

AI를 활용한 무인기 표적과 새 판단 방법

2023.08.23

LIG넥스원 / 레이다연구소 4팀
조인철 선임연구원

Table of Contents

- 1 연구 목적
 - 2 학습 데이터 특성
 - 3 전처리
 - 4 딥 러닝 모델 구조
 - 5 학습 결과
 - 6 정리
-

1 연구목적

- 증가하는 무인기 위협에 대응하기 위한 정확한 탐지 / 식별 수단 필요

시사주간

北 무인기에 놀란 軍...이번엔 쫓다보니 '새떼'

[시사주간=양승진 북한 전문기자] 군 당국은 27일 인천 강화 지역 상공에 나타났다는 '무인기'에 대해 새떼를 오인한 것이라고 밝혔다.



KBS 뉴스

북한 무인기 출현 소동...'새떼'로 판명

[앵커] 안녕하십니까. 오늘(27일) 북한 무인기가 또 다시, 넘어온 것 같다는 소식에 급하게 전투기 여러대...



프레시안

북한 무인기, 용산 대통령실 포함 비행금지구역 상공도 지나갔다

지난해 말 군사분계선(MDL)을 넘어 남하한 북한의 무인기가 용산 대통령실 상공을 포함한 비행금지구역을 지나간 것으로 확인됐다. 비행금지구역 통과



서울경제

北무인기에 서울까지 뚫렸다...100발 사격에도 격추 못해 - 서울경제

정치 > 대통령실 뉴스: 북한의 대남 군사 도발이 도를 넘어서고 있다. 지난 수개월 사이 북방한계선(NLL) 이남의 공해상 등으로 탄도미사일이나 포탄...



뉴스시스

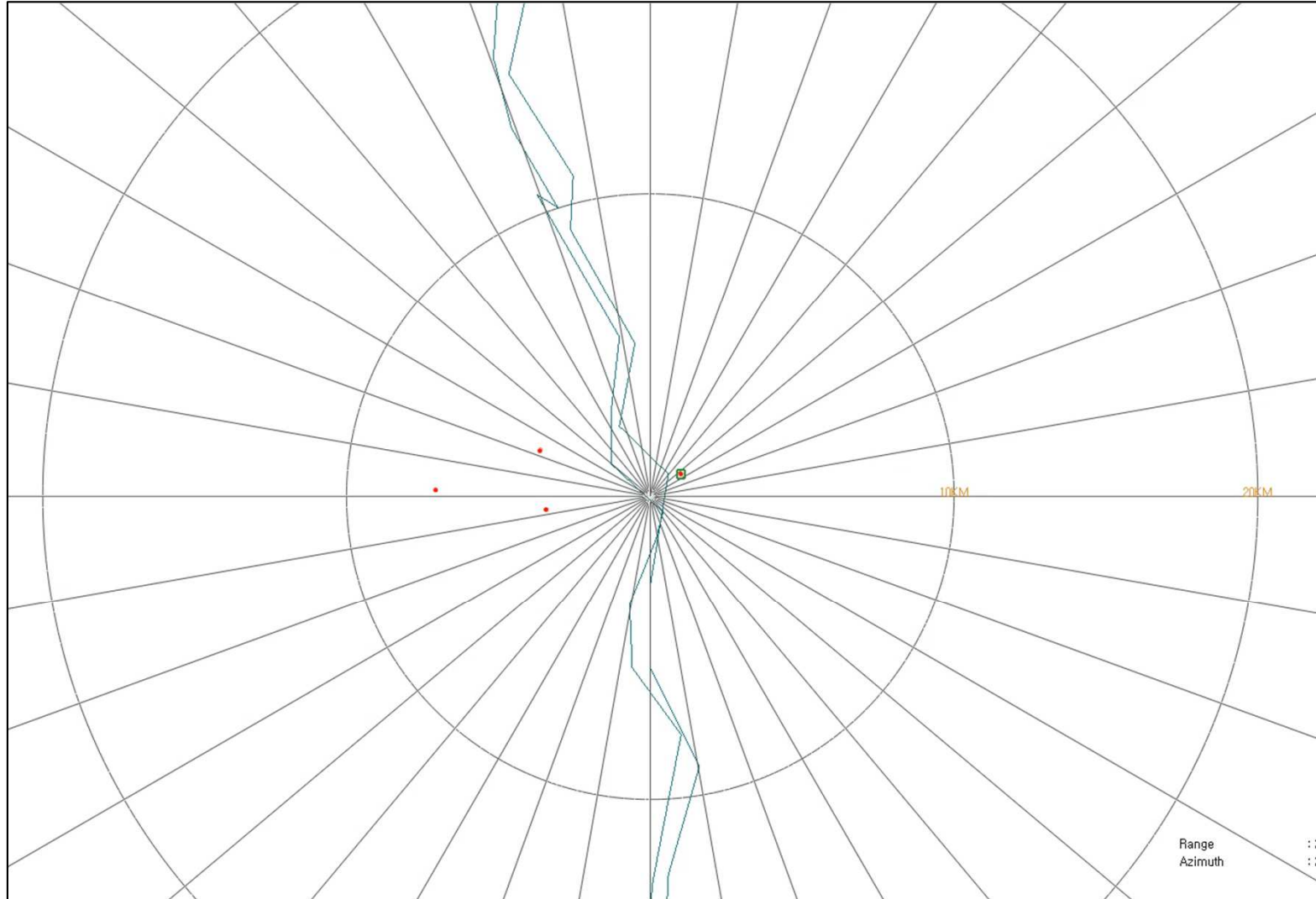
북한 무인기 놓친 軍, '새떼'에 대응전력 출격

[서울=뉴스시스] 하종민 기자 = 전날 북한의 무인기가 남하한 것에 놀란 군이 새떼를 무인으로 오인해 대응전력이 출격하는 해프닝이 발생했다.



