글로벌-로컬 게이트 특징을 활용한 인코더 기반 드론 키포인트 추출 연구

황서빈[1], 조영준[2]
[1]전남대학교 인공지능융합과 석박통합과정
[2]전남대학교 인공지능융합과 부교수
cnu.cvl.hsb@gmail.com, yj.cho@jnu.ac.kr

A Study on Encoder-based Drone Key-point Extraction with Gated Global-Local Features

Seo-Bin Hwang[1], Yeong-Jun Cho[2]
[1,2]Dept. of AI Convergence, Chonnam National University

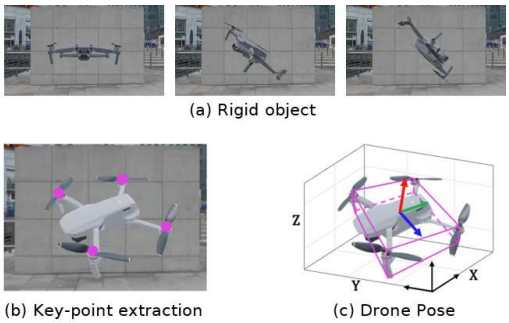
요 약

최근 불법적인 정보 탈취을 목적으로 드론의 사용이 증가하고 있으며, 이를 대응하기 위해 다양한 기술적 해결책이 개발되고 있다. 특히, 드론의 정확한 자세 추정이 필요하지만, 드론의 자유로운 움직임과 회전 때문에 기술적 구현이 어렵다. 따라서 본 논문은 트랜스포머의 인코더 구조를 사용한 키포인트 추출 방법을 제안해 기존 문제를 해결한다. 이 방법은 로컬과 글로벌 특징을 결합하는 게이트 메커니즘을 도입하여, 다양한 각도에서도 일관된 키포인트 추정을 가능하게 하고, 드론의 움직임과 상태를 보다 정확하게 분석할 수 있다. 논문은 드론 2차원 키포인트 데이터셋을 제공하고, End-to-end 방법론을 제안하여 드론의 키포인트를 정확히 추정하며, 이를 통해 불법 드론 활동을 실시간으로 탐지하고 대응할 수 있는 기술적 기반을 마련한다.

1. 서론

최근 불법적인 정보 탈취를 목적으로 운영되는 드론(drone)의 사용이 증가하고 있다. 이에 대응하기 위해 국가와 민간 부문에서는 다양한 기술적 해결책을 개발 중이다[1]. 예를 들어, 실시간 모니터링 시스템을 통해 불법 드론의 비행 경로를 추적하거나, 의도적인 경로 변경 및 비정상적인 행동을 감지하는 기술이 적용되고 있다. 이러한 문제를 효과적으로 해결하기 위해서는 드론의 자세를 정확하게 추정하는 것이 필수적이다. 그러나 드론은 상공에서 자유롭게 움직이며, 각 축을 중심으로 회전이 가능하기 때문에 이 과정은 기술적으로 도전적인 과제다.

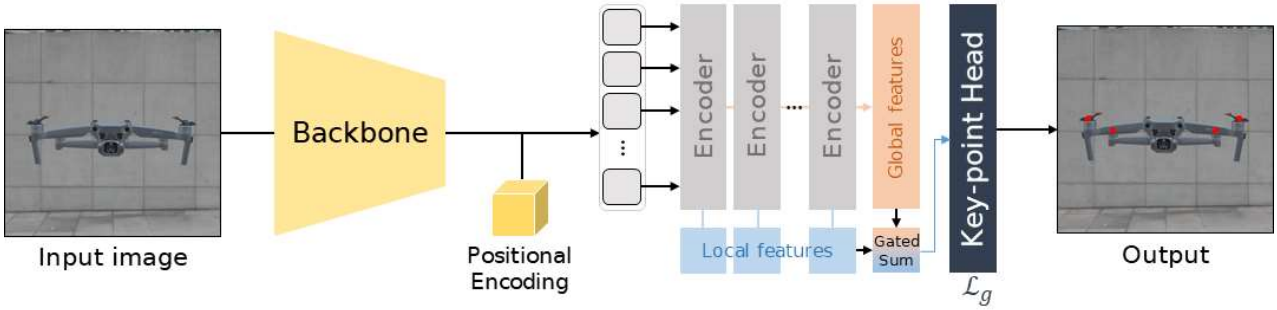
드론의 자세를 바로 예측하기보다는, 먼저 움직임이 적은 사전 정보(e.g., 키포인트(key-point))를 추정하는 것이 더 효율적일 수 있다. 그림 1에서처럼 드론은 프로펠러와 같은 경직(rigid)된 구조를 가지고 있으며, 이러한 고정된 구조는 신뢰할 수 있는 키포인트로 사용할 수 있다. 모든 드론은 유사한 모양을 가지고 있어 프로펠러와 같은 고유한 특징을 바탕으로 키포인트를 추정할 수 있지만, 정적 형태로 인해 다양한 각도에서 촬영된 이미지가 유사하게



