특집논문 (Special Paper)
방송공학회논문지 제31권 제1호, 2026년 1월 (JBE Vol.31, No.1, January 2026)
https://doi.org/10.5909/JBE.2026.31.1.2
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

영상 압축 환경에서의 압축 강도를 고려한 Denoising 기반 VGGT 성능 분석

김 제 희[a)]*, 김 동 휘[a)]*, 문 채 원[a)], 이 윤 호[a)], 김 은 지[a)], 정 진 우[b)], 박 상 효[a)‡]

Performance Analysis of VGGT with Denoising under Various Compression Levels in Lossy Compression

Jehee Kim[a)]*, Dong-hwi Kim[a)]*, Chaewon Moon[a)], Yoonho Lee[a)], Eunji Kim[a)], Jinwoo Jeong[b)], and Sang-hyo Park[a)‡]

요 약

최근 VGGT를 비롯한 대규모 비전 트랜스포머 기반 모델들이 등장하며, 복잡한 3D 장면을 정밀하게 해석하는 기술이 발전하고 있다. 그러나 이러한 발전은 주로 비압축 영상을 전제로 한 학습·평가에 기반하기 때문에, 실제 환경에서 흔히 사용되는 손실 압축 영상의 영향을 충분히 고려하지 않는다. 손실 압축으로 발생하는 압축 아티팩트는 3D 비전 모델의 장면 해석 성능을 저하시킬 수 있다. 본 논문에서는 대표적인 3D 장면 해석 모델인 VGGT를 대상으로, 압축 아티팩트가 추론 성능에 미치는 영향을 분석하여 실제 환경 적용 시의 안정성 확보에 기여하고자 한다. 이를 위해 CO3Dv2 데이터셋에서 9개의 카테고리를 선정하고 범용적으로 사용되는 이미지 코덱인 JPEG과 비디오 코덱인 AVC를 사용하여 다양한 압축 강도로 데이터셋을 구성하여 추론 성능의 하락을 확인하였다. 이후 CO3Dv2 데이터셋의 모든 카테고리에서 압축 이미지와 노이즈 제거된 이미지를 Ground Truth와 비교하여 압축 강도 변화에 따른 VGGT 모델의 카메라 포즈, 깊이, 포인트 맵 추론 능력을 평가하였다. 그 결과 AVC 코덱은 QP 42 이상, JPEG은 Quality 20 이하부터 성능 저하가 뚜렷하게 나타났다.

Abstract

Recent advancements in large-scale vision transformer models, including VGGT, have significantly improved the ability to interpret complex 3D scenes. However, most of these models are trained and evaluated using uncompressed data, overlooking the impact of lossy compression commonly present in real-world scenarios. Compression artifacts caused by lossy compression can degrade the scene understanding performance of 3D vision models. In this study, we investigate the impact of compression artifacts on VGGT, a representative model for 3D scene understanding, to assess its robustness in practical environments. We select nine categories from the CO3Dv2 dataset and apply various compression levels using commonly used codecs: JPEG for images and AVC for video. The model's performance is evaluated by comparing its camera pose estimation, depth estimation, and point map reconstruction accuracy against the original ground truth data. Experimental results indicate that degradation becomes noticeable when using AVC at QP $\geq$ 42 and JPEG at Quality $\leq$ 20. These findings suggest the importance of considering lossy compression distortion when deploying transformer-based 3D vision models in real-world applications.

Keyword : 3D reconstruction, 3D scene understanding, Video/Image Compression

Ⅰ. 서 론

최근 VGGT (Visual Geometry Grounded Transformer)[1]와 같은 3D 비전 모델의 발전은 복잡한 현실 장면을 세밀하게 재구성하고 이해하는 능력을 크게 향상시켰다. VGGT는 단일 이미지 입력만으로 카메라 파라미터, 깊이 맵 및 포인트 맵을 추론할 수 있는 모델로, 단순한 구조와 사전학습 된 신경망을 기반으로 3D 재구성 분야에서 현존 최고 수준의 성능을 보여준다. 그러나 현실적으로는 이러한 모델을 실제 응용 (3D 객체 탐지, 장면 복원 등)에 적용하기 위해서는 전송 대역폭의 한계, 저장공간의 부족, 실시간 처리 속도 요구와 같은 제약을 직면하게 된다. 이러한 이유로, 손실 압축이 적용된 데이터를 입력으로 사용하는 경우가 일반적이다. 손실 압축은 주파수 도메인 변환 과정에서 고주파 영역의 정보를 일부 포기하는 대신 비트 전송률을 줄인다. 하지만 정보를 포기하면서 발생하는 압축 아티팩트는 인간의 영상 인지에 영향을 줄 뿐만 아니라, 컴퓨터 비전 모델의 장면 해석 정확도를 저하시키는 요인으로 작용할 가능성이 있다.

2D 비전 분야에서는 압축 아티팩트가 모델 성능에 미치는 영향을 다룬 연구[2,3]가 이루어져 왔으나, 3D 비전 분야에서는 압축 강도와 모델 성능 간의 관계를 체계적으로 분석한 연구가 거의 없는 실정이다. 이러한 연구의 부재는 3D 비전 모델을 현실적으로 적용함에 있어, 압축 영상을 신뢰성 있게 활용하는 데에 위험요소로 작용한다. 예를 들어, 네트워크 환경으로 인해 압축된 이미지가 입력으로 사용될 경우, 모델이 예기치 못한 오차를 내거나 재구성 품질이 급격히 저하될 수 있다. 결국, 압축 강도에 따른 입력 이미지의 품질 저하가 3D 비전 모델의 성능에 미치는 영향을 명확히 규명하는 것이 본 연구에서 제안하는 압축 영상 입력 기반 3D 비전 모델의 활용 가능성을 보장하는 핵심 과제이다.

본 연구에서는 현재 최고 수준의 성능 모델인 VGGT에서 입력 이미지의 압축 강도의 변화 따른 성능 변화를 정성적, 정량적으로 분석하였다. 이를 위해 CO3Dv2 데이터셋[4]을 기반으로, 대표적 이미지 코덱인 JPEG[5]과 대표적 동영상 코덱인 AVC (H.264)[7]에 대해 다양한 압축 강도로 구성된 압축 이미지 데이터셋을 구성하여 압축으로 인한 성능 하락을 관찰하였다. 또한, HEVC, VVC로 압축된 검증용 데이터셋을 구성하였다. 검증용 압축 데이터셋에 대해 동영상 기반 압축 노이즈 제거 모델 (STDF[29], RFDA[30], 그리고 STFF[31])을 적용하여 노이즈 제거를 실시하였고, 압축 영상 및 노이즈 제거된 영상의 VGGT 모델의 추론을 실시하여 평가를 실시하였다. 정성적, 정량적 결과를 시각적 비교 및 포인트 맵의 포인트 수, Chamfer Distance[9], Area Under Curve (AUC)를 평가지표로 하여 압축 강도에 따른 VGGT 성능을 평가하였다. 전반적인 연구의 흐름은 그림 1과 같다.