1 연구목적

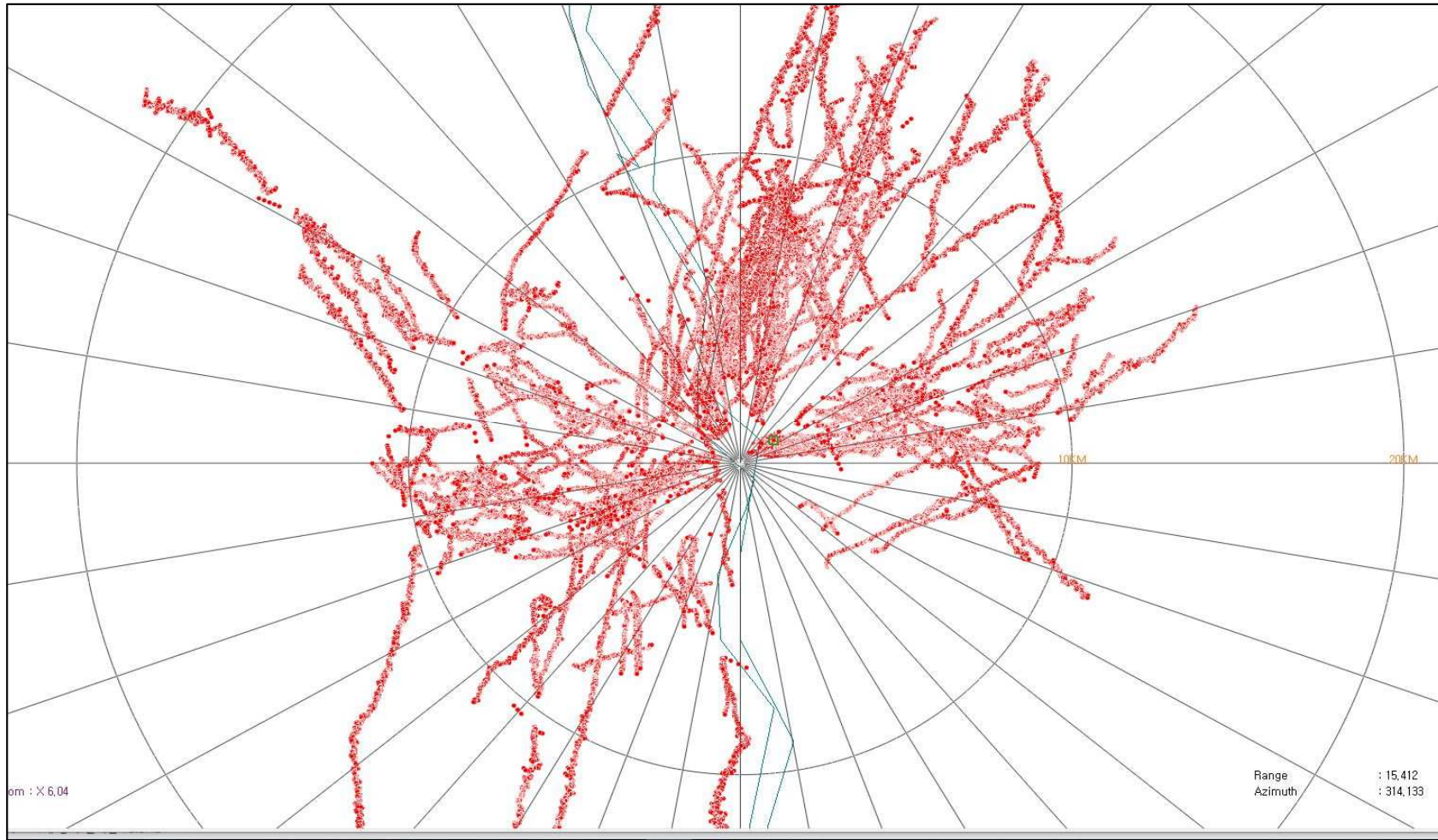
- 레이더 탐지/추적/식별 결과가 무인기인지 조류의 비행인지 판단하기 어려움



식별이 어려운 추적결과

1 연구목적

- 레이더 탐지/추적/식별 결과가 무인기인지 조류의 비행인지 판단하기 어렵기 때문에 인공지능을 활용하여 표적 식별 성능을 향상시키고자 함.



식별이 어려운 추적결과

2 학습 데이터 특성

- 획득한 추적 결과에 따라 식별 가능한 항공기 / 무인기 / 새 표적으로 라벨링(전처리)
 - 식별 가능한 항공기 : 피아식별코드가 부여된 표적
 - 무인기 : 드론 및 초소형 비행체로 군/민간 업체에서 입수한 비행시나리오로 대조하여 식별한 표적
 - 새 : 운용자 / 기타 센서 탐지 결과에 의해 식별한 표적

데이터 필드	설명	단위
SlantR(r)	레이다와 표적 간의 Slant Range	m
GroundR(r)	레이다와 표적 간의 Ground Range	m
Azimuth(d)	레이다 기준 표적의 방위각	degree
Elevation(d)	레이다 기준 표적의 고각	degree
Height(h)	레이다 기준 표적의 고도	feet
Velocity(v)	표적의 속도	m/s
Heading(d)	표적의 이동 방향	degree
PlotPower(dB)	Track 구성 시 연관되는 최대 5개의 Plot 중 사용할 Plot 선정에 활용된 신호의 세기	dB
Hit	Track 구성시 활용되는 Subplot의 Hit 수	개