(그림 1) (a)는 경직된 객체의 특징을 가진 드론을 보여준다. (b)는 드론의 프로펠러를 키포인트로 설정해 추출한 결과이며, (c)는 드론의 3차원상에서 자세를 정밀히 추정해 재현한 결과물이다.

보이는 문제가 발생할 수 있다. 이로 인해 키포인트를 정확하고 일관되게 추출하는 것은 여전히 어려운 과제다.

이러한 문제를 해결하기 위해, 본 논문은 트랜스포머(Transformer)[2]의 인코더(Encoder) 구조를 활용한 키포인트 추출 방법을 제안한다. 트랜스포머는 이미지 내에서 전역적인 관계를 모델링하는 데 뛰어난 성능을 보이며, 특히 고정된 구조를 가진 드론의 키포인트를 추출하는 데 적합하다. 본 연구에서는



(그림 2) 제안 방법의 아키텍처.

로컬(local) 및 글로벌(global) 특징을 결합하는 게이트 합(Gated sum) 메커니즘을 도입하여 다양한 각도에서 일관된 키포인트 추정을 가능하게 하며, 이를 통해 드론의 움직임과 상태를 보다 정확하게 분석할 수 있는 모델을 구축했다.

본 논문의 구성은 다음과 같이 구성된다. 먼저 제2장에서는 객체 위치 추정과 관련된 기존 연구들을 검토한다. 이어지는 제3장에서는 본 논문의 핵심이 되는 방법론을 단계별로 상세히 기술한다. 제4장에서는 검증에 활용된 생성 데이터셋과 실험결과에 대해 설명한다. 마지막 장에서는 차후 연구 방향을 제시하며 마무리한다.

본 연구는 드론 키포인트 추출 연구 분야에 여러 중요한 기여를 하고 있다. 첫째, 우리는 드론 2차원 키포인트 데이터셋 제공한다. 이는 새로운 3차원 합성 데이터로 구성된다. 제공 데이터의 경우, 실제 360도 카메라로 촬영한 배경에 드론을 정교하게 합성하여 실세계(Real-world) 데이터와 매우 유사한 특성을 지닌다. 이러한 고품질 데이터셋은 드론 관련 연 검증에 큰 도움이 될 것이다. 둘째, 우리는 End-to-end 방법론을 제안한다.

이러한 방법론은 불법적인 드론 활동을 감지하고 대응하는 시스템의 기술적 기반을 마련할 수 있다. 키포인트 추출을 통해 타 방법론과 결합하여 드론의 상태와 비행 패턴을 예측하고, 비정상적인 행동을 실시간으로 탐지하여 빠르게 대응하는 것이 가능해질 것이다.

2. 배경 연구

사람의 자세와 관련된 연구[3]에서는 히트맵 회귀(heatmap regression) 기반 방법과 좌표 회귀(coordinate regression) 방법을 비교했다. 히트맵 회귀는 정답과 유사한 히트맵을 생성함으로써 고정밀도 키포인트 예측을 가능하게 하지만, 이를 생성하는 과정은 매우 복잡하고 계산 자원이 많이 필요하다. 반면, 좌표 회귀 기반 방법은 키포인트의 좌표를 직접 예측하는 방식으로, 모델 구조가 단순해지고 계산 효율성도 높다. 그러나 키포인트의 공간적 관계를 충분히 반영하지 못해 예측의 정확도가 낮은 문제가 있다.

