김시원
AI Researcher

+82 10-2007-4229
kimsiw42@ajou.ac.kr

🏠 kimww42.github.io

Fields of Interest: Image Classification, Diffusion, Image Restoration, 3D Vision, etc.

Education
아주대학교
주전공: 소프트웨어학
마이크로전공: 인공지능, 의료인공지능
(2019.03 ~ 2025.08)

Work Experience
- KAIST DAVIAN Lab 학부연구생(現)
- Ajou University CI Lab 학부연구생(前)
- Seoul National University Bundang Hospital, AI Center Research Intern(前)
- Insilicogen, Inc. (BioInformatics Intern, Backend Freelance) (前)

Publications
- ForestSplats: Deformable transient field for Gaussian Splatting in the Wild / arXiv 2025
- Efficient Deep Learning Approaches for Processing Ultra-Widefield Retinal Imaging / MICCAIW 2024
- Diagnosis for Tubal Patency with Contrast Medium in Hysterosalpingography Images Using Asymmetric Contrastive Learning / CKAIA 2024
- A Real-Time Eye Gaze Tracking Based Digital Mouse / IMIS 2024
- 근육병 환자를 위한 단일 카메라 기반 시선 추적 연구 / KIISS 2024

Honors & Awards

Challenge

2024	**3rd Award**, Ajou Softcon
2023	**8th Award**, Medical AI Idea Challenge(MOHW, KHIDI)
2023	**2nd Award**, SNUBH COVID-19 Datathon
2023	**1st Award**, K-ium Medical AI Competition(PNUH, KNUH, CNUH)
2021	**1st Award**, R.O.K 11Div Internal Security Threat Detection Contest
2018	**Finalist**, 2018 KOI(Korea Olympiad in Informatics) National Final
2018	**Encouragement Award**, 2018 KOI(Korea Olympiad in Informatics) Busan Regional Preliminary
2017	**Encouragement Award**, 2017 KOI(Korea Olympiad in Informatics) Busan Regional Preliminary

Scholarship

2024	**Da-San Scholarship**, students showed excellent grades in the department.
2023	**Won-Cheon Scholarship**, students showed excellent grades in the department.
2019	**SW Competency Scholarship**, students showed excellent SW competency.
2019	**SW Excellent Talent Scholarship**, students entered the school through SW Specialist Screening

Research / Projects

(1) Long-CLIP without Hallucination and Concept Association Bias

#Vision Language Model #Hallucination #CLIP

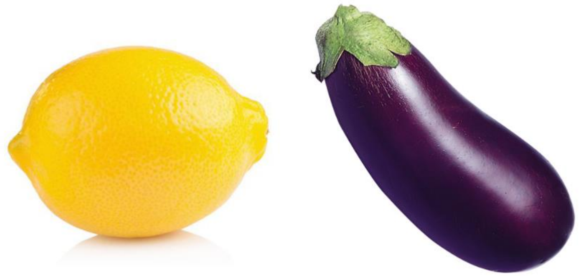
지도교수: 아주대학교 소프트웨어학과 조현석 교수
수행기관: 학부 AI집중교육2 과목 프로젝트로 수행
수행기간: 2024.11 ~ 2024.12
프로젝트 내 수행 역할: 프로젝트 리딩, 문제 제시, 방법론 제시, 실험 및 결과 정리

해결하고자 하는 문제
Long-CLIP[1] 모델에서 발생하는 Concept Association Bias 문제 및 Hallucination 문제를 개선하고자 함

해결 방법
1) 긴 문장을 기반으로 Hard Negative Text를 샘플링하기 위한 방법론 사용 -> 긴 문장을 랜덤한 순서로 섞고, 문장 내의 단어 순서들을 랜덤하게 샘플링
2) Positive Text와 Hard Negative Text간의 유사도를 대조학습 Loss function에 반영한 Text Distance Margin Triplet Loss 제안

실험 결과 및 결론
긴 문장을 기반으로 Hard Negative를 샘플링하기 위한 방법과 유사도 기반 Loss Function을 새롭게 제안
Long-CLIP 모델 대비 Concept Association Bias 와 Hallucination 문제가 개선된 것을 확인
일반 CLIP 모델에 Hallucination 개선을 위한 선행 연구 대비로는 부족한 성능을 보임

Supplementary -> Page 08



CLIP: "In this picture, the color of the lemon is purple."