기본 구성 필드

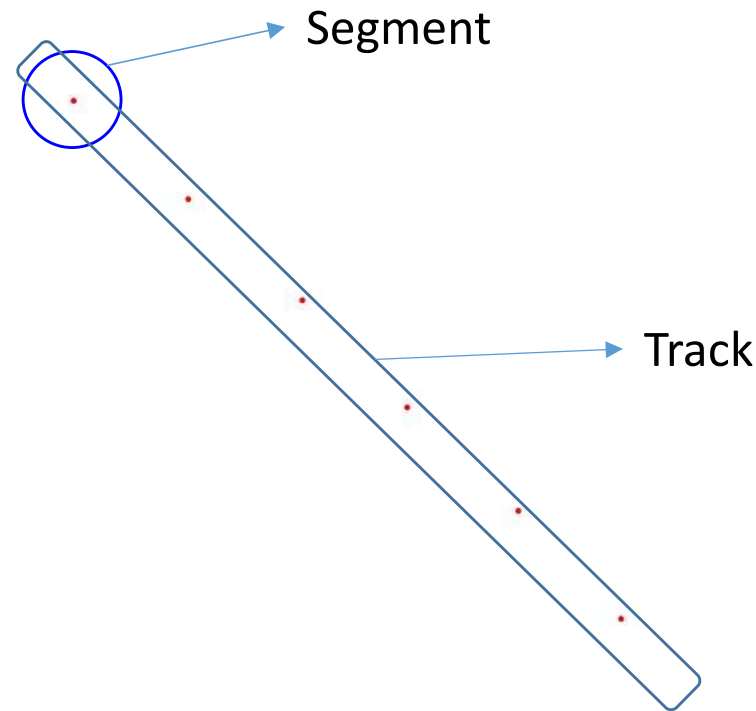
추가 구성 필드 : 필드가 추가 됨에 따라 학습 결과에 영향을 미치는 지 판단하기 위해 구성함

2 학습 데이터 특성

- Track / Segment

- Track : 일정 추적 요건을 만족하여 레이다 운용 콘솔에 전시되는 표적의 항적(추적 개시 - 소실)

- Segment : Track 데이터를 구성하고 있는 하나의 포인트



2 학습 데이터 특성

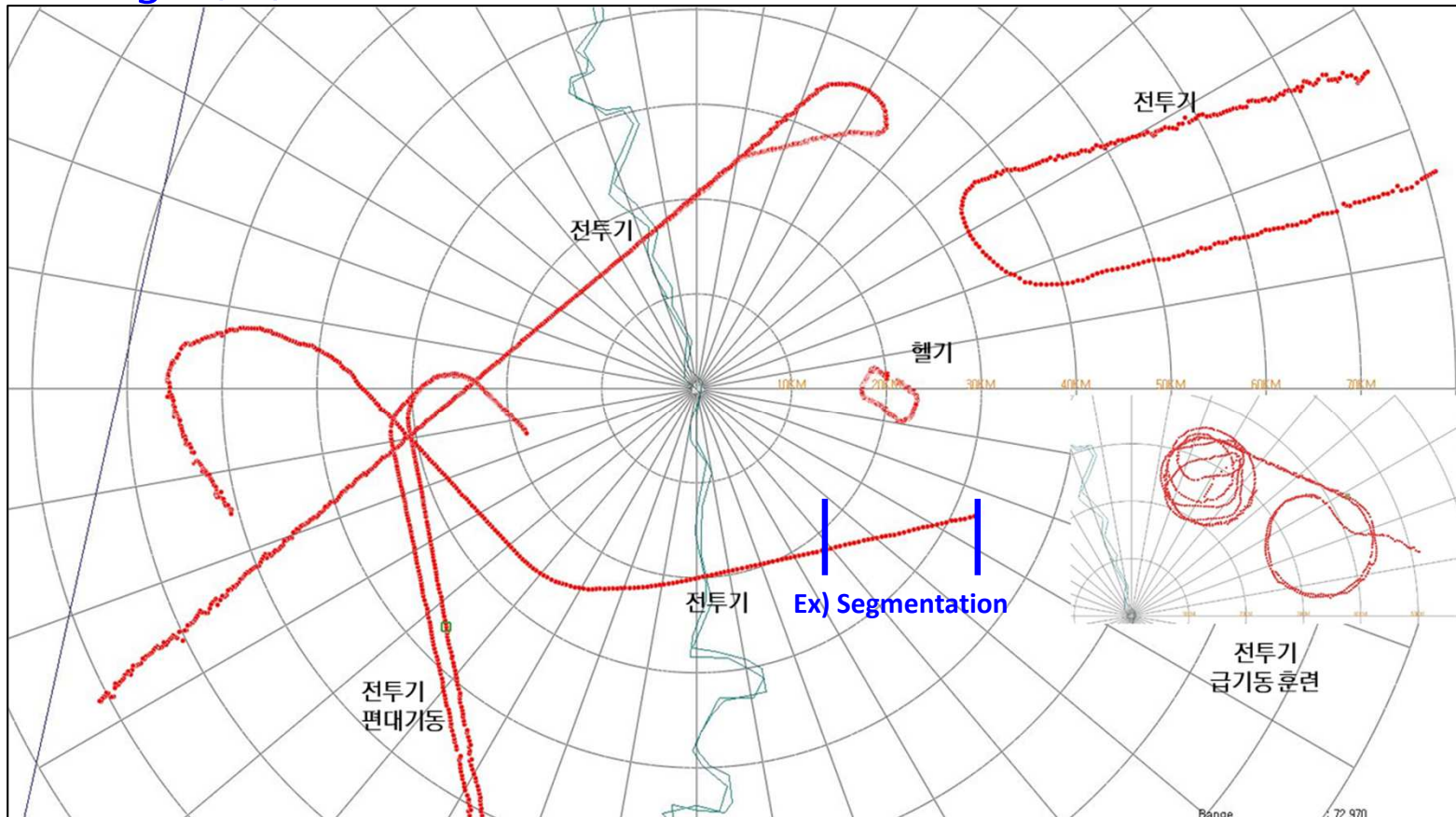
- 학습 데이터 제한 사항
 - 무인기 데이터의 개수가 현저히 부족
 - 각 Track 데이터의 유지도와 측정 시간이 각각 다름 -> 일정한 간격 / 길이로 정규화

필요성 존재

표적	Class Label	Track 개수	Segment 개수
새	0	1733	407,693
무인기	1	8	3,369
식별 가능 항공기	2	1085	201,615

3 전처리

- 라벨링이 완료된 데이터를 분할함(Segmentation)
 - 각 Track 마다 길이가 상이하여 Track을 일정 길이로 잘라 입력 데이터로 사용
 - Segment Length(SL) : 분할할 데이터의 길이를 결정함



