# Upsampling

- **DFU**
  - Core Methodology
    - Sparse depth map을 encoder-decoder 네트워크의 중간에서 dense feature로 보강하여 high resolution의 정밀한 depth map으로 upsampling하는 Module
    - 기존 ED-Net 방식의 'dense→sparse' 과정에서의 정보 손실 문제를 해결하여, 저해상도 depth feature를 밀도 있고 완전한 고해상도 feature로 변환
  - Merit/Demerit
    - Jetson에서 실시간성은 해봐야 알듯함.
  - Feature
    - skip-connection으로 encoder feature를 decoder에 넘기면서, Confidence-aware Guidance Module (CGM)을 통해 adaptive receptive field를 적용해 업샘플링 성능을 최적화
    - 이렇게 완성된 dense depth map은 후속 grasp pose 예측 정확도를 크게 향상
- **OGNI-DC**
  - Core Methodology
    - OGNI-DC는 희소한 깊이 입력으로부터 조밀하고 정확한 깊이 맵을 복원하기 위해 **최적화 기반 접근 방식과 심층 신경망을 결합**한 반복적인 깊이 완성 프레임워크
    - 핵심 아이디어는 전통적인 최적화 방법(예: Variational methods)의 견고함과 심층 신경망의 강력한 표현 학습 능력을 융합하는 것입니다. 각 반복 단계에서, 신경망은 이전 단계의 깊이 추정치와 희소 깊이 입력을 받아 **업데이트 방향(update direction) 또는 잔차(residual)를 예측**합니다. 이 예측은 최적화 문제의 해를 점진적으로 개선하는 데 사용
    - 기존의 단순한 end-to-end 학습 방식과 달리, OGNI-DC는 **모델 기반 최적화(model-based optimization) 과정을 신경망 내부에 명시적으로 통합**함으로써, 희소하고 노이즈가 많은 입력에 대해서도 강인한 깊이 완성을 목표로 합니다. 이는 마치 최적화 알고리즘의 각 반복 스텝을 학습 가능한 신경망 레이어로 펼쳐놓은(unrolling) 형태와 유사
    - 구체적으로는, 각 반복 k에서 현재의 깊이 추정치 $\mathbf{D_{k}}$ 와 희소 깊이 입력 $\mathbf{D_{s}}$ 를 사용하여 신경망 $\mathbf{f_{\theta}}$ 가 업데이트 $\mathbf{\Delta D_{k}}$ 를 예측하고, 이를 통해 다음 단계의 깊이 추정치 $\mathbf{D_{k+1}}=\mathbf{D_{k}}+\Delta D_{k}$ (또는 다른 형태의 업데이트 규칙)를 얻습니다. 이 과정은 사전 정의된 반복 횟수만큼 또는 수렴 조건이 만족될 때까지 수행
  - Merit
    - **강인성 (Robustness):** 최적화 기반 접근 방식의 통합으로 인해 입력 데이터의 노이즈나 희소성 변화에 상대적으로 강인한 성능을 보입니다. 신경망이 데이터 분포에 과적합되는 것을 완화하고, 물리적 또는 기하학적 제약 조건을 간접적으로 학습할 수 있는 구조를 제공합니다.
    - **해석 가능성 (Interpretability):** 반복적인 개선 과정은 각 단계에서 깊이 맵이 어떻게 정제되는지에 대한 직관적인 이해를 제공할 수 있습니다. 이는 완전한 블랙박스 형태의 end-to-end 모델에 비해 디버깅이나 성능 분석에 유리할 수 있습니다.
    - **정확도 향상:** 최적화 과정을 통해 점진적으로 해를 개선하므로, 특히 복잡한 장면이나 정밀한 깊이 정보를 요구하는 경우 높은 정확도를 달성할 잠재력이 있습니다. 신경망은 데이터로부터 복잡한 패턴을 학습하고, 최적화는 이를 기반으로 일관성 있는 해를 찾도록 유도합니다.
    - **유연성 (Flexibility):** 최적화 문제의 정식화(formulation)나 신경망 구조를 특정 애플리케이션의 요구사항이나 사용 가능한 센서 특성에 맞게 조정할 여지가 있습니다. 예를 들어, 다른

종류의 정규화 항(regularization term)을 최적화 과정에 통합하거나, 반복 단계별로 다른 네트워크 구조를 사용할 수도 있습니다.