Ⅱ. 관련 연구

1. 영상 압축 코덱

영상 압축 기술은 한정된 전송 대역폭과 저장공간을 효율적으로 사용하기 위한 핵심 기술로, 이미지 기반 압축 기법으로는 JPEG[5], JPEG2000[6], 동영상 기반 압축으로는 AVC[7], HEVC[8], VVC[26] 등 다양한 표준 손실 압축 코덱이 널리 사용되고 있다. 그중 이미지 압축에는 JPEG이, 동영상 압축에는 HEVC가 가장 범용적으로 사용된다. 이러한 손실 압축은 주파수 도메인 변환 과정에서 고주파 영역의 정보를 일부 제거하는 방식으로 비트 전송률을 줄이는 방식을 채택하였다. 부호화 과정에서 블로킹, 링잉, 블러링 등의 압축 아티팩트[20]가 나타나게 되는데, 이는 인간의 영상 인지에는 둔감히 반응하도록 설계되었으나 영상 내의 지역적인 정보가 왜곡되므로 컴퓨터 비전 모델이 정확한 정보

a) 경북대학교 컴퓨터학부(Kyungpook National University)
b) 한국전자기술연구원(Korea Electronics Technology Institute)
* Equal contribution
‡ Corresponding Author : 박상효(Sang-hyo Park)
E-mail: s.park@knu.ac.kr
Tel: +82-53-950-6373
ORCID: https://orcid.org/0000-0002-7282-7686
※ This research was supported by the Regional Innovation System & Education(RISE) Glocal 30 program through the Daegu RISE Center, funded by the Ministry of Education(MOE) and the Daegu, Republic of Korea.(2025-RISE-03-001)
· Manuscript November 26, 2025; Revised January 9, 2026; Accepted January 9, 2026.

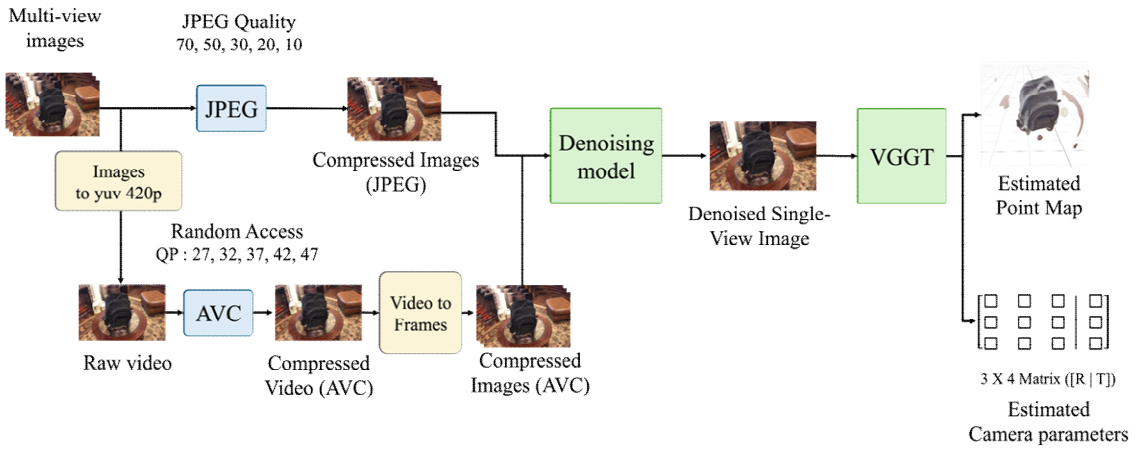

그림 1. 본 논문에 사용된 전반적인 프레임워크. 이미지, 동영상 코덱 기반의 압축 방법을 사용하였으며, 각 프레임은 VGGT의 입력으로 사용되어 추론 결과를 통해 포인트 맵, 카메라 포즈에 대한 결과를 도출

Fig. 1. Overall framework used in this paper. The input images are compressed using standard codecs: JPEG for images and AVC for videos. Each frame is fed into VGGT to infer point maps and camera poses.

를 해석하는 데 어려움을 주는 요인으로 작용한다.

2. 딥러닝 기반 노이즈 제거

노이즈 제거 (Denoising)는 이미지 품질 복원 분야에서 가장 오래 연구된 문제 중 하나로, 고전적인 필터링 기반 접근 방식에서 시작하여 현재는 딥러닝 기반 모델을 통한 노이즈 제거로 빠르게 발전해 왔다. 압축 아티팩트 제거는 일반적인 노이즈 제거보다 더 복잡한 구조적 노이즈를 다루어야 하므로, 다양한 신경망 구조가 제안되었다. 초기 딥러닝 기반 복원 모델인 ARCNN[23]은 CNN을 활용해 픽셀 주변의 지역적인 패턴을 학습하였고, 이후 잔차 학습을 추가한 DnCNN[24], U-Net 구조를 채택한 DRUNet[25]으로 발전하였다.

최근의 신경망 구조의 이미지 노이즈 제거 연구에서는 트랜스포머 구조를 활용한 JDEC[14], PromptCIR[15] 등이 제안되며, 장거리 의존성 학습을 통한 전역적 맥락 파악 능력을 활용해 반복적 패턴이나 블록 구조와 같은 압축 아티팩트를 보다 정밀하게 제거하는 성능을 보이고 있다. 또한 Diffusion 기반 모델은 확률적 노이즈 과정을 모방해 이미지를 점진적으로 또는 단일 단계로 복원하는 방식으로 동작한다. 특히 one-step Diffusion 모델은 기존 Diffusion 방식 대비 복원 속도를 크게 개선하면서 복원 품질을 유지할 수 있어 CODiff[16], StableCodec[17], OSCAR[18], OSEDiff[19]와 같은 모델들이 주목받고 있다.

멀티 프레임 기반 노이즈 제거 연구에서는 MFQE[27]가 최초로 모션 보상 기반 프레임 정렬과, peak quality frame (PQF) 검출을 통해 압축으로 열화된 프레임의 화질을 복원하였다. 이후, 옵티컬 플로우를 대체하여 deformable convolution (DCN)[28]을 활용하여 인접 프레임을 정렬 및 융합하는 STDF[29] 모델이 제안되어 멀티 프레임 노이즈 제거 연구의 효율성을 끌어올렸다. 이후 RFDA[30]는 재귀적 융합과, 어텐션 메커니즘을 통해 압축 아티팩트가 심한 영역에 복원을 집중시키는 방식으로 기존 정렬 및 융합의 한계를 보완하였으며, 최근의 STFF[31] 모델은 정렬, 융합 축에 주파수 정보를 함께 결합하는 퓨전 방식으로 디테일 복원 능력을 강화하는 방향으로 발전하였다. 요컨대, 노이즈 제거 및 압축 아티팩트 복원 기법은 충분히 축적되어 왔음에도 불구하고, 압축 입력을 사용하는 3D 비전 모델에 대한 적용 가능성을 분석한 연구는 아직 제시되지 않았다. 본 연구에서는 STDF, RFDA, STFF를 동영상 복원 기법으로 채택하여 기존 연구의 공백을 해소하고자 한다.

3. 다중 시점 재구성

다중 시점 재구성 (Multi-view Stereo, MVS) 기법은 여

러 시점에서 촬영된 영상들로부터 기하학적 정보를 추측하여 3차원 구조를 복원하는 기술이다. 초기의 연구들은 Structure-from-Motion (SfM)[21]과 같이 기하학적 최적화 기반 접근이 주류를 이루었으며, 영상 간의 대응점을 추출하고 이를 활용해 기하학적 정보를 추론해 내는 방법을 사용했다. 이후 DUSt3R[22]과 같이 트랜스포머 기반 신경망을 학습하여 복잡한 최적화 과정 없이도 빠르게 다중 시점 간의 기하학적 관계를 빠르게 추론할 수 있는 방법으로 발전해 왔다. 최근 제안된 VGGT[1]는 단일 이미지 입력만으로도 빠르게 카메라 파라미터, 깊이, 포인트 맵을 예측할 수 있는 3D 비전 모델로, 현존 최고 성능을 보인다. 그러나, 기존 제시된 모델의 한계점은 대부분 압축된 프레임을 고려하지 않았으며, 대부분의 결과를 고품질 입력 영상을 전제로 평가되었으므로 손실 압축으로 인한 입력 영상 품질 저하가 3D 재구성 성능에 미치는 영향에 대해서는 충분히 분석되지 않은 실정이다.