Track의 길이가 상이한 데이터

3 전처리

- 세그먼트 길이에 따른 데이터 비율
 - 세그먼트 길이에 따라 모델 훈련에 활용 되는 입력 데이터의 개수가 달라짐
 - 세그먼트 길이에 따라 훈련 결과가 달라질 수 있다는 것을 의미함
 - 본 시험에서는 훈련 80%, 검증 10%, 테스트 10% 비율로 구성함

[기본 필드 데이터]

표적	총 개수 (SL:30)	훈련	검증	테스트
새	14712	11788	1462	1462
무인기	110	90	10	10
식별 가능 항공기	7439	5931	754	754

표적	총 개수 (SL:50)	훈련	검증	테스트
새	9230	7411	913	906
무인기	67	51	8	8
식별 가능 항공기	4612	3586	469	477

표적	총 개수 (SL:100)	훈련	검증	테스트
새	5146	4129	508	509
무인기	32	21	8	3
식별 가능 항공기	2467	1866	249	253

[필드를 추가한 데이터]

표적	총 개수 (SL:30)	훈련	검증	테스트
새	12845	10247	1313	1285
무인기	113	82	13	18
식별 가능 항공기	6822	5495	652	675

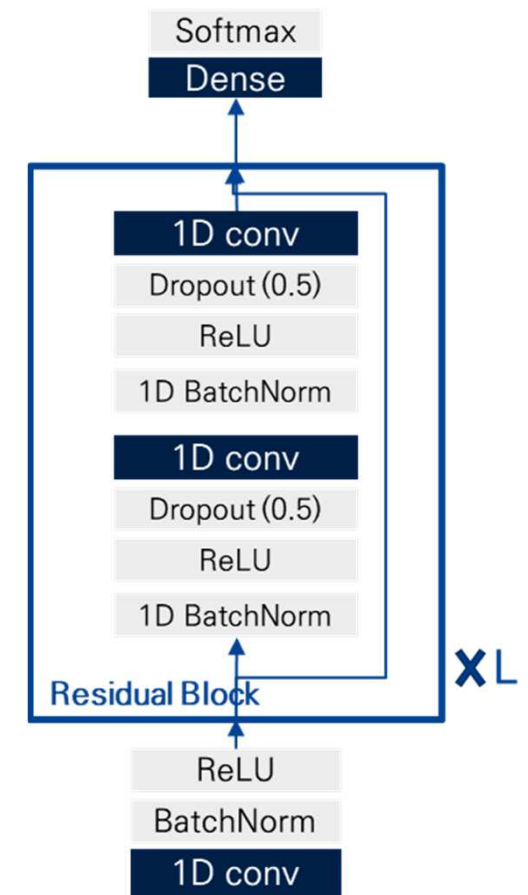
표적	총 개수 (SL:50)	훈련	검증	테스트
새	8057	6479	776	802
무인기	67	50	8	9
식별 가능 항공기	4222	3348	450	424

표적	총 개수 (SL:100)	훈련	검증	테스트
새	4475	3582	437	456
무인기	33	23	6	4
식별 가능 항공기	2271	2819	234	218

SL에 따른 훈련 / 검증 / 테스트 데이터의 변화

4 딥러닝 모델 구조

- Convolution Layer를 기반으로 한 1D ResNet 모델
 - 각 필드를 채널 정보로 간주하여 시간도메인의 다채널 1D 신호로 취급 가능
 - 1D Convolution Layer를 이용하여 필드 정보를 학습
 - Residual Block(L) 개수는 4로 설정
 - > 블록마다 Residual을 더하여 Vanishing Gradient를 방지
- 훈련 파라미터
 - Dropout ratio : 0.5
 - # of residual block : 4
 - # of 1D convolution filter : 64
 - Batch Size : 128 / Epoch : 2000
 - Optimizer : Adam



1D ResNet 모델 구조

4 딥러닝 모델 구조

- 불균형 데이터 학습 방법

표적	Class Label	Track 개수	Segment 개수
새	0	1733	407,693
무인기	1	8	3,369
식별 가능 항공기	2	1085	201,615

- 무인기 데이터가 다른 Class에 비해 현저히 적기 때문에 클래스 별 손실 함수에 가중치를 설정함

- 손실함수는 멀티 클래스 손실함수를 사용함

$$CCE = - \sum_{i=1}^C t_i p_i \log(f(s)_i) \quad \text{where } f(s)_i = \frac{e^{s_i}}{\sum_{j=1}^C e^{s_j}}$$

CCE 손실함수

$$t_i = \frac{N_{trunmanned}}{N_{tri}}$$

$N_{trunmanned}$: # of training data of unmanned plane
 N_{tri} : # of training data of class i where $i = \{1, 2, \dots, C\}$

클래스 별 손실함수의 가중치

5 학습 결과

- 학습 결과를 Metric / Confusion Matrix / t-SNE 관점에서 분석함

5 학습 결과 - Metric

- 손실 함수(CCE)에 의한 훈련 성능 평가
 - 클래스별 데이터 불균형으로 인해 새 데이터로 오분류하는 확률이 높은 경우 모델의 성능이 높은 것으로 평가됨
 - 따라서 클래스 별 가중치를 부여하지 않았을 때의 모델을 높은 성능으로 평가
 - 모델 성능 평가에 적절한 기준이 아님

SL	CCE 클래스별 가중치	초기 데이터		필드 추가 데이터	
		검증	테스트	검증	테스트
30	-	0.5741	0.5725	0.5808	0.5878
	O	0.6214	0.6165	0.6352	0.6348
50	-	0.5723	<u>0.5821</u>	<u>0.5890</u>	<u>0.5886</u>
	O	0.6885	0.7025	0.6236	0.6372
100	-	0.5914	0.5936	0.5945	0.6132
	O	0.6935	0.7202	0.6589	0.6877

CCE 훈련 성능 평가 Metric

5 학습 결과 - Metric

- Weighted-F1 점수에 의한 훈련 성능 평가

- Precision, Recall의 조화 평균
- 분류 모델의 성능 평가 기준으로
- 클래스별 데이터 개수가 고려된 F1 점수 활용
- 여전히 데이터 개수가 많은 새 데이터로 오분류가 높은 경우 모델 성능을 높게 평가
- 적절한 모델 성능 평가 기준이 될 수 없음

$$F_{1_weighted} = \sum_{i=1}^C \alpha_i F_{1_i} \quad \text{where } \alpha_i = \frac{N_{class_i}}{N_{total}}$$