- Demerit
  - **계산 비용 (Computational Cost):** 반복적인 구조는 추론 과정에서 단일 패스(single-pass) 네트워크에 비해 더 많은 계산량을 요구할 수 있습니다. 각 반복마다 신경망 연산이 수행되므로, 반복 횟수가 증가하면 전체 추론 시간이 길어집니다.
  - **실시간성 제약 (Real-time Constraints):** 높은 계산 비용은 Jetson과 같은 임베디드 시스템에서의 실시간 처리에 제약이 될 수 있습니다. DFU와 마찬가지로, 실제 하드웨어에서의 실시간성 확보 여부는 모델의 복잡도, 반복 횟수, 최적화 수준 등 다양한 요인에 따라 달라지므로 실험적인 검증이 필요합니다.
  - **학습의 복잡성 (Training Complexity):** 반복적인 구조와 최적화 과정의 통합은 학습 과정을 더 복잡하게 만들 수 있습니다. 적절한 손실 함수 설계, 안정적인 학습을 위한 초기화 전략, 반복 단계 간의 정보 흐름 제어 등이 중요하며, 이는 하이퍼파라미터 튜닝을 어렵게 만들 수 있습니다.
  - **수렴 보장 (Convergence Guarantee):** 학습된 신경망에 의해 업데이트가 결정되므로, 전통적인 최적화 알고리즘처럼 항상 전역 최적해(global optimum)나 특정 지점으로의 수렴을 이론적으로 보장하기 어려울 수 있습니다. 실제로는 잘 작동하는 경우가 많지만, 특정 입력에 대해서는 발산하거나 만족스럽지 못한 지역해(local minimum)에 수렴할 가능성도 배제할 수 없습니다.
- Feature
  - **최적화 단계의 학습 (Learned Optimization Steps):** OGNI-DC의 가장 큰 특징은 전통적인 최적화 알고리즘의 반복 스텝 자체를 학습 가능한 신경망 모듈로 대체했다는 점입니다. 이를 통해 데이터로부터 최적의 업데이트 전략을 학습하며, 문제 특화된 최적화기를 설계하는 효과를 얻습니다.
  - **반복적 정제 (Iterative Refinement):** 초기 추정치에서 시작하여 점진적으로 깊이 맵을 개선해나가는 반복적 정제 과정을 통해, 단일 단계 추론 방식보다 더 정교하고 일관성 있는 결과를 생성할 수 있습니다. 이는 특히 희소성이 매우 높거나 객체의 경계가 복잡한 영역에서 유리하게 작용할 수 있습니다.
  - **가이드된 신경망 (Guided Neural Network):** 각 반복 단계에서 신경망은 이전 단계의 출력과 원본 희소 입력을 모두 활용하여 다음 업데이트를 예측합니다. 이는 마치 최적화 알고리즘이 현재 해와 문제 정의를 바탕으로 다음 탐색 방향을 결정하는 것과 유사하며, 신경망이 맹목적으로 출력을 생성하는 것이 아니라 명확한 가이드라인 하에 작동하도록 합니다.
  - **다양한 최적화 기법과의 잠재적 결합:** 논문에서 제시된 특정 형태 외에도, 다양한 최적화 알고리즘(예: ADMM - Alternating Direction Method of Multipliers, Primal-Dual methods 등)의 반복 스텝을 신경망으로 대체하거나 결합하는 형태로 OGNI-DC의 아이디어를 확장할 수 있는 잠재력이 있습니다.