Ⅲ. 본 론

본 연구에서는 압축 영상이 VGGT 성능에 미치는 영향을 분석 관점과 복원 적용 관점에서 구분하여 접근한다. 먼저, 압축 강도가 3D 추론 성능에 미치는 영향을 정밀하게 분석하기 위해, 압축 파라미터를 비교적 안정적으로 제어할 수 있는 AVC 기반 압축을 분석용 기준으로 사용하였다. 이는 특정 코덱 구조나 참조 프레임 설정에 따른 영향을 최소화하고, 압축 강도 변화에 따른 성능 추이를 관찰하기 위함이다. 영상 복원 실험에서는 실제 전송 및 저장 환경에서 널리 사용되는 HEVC 및 VVC 코덱을 적용하여, 복원 기법들이 표준 비디오 압축 환경에서도 VGGT의 추론 성능을 개선할 수 있는지를 검증하였다.

1. 실험 데이터셋 구성

데이터셋은 CO3Dv2 데이터셋에서 9개 카테고리를 선정하여 정성적으로 신뢰도가 높은 하나의 시퀀스를 선정하였다. 압축은 JPEG과 AVC 및 HEVC, VVC 코덱을 사용하여 진행하였다. JPEG 압축은 python openCV 라이브러리를 사용하였고 Quality 값은 70, 50, 30, 20, 10으로 설정하였다. AVC 코덱은 ffmpeg을 사용해서 이미지 형식을 yuv420p로 변환하여 하나의 시퀀스로 이어 붙인 뒤, QP를 27, 32, 37, 42, 47로 설정하여 압축 후 프레임을 추출하여 데이터셋을 구성하였다. 데이터셋 구성 과정은 그림 1에 시각적으로 표현되어 있다. 평가를 위하여 HEVC, VVC를 각각 All intra 모드에 대해 QP 27, 32, 37을 통해 압축하였고, 각각의 압축 프레임 결과를 STDF, RFDA, STFF를 통해 노이즈 제거된 프레임을 구성하였다.

2. 실험 환경 및 평가 방법

3.1에서 구성한 압축된 데이터셋으로 VGGT 모델을 사용하여 NVIDIA RTX 3080 환경에서 실험하였다. 입력 영상과 모델의 추론 성능 사이의 관계성을 찾기 위해 정성적 평가로 포인트 맵 추정의 시각적 비교, 정량적 평가로 추론된 포인트 맵의 포인트 수, Chamfer Distance와 Area Under Curve (AUC)를 사용하여 포인트 맵의 유사도와 카메라 포즈 추론 능력을 정량적으로 평가하였다. 해당 실험의 경우, 각 프레임을 개별적으로 VGGT의 입력으로 사용하였다. Chamfer Distance와 AUC의 평가 기준으로 CO3Dv2 데이터셋에서 제공하는 Ground truth (GT)를 사용하였다.

Ⅳ. 실험 결과

1. 포인트 맵 추정

그림 2, 그리고 그림 3은 서로 다른 압축 강도의 이미지를 입력으로 주었을 때 모델이 추론한 포인트 맵의 변화 시각화 결과이다. 추론이 완전히 실패한 경우가 아니라면 대부분 중심 물체의 두드러지는 시각적 차이는 나타나지 않았고, 배경에 대한 포인트 수가 감소하는 경향을 보였다. 그러나 AVC 코덱의 QP 42, 47에서 중심 물체의 포인트 예측에 대해서도 성능 저하가 나타났다. 이는 그림 3 하단에서 확인할 수 있다. 각 카테고리별 압축 강도에 따른 예측 포인트 맵은 그림 4에 나타나 있다.

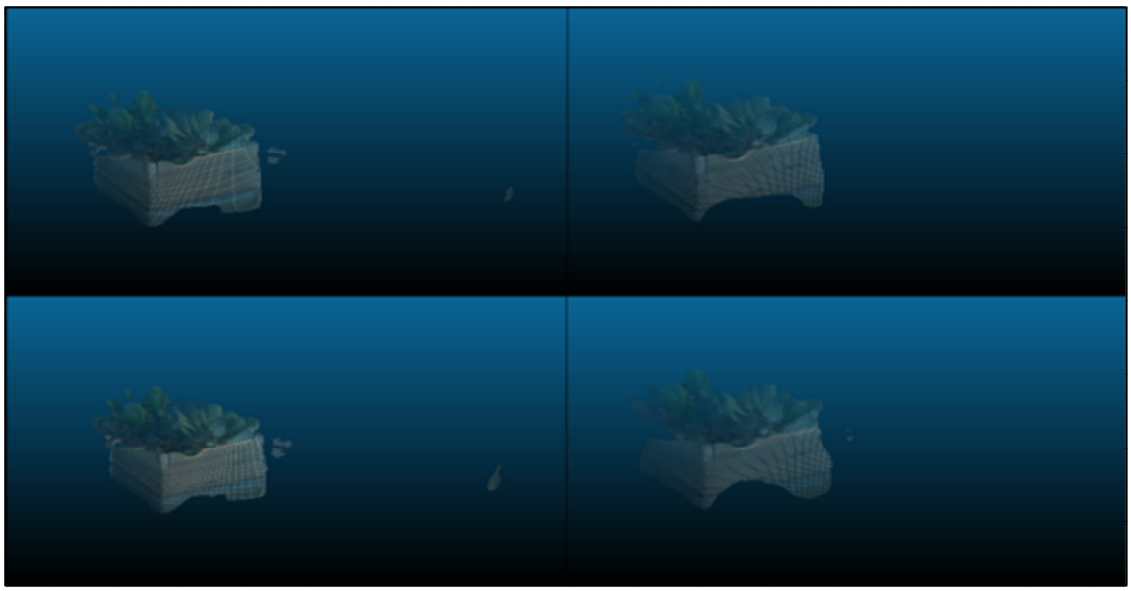

그림 2. CO3Dv2 데이터셋의 plant 카테고리에 대한 VGGT의 포인트 맵 추론 후 시각화 결과
(위: JPEG Quality 70, 10, 아래: AVC QP 27, 47)
Fig. 2. Inferred point maps by VGGT for the Plant category on the CO3Dv2 dataset
(Top: JPEG Quality 70 (left) and 10 (right); Bottom: AVC QP 27 (left) and 47 (right))

그림 3. CO3Dv2 데이터셋의 backpack 카테고리에 대한 VGGT의 포인트 맵 추론 후 시각화 결과
(위: JPEG Quality 70, 10, 아래: AVC QP 27, 47)
Fig. 3. Inferred point maps by VGGT for the backpack category on the CO3Dv2 dataset
(Top: JPEG Quality 70 (left) and 10 (right); Bottom: AVC QP 27 (left) and 47 (right))

그림 4. 각 카테고리별 압축 강도에 따른 예측된 포인트 맵, 예측이 제대로 이루어지지 않은 장면은 공백 처리
Fig. 4. Inferred Point maps by VGGT for each category at different compression levels; scenes with failed predictions are left empty.