N_{class_i} : Number of Segments in class i
 N_{total} : Total number of segments

SL	CCE 클래스별 가중치	초기 데이터		필드 추가 데이터	
		검증	테스트	검증	테스트
30	-	0.9754	0.9763	0.9675	0.9582
	O	0.9475	0.9571	0.9329	0.9319
50	-	0.9753	0.9664	0.9596	0.9593
	O	0.9047	0.8905	0.9385	0.9316
100	-	0.9545	0.9551	0.9545	0.9358
	O	0.8889	0.8729	0.9144	0.8949

Weighted-F1 훈련 성능 평가 Metric

5 학습 결과 - Metric

- Macro-F1 점수에 의한 훈련 성능 평가
 - 각 Class 별로 F1 점수를 계산한 후, 데이터의 개수와 무관하게 hard-average 한 값
 - 불균형 데이터 클래스를 모두 동일한 가중치로써 모델의 성능을 평가하고 싶은 경우 적합
 - F1 점수 자체는 낮아지는 듯 하나, 각 모델의 경향성을 잘 반영함
 - 필드 추가 데이터를 사용할 때 성능이 향상 됨
 - SL이 30일 때 성능 향상

SL	클래스별 가중치	초기 데이터		추가 데이터	
		검증	테스트	검증	테스트
30	-	0.6522	0.6529	0.6480	0.6434
	O	0.6865	0.6795	<u>0.6916</u>	0.7033
50	-	0.6540	0.6468	0.6429	0.6428
	O	0.6452	0.6343	0.7199	<u>0.6976</u>
100	-	0.6414	0.6364	0.6406	0.6234
	O	0.6616	0.6125	0.6877	0.6445

