2월: 동국대학교 컴퓨터공학과 학사
1997년 2월: 포항공과대학교 컴퓨터공학과 석사
2009년 1월: 한국과학기술원 전산학과 박사
1997년 1월~1999년 6월: 현대정보기술 선임
1999년 7월~2015년 8월: 한국전자통신연구원 인증기술연구실 실장/책임연구원
2015년 9월~2020년 8월: 공주대학교 의료정보학과 부교수
2020년 9월~현재: 숭실대학교 소프트웨어학부 교수
2016년~현재: 정보보호학회 이사
〈관심분야〉 인증, 개인정보보호, AI 보안