트랜스포머를 활용한 타 방법론[4]에서도 U-Net[5]과 같은 구조를 차용하여 히트맵을 생성하고 비교하는 방식을 사용하고 있다. 이러한 방법에서는 일반적으로 인코더에서 직접 히트맵을 생성하지 않고, 별도의 네트워크를 통해 히트맵을 생성하고 이를 활용하여 키포인트 예측을 수행한다. 이와 같은 접근법은 히트맵의 공간적 정보를 활용하여 키포인트의 정확한 위치를 추정하는 데 도움을 줄 수 있지만, 생성 과정에서의 복잡성은 여전히 존재한다. 앞서 언급된 문제들을 해결하기 위해, 본 논문은 트랜스포머의 인코더 구조를 활용하여 로컬과 글로벌 특징을 결합한 게이트 메커니즘으로 드론의 키포인트를 정밀하게 추출하는 방법을 제안한다.

3. 제안 방법

본 연구에서는 드론 이미지에서 프로펠러와 같은 키포인트를 정확히 추정하는 모델을 제안한다(그림). 제안된 모델은 다음과 같은 과정으로 구성된다. 입력된 드론 이미지는 먼저 ResNet[6] 기반의 Convolutional Neural Networks (CNN)[7] 백본을 통해 처리된다. 이 과정에서 이미지는 패치별로 나뉘며, 각 패치는 CNN을 통해 특징을 추출하게 된다. ResNet에서 추출된 특징들은 위치 정보를 포함한 Positional Encoding을 추가한 후 인코더에 입력된다. 인코더는 여러 개의 헤드별 셀프 어텐션(multi-head self-attention)으로 구성되며, 각 레이어는 키포인트의 로컬 특성 정보를 학습한다. 인코더는 1층부터 N층까지 존재하며, 각 레이어는 로컬 특징을 학습하고, 이들 특징들은 모든 인코더 출력의 평균값인 글로벌 특성과 함께 게이팅된 합(sum) 연산을 통해 통합된다. 이 과정에서 중요한 키포인트 정보가 강화되며, 최종적으로 키포인트 특성이

생성된다. 통합된 정보는 최종적으로 키포인트 헤드로 전달되어, 드론의 키포인트를 예측한다. 모델의 수렴을 개선하고 예측 정확도를 높이기 위해 가우시안 손실(Gaussian Loss) 함수를 사용하였다.

가. 인코더

드론 이미지에서 프로펠러와 같은 키포인트 영역을 추정하기 위해 인코더 구조를 사용한다. 인코더는 이미지를 처리하여 드론의 키포인트를 추출하며, 고차원 특징 맵에서 키포인트 정보를 얻는다.

본 연구에서는 ResNet-50을 백본으로 사용하였으며, 이 네트워크는 이미지로부터 C차원의 고차원 특징 맵을 생성한다. 특징 맵의 크기는 입력 이미지 크기의 1/32로 축소되며, 입력 이미지의 높이 I_h와 너비 I_w가 각각 32로 나누어진 크기의 특징 맵을 얻을 수 있다. 이는 드론 이미지 내에서 특징적인 요소를 추출하는 데 충분한 해상도를 제공한다.

인코더는 입력 이미지의 특징을 추출하고 맥락 정보를 학습한다. 백본에서 추출된 특징들은 위치 정보를 포함하기 위해 포지셔널 인코딩을 거친 후 인코더에 입력된다. 인코더는 여러 개의 헤드 셀프-어텐션 층으로 구성되며, 입력 특징들 간의 복잡한 관계를 학습하는 데 효과적이다. 이를 통해 인코더는 이미지 내의 드론 키포인트와 관련된 중요한 정보를 점진적으로 추출한다. 인코더의 최종 출력은 전체 이미지에 대한 글로벌 특징와 각 인코더 층에서 생성된 로컬 특징들로 구성된다. 이 출력은 다음 단계에서 키포인트 추정을 위한 중요한 기반이 된다.