Macro-F1 훈련 성능 평가 Metric

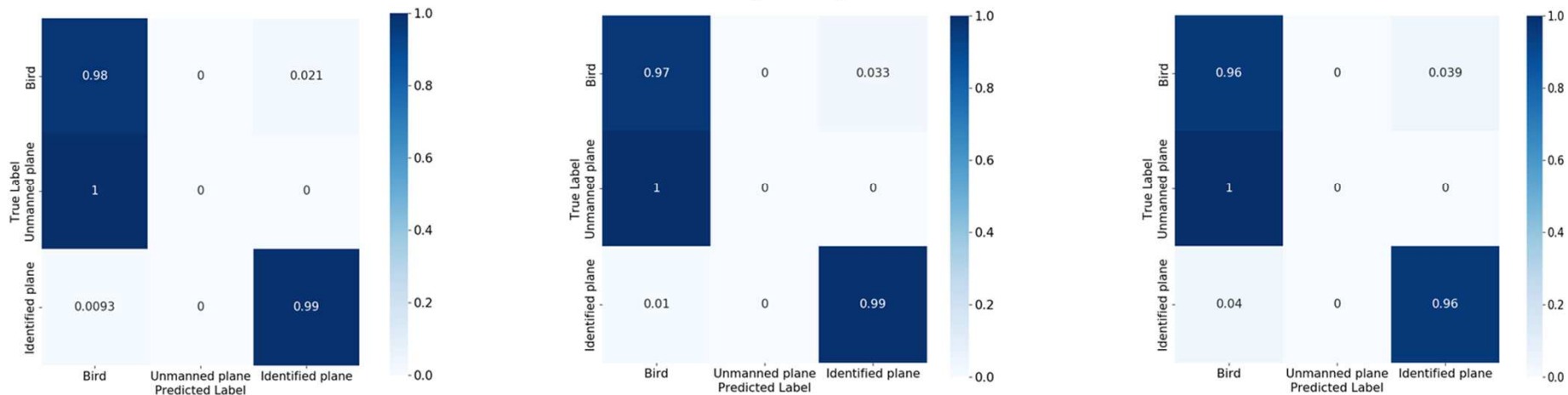
5 학습 결과 - Confusion Matrix

- 초기 데이터 (클래스별 가중치 X)
 - 모든 클래스를 동일한 가중치로 학습할 경우, 데이터의 개수가 많은 클래스의 정확도를 높이는데

치중

- 모든 무인기 데이터를 새 데이터로 분류
 - 무인기 데이터가 식별가능 항공기보다 새 데이터와 유사한 특징을 가지고 있음을

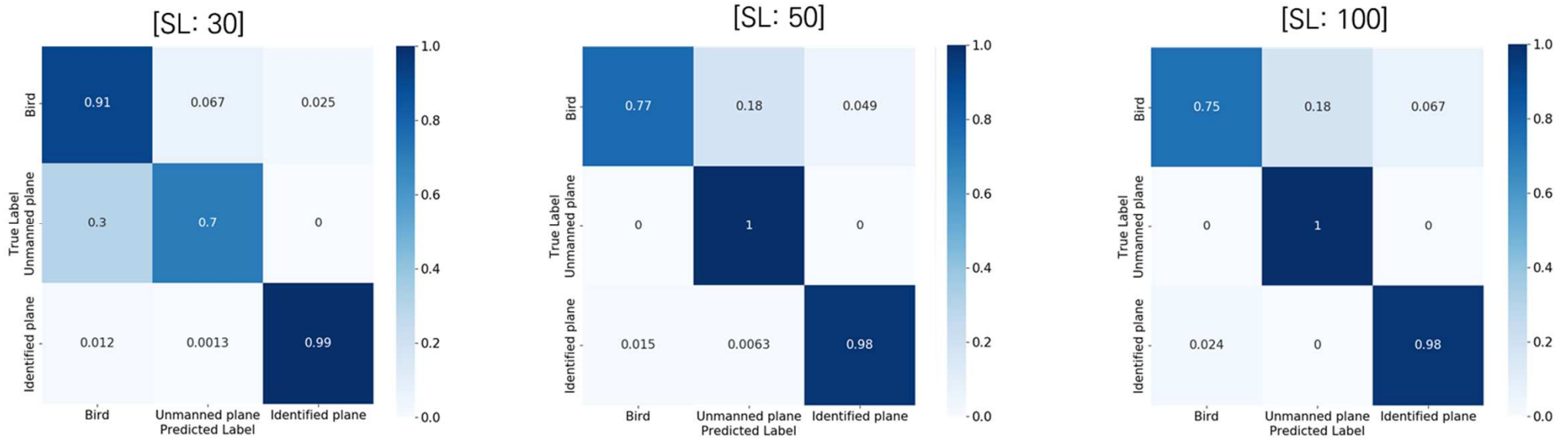
확인



SL별 Confusion Matrix

5 학습 결과 - Confusion Matrix

- 초기 데이터 (클래스별 가중치 0)
 - 무인기 데이터의 가중치를 크게 설정하여 학습할 경우, 새 데이터의 오분류 증가
 - 식별 가능 항공기의 성능은 유지됨.



SL별 Confusion Matrix

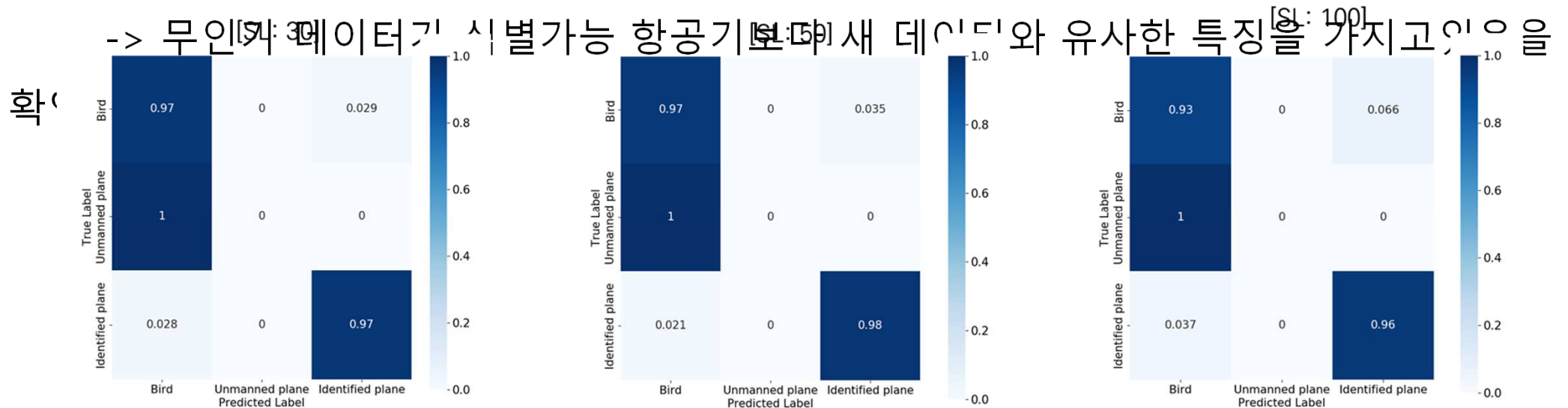
5 학습 결과 - Confusion Matrix

- 필드 추가 데이터 (클래스별 가중치 X)
 - 모든 클래스를 동일한 가중치로 학습할 경우, 데이터의 개수가 많은 클래스의 정확도를 높이는데

치중

- 모든 무인기 데이터를 새 데이터로 분류

-> 무인기 데이터가 식별가능 항공기보다 새 데이터와 유사한 특징을 가지고 있을

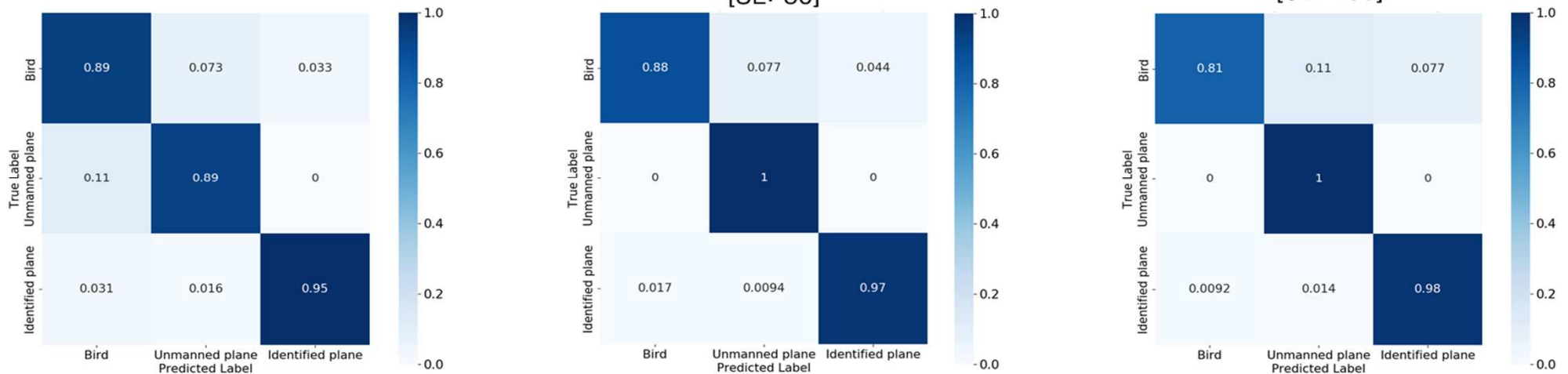


SL별 Confusion Matrix

5 학습 결과 - Confusion Matrix

- 필드 추가 데이터 (클래스별 가중치 0)
 - [초기 데이터 + 클래스 가중치 0]의 경우에 비해 새 데이터의 성능 감소 완화
 - 추가된 필드 정보(Plot 신호세기 및 Hit 개수)가 새와 무인기 분류에 유용
 - Macro-F1 성능 기준으로 SL 30에서 성능이 가장 좋은 듯 하나, Confusion matrix를 기준으로

