

Jetson AGX Orin 기반 nanoOWL + nanoSAM 실시간 그래스핑 시스템: 최신 연구 동향

- 2023년 이후 Jetson AGX Orin 64GB 플랫폼에서는 nanoOWL과 nanoSAM을 결합한 실시간 자연어 기반 객체 감지 및 세그멘테이션이 빠르게 상용화 및 연구적으로 확장되고 있습니다. nanoOWL은 OWL-ViT를 NVIDIA TensorRT로 최적화해 Jetson 시리즈에서 텍스트 임베딩 기반 개방어휘 객체 탐지(Zero-shot Open-Vocabulary Detection)를 실시간으로 수행하며, nanoSAM은 기존 SAM(Segment Anything Model)의 경량화(ResNet18 인코더 기반) 버전으로, Jetson Orin에서도 실시간 인스턴스 세그멘테이션이 가능합니다. 두 모델 모두 Jetson AGX Orin의 Ampere GPU와 TensorRT 최적화 엔진의 강점을 적극 활용하여 95fps(OWL-ViT B/32, 768 입력 크기 기준) 이상의 실제 성능을 달성하며, Python API와 Docker container 기반 데모 및 튜토리얼이 공식적으로 제공되고 있습니다. 개발자는 "find a red cup"과 같은 자연어 명령을 입력하면 nanoOWL이 텍스트 임베딩을 기반으로 해당 객체의 위치를 탐지하고, 그 위치 정보를 nanoSAM의 세그멘테이션 입력으로 활용하여 매우 정밀한 마스크를 실시간으로 획득할 수 있습니다. 이 과정은 Jetson 플랫폼의 강력한 병렬 연산 능력과 최적화된 추론 파이프라인(ONNX-Trt 변환, FP16 또는 FP32 정밀도 활용) 덕분에 엣지 디바이스 수준에서도 고성능을 보장합니다.
- 최신 연구에서는 LGrasp6D, Grasp-Anything++와 같이 비전-언어 통합 파이프라인을 활용한 자연어 지시 기반 6-DoF 그래스핑 및 멀티모달 세그멘테이션 응용이 두드러지고 있습니다. 주요 흐름은 자연어 명령으로 원하는 객체나 속성(예: "좌측 끝의 푸른컵")을 지정하면, 비전-언어 모델이 해당 영역 위치 및 마스크를 출력하고, 이를 실제 로봇 조작으로 연동해 복잡한 실내 환경에서도 사용자의 의도를 즉각적으로 반영하는 지능형 조작이 가능하다는 점에 있습니다. 이러한 연구들의 실제 코드는 공식 GitHub 저장소(nanoOWL, nanoSAM 등)에 공개되어 있어 모듈별로 쉽고 빠르게 적용, 확장할 수 있습니다.

세그멘테이션 마스크와 Depth 정보 결합, 강화학습(RL) 적용 최신 사례

- 세그멘테이션 마스크와 depth 카메라 정보를 결합하고 강화학습(특정 RL 알고리즘에 제한 없음)을 활용한 그래스핑 응용은 로봇 조작 및 복잡 환경 인지 분야의 핵심 방법론으로 부상하고 있습니다. 특히 2024년 발표된 PointPatchRL(PPRL)은 세분화된 마스크와 포인트 클라우드(Depth 기반 3D 정보)를 통합하여, 트랜스포머 기반 인코더로 환경의 기하학적 정보를 정확하게 임베딩하여 RL정책 학습에 제공하는 접근법을 선보였습니다. PPRL은 포인트 클라우드 데이터를 여러 패치로 분할하고(transformer+masking), 이를 마스크 재구성 목표와 함께 학습함으로써 소수의 PointNet/PointNet++ 계열 모델보다 훨씬 뛰어난 성능과 표준 이미지 기반 방법 대비 높은 견고성을 보입니다. 실제로 여러 뷰포인트, 다양한 종류의 객체(특히 밀집·부분 가림 환경, 변형 물체)에서 이미지 기반 DrQ-v2, Dreamer 등 SOTA 강화학습 알고리즘보다 조작 성공률이 크게 높아지는 것이 입증되었습니다.
- PPRL은 Soft Actor Critic(SAC)을 포함해 다양한 강화학습 스킴과도 쉽게 통합되며, Depth를 포함한 실제 3D 포인트 클라우드와 이와 연계된 마스크(세그멘테이션 정보)가 RL 정책 상태 공간으로 투입된다는 점에서 Jetson 환경 기반 실시간 그래스핑 파이프라인과 직접 연동 가능합니다. 본 방법론은 대규모 실내 실험 및 다양한 작업에 general하게 적용되며, 표준 코드와 데모가 오픈되어 있어 실제 시스템 개발의 기초가 됩니다.