CLIP 모델의 Concept Association Bias (Hallucination) 문제의 예시 [2]

[1] Zhang et al., Long-CLIP: Unlocking the Long-Text Capability of CLIP (ECCV 2024)
[2] Yamada et al., When are Lemos Purple? The Concept Association Bias of Vision-Language Models (EMNLP 2023)

(2) Text based Image Restoration for Composite Degradation

#Image Restoration #Image-Text Alignment #Diffusion

지도교수: 아주대학교 소프트웨어학과 유종빈 교수
수행기관: 학부 AI집중교육1 과목 프로젝트로 수행 -> 성능 고도화 연구 진행 중
수행기간: 2024.09 ~ Present
프로젝트 내 수행 역할: 프로젝트 리딩, 문제 제시, 방법론 제시, 실험 및 결과 정리

해결하고자 하는 문제
여러 손상이 복합된 이미지를 Text 기반의 지시를 통해 한 번에 복원하고자 함

해결 방법
1) AirNet[1] 아키텍쳐를 기반으로 Text 입력에 대해 추가 정보를 받을 수 있는 Text Image Combined SFT Layer 제안
2) Text 지시를 Projection하여 유사한 손상 종류 프롬프트끼리 묶는 Projection Layer 추가

실험 결과 및 결론
Text 프롬프트의 지시에 맞는 손상이 잘 복원되는 것을 확인
적은 양의 데이터셋에서는 SOTA 성능을 달성, 그러나 대량의 데이터셋에서는 SOTA 달성 실패
사용한 기반 아키텍쳐의 한계로 추정하여, **Conditional Diffusion 모델 기반의 성능 고도화 진행 중**

Supplementary -> Page 12
Repository -> github.com/kimww42/ICDR


Haze Image


"**Erase the fog** to enhance the scene's clarity."

[1] Li et al., All-In-One Image Restoration for Unknown Corruption (CVPR 2022)

(3) 3D Mesh Stylization with Multiple Prompting

#3D Vision #Mesh Stylization #CLIP

지도교수: 아주대학교 소프트웨어학과 조현석 교수
수행기관: 학부 자기주도연구1 과목 프로젝트로 수행
수행기간: 2024.07 ~ 2024.08
프로젝트 내 수행 역할: 단일 수행 연구

해결하고자 하는 문제
CLIP 기반의 3D Mesh Stylization 선행연구가 여러 Text 지시사항을 복합해서 반영하지 못하는 문제를 개선하기 위함

해결 방법
1) 입력된 Text Prompt를 지시사항 유형 별로 분리하여 Loss 계산에 사용

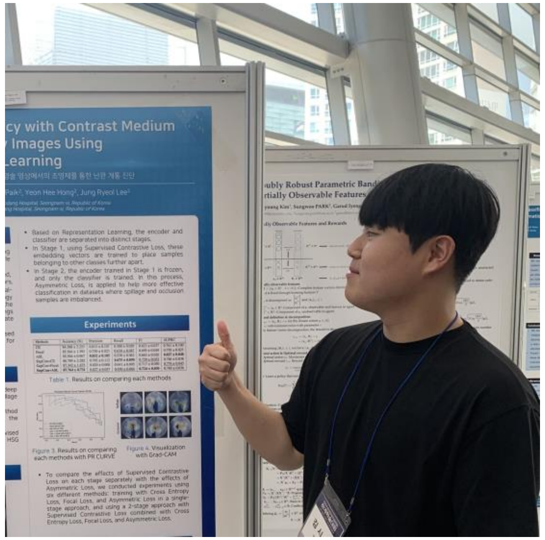
실험 결과 및 결론
User Study 결과, 선행 연구 대비 여러 지시사항을 잘 반영하고 더 높은 퀄리티를 달성할 수 있었음
Metric 기반의 평가에서는 선행 연구 대비 부족한 성능을 보임 (성능을 제대로 평가할 수 있는 평가 방법의 부재)

Supplementary -> Page 18
Repository -> github.com/kimww42/3DS-MP


Majestic Dragon with Fire Wings. Ironman with Green suit. Fluffy squirrel with a =striped tail=.