# Depth Image

- **GG-CNN / GG-CNN2**
  - Core Methodology
    - Depth image를 input으로 받아, 각 픽셀 위치에서 Antipodal Grasp (평행 파지)의 품질(Quality), 각도(Angle), 파지 폭(Width)을 예측하는 FCN
    - GG-CNN2는 GG-CNN의 개선 버전
    - Dilated Convolution 등을 활용하여 성능을 향상
  - Merit
    - 매우 빠르고 경량이며, 실시간 및 Closed-loop 제어에 적합하다. 구현이 비교적 용이

    - Demerit
        - Depth Image 기반의 2D 처리 방식으로 인해 3D 공간 정보 활용에 한계가 있다. Planar grasp (주로 Top-down grasp)에 적합하며, 복잡한 6-DOF 파지나 특정 방향에서의 접근이 필요한 경우 적용이 어렵다
    - Feature
        - 실시간 성능 요구사항 (특히 5 FPS 이상)을 충족할 가능성이 매우 높음
        - DFU와의 조합을 통해 NanoSAM 마스크 영역 내 객체 파지에 적용하기에 적합
        - 초기 단순 환경 실험에 우선적으로 고려할 수 있는 강력한 후보

## Depth Pointcloud

- **PointNetGPD**
    - Core Methodology
        - PointNet 아키텍처를 기반으로 한 End-to-end Grasp 'Evaluation' 모델이다. 로봇 Gripper가 닫힐 영역 내의 3D Point Cloud를 직접 입력받아 해당 Grasp Candidate의 품질(Quality)을 평가
        - Sparse Point Cloud에서도 강인하게 작동하도록 설계
    - Merit
        - Point Cloud 데이터를 직접 처리하며, Sparse 데이터에 강인하고 모델이 비교적 경량
    - Demerit
        - End-to-end Grasp 'Generation'이 아닌 'Evaluation' 모델
        - 별도의 Grasp Candidate Sampling 또는 Generation 메커니즘이 필요
        - 6-DOF Grasp Pose를 직접 생성하지 않음
    - Feature
        - Point Cloud 기반 접근 방식을 선호할 경우 고려할 수 있음
        - 실시간 성능 잠재력은 있으나, 전체 시스템 Latency는 Candidate 생성 부분에 크게 의존
- **REGNet / REGNet-V2**
    - Core Methodology
        - Color image도 추가로 가능, pointcloud(XYZ or XYZRGB)도 사용 가능
        - REgion-based Grasp Network의 약자로, Point Cloud를 입력받아 End-to-end 방식으로 6-DOF Grasp를 검출
        - 일반적으로 3단계(Score Network (SN), Grasp Region Network (GRN), Refine Network (RN))로 구성
        - SN은 파지에 적합한 점들을 필터링하고, GRN은 선택된 점들을 기반으로 Grasp Proposal을 생성하며, RN은 이를 정제
        - REGNet-V2는 Multi-layer 구조를 도입하여 성능을 개선한 버전
    - Merit
        - End-to-end 방식으로 Point Cloud로부터 6-DOF 파지를 직접 검출
    - Demerit
        - 실시간 성능이 사용자의 목표치에 비해 다소 부족
        - 후속 연구로 *HGGD* , *RegionNormalizedGrasp* 등이 발표되어 성능 및 효율성이 개선되었을 가능성
    - Feature
        - 6-DOF 파지가 가능한 Point Cloud 기반 옵션 중 하나
        - 하지만 실시간 성능 제약 가능성이 상대적으로 높으므로, 후속 연구인 HGGD나 RNGNet을 우선적으로 검토하는 것 추천
- **Contact-GraspNet**

    - Core Methodology
        - Depth image에서 pointcloud로 변경하여 입력을 받음, Segmentation Mask 정보를 같이 사용
        - 6-DOF Grasp Pose의 분포(Distribution)를 직접 예측
        - 객체 분할(Segmentation) 정보를 활용하는 것이 강력히 권장
    - Merit
        - Point Cloud를 직접 처리하여 6-DOF Grasp 분포를 예측
        - Segmentation 정보를 효과적으로 활용하여 성능을 향상
    - Demerit
        - 공식 PyTorch 구현이 없음
        - 비공식 구현체의 성능 및 안정성 검증이 필요
        - 실시간 성능에 대한 명확한 정보가 부족
    - Feature
        - 6-DOF 파지가 가능하고 Point Cloud 기반 접근 방식을 사용하며, NanoSAM과의 통합이 용이하다는 점에서 매력적인 후보
        - 하지만 실시간 성능과 PyTorch 구현체의 신뢰성 확인이 선행