객체 형상 특성 고려 최적 그래스핑 포인트 및 접근 방향 결정 알고리즘

- 객체의 3D 형상과 표면 특성을 적극 고려한 최적 그래스핑 지점 및 접근 방향 산정은 2023년 이후 AI로보틱스 분야의 핵심 연구주제로 자리잡았습니다. 최신 방법들은 세그멘테이션 마스크에 기반한 객체 경계 추출과 깊이/포인트 클라우드 정보를 융합하여, 실제 3D 표면의 불균일성 및 shape 완전성을 분석한 후 최적의 grasp point를 도출합니다. 대부분 딥러닝 기반 shape 특성 추정(pre-trained feature extractor,

포인트 클라우드 트랜스포머, 강화학습/최적화)과 전통적 그래스핑 안정성 지수(예: force-closure, grasp quality metric) 병합 방식을 사용합니다. 제약 최적화 문제로 모델링하여, 로봇 엔드 이펙터(그리퍼)의 접근 방향 역시 주변 장애물 맥락, 객체 중심점(centroid) 및 표면 normal, 마스크 연속성 등을 통합 분석하여 계산합니다.

- 또한, 강화학습 정책이 직접 pose candidate set에서 성공률이 높은 grasping point 및 approach direction을 선택하도록 reward 설계를 적용하는 트렌드가 늘고 있습니다. 이때 shape complexity, occlusion, 주변 환경 특성이 고려된 reward function 및 policy 네트워크 설계가 일반적입니다.

복잡 환경(밀집, 가려짐 등)에서 그래스핑 성공률 향상 방식

- 복잡(밀집/부분 가림/비정형) 환경에서 그래스핑 성공률을 높이기 위한 최신 방법론은 세 가지 큰 흐름으로 분류됩니다. 첫째, Mask 기반 Segmentation과 depth를 통합해 객체별 상태와 pose를 추출한 뒤(멀티 타겟 세그멘테이션), multi-target push-grasping 등 협동 조작 전략을 병용하여 효율성을 높입니다. 예를 들어, Wu 외 연구(2023)는 각각의 target 객체에 대해 graspable probability(잡힐 확률)를 정의하고, pushing(밀기)과 grasping(잡기) 정책을 강화학습 기반으로 동시에 훈련하여, 여러 개의 객체가 겹치거나 가려진 상황에서도 불필요한 동작을 줄이고 성공률을 극대화했습니다.
- 둘째, 비전-언어 모델 기반 ThinkGrasp 같은 최신 시스템은 GPT-4o 등 대형 언어모델의 문맥 추론력과 LangSAM 및 VL-파트 기반 세그멘테이션을 결합해, 복잡하게 가려진 목표에 단계적으로 접근하도록 하여 장애물 제거/순차적 탐색을 통해 최종 목적을 달성할 수 있게 합니다. 이 방식은 실제 실험에서 기존 SOTA 대비 약 98% 이상의 성공률, 강력한 일반화 능력 및 시물-투-리얼 성능을 입증했습니다. 셋째, 실시간 multi-view 3D object grasping, 6DoF pose/candidate refinement, 아래에서 위로(bottom-up) 방식의 masking 등 다양한 환경에서 강인성을 보장하기 위한 알고리즘이 일반적으로 병용됩니다.

마스크 연속성 체크, 그리퍼 너비 최적화, 객체 중심점 산출 등 그래스핑 기법

- 세그멘테이션 마스크 정보의 연속적 일관성을 평가하는 것은 객체 추적 및 로봇 조작에서 필수적입니다. 최신 연구에서는 MOTS(Multi Object Tracking and Segmentation) 기법을 실시간 적용하여 프레임 별 마스크 연속성을 검증하고, 객체 분할이 일관적으로 유지되는지 check합니다. 그리퍼 너비 최적화는 마스크 경계 정보와 depth map(또는 3D mesh)에서 주요 contact point(그립 위치)까지의 유효 span을 동적으로 분석해, 객체 손상 없이 파지력을 극대화하는 adaptive gripper design 및 제어 알고리즘으로 구현됩니다. 객체 중심점 계산은 세그멘테이션 결과를 기반으로 centroid 추출, surface normal estimation, minimal enclosing box 등 컴퓨터 비전 기법과 딥러닝 기반 피쳐 추출을 결합한 접근이 점차 확산 중입니다. 제약 최적화 및 실시간 하드웨어-소프트 통합(예: 전기/공압 복합, 3D 프린팅 기반 경량 그리퍼) 역시 최신 보고에서 눈에 띕니다.