(4) Diagnosis for Tubal Patency Using Contrastive Learning

#Image Classification #Contrastive Learning #Imbalanced Classification



지도교수: 분당서울대학교병원 산부인과 이정렬 교수 / 의료인공지능센터 김명주 선임연구원
수행기관: 분당서울대학교병원 의료인공지능센터의 산부인과 과제로 수행
수행기간: 2024.03 ~ Present
프로젝트 내 수행 역할: 데이터 전처리, 방법론 확립, 실험 및 결과 정리, 논문 작성

해결하고자 하는 문제
자궁난관조영술 X-ray 이미지 기반으로 자궁 내 질병 및 난관 상태 진단
선행연구 없음, 부족하고 불균형한 데이터 양

해결 방법
1) Supervised Contrastive Learning 을 사용하여 적은 데이터 양에서도 분별력을 더욱 높임
2) Asymmetric Loss 를 사용하여 불균형한 Class에서의 분별력을 더욱 높임

실험 결과 및 결론
선행 연구가 없었지만, 딥러닝 기반의 방법론으로 자궁난관조영술 진단의 가능성을 확인
Supervised Contrastive Learning + Asymmetric Loss의 조합이 불균형한 메디컬 이미지 분류에서 효과적임을 확인

실험 결과를 바탕으로 2024 한국인공지능학회 하계학술대회 발표 -> 추가 데이터 확보 후 저널 게재를 위한 추가 연구 진행 중

Supplementary -> Page 22

김시원
AI Researcher

+82 10-2007-4229
kimsiw42@ajou.ac.kr

🏠 kimww42.github.io

감사합니다

언제나 새로운 시각으로 끊임없이 노력하는 AI 연구자가 되겠습니다

Supplementary #1

Long-CLIP without Hallucination and Concept Association Bias

Long-CLIP without Hallucination and Concept Association Bias

Problem: LongCLIP 연구의 Fig. 2(b)에 제시된 Concept Association Bias, Hallucination 문제가 실제로는 크게 개선되지 않음을 확인

Model	ARO		VALSE							
	Relation	Attribute	Existence	Plurality	Counting	Sp.rel.	Actions	Coreference	Foil-it	Avg.
CLIP	59.3	62.9	68.7	57.1	61.0	65.4	74.0	52.5	89.8	65.3
BLIP	59.0	**88.0**	**86.3**	73.2	**68.1**	71.5	69.1	51.0	93.8	70.0
BLIP2	41.2	71.3	41.2	71.3	55.5	71.5	66.0	50.3	**95.9**	65.4
LongCLIP	59.7 + 0.4	63.6 + 0.7	68.7	66.0	65.8	62.2	76.9	56.7	91.0	67.8 + 2.5
Hard Negative based method										
NegCLIP	80.2	70.5	76.8	71.7	65.0	72.9	83.0	55.2	91.9	71.6
CE-CLIP	**83.6** + 24.3	77.1 + 14.3	84.5	**79.2**	67.8	**76.4**	**86.3**	**67.2**	94.7	**76.7** +11.4

Results (%) on ARO and VALSE benchmark. The best scores for each section are highlighted in bold. +, - scores compared to CLIP.