2. 추정된 포인트 맵의 포인트 수

VGGT가 추론한 포인트 맵의 포인트 수를 확인한 결과, 압축 강도에 따라 변화가 나타났다. 압축 강도가 높아질수록 추론된 포인트 수가 감소하는 경향을 보였다. 이는 부분적인 압축 아티팩트로 인해 특정 픽셀에 대한 깊이 예측이 어려워졌기 때문으로 보인다. 압축이 적용된 프레임의 VGGT 추론 후 포인트 수의 변화는 표 1과 표 2에 제시하였다.

표 1. VGGT의 입력을 JPEG 압축된 이미지로 사용하였을 때의 압축 강도에 따른 포인트 맵의 포인트 수 평균. 빨간 글씨는 하락이 가장 작은 구간이며, 파란 글씨는 하락이 가장 큰 구간을 의미

Table 1. The mean of the number of points in the point map when JPEG-compressed images are used as input to VGGT. Red text indicates the QF with the highest mean number of points, while blue text indicates the QF with the lowest mean number of points.

Methods		Mean of number of points
RAW		26,615
JPEG	Q 70	25,235
	Q 50	22,855
	Q 30	20,013
	Q 20	16,038
	Q 10	15,356

표 2. VGGT의 입력을 AVC 압축된 이미지로 사용하였을 때의 압축 강도에 따른 포인트 맵의 포인트 수 평균. 빨간 글씨는 하락이 가장 작은 구간이며, 파란 글씨는 하락이 가장 큰 압축 구간을 의미

Table 2. The mean of the number of points in the point map when AVC-compressed images are used as input to VGGT. Red text indicates the QP with the highest mean number of points, while blue text indicates the QP with the lowest mean number of points.

Methods		Mean of number of points
RAW		26,615
AVC	QP 27	26,174
	QP 32	24,492
	QP 37	22,597
	QP 42	20,149
	QP 47	16,450

표 3. 이미지 코덱 압축 데이터를 VGGT 입력으로 사용했을 때, 압축 강도에 따른 Chamfer distance 평균. 빨간 글씨는 하락이 가장 적은 구간이며, 파란 글씨는 하락이 가장 큰 압축 구간을 의미

Table 3. The mean of the chamfer distance in the point map when JPEG-compressed images are used as input to VGGT. Red text indicates the QF with the highest mean chamfer distance, while blue text indicates the QF with the lowest mean of chamfer distance.

Methods		Chamfer Distance
RAW		3.4801
JPEG	Q 70	3.4801
	Q 50	3.5059
	Q 30	3.5075
	Q 20	3.5226
	Q 10	3.5400

표 4. 동영상 코덱 압축 데이터를 VGGT 입력으로 사용했을 때, 압축 강도에 따른 Chamfer distance 평균. 빨간 글씨는 하락이 가장 적은 구간이며, 파란 글씨는 하락이 가장 큰 압축 구간을 의미

Table 4. The mean of chamfer distance in the point map when AVC-compressed images are used as input to VGGT. Red text indicates the QP with the highest mean chamfer distance, while blue text indicates the QP with the lowest mean of chamfer distance.

Methods		Chamfer Distance
RAW		3.4801
AVC	QP 27	3.4721
	QP 32	3.4923
	QP 37	3.5279
	QP 42	3.5615
	QP 47	3.5884

3. 추정된 포인트 맵의 유사도

한 장의 이미지를 넣었을 때 VGGT 모델이 생성하는 포인트 맵의 유사도를 평가하기 위해서 Chamfer Distance를 평가지표로 사용하였다. Chamfer Distance는 두 점 집합 $S_1, S_2 \subseteq \mathbb{R}^3$의 차이를 측정하는 지표로 다음의 식으로 정의된다:

$$d_{CD}(S_1, S_2) = \frac{1}{|S_1|}\sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \frac{1}{|S_2|}\sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2. \quad (1)$$

식 (1)은 차이가 작을수록 두 집합이 유사함을 의미한다. 표 3, 4에서 볼 수 있듯이, 평균적으로 압축 강도가 강한 입력에 대해 Ground Truth와 차이가 있는 점 집합을 생성하는 경향을 보였다. 그러나 일부 카테고리에 대해선 이와 반대되는 결과를 보이기도 하였다. Laptop 카테고리의 경우엔 AVC QP 47, JPEG Quality 10과 같이 강한 압축이 적용된 입력에서 Chamfer Distance가 오히려 감소했다. 이는 VGGT를 포함한 3D 장면을 추론하는 모델이 유리나 거울같이 빛이 반사되는 입력이 들어왔을 때 반사된 영역에 대한 깊이를 정확하게 예측하지 못하는 한계[9,10] 때문으로 해석된다. 그림 5에서 보이듯 Laptop 카테고리는 노트북 화면이나 책상에 물체가 반사되는 입력이 존재하여 저품질 포인트 맵을 추론한 것이 확인되었다. GT가 추론한 포인트 맵의

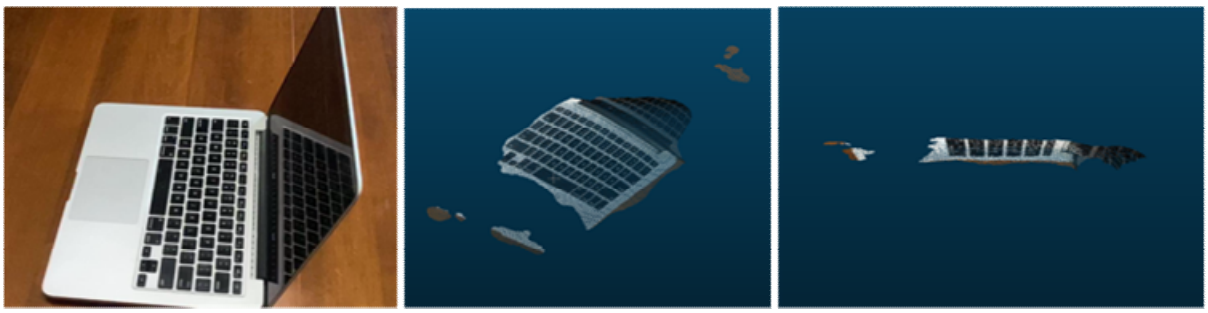

그림 5. Laptop 이미지와 이에 대한 포인트 맵 추론 결과 (순서대로 입력 이미지, 포인트 맵 상단 시점, 옆면 시점)
Fig. 5. Input images of the Laptop category and the corresponding point maps inferred by VGGT (from left to right: input image, top-view of the point map, and side-view of the point map)

포인트 수를 확인한 결과, 압축 강도에 따라 변화가 나타났다. 압축 강도가 높아질수록 추론된 포인트 수가 감소하는 경향을 보였다. 이는 부분적인 압축 아티팩트로 인해 특정 픽셀에 대한 깊이 예측이 어려워졌기 때문으로 보인다.

4. 카메라 포즈 추정

VGGT가 추론한 카메라 포즈의 정확도를 평가하기 위해 Area Under the curve lower than a fixed threshold (AUC@θ) 평가지표를 활용하였으며 AUC@30, 15, 5, 3을 사용하였다. 해당 평가에 대한 식은 아래의 식으로 나타낼 수 있다.

$$A(\theta) = \frac{1}{N}\sum_{i=1}^{N} 1(e_i \le \theta),\ AUC@\theta = \frac{1}{\theta_{max}}\int_0^{\theta} A(\theta)\, d\theta. \quad (2)$$

예측된 카메라 포즈가 GT 포즈와의 오차(e_i)가 θ 이하일 때 정답으로 간주되는 비율을 의미한다. θ는 평가 기준의 엄격함을 의미하고 작을수록 엄격한 기준을 적용한다. 또한 AUC 값이 클수록 카메라 포즈 예측이 기준을 통과한 비율이 높음을 의미한다. 압축 강도에 따른 카메라 포즈 예측 AUC는 표 5, 6을 통하여 보고한다.