Jetson AGX Orin 등 Jetson 플랫폼의 실시간성 보장 최적화 전략

- Jetson AGX Orin 64GB와 같이 엣지 AI 하드웨어의 실시간성과 효율성 최적화는 대규모 실내 환경, 네트워크가 불안정한 현장, 고속/고정확 추론 등이 요구되는 현대 로봇틱스 응용에 가장 중요한 과제입니다. Jetson AGX Orin은 최대 275 TOPS의 연산성능, 2048코어 Ampere GPU, 64GB LPDDR5 메모리 등 최첨단 리소스를 바탕으로 다양한 딥러닝·비전·로봇틱스 워크로드를 실시간으로 처리할 수 있게 설계되었습니다. NVIDIA의 TensorRT와 Deep Learning Accelerator(DLA), 하드웨어별 mixed precision 연산(FP16/INT8), 온디맨드 전력 관리 등이 적극 활용되어, GPU/DLA 간 워크로드 분산 및 레이턴시 감소(저지연성)를 보장합니다.

- 실시간성을 위해서는 딥러닝 모델 경량화(압축 및 pruning), inference 엔진 최적화, 메모리/대역폭 관리 개선, 멀티스레드 파이프라인(Async/Stream 방식), 각종 인터페이스(Bus, Camera, Sensor I/O) 최적화가 필수적입니다. Jetson 컨테이너 기반 배포, ONNX 교차호환, Edge 특화된 신경망 구조(SNN, 경량 CNN 등)가 실제 적용사례에서 검증되고 있습니다.
- 아울러, 대규모 실내 환경에서도 네트워크 장애, 대량 데이터 연산, 수십 개 센서 동시 운영 등 다양한 변수를 고려한 시스템 동적 자원 관리(Edge AI OS, 온디바이스 협업, Task Offloading Architecture)와 실제 산업 로봇/스마트팩토리/스마트빌딩 분야 적용 사례가 활발히 탐구되고 있습니다.

Github 오픈 리소스 정리 (최신 공식/참고 자료)

| 프로젝트명 | 주요 내용 및 기능 | Github 링크 (공식) |
|--------------------|----------------------------|---------------------------------------|
| nanoOWL | OWL-ViT 실시간 Jetson 추론 | NVIDIA-AI-IOT/nanoowl |
| nanoSAM | Segment Anything Model 경량화 | NVIDIA-AI-IOT/nanosam |
| PointPatchRL(PPRL) | 포인트클라우드+마스크 기반 RL | arXiv 페이지/논문 내 공식 링크 제공 |
| ThinkGrasp | 비전-언어 기반 전략적 그래스핑 | H-Freax/ThinkGrasp |
| 기타 | MOTS, MaskRCNN, GCN 등 | 각 논문 내 깃허브 확인 |

종합 정리 및 확장성 관련 제언

- Jetson AGX Orin 환경에서 nanoOWL+SAM을 활용한 자연어 기반 실시간 그래스핑 시스템은 최신 논문과 상업적/오픈소스 지원을 기반으로 곧바로 연구 및 적용이 가능합니다. Depth 카메라 데이터를 결합한 3D 그래스핑 및 RL 기반 제어나, 복잡한 다목적 실내 작업 지원 등이 이미 다수 사례에서 입증되었습니다. 일반화 성능, 실내 복수 작업, 대규모 multi-agent-multi-camera 시스템 등으로 확장할 때도 분산 Edge AI와 고효율 하드웨어-소프트웨어 통합 전략이 병행 적용되고 있습니다. 세그멘테이션-그래스핑-강화학습 전체 파이프라인 및 다양한 하드웨어 가속지원은 최첨단 연구 기반 Github 오픈소스가 나날이 풍부해지는 추세이니 이를 적극적으로 실험 및 개발에 활용하시길 권장합니다.