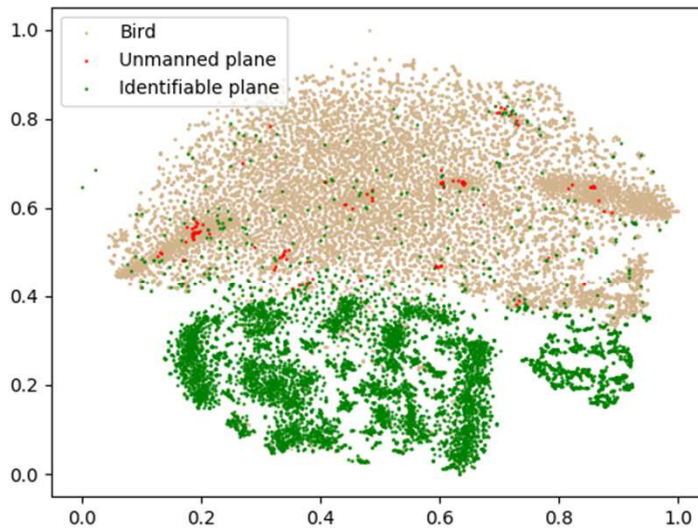
SL이 50일 때가 30일 때 보다 무인기 데이터의 분류 성능이 높음



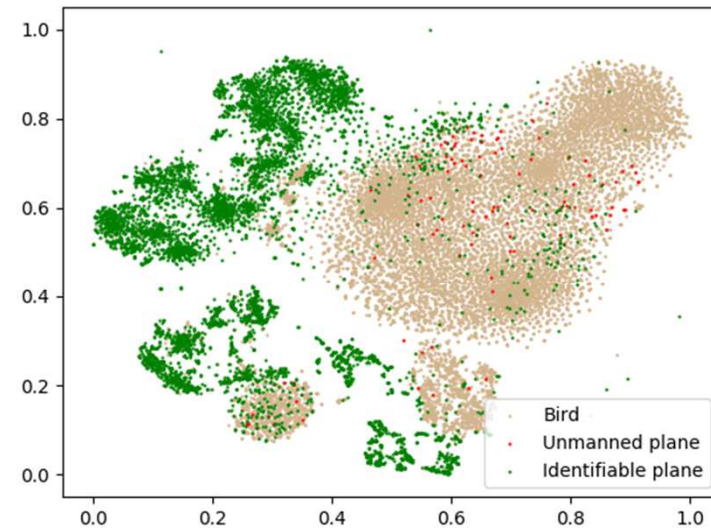
SL별 Confusion Matrix

5 학습 결과 - t-SNE

- 전처리 데이터(라벨링 및 Segmentation)
 - 모델을 거치지 않은 전처리 데이터의 t-SNE Plot 도시화
 - 데이터 클래스간 경계가 불분명
 - 3개의 클래스가 혼재되어 있음
 - 식별가능 항공기와 새 데이터 간의 경계 역시 모호함



[초기 데이터의 전처리 후 t-SNE]

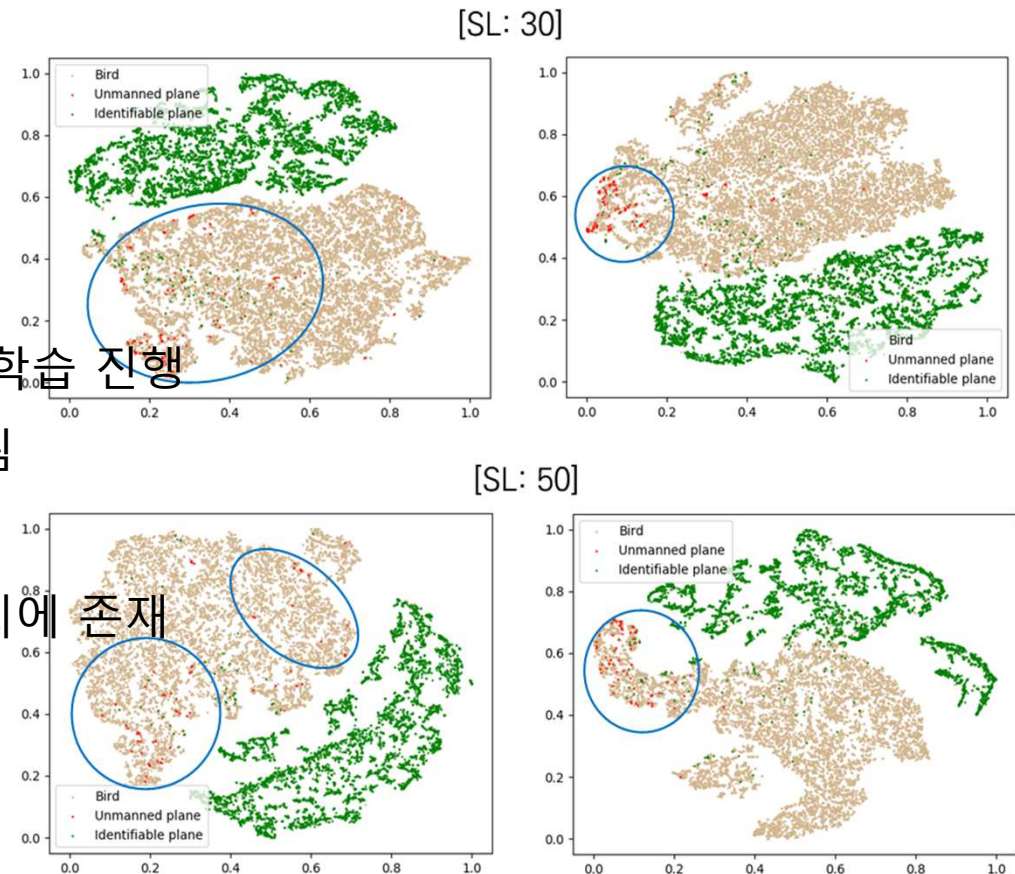


[추가 데이터의 전처리 후 t-SNE]