표 5, 그리고 6에서 확인할 수 있듯이 압축 영상의 경우, 압축 강도가 커질수록 카메라 포즈 예측 성능이 전반적으로 저하되는 경향을 보였다. 반면, 표 7에 나타낸 평균 PSNR 저하가 AVC보다 JPEG에서 더 컸음에도 불구하고 JPEG 코덱에서는 AUC@30 및 AUC@15 기준에서 압축

표 5. VGGT의 입력을 JPEG 압축된 이미지로 사용하였을 때의 압축 강도에 따른 AUC 평가의 변화와 평균. 빨간 글씨는 하락이 가장 작은 구간이며, 파란 글씨는 하락이 가장 큰 구간을 의미
Table 5. The mean of AUC when JPEG-compressed images are used as input to VGGT. Red text indicates the QF with the highest mean of AUC, while blue text indicates the QF with the lowest mean of AUC.

Methods		AUC@30 (↑)	AUC@15 (↑)	AUC@5 (↑)	AUC@3 (↑)
RAW		0.9745	0.8153	0.7327	0.6624
JPEG	Q 70	0.9749	0.8155	0.7314	0.6603
	Q 50	0.9756	0.8173	0.7365	0.6656
	Q 30	0.9732	0.8135	0.7257	0.6540
	Q 20	0.9716	0.8131	0.7168	0.6402
	Q 10	0.9679	0.8078	0.7054	0.6212

표 6. VGGT의 입력을 AVC 압축된 이미지로 사용하였을 때의 압축 강도에 따른 AUC 평가의 변화와 평균. 빨간 글씨는 하락이 가장 작은 구간이며, 파란 글씨는 하락이 가장 큰 구간을 의미
Table 6. The mean of AUC when AVC-compressed images are used as input to VGGT. Red text indicates the QP with the highest mean of AUC, while blue text indicates the QP with the lowest mean of AUC.

Methods		AUC@30 (↑)	AUC@15 (↑)	AUC@5 (↑)	AUC@3 (↑)
RAW		0.9745	0.8153	0.7327	0.6624
AVC	QP 27	0.9604	0.7941	0.6806	0.6032
	QP 32	0.9576	0.7920	0.6698	0.5862
	QP 37	0.9509	0.7788	0.6368	0.5386
	QP 42	0.9405	0.7655	0.6076	0.5111
	QP 47	0.9210	0.7312	0.5441	0.4370

표 7. 입력 이미지의 압축 강도에 따른 PSNR 평균
Table 7. Average PSNR as a function of compression level of input image

Codec	Compression Level	PSNR ↑
AVC	QP 27	38.49
	QP 32	36.35
	QP 37	34.04
	QP 42	31.56
	QP 47	28.68
JPEG	Quality 70	47.33
	Quality 50	40.46
	Quality 30	38.50
	Quality 20	35.95
	Quality 10	31.59

강도의 변화가 예측 성능에 유의미한 영향을 주지 않았다. 즉, JPEG 코덱에서는 압축이 강해져도 대략적인 카메라 포즈를 비교적 정확히 예측했다고 볼 수 있다. 이는 JPEG의 세부적인 디테일을 저하시키면서 압축하는 방식이 이미지를 전역적으로 해석하는 Vision Transformer[12,13] 기반의 VGGT에서는 큰 영향을 주지 않은 것으로 추정된다. 한편, Laptop 카테고리의 경우 임계 값(θ)과 압축 강도에 관계 없이 카메라 포즈 추정이 안정적으로 이루어지지 않았다. 이는 앞서 언급한 반사에 의한 시각적 왜곡이 모델의 카메라 포즈 추론 과정에 영향을 준 것으로 보인다. 전반적인 연구의 결과, AVC에서 QP 증가에 따라 성능 저하가 누적되며, 특히 높은 QP 구간에서 포인트 수 감소와 AUC 하락이 더 두드러지는 경향을 보였다 (그림 6).

5. 노이즈 제거 모델 활용 연구 결과

본 절에서는 압축 영상 입력 환경에서의 3D 비전 모델성능 변화를 분석하기 위해, CO3Dv2 데이터셋의 모든 카테고리에 대해 1개씩의 대표 씬을 HEVC, VVC를 각각 QP 27, 32, 37을 적용하여 압축 테스트 데이터셋을 구성하였다. 이후, 대표적 멀티 프레임 기반 노이즈 제거 모델인 STDF, RFDA, STFF를 활용하여 부호화 영상의 압축 아티팩트를 감소시킨 뒤, 해당 프레임들을 VGGT 모델의 입력으로 사용하여 카메라 포즈 예측 결과인 AUC@30, AUC@15, AUC@5, AUC@3, 그리고 포인트 맵의 예측 결

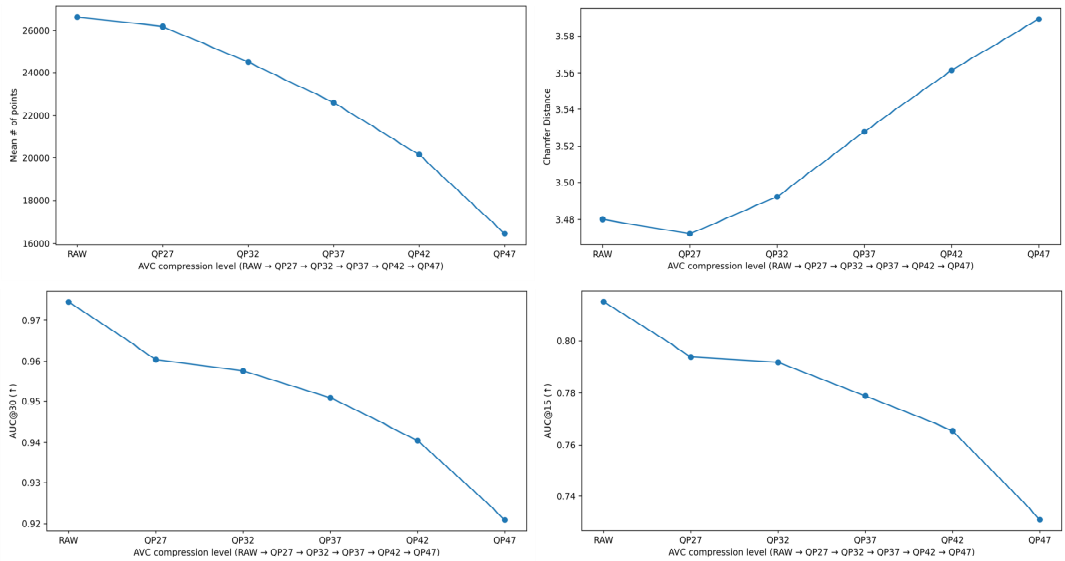

그림 6. AVC 압축을 이용한 VGGT의 추론 성능 변화의 추이. 대표 9개 카테고리에 대해 RAW 및 AVC (QP 27, 32, 37, 42 ,47) 입력을 사용하여 평균 포인트 개수, Chamfer distance, AUC@30, 그리고 AUC@15의 변화를 나타낸다. 표의 내용과 같이 강한 압축이 수반될 경우, (QP 42 ~ 47구간) 성능 하락의 폭이 더 커지는 것을 확인할 수 있다.