출처:

[1] Tutorial - Introduction - NVIDIA Jetson AI Lab, <https://www.jetson-ai-lab.com/tutorial-intro.html>

[2] NanoOWL - NVIDIA Jetson AI Lab, https://www.jetson-ai-lab.com/vit/tutorial_nanoowl.html

[3] NVIDIA-AI-IOT/nanosam: A distilled Segment Anything (SAM) model ..., <https://github.com/NVIDIA-AI-IOT/nanosam>

[4] rhett-chen/Robotic-grasping-papers - GitHub, <https://github.com/rhett-chen/Robotic-grasping-papers>

[5] NVIDIA-AI-IOT/nanoowl: A project that optimizes OWL-ViT ... - GitHub, <https://github.com/NVIDIA-AI-IOT/nanoowl>

[6] Abstract arXiv:2303.08774v6 [cs.CL] 4 Mar 2024, <https://arxiv.org/pdf/2303.08774>

[7] 그리퍼 제조사들의 개발동기와 적용 기술 및 제품 특징 < 로봇기술 < 인더스트리4.0 < 기사본문 - MM 산업뉴스, <https://www.mmkorea.net/news/articleView.html?idxno=10194>

- [8] 이미지 Segmentation 문제와 딥러닝: GCN으로 개 고양이 분류 및 분할하기, <https://www.cognex.com/ko-kr/blogs/deep-learning/research/deep-learning-segmentation-finding-marking-dogs-cats-gcn>
- [9] FA저널 모바일 사이트, [칼럼] 제조 AI, 공정 특성에 적합한 알고리즘 선택이 성공의 열쇠, <http://m.fajournal.com/news/articleView.html?idxno=14550>
- [10] Masked Reconstruction Improves Reinforcement Learning on Point ..., <https://arxiv.org/html/2410.18800v1>
- [11] Efficient push-grasping for multiple target objects in clutter ... - Frontiers, <https://www.frontiersin.org/journals/neurobotics/articles/10.3389/fnbot.2023.1188468/full>
- [12] A Vision-Language System for Strategic Part Grasping in Clutter, <https://openreview.net/forum?id=MsCbblqHRA¬elId=MsCbblqHRA>
- [13] ThinkGrasp: A Vision-Language System for Strategic Part Grasping ..., <https://arxiv.org/html/2407.11298v1>
- [14] 차세대 로봇틱스를 위한 Jetson AGX Orin - NVIDIA, <https://www.nvidia.com/ko-kr/autonomous-machines/embedded-systems/jetson-orin/>
- [15] 최신 로드맵 자료보고서 | 중소기업 전략기술 로드맵, <https://smroadmap.smttech.go.kr/mpsvc/dtrprt/mpsvcDtrprtDetail.do?cmYyyy=2023&cmIdx=3615>
- [16] NanoOwl Tutorial on jetson-ai-lab has significantly lower frame rate ..., <https://forums.developer.nvidia.com/t/nanoowl-tutorial-on-jetson-ai-lab-has-significantly-lower-frame-rate-than-shared-on-github/311248>
- [17] Language-Driven 6-DoF Grasp Detection Using Negative Prompt ..., https://dl.acm.org/doi/10.1007/978-3-031-72655-2_21
- [18] [PDF] Language-Driven 6-DoF Grasp Detection Using Negative Prompt ..., https://www.ecva.net/papers/eccv_2024/papers_ECCV/papers/02933.pdf
- [19] [PDF] Language-driven Grasp Detection - CVF Open Access, https://openaccess.thecvf.com/content/CVPR2024/papers/Vuong_Language-driven_Grasp_Detection_CVPR_2024_paper.pdf
- [20] [PDF] 건설현장 증강현실 기반 인간 로봇 - 더브이씨, [https://grant-documents.thevc.kr/251972_\(%EA%B8%B0%ED%9A%8D%EB%B3%B4%EA%B3%A0%EC%84%9C\)+23%EB%85%84%EB%8F%84+%EB%8B%A4%EB%B6%80%EC%B2%98%EA%B3%B5%EB%8F%99%EA%B8%B0%ED%9A%8D+%EB%B3%B4%EA%B3%A0%EC%84%9C.pdf](https://grant-documents.thevc.kr/251972_(%EA%B8%B0%ED%9A%8D%EB%B3%B4%EA%B3%A0%EC%84%9C)+23%EB%85%84%EB%8F%84+%EB%8B%A4%EB%B6%80%EC%B2%98%EA%B3%B5%EB%8F%99%EA%B8%B0%ED%9A%8D+%EB%B3%B4%EA%B3%A0%EC%84%9C.pdf)
- [21] Hanyang - 한양대학교 ERICA학술정보관, <https://information.hanyang.ac.kr/>
- [22] [논문 정리] Mask R-CNN (1) - Give all that you can ⚡, <https://rladuddms.tistory.com/10>
- [23] [논문] AI 기반의 주조 공정 파라미터 최적화를 통한 알고리즘 개선, <https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=JAKO202318443298962>
- [24] [XLS] 주제분류별 상위 300개 논문 건수, https://school.jbedu.kr/_cmm/fileDownload/yanghyeon-h/M010903/3784e1782bc211c28b6ef82ac82c7f4e