Model	REPLACE				SWAP			ADD		
	Object	Attribute	Relation	Avg.	Object	Attribute	Avg.	Object	Attribute	Avg.
CLIP	90.9	80.0	69.2	80.2	61.4	64.0	62.7	77.2	68.2	72.7
BLIP2	-	-	-	86.7	-	-	69.8	-	-	86.5
LongCLIP	**94.3**	85.0	74.0	84.4 + 4.2	66.9	70.2	68.5 + 5.8	83.9	75.0	79.5 + 6.8
Hard Negative based method										
NegCLIP	92.7	85.9	76.5	85.0	75.2	75.4	75.3	88.8	82.8	85.8
CE-CLIP	93.8	**90.8**	**83.2**	**89.3** + 9.1	**76.8**	**79.3**	**78.0** + 15.3	**93.8**	**94.9**	**94.4** + 21.7

Results (%) on Sugar-Crepe benchmark. The best scores for each section are highlighted in bold. +, - scores compared to CLIP.

기존의 CLIP과 비교하였을 때
약 5%정도의 성능향상만이 존재

Hallucination을 제거하기 위한
직접적인 방법론인 **Hard Negative based 방법론보다는 낮은 성능을 보임**

Long-CLIP without Hallucination and Concept Association Bias

Methods

1) Long Text의 특성에 맞는 Hard Negative Text Augmentation 방법
2) Text Distance Margin Triplet Loss

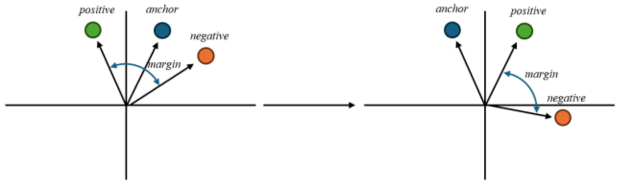
"The lemon on the **left** is yellow and the eggplant on the **right** is purple. The bright yellow of the lemon contrasts vividly with the deep purple of the eggplant. While the lemon's smooth, shiny surface reflects light, the eggplant has a slightly matte and textured finish. The lemon exudes a tangy, citrus aroma, whereas the eggplant has a more subtle, earthy scent. Both are fresh and vibrant, perfectly capturing the essence of their natural colors. Together, they create a visually appealing balance of warm and cool tones on the table."

Sampling Long Caption's Sequence →

"The bright yellow of the lemon contrasts vividly with the deep purple of the eggplant. The lemon on the **left** is purple and the eggplant on the **right** is yellow. While the lemon's smooth, shiny surface reflects light, the eggplant has a slightly matte and textured finish. Both are fresh and vibrant, perfectly capturing the essence of their natural colors. The lemon exudes a tangy, citrus aroma, whereas the eggplant has a more subtle, earthy scent. Together, they create a visually appealing balance of warm and cool tones on the table."

x^a : *anchor (image)*
x^p : *positive (caption)*
x^n : *negative (caption)*

$$\mathcal{L}_{TDMtriplet} = \max(0, \sum_{i}^{N} [\| f(x_i^a) - f(x_i^p) \|_2^2 - \| f(x_i^a) - f(x_i^n) \|_2^2 + \| f(x_i^p) - f(x_i^n) \|_2^2])$$



Method 1
개별 문장 내에서 단어 순서를 바꾸는 Augmentation와
전체 단락 내에서 문장 순서를 바꾸는 Augmentation를 동시에 적용함

증강한 Hard Negative Text를 Contrastive Learning을 통해 밀어내도록 학습

Method 2
Negative Text를 단순히 0으로 밀어내는 것이 아닌, anchor Text와의 의미적 유사도를
반영하여 해당 유사도만큼만 멀어지도록 학습

유사한 의미의 Long Text가 부적절하게 너무 멀어지는 것을 방지

Long-CLIP without Hallucination and Concept Association Bias

Conclusion

1) Long Text의 특성에 맞는 증강 방법론과 Loss 제안
2) 기존 LongCLIP에 비해 Hallucination 문제가 조금 개선되었으나, 다른 hard negative 방법론에 비해 부족한 성능을 보임