나. Global-Local 게이팅 합

키포인트가 게이팅된 합(sum) 결과물은 모델이 학습한 고차원 특징 벡터를 사용하여 드론의 프로펠러 위치를 정밀하게 예측하는 역할을 한다. 본 연구에서는 여러 개의 인코더 각 층에서 키포인트 정보를 추출하였다. 여러 추출된 정보와 전체 인코더의 평균 정보에 게이팅된 합 기법으로 결합하였다. 이로써 다양한 관점에서 예측된 키포인트 위치를 하나의 정밀한 예측 값으로 통합하였다. 통합된 정보는 0~1 사이의 값으로 정규화하기 위해 시그모이드(sigmoid)를 적용한다.

다. 가우시안 손실 함수

본 연구에서는 좌표 회귀 기반 방법을 채택하여, 영상 입력 후 키포인트의 좌표를 직접 예측하는 접근법을 사용하였다. 이러한 방법을 선택한 이유는 키포인트의 위치를 직접적으로 추정함으로써 예측의 복잡성을 줄이고 학습을 효율적으로 할 수 있기 때문이다. 또한, 좌표 회귀 방법은 히트맵 방식에 비해 메모리 사용량이 적고 추론 속도가 빠르다는 장점이 있다. 그러나 좌표 회귀 방법의 경우, 키포인트의 실제 크기가 픽셀 단위로 매우 작아 정확한 예측이 어렵다. 이를 보완하기 위해, 우리는 가우시안 분포를 이용하여 특정 지점으로 수렴할 수 있도록 하는 가우시안 손실 함수를 도입하였다.

가우시안 손실은 이러한 가우시안을 기반으로 학습이 이루어진다. 가우시안 손실 함수는 다음과 같다:

$$L_g = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2\pi\sigma^2}\exp\left(-\frac{(x_i-\hat{x_i})^2+(y_i-\hat{y_i})^2}{2\sigma^2}\right)\right), \quad (1)$$

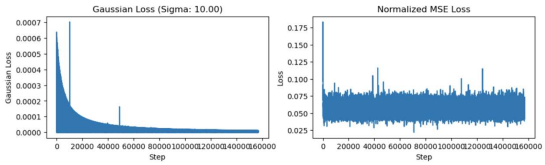
여기서 $\hat{x_i}, \hat{y_i}$는 예측된 키포인트 좌표이고, x_i, y_i는 ground truth (GT) 좌표이다. σ는 가우시안 분포의 표준편차로 GT로부터 분산을 의미한다. 이때 σ값은 0.1로 설정 후에 진행했다. 해당 손실함수로 좌표 회귀 작업에서도 히트맵 회귀처럼 더 빠르고 잘 수렴하며 정확한 학습이 가능해진다.

4. 실험 결과

가. 데이터셋

본 연구에서 사용된 데이터셋은 복잡한 배경과 다양한 드론 객체를 포함하도록 설계되었다. 데이터셋 생성은 Blender라는 3D 프로그램을 통해 이루어졌으며, 실제 360도 카메라로 촬영된 배경에 드론 객체를 삽입하여 렌더링한 결과물이다(그림 3). 사용된 배경은 5종의 복잡한 환경을 포함하며, 드론 모델로는 DJI의 mini2와 air2s가 사용되었다. 시나리오당 1,000개의 프레임을 생성하여, 총 10,000개의 이미지를 확보하였다. 데이터셋은 학습, 검증, 테스트 세트로 나뉘며, 각 세트는 7:2:1로 구성된다. 생성된 데이터셋은 이미지와 드론의 프로펠러에 대응하는 키포인트로 구성되어 있으며, 이 키포인트는 드론의 움직임 및 상태 분석을 위한 중요한 정보를 제공한다.

(그림 3) 생성한 드론 키포인트 데이터셋의 렌더링 결과물



(그림 4) 손실함수 수렴 실험 결과 그래프.
(왼쪽) 가우시안 손실, (오른쪽) 일반적인 MSE 손실.

나. 손실함수 수렴 실험

실험을 위해 먼저 랜덤하게 20,000개의 포인트를 생성하여, 제안된 가우시안 손실 함수와 일반적인 MSE 손실 함수의 수렴 성능을 비교하였다. 이때 σ는 10으로 설정했다. 두 손실 함수의 수렴 과정을 시각화한 결과, 가우시안 손실 함수가 MSE 손실 함수에 비해 더욱 빠르고 안정적으로 수렴하는 것을 확인할 수 있었다.