- [25] 시각 기반 그리핑 - SCHUNK, <https://schunk.com/kr/ko/jeog-yong-bun-ya-applications-/bijyeon-giban-geuliping>
- [26] [PDF] 대학수학능력시험·EBS·평가원 모의평가 철저분석 수능빈출어휘 - h, https://school.cbe.go.kr/_cmm/fileDownload/os-h/ M010603/ed4f7415ec93ca66bd3c6b2d0ca03bb9
- [27] BMW가 생산을 위해 그리퍼를 최적화하는 방법 < 소재장비 > 기사본문 - MM 뉴스룸, <https://www.mmkorea.net/news/articleView.html?idxno=22013>
- [28] Focus-Then-Decide: Segmentation-Assisted Reinforcement Learning, https://www.researchgate.net/publication/379288534_Focus-Then-Decide_Segmentation-Assisted_Reinforcement_Learning
- [29] MaDi: Learning to Mask Distractions for Generalization in Visual ..., <https://arxiv.org/html/2312.15339v1>
- [30] [PDF] Focus-Then-Decide: Segmentation-Assisted Reinforcement Learning, <https://ojs.aaai.org/index.php/AAAI/article/view/29002/29902>
- [31] Objects matter: object-centric world models improve reinforcement..., <https://openreview.net/forum?id=Q2hkp8WIDS>
- [32] Sequential conditional reinforcement learning for simultaneous ..., <https://www.sciencedirect.com/science/article/abs/pii/S1361841520302255>
- [33] GraspClutter6D: A Large-scale Real-world Dataset for Robust ..., <https://arxiv.org/html/2504.06866v1>
- [34] Real-time multi-view 3D object grasping in highly cluttered ..., <https://www.sciencedirect.com/science/article/pii/S0921889022002020>
- [35] Reactive Grasping in Heavily Cluttered Environment - ResearchGate, https://www.researchgate.net/publication/361285313_Grasping_as_Inference_Reactive_Grasping_in_Heavily_Cluttered_Environment
- [36] DLA로 NVIDIA Jetson Orin에서 딥 러닝 성능 극대화하기, <https://developer.nvidia.com/ko-kr/blog/maximizing-deep-learning-performance-on-nvidia-jetson-orin-with-dla/>
- [37] [PDF] 엣지 딥 러닝 가속기의 추론 성능 분석 - ETRI KSP, <https://ksp.etri.re.kr/ksp/article/file/68530.pdf>
- [38] NVIDIA Jetson Orin을 통해 최첨단 서버급 성능 제공하기, <https://developer.nvidia.com/ko-kr/blog/delivering-server-class-performance-at-the-edge-with-nvidia-jetson-orin/>
- [39] 로봇틱스와 엣지 AI 위한 NVIDIA Jetson AGX Orin | 비맥스테크놀로지, <https://blog.naver.com/bemax00/223165377765>
- [40] [PDF] MaDi: Learning to Mask Distractions for Generalization in Visual ..., <https://www.ifaamas.org/Proceedings/aamas2024/pdfs/p733.pdf>
- [41] [PDF] Adaptive-Masking Policy with Deep Reinforcement Learning for Self ..., https://zhenghuaxu.info/files/2023_ICME.pdf
- [42] A grasp planning algorithm under uneven contact point distribution ..., <https://www.sciopen.com/article/10.1016/j.cja.2023.02.026>

[43] Learn, detect, and grasp objects in real-world settings, <https://link.springer.com/article/10.1007/s00502-020-00817-6>

[44] Beyond Top-Grasps Through Scene Completion | Request PDF, https://www.researchgate.net/publication/344982727_Beyond_Top-Grasps_Through_Scene_Completion