Fig. 6. Trends in VGGT inference performance under AVC compression. For nine representative categories, changes in the mean number of points, Chamfer Distance, AUC@30, and AUC@15 are shown using RAW and AVC inputs (QP 27, 32, 37, 42, and 47). Stronger compression (QP 42-47) leads to larger performance degradation.

표 8. HEVC 코덱의 QP 27 압축 프레임 및 노이즈 제거 모델에 대한 카메라 포즈 예측, 포인트 맵 예측 결과. 빨간색은 RAW에 대한 VGGT 추론 결과를 제외한 최고 성능을 의미

Table 8. Camera Pose estimation and point map prediction results for HEVC-compressed frames at QP 27 and for the denoising model. Red indicates the best performance excluding VGGT inference result on RAW input.

Method	Camera pose estimation				Point map estimation		
	AUC@30 (↑)	AUC@15 (↑)	AUC@5 (↑)	AUC@3 (↑)	Accuracy (↓)	Completeness (↓)	Chamfer Distance (↓)
RAW	0.8878	0.8232	0.6713	0.5742	0.1674	1.2470	0.7072
Compressed (HEVC, QP 27)	0.8391	0.7380	0.5089	0.3824	0.1686	1.2473	0.7080
STDF	0.8358	0.7497	0.5388	0.4155	0.1627	1.2480	0.7053
RFDA	0.8258	0.7187	0.4792	0.3547	0.1689	1.2380	0.7035
STFF	0.8252	0.7148	0.4710	0.3383	0.1700	1.2376	0.7038

표 9. HEVC 코덱의 QP 32 압축 프레임 및 노이즈 제거 모델에 대한 카메라 포즈 예측, 포인트 맵 예측 결과. 빨간색은 RAW에 대한 VGGT 추론 결과를 제외한 최고 성능을 의미

Table 9. Camera Pose estimation and point map prediction results for HEVC-compressed frames at QP 32 and for the denoising model. Red indicates the best performance excluding VGGT inference result on RAW input.

Method	Camera pose estimation				Point map estimation		
	AUC@30 (↑)	AUC@15 (↑)	AUC@5 (↑)	AUC@3 (↑)	Accuracy (↓)	Completeness (↓)	Chamfer Distance (↓)
RAW	0.8878	0.8232	0.6713	0.5742	0.1674	1.2470	0.7072
Compressed (HEVC, QP 32)	0.8725	0.7853	0.5808	0.4587	0.1604	1.2789	0.7197
STDF	0.8669	0.7809	0.5851	0.4704	0.1559	1.2786	0.7172
RFDA	0.8640	0.7756	0.5773	0.4618	0.1609	1.2770	0.7189
STFF	0.8628	0.7722	0.5696	0.4536	0.1608	1.2773	0.7190

표 10. HEVC 코덱의 QP 37 압축 프레임 및 노이즈 제거 모델에 대한 카메라 포즈 예측, 포인트 맵 예측 결과. 빨간색은 RAW에 대한 VGGT 추론 결과를 제외한 최고 성능을 의미

Table 10. Camera Pose estimation and point map prediction results for HEVC-compressed frames at QP 37 and for the denoising model. Red indicates the best performance excluding VGGT inference result on RAW input.

Method	Camera pose estimation				Point map estimation		
	AUC@30 (↑)	AUC@15 (↑)	AUC@5 (↑)	AUC@3 (↑)	Accuracy (↓)	Completeness (↓)	Chamfer Distance (↓)
RAW	0.8878	0.8232	0.6713	0.5742	0.1674	1.2470	0.7072
Compressed (HEVC, QP 37)	0.8368	0.7412	0.5160	0.3898	0.1526	1.3131	0.7328
STDF	0.8414	0.7417	0.5086	0.3834	0.1493	1.3050	0.7271
RFDA	0.8506	0.7559	0.5281	0.4060	0.1552	1.3030	0.7291
STFF	0.8430	0.7538	0.5482	0.4322	0.1561	1.3055	0.7308

표 11. VVC 코덱의 QP 27 압축 프레임 및 노이즈 제거 모델에 대한 카메라 포즈 예측, 포인트 맵 예측 결과. 빨간색은 RAW에 대한 VGGT 추론 결과를 제외한 최고 성능을 의미

Table 11. Camera Pose estimation and point map prediction results for VVC-compressed frames at QP 27 and for the denoising model. Red indicates the best performance excluding VGGT inference result on RAW input.

Method	Camera pose estimation				Point map estimation		
	AUC@30 (↑)	AUC@15 (↑)	AUC@5 (↑)	AUC@3 (↑)	Accuracy (↓)	Completeness (↓)	Chamfer Distance (↓)
RAW	0.8878	0.8232	0.6713	0.5742	0.1674	1.2470	0.7072
Compressed (VVC, QP 27)	0.8558	0.7658	0.5461	0.4246	0.1694	1.2468	0.7081
STDF	0.8562	0.7695	0.5673	0.4455	0.1625	1.2445	0.7035
RFDA	0.8434	0.7441	0.5156	0.3933	0.1683	1.2419	0.7051
STFF	0.8396	0.7387	0.5051	0.3785	0.1682	1.2397	0.7039

표 12. VVC 코덱의 QP 32 압축 프레임 및 노이즈 제거 모델에 대한 카메라 포즈 예측, 포인트 맵 예측 결과. 빨간색은 RAW에 대한 VGGT 추론 결과를 제외한 최고 성능을 의미
Table 12. Camera Pose estimation and point map prediction results for VVC-compressed frames at QP 32 and for the denoising model. Red indicates the best performance excluding VGGT inference result on RAW input.

Method	Camera pose estimation				Point map estimation		
	AUC@30 (↑)	AUC@15 (↑)	AUC@5 (↑)	AUC@3 (↑)	Accuracy (↓)	Completeness (↓)	Chamfer Distance (↓)
RAW	0.8878	0.8232	0.6713	0.5742	0.1674	1.2470	0.7072
Compressed (VVC, QP 32)	0.8697	0.7868	0.5940	0.4794	0.1612	1.2795	0.7203
STDF	0.8655	0.7814	0.5921	0.4780	0.1557	1.2810	0.7184
RFDA	0.8669	0.7817	0.5873	0.4748	0.1604	1.2789	0.7196
STFF	0.8624	0.7740	0.5766	0.4579	0.1611	1.2778	0.7195

표 13. VVC 코덱의 QP 37 압축 프레임 및 노이즈 제거 모델에 대한 카메라 포즈 예측, 포인트 맵 예측 결과. 빨간색은 RAW에 대한 VGGT 추론 결과를 제외한 최고 성능을 의미
Table 13. Camera Pose estimation and point map prediction results for VVC-compressed frames at QP 37 and for the denoising model. Red indicates the best performance excluding VGGT inference result on RAW input.