5 학습 결과 - t-SNE

- 초기 데이터

- 훈련된 모델의 마지막 Layer의 Feature 이용
- 새와 식별가능 항공기 사이가 구분이 가능함
- (좌) 클래스별 가중치 X
 - > 데이터 개수가 많은 클래스에 초점을 맞춰 학습 진행
 - > 무인기 데이터는 새 데이터의 군집에 포함됨
- (우) 클래스별 가중치 O
 - > 무인기 데이터는 새 데이터 분포의 가장자리에 존재



5 학습 결과 - t-SNE

- 초기 데이터

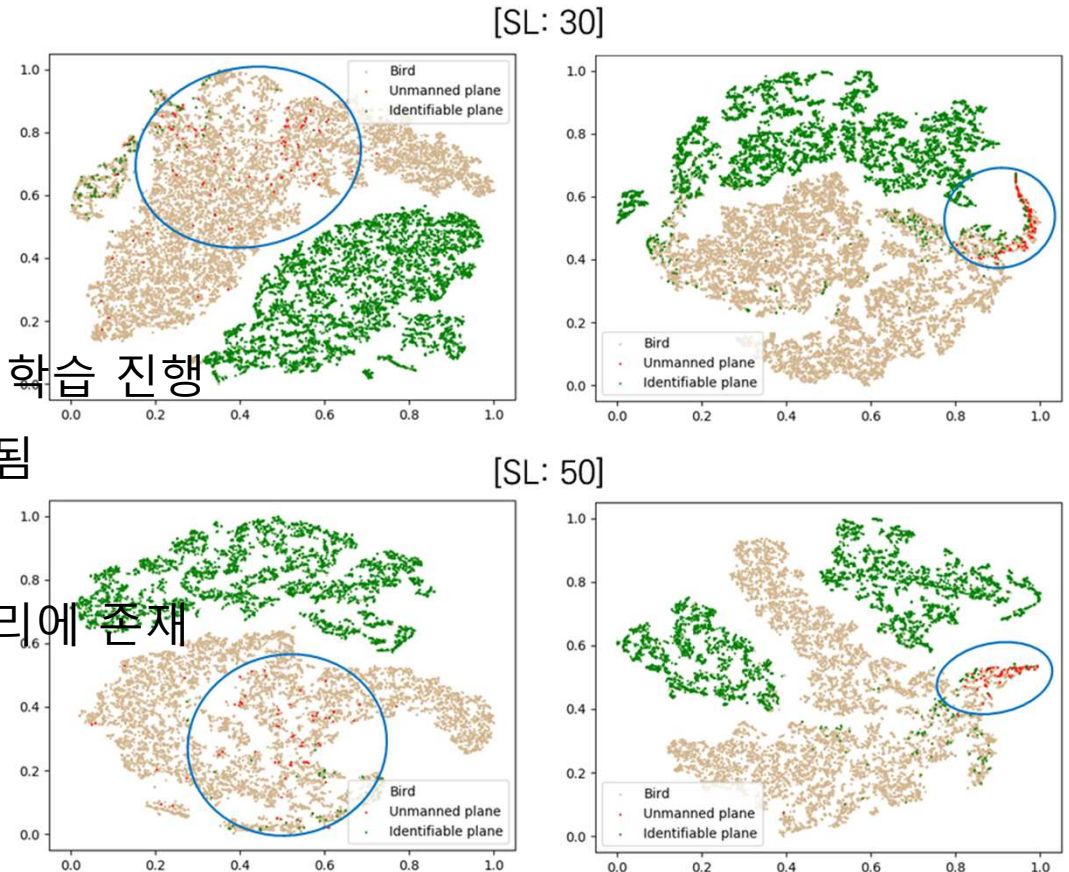
- 훈련된 모델의 마지막 Layer의 Feature 이용
- 새와 식별가능 항공기 사이가 구분이 가능함

- (좌) 클래스별 가중치 X

- > 데이터 개수가 많은 클래스에 초점을 맞춰 학습 진행
- > 무인기 데이터는 새 데이터의 군집에 포함됨

- (우) 클래스별 가중치 O

- > 무인기 데이터는 새 데이터 분포의 가장자리에 존재



연구 요약

- 데이터 처리
 - 레이다 Track 필드 정보를 다채널 시계열 데이터로 간주
 - Track 정보의 특성을 반영한 전처리 및 각 필드의 도메인에 따른 정규화 진행
- 모델 학습
 - 1D ResNet 구조 이용 → 세가지 클래스로 분류하도록 학습
 - 클래스간 불균형 고려 → 훈련 데이터의 가중치에 따라 손실 함수의 가중치 설정
- 모델 검토
 - 불균형 데이터로 학습된 모델의 성능 평가에 적합한 metric 검토 (Macro-F1)
 - Confusion matrix, t-SNE를 활용한 모델 성능 검토

결론

- 데이터 불균형 가중치 설정 하지 않은 경우
 - 새와 식별가능항공기는 각각 98%, 99%의 확률로 분류 성능 확보.
- 데이터 불균형 극복 위해 class 별 손실함수의 가중치를 다르게 할 경우
 - 무인기 분류 성능은 향상하지만 새의 분류 성능은 감소.
 - 추가된 필드 정보 (track선정을 위한 신호 세기 및 Hit 개수)가 새와 무인기 분류에 유용함 확인.
 - 세그먼트 길이에 따른 성능 차이 존재. (SL:50 에서 새 데이터 분류 성능 감소 저하)

검토 사항

- 무인기 데이터의 불균형으로 인한 문제
 - 현재 무인기 데이터가 새와 식별가능 항공기에 비해 현저히 적기 때문에, 최종 목표를 위해 무인기 데이터를 확보한 후 현재 필드 정보의 유효성 검토가 요구됨.
 - 추가 무인기 데이터 (초기 필드) 확보 → 무인기 데이터 개수 증가 시 성능 비교 후 검토 계획.
- 사용되는 필드 정보 (Feature)의 유효성 검토
 - 추가 Track 정보 또는 Track 정보 추출 이전 단계의 정보 활용 방법에 대한 검토 필요.