그림 4에서처럼, 가우시안 손실 함수는 학습 초기부터 급격히 감소하며, 빠른 수렴을 보였다. 반면, MSE 손실 함수는 상대적으로 일정한 값을 유지하며 학습 후반부에도 일정한 수준의 오차를 보였다. 특히, 가우시안 손실 함수는 예측 정보가 GT에 더 잘 수렴하는 경향을 보였으며, 이는 좌표 회귀 기반 키포인트 예측 모델의 성능 향상에 기여한다. 결론적으로, 가우시안 손실 함수는 좌표 예측 작업에서 보다 효과적인 수렴 성능을 보이며, 정확한 키포인트 예측을 가능하게 함을 실험을 통해 확인하였다.

다. 키포인트 추출 결과

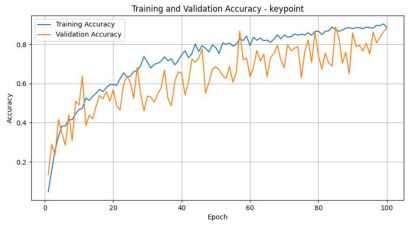
RTX 3060 Ti, Ubuntu 환경에서 제안 모델을 사용해 드론의 각 프로펠러에 대한 키포인트를 성공적으로 추출했다. 학습은 초기 학습률 0.0001을 사용하여 100 에포크(epoch) 동안 진행되었으며, 가우시안 손실 함수를 적용하였다. 모델은 다양한 배경과 드론 유형에 대해 안정적으로 동작했고, 키포인트 추출의 정확도를 높이는 데 기여했다. 그림 5는 실제 추출된 키포인트를 보여준다. 예측된 키포인트는 GT와 비교하여 매우 정확하게 일치하며, 제안된 아키텍처의 우수성을 시각적으로 확인할 수 있다.

또한, 학습 과정에서의 손실 함수 수렴 그래프(그림 6)에서 볼 수 있듯이, 학습의 안정성과 성능 향상을 확인할 수 있다. 제안된 모델은 학습 초기부터 빠르게 수렴하였으며, 예측 정확도를 높이는 데 효과적임을 입증하였다.

5. 결론

본 연구에서는 트랜스포머 인코더 구조를 활용해 드론의 키포인트를 추출하는 방법을 제안하였다. 로

(그림 5) ●는 예측된 키포인트 결과.



(그림 6) 학습과 검증 정확도 그래프.

컬과 글로벌 특징을 결합하는 게이트 메커니즘을 통해 다양한 각도에서도 일관된 예측이 가능했다. 제안된 모델은 드론의 자유로운 움직임에도 높은 정확도의 키포인트 추정을 달성했으며, 이를 통해 불법 드론 활동을 실시간으로 탐지하고 대응할 수 있는 기술적 기반을 제공하였다. 차후 연구에서는 본 방법론을 활용해 드론의 자세까지 추정하는 연구로 발전시키고자 한다.

감사의 글

본 논문은 과학기술정보통신부 및 정보통신기획평가원의 인공지능융합혁신인재양성사업 연구 결과로 수행되었으며(IITP-2023-RS-2023-00256629), 농림축산식품부의 재원으로 농림식품기술기획평가원의 농식품과학기술융합형연구인력양성사업의 지원을 받아 연구되었음(RS-2024-00397026).

참고문헌

[1] 이인재, 불법 드론 대응을 위한 저고도 드론 탐지 기술 동향, (ETRI) 인공지능 서비스 및 인프라 기술, 37권, 1호, 10-20쪽.
[2] VASWANI A, Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
[3] Wang Yanxia, DB-HRNet: Dual Branch High-Resolution Network for Human Pose Estimation, *IEEE Access,* 11, 2023, 120628-120641.
[4] HAMPALI, Keypoint transformer: Solving joint identification in challenging hands and object interactions for accurate 3d pose estimation, CVPR, 2022, p.11090-11100.
[5] Ronneberger, U-net: Convolutional networks for biomedical image segmentation, *MICCAI, Germany,* 2015, p.234-241.
[6] He Kaiming, Deep residual learning for image recognition, CVPR, 2016, p.770-778.
[7] Krizhevsky, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems,* 2012.