## RGB-D Image

- **GR-ConvNet**
    - Core Methodology
        - GG-CNN과 유사하게 n-channel image (일반적으로 RGB-D)를 입력받아 픽셀 단위로 Antipodal Grasp (Quality, Angle, Width)를 예측
        - GG-CNN과의 주요 차이점은 Residual Network 구조를 사용하여 성능 향상을 도모
    - Merit
        - GG-CNN 대비 Residual 구조를 통해 성능(정확도) 향상을 기대
        - Real-time 성능을 제공
        - RGB 정보를 활용하여 Depth 정보만 사용하는 것보다 더 풍부한 특징 추출이 가능
    - Demerit
        - GG-CNN과 동일하게 Planar grasp의 한계
    - Feature
        - GG-CNN2의 좋은 대안이 될 수 있으며, 특히 RGB 정보를 함께 활용하고자 할 때 유리
        - 실시간 성능 요구사항 충족 가능성이 높음
- **HGGD**
    - Core Methodology
        - 효율적인 6-DOF 파지 검출을 위해 Heatmap Guidance 방식
        - RGB-D 이미지를 입력받아 2D CNN으로 Grasp Location Heatmap을 예측하고, 이 Heatmap을 가이드 삼아 파지 가능성이 높은 Local 영역의 Point Cloud에 집중하여 Grasp Pose를 생성 (Global-to-local, Semantic-to-point 접근)
        - Non-uniform anchor sampling 기법으로 정확도와 다양성을 향상
    - Merit
        - 효율성과 실시간 성능을 크게 강조
        - Heatmap Guidance를 통해 파지 가능 영역에 집중하여 정확도와 다양성 증가
    - Demerit
        - RGB-D 입력이 모두 필요할 가능성
        - 구체적인 추론 속도(FPS) 정보가 부족
    - Feature

        - REGNet의 후속 연구로서, 실시간 6-DOF 파지를 위한 매우 유망한 후보
        - DFU와의 결합 가능성도 탐색해볼 가치
        - 후속 연구인 RNGNet의 성능(50 FPS)은 Jetson Orin에서의 실현 가능성을 높게 시사

## Multi-View

- **VGN**
    - Core Methodology
        - 3D TSDF(Multi-View Depth image fusion)
        - 3D Convolutional Neural Network (3D CNN)를 사용하여 Truncated Signed Distance Function (TSDF) 형태로 표현된 3D 볼륨 입력을 받아, 각 Voxel 위치에서 6-DOF Grasp (Quality, Orientation, Width)를 직접 예측
        - End-to-end 방식으로 학습하며, 전체 3D 장면 정보를 활용하여 암시적으로 충돌 없는 (Collision-free) 파지를 학습하는 것을 목표
    - Merit
        - Real-time 6-DOF 파지 생성이 가능
        - 전체 3D 장면 정보를 활용하여 Clutter 환경에서의 파지 성공률이 높다
        - 단일 Forward Pass로 전체 작업 공간에 대한 파지 예측이 가능
    - Demerit
        - 입력으로 TSDF 생성이 필수적이며, 이는 일반적으로 Multi-view Depth 데이터와 추가적인 처리 시간을 요구
        - Voxel 해상도에 따라 메모리 사용량과 계산 비용이 증가하며, 이는 파지 정확도와 Trade-off 관계
    - Feature
        - 6-DOF 파지가 가능하며 실시간 성능 잠재력이 매우 높음
        - NanoSAM으로 분할된 객체 영역에 대해 TSDF를 생성하고 VGN을 적용하는 파이프라인을 고려할 수 있음
        - 단일 Realsense 카메라 사용 시 TSDF 생성을 위한 로봇의 스캔 동작이 필요할 수 있으며, 이는 전체 Latency를 증가시키는 요인