Model	ARO		VALSE							
	Relation	Attribute	Existence	Plurality	Counting	Sp.rel.	Actions	Coreference	Foil-it	Avg.
CLIP	59.3	62.9	68.7	57.1	61.0	65.4	74.0	52.5	89.8	65.3
BLIP	59.0	**88.0**	**86.3**	73.2	**68.1**	71.5	69.1	51.0	93.8	70.0
LongCLIP	59.7	63.6	68.7	66.0	65.8	62.2	76.9	56.7	91.0	67.8
Hard Negative based method										
NegCLIP	80.2	70.5	76.8	71.7	65.0	72.9	83.0	55.2	91.9	71.6
CE-CLIP	**83.6** + 23.9	77.1 + 13.5	84.5	**79.2**	67.8	**76.4**	**86.3**	**67.2**	**94.7**	**76.7** +8.9
Ours	64.0 + 4.3	65.7 + 2.1	73.6	71.8	65.5	65.4	78.0	48.9	92.3	70.8 + 3.0

Results (%) on ARO and VALSE benchmark. The best scores for each section are highlighted in bold. +, - scores compared to LongCLIP.

Model	REPLACE				SWAP			ADD		
	Object	Attribute	Relation	Avg.	Object	Attribute	Avg.	Object	Attribute	Avg.
CLIP	90.9	80.0	69.2	80.2	61.4	64.0	62.7	77.2	68.2	72.7
LongCLIP	94.3	85.0	74.0	84.4	66.9	70.2	68.5	83.9	75.0	79.5
Hard Negative based method										
NegCLIP	92.7	85.9	76.5	85.0	75.2	75.4	75.3	88.8	82.8	85.8
CE-CLIP	93.8	**90.8**	**83.2**	**89.3** + 4.9	**76.8**	**79.3**	**78.0** + 9.5	**93.8**	**94.9**	**94.4** + 14.9
Ours	**94.7**	84.1	77.2	85.3 + 0.9	71.8	72.2	72.0 + 3.5	93.2	88.2	90.7 + 11.2

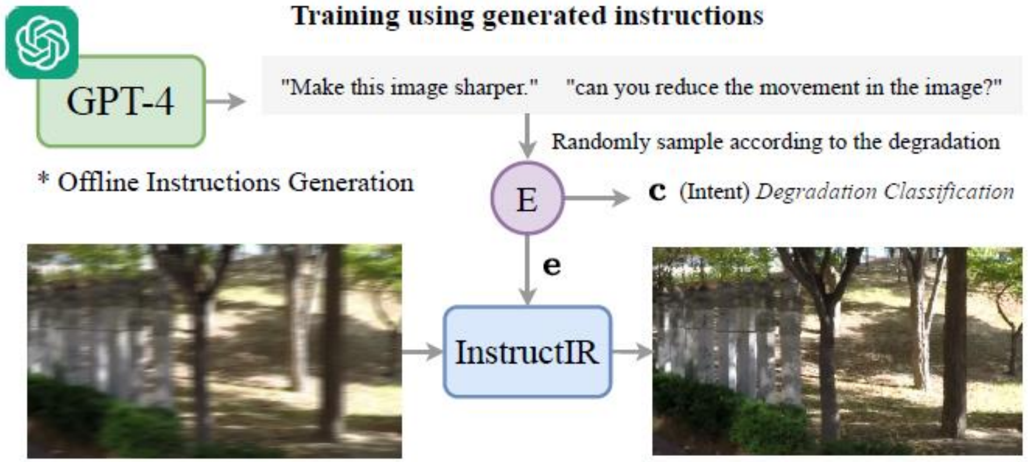
Results (%) on Sugar-Crepe benchmark. The best scores for each section are highlighted in bold. +, - scores compared to LongCLIP.

Supplementary #2

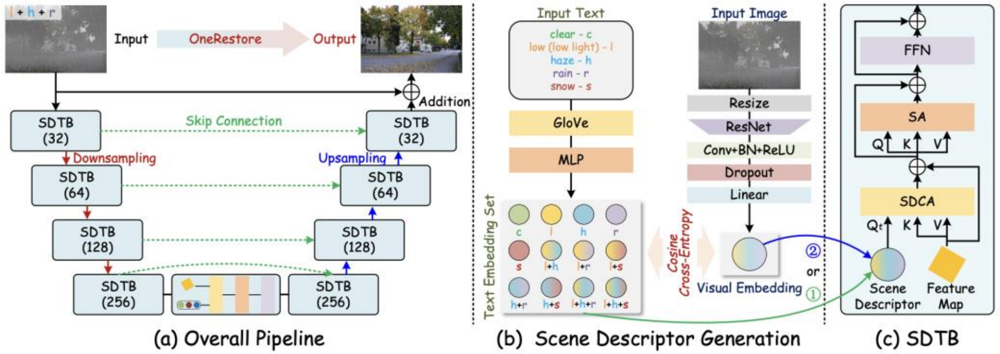
Text based Image Restoration for Composite Degradation