Methods	Camera pose estimation				Point map estimation		
	AUC@30 (↑)	AUC@15 (↑)	AUC@5 (↑)	AUC@3 (↑)	Accuracy (↓)	Completeness (↓)	Chamfer Distance (↓)
RAW	0.8878	0.8232	0.6713	0.5742	0.1674	1.2470	0.7072
Compressed (VVC, QP 37)	0.8440	0.7494	0.5247	0.3984	0.1548	1.3071	0.7309
STDF	0.8481	0.7523	0.5277	0.4060	0.1502	1.3019	0.7261
RFDA	0.8459	0.7577	0.5474	0.4250	0.1560	1.3025	0.7292
STFF	0.8388	0.7519	0.5426	0.4279	0.1572	1.2996	0.7284

과인 accuracy, completeness, chamfer distance에 대한 결과를 표 8-13에서 보고한다. 각각의 QP에 맞게 MFQEv2 데이터셋으로 기 학습된 가중치를 활용하여 노이즈 제거 모델의 추론 결과를 활용하였다.

적용 결과, HEVC, VVC에서도 압축 강도가 증가할수록 예측 성능이 일관되게 저하되는 경향을 보임을 알 수 있다. 노이즈 제거 모델을 적용한 경우, HEVC 압축 데이터에 대해서는 (표 8-10) STDF가 낮은 강도의 QP 설정에서 카메라 포즈 추정과 포인트 맵 정확도 측면에서 가장 안정적인 성능 향상 기여를 보였으며, 높은 압축 강도에서는 STFF 모델이 카메라 포즈 예측에서의 가장 높은 향상도를 달성하였다. VVC 압축 데이터에 대해서는 (표 11-13) 낮은 압축 강도에 대해서는 STDF가 비교적 카메라 포즈 및 포인트 맵 예측 성능에서 높은 회복세를 이루는 결과를 확인할 수 있었다.

V. 결 론

본 연구에서는 CO3Dv2 기반 데이터셋에 JPEG과 AVC 코덱을 적용하여 다양한 압축 강도의 데이터셋을 구성하고, VGGT의 성능 변화를 분석하였다. 정량적으로, 압축 강도가 높아질수록 추정된 포인트 수 감소, 추론된 포인트 맵 유사도 저하를 확인할 수 있었다. 카메라 포즈 예측은 AVC는 모든 임계 값에서 성능 저하가 나타났다. 특히, AVC 코덱은 QP 42 이상, JPEG은 Quality 20 이하부터 성능 저하가 급격히 발생하였다. 정성적 평가에서도 예측 성능 저하를 확인할 수 있었다. 이로써, 압축 유무를 확인하지 않고 기존 3D 비전 모델의 입력으로 활용할 경우 성능의 하락을 피할 수 없음을 확인할 수 있었다. 본 결과는 강한 압축 환경에서도 강건한 3D 재구성 모델 설계 필요성을 시사한다. 노이즈 제거 모델을 거친 이미지의 VGGT 추론 결과, 현존

하는 멀티 프레임 노이즈 제거 모델을 활용할 경우 카메라 포즈 예측 및 포인트 맵 예측에 대한 보완이 가능함을 알 수 있었다. 향후 연구에서는 압축된 영상에 앞서 언급하였으나 활용되지 않은 노이즈 제거 모델[14-19]을 적용해 데이터셋을 확장하여 구축하고, 확장된 데이터셋을 활용하여 다양한 3D 비전 모델의 성능을 분석할 예정이다.

참 고 문 헌 (References)

[1] J. Wang, M. Chen, N. Karaev, A.Vedaldi, C. Rupprecht, and D. Novotny, “VGGT: Visual Geometry Grounded Transformer,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, USA, pp.5294-5306, June 2025.
doi: https://doi.org/10.1109/CVPR52734.2025.00499

[2] M. Ehrlich, L. Davis, S.-N. Lim, and A. Shrivastava, “Analyzing and Mitigating JPEG Compression Defects in Deep Learning,” *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, Canada, pp.2357-2367, Oct. 2021.
doi: https://doi.org/10.1109/ICCVW54120.2021.00267

[3] M. O’Byrne, M. Sugrue, Vibhoothi, and A. Kokaram, “Impact of Video Compression on the Performance of Object Detection Systems for Surveillance Applications,” *Proceedings of 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Madrid, Spain, pp.1-8, Dec. 2022.
doi: https://doi.org/10.1109/AVSS56176.2022.9959476

[4] J. Reizenstein, R. Shapovalov, P. Henzler, L. Sbordone, P. Labatut, and D. Novotny, “Common Objects in 3D: Large-Scale Learning and Evaluation of Real-Life 3D Category Reconstruction,” *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, Canada, pp.10901-10911, Oct. 2021.
doi: https://doi.org/10.1109/ICCV48922.2021.01072

[5] G. K. Wallace, “The JPEG still picture compression standard,” *IEEE Transactions on Consumer Electronics,* Vol.38, No.1, pp.xviii - xxxiv, Feb. 1992.
doi: https://doi.org/10.1109/30.125072

[6] ISO/IEC, *JPEG 2000 Image Coding System — Part 1: Core Coding System,* ISO/IEC 15444-1 (also published as ITU-T Rec. T.800), 2nd ed., 2004.

[7] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC Video Coding Standard,” *IEEE Transactions on Circuits and Systems for Video Technology,* Vol.13, No.7, pp.560 - 576, Jul. 2003.
doi: https://doi.org/10.1109/TCSVT.2003.815165

[8] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand. “Overview of the high efficiency video coding (HEVC) standard,” *IEEE Transactions on Circuits and Systems for Video Technology,* Vol.22, No.12, pp.1649 - 1668, Dec. 2012.
doi: https://doi.org/10.1109/TCSVT.2012.2221191

[9] H. Fan, H. Su, and L. J. Guibas, “A Point Set Generation Network for 3D Object Reconstruction from a Single Image,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Honolulu, USA, pp.605 - 613, Jul. 2017.
doi: https://doi.org/10.1109/CVPR.2017.264

[10] J.Wu, Z. Wang, I. Laina, and V. A. Prisacariu, “Reflect3r: Single-View 3D Stereo Reconstruction Aided by Mirror Reflections,” *arXiv preprint arXiv:2509.20607,* 2025.
doi: https://doi.org/10.48550/arXiv.2509.20607

[11] Y. Chen, H. Xu, C. Zheng, B. Zhuang, M. Pollefeys, A. Geiger, T.-J. Cham, and J. Cai, “MVSplat: Efficient 3D Gaussian Splatting from Sparse Multi-View Images,” *Proceedings of the European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science,* Vol.15079, Milan, Italy, pp.370-386, 2024.
doi: https://doi.org/10.1007/978-3-031-72664-4_21

[12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16×16 words: Transformers for image recognition at scale,” *Proceedings of the International Conference on Learning Representations(ICLR),* 2021.
doi: https://doi.org/10.48550/arXiv.2010.11929

[13] S. Paul, and P.-Y. Chen, “Vision Transformers are Robust Learners,” *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI),* Vol.36, No.2, pp.2071 - 2081, 2022.
doi: https://doi.org/10.1609/aaai.v36i2.20103

[14] W. K. Han, S. Im, J. Kim, and K. H. Jin, “JDEC: JPEG Decoding via Enhance Continuous Cosine Coefficients,” *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* Seattle, USA, pp. 2784-2793, June 2024.
doi: https://doi.org/10.1109/cvpr52733.2024.00269

[15] B. Li, X. Li, Y. Lu, R. Feng, M. Guo, S. Zhao, L. Zhang, and Z. Chen, “PromptCIR: Blind Compressed Image Restoration with Prompt Learning,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* Seattle, USA, pp.6442-6452, 2024.
doi: https://doi.org/10.1109/CVPRW63382.2024.00645

[16] J. Guo, Z. Chen, W. Li, Y. Guo, and Y. Zhang, “Compression-Aware One-Step Diffusion Model for JPEG Artifact Removal,” *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV),* Honolulu, Hawaii, pp.14930-14939, Oct. 2025.
doi: https://doi.org/10.48550/arXiv.2502.09873