Image Restoration for Composite Degradation

Problem: 복합 손상이 있는 이미지를 Text 지시 형태로 한번에 복원하기 위한 방법을 연구하고자 함



InstructIR (ECCV 2024)
Text 지시 형태로 이미지를 복원하기 위한 방법을 제안
그러나, 여러 복합 손상이 있는 경우가 연구되지 않음



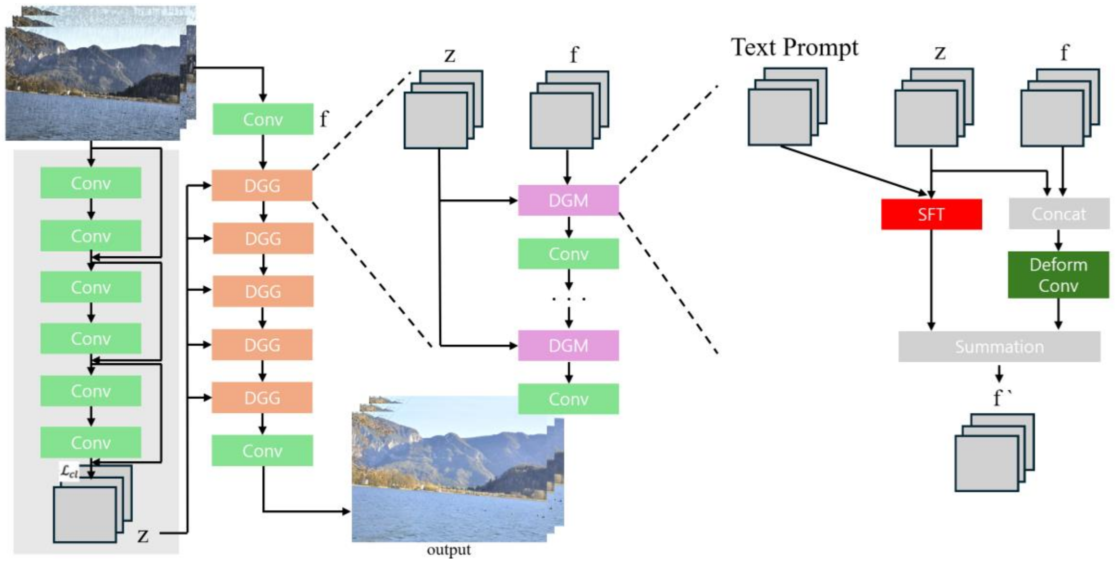
OneRestore (ECCV 2024)
복합 손상이 있는 경우 이미지를 복원하기 위한 방법을 제안
그러나, 실제 Text Prompt형태의 지시로 복원하기 위한 방법이 사용되지 않음