[17] T. Zhang, X. Luo, L. Li, and D. Liu, “StableCodec: Taming One-Step Diffusion for Extreme Image Compression,” *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV),* Honolulu, Hawaii, pp.17379-17389, Oct. 2025.
doi: https://doi.org/10.48550/arXiv.2506.21977

[18] J. Guo, Y. Ji, Z. Chen, K. Liu, M. Liu, W. Rao, W. Li, Y. Guo, and Y. Zhang, “OSCAR: One-Step Diffusion Codec Across Multiple Bitrates,” *Proceedings of the Neural Information Processing Systems (NeurIPS),* San Diego, USA, Dec. 2025.
doi: https://doi.org/10.48550/arXiv.2505.16091

[19] R. Wu, L. Sun, Z. Ma, and L. Zhang, “One-Step Effective Diffusion Network for Real-World Image Super-Resolution,” *Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS),* Vancouver, Canada, pp.92529-92553, Dec. 2024.

doi: https://doi.org/10.52202/079017-2938
[20] M.-Y. Shen, and C.-C.J. Kuo, “Review of Postprocessing Techniques for Compression Artifact Removal,” *Journal of Visual Communication and Image Representation,* Vol.9, No.1, pp.2-14, Mar. 1998.
doi: https://doi.org/10.1006/jcvi.1997.0378
[21] O. Ozyesil, V. Voroninski, R. Basri, and A. Singer, “A Survey of Structure from Motion,” *Acta Numerica,* Vol.26, pp.305-364, May, 2017.
doi: https://doi.org/10.1017/S096249291700006X
[22] S. Wang, V. Leroy, Y. Cabon, B. Chidlovskii and Jerome Revaud, “DUSt3R: Geometric 3D Vision Made Easy,” *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, pp 20697-20709, June 2024.
doi: https://doi.org/10.1109/CVPR52733.2024.01956
[23] C. Dong, Y.Deng, C. C. Loy, and X. Tang, “Compression Artifacts Reduction by a Deep Convolutional Network,” *Proceedings of the International Conference on Computer Vision (ICCV),* Santiago, Chile, pp.576-584, Dec. 2015.
doi: https://doi.org/10.1109/ICCV.2015.73
[24] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising,” *IEEE Transaction on Image Processing,* Vol.26, No.7, pp.3142-3155, July 2017.
doi: https://doi.org/10.1109/TIP.2017.2662206
[25] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. V. Gool, and R. Timofte, “Plug-and-Play Image Restoration With Deep Denoiser Prior,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol.44, No.10, pp.6360-6376, Oct. 2022.
doi: https://doi.org/10.1109/TPAMI.2021.3088914
[26] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, “Overview of the versatile video coding (VVC) standard and its applications,” *IEEE Transactions on Circuits and Systems for Video Technology,* Vol.31, No.10, pp.3736-3764, Oct. 2021.
doi: https://doi.org/10.1109/TCSVT.2021.3101953
[27] R. Yang, M. Xu, Z. Wang, and T. Li, “Multi-frame quality enhancement for compressed video,” *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR),* Salt Lake city, USA, pp. 6664-6673, June 2018.
doi: https://doi.org/10.1109/CVPR.2018.00697
[28] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, “Deformable convolutional networks,” *Proceedings of the IEEE international conference on computer vision (ICCV),* Venice, Italy, pp.764-773, Oct. 2017.
doi: https://doi.org/10.1109/ICCV.2017.89
[29] J. Deng, L. Wang, S. Pu, and C. Zhuo, “Spatio-temporal deformable convolution for compressed video quality enhancement,” *Proceedings of the AAAI conference on artificial intelligence,* Vol.34, No.07, pp.10696-10703, April 2020.
doi: https://doi.org/10.1609/aaai.v34i07.6697
[30] M. Zhao, Y. Xu, and S. Zhou, “Recursive fusion and deformable spatiotemporal attention for video compression artifact reduction,” *Proceedings of the 29th ACM international conference on multimedia,* pp.5646-5654, Oct. 2021.
doi: https://doi.org/10.1145/3474085.3475710
[31] M. Wang, Y. Liao, W. Chen, L. Lin, and T. Zhao, “STFF: Spatio-Temporal and Frequency Fusion for Video Compression Artifact Removal,” *IEEE Transactions on Broadcasting,* Vol.71, No.2, pp.542-554, June 2025.
doi: https://doi.org/10.1109/TBC.2025.3550018

저 자 소 개

김 제 희

- 2021년 3월 ~ 현재 : 경북대학교 컴퓨터학부 학석사연계과정
- ORCID : https://orcid.org/0009-0008-3778-241X
- 주관심분야 : 컴퓨터 비전, 딥 러닝, 이미지 프로세싱

김 동 휘

- 2021년 2월 : 대전대학교 전자정보통신공학 학사
- 2023년 2월 : 경북대학교 컴퓨터학부 석사
- 2023년 3월 ~ 현재 : 경북대학교 컴퓨터학부 박사과정
- ORCID : https://orcid.org/0000-0002-5188-8834
- 주관심분야 : 컴퓨터 비전, 영상처리, 3차원 재구성, 생성 모델

저 자 소 개

문 채 원

- 2025년 2월 : 경북대학교 컴퓨터학부 학사
- 2025년 3월 ~ 현재 : 경북대학교 컴퓨터학부 석사과정
- ORCID : https://orcid.org/0009-0001-3492-3626
- 주관심분야 : 3D 객체 탐지, 3D 장면 이해, 컴퓨터 비전

이 윤 호

- 2022년 3월 ~ 현재 : 경북대학교 컴퓨터학부 심화컴퓨터전공 학사과정
- ORCID : https://orcid.org/0009-0004-8595-1493
- 주관심분야 : Computer VIsion, Deep Learning, Video Processing, 3D Point Cloud

김 은 지

- 2025년 8월 : 경북대학교 컴퓨터학부 학사
- 2025년 9월 ~ 현재 : 경북대학교 컴퓨터학부 석사과정
- ORCID : https://orcid.org/0009-0008-2319-0325
- 주관심분야 : Video Compression, LLM, Model Compression, 3D reconstruction

정 진 우

- 2004년 2월 : 연세대학교 전기전자공학부 공학사
- 2011년 8월 : 연세대학교 전기전자공학부 공학박사
- 2011년 ~ 2015년 : 삼성전자 VD사업부 책임
- 2016년 ~ 현재 : 한국전자기술연구원 책임
- ORCID : https://orcid.org/0000-0003-0528-8755
- 주관심분야 : 딥러닝 경량화, 컴퓨터 비젼, 영상처리

박 상 효

- 2011년 2월 : 한양대학교 컴퓨터전공 학사
- 2017년 8월 : 한양대학교 컴퓨터·소프트웨어학과 박사
- 2017년 5월 ~ 2018년 2월 : 전자부품연구원 지능형영상처리센터 Post-doc
- 2018년 3월 ~ 2018년 12월 : 연세대학교 바른ICT연구소 연구원
- 2019년 2월 ~ 2020년 1월 : 이화여자대학교 전자전기공학과 박사후연구원
- 2020년 3월 ~ 현재 : 경북대학교 컴퓨터학부 부교수
- ORCID : https://orcid.org/0000-0002-7282-7686
- 주관심분야 : VVC, 모델 압축/경량화, 생성모델, 3차원 비전, 영상처리