Image Restoration for Composite Degradation

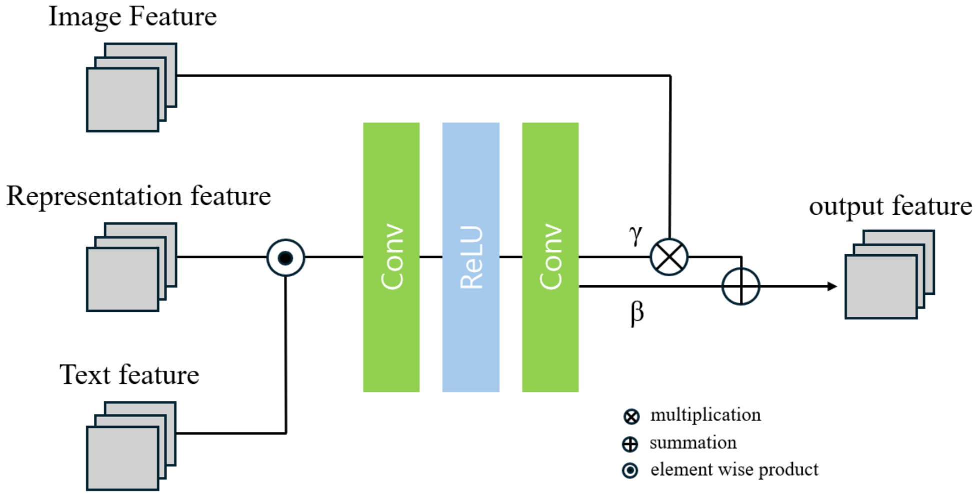
Methods

1) One to Many Restoration 모델인 AirNet 아키텍쳐를 base로 하여 Text Image Combined SFT Layer 제안

2) GPT로 생성한 손상종류별 Text Prompt를 클래스별로 Projection하는 layer를 추가



AirNet Architecture에 제안하는 Text-Image Conbined SFT Layer를 추가



Text-Image Conbined SFT(Spatial Feature Transform) Layer

Image Restoration for Composite Degradation

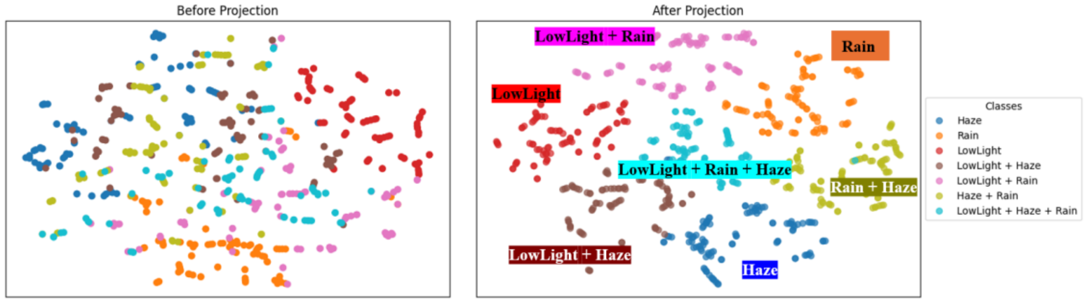
Methods

1) One to Many Restoration 모델인 AirNet 아키텍쳐를 base로 하여 Text Image Combined SFT Layer 제안

2) GPT로 생성한 손상종류별 Text Prompt를 클래스별로 Projection하는 layer를 추가

Degradation	Prompts
Deraining	"Please remove the raindrops from the image." "Clean up the rain effect to make the image clear." "Remove the rain from the scene to reveal more details."
Dehazing	"Eliminate the haze to make the image clearer." "Restore the image by removing the haze." "Remove the fog to reveal the hidden details."
Low-Light	"Increase the brightness to reveal more details." "Make the dark areas of the image clearer." "Improve the clarity by increasing the brightness."
Deraining + Dehazing	"Please remove the haze and rain from the image." "Eliminate both rain and fog to enhance the image clarity." "Clear the rain and haze for a sharper image.",
Deraining + Low-Light	"Eliminate the rain and brighten the dark regions for a clearer image." "Clear the rain and increase brightness to make the image sharper." "Remove the rain and brighten the shadows for better clarity."
Dehazing + Low-Light	"Increase brightness and remove the haze for better visibility." "Dissipate the fog and brighten the dim regions of the image." "Clear the haze and brighten up the shadows for a clearer image."
Deraining + Dehazing + Low-Light	"Clear the rain, haze, and shadows for improved visibility." "Remove the rain, fog, and lighten the image for clearer details." "Dissipate both rain and fog while brightening the shadows."

GPT-4o 모델을 통해 손상종류별 복원 Text Prompt를 생성한 예시



(a) t-SNE of embeddings before training (frozen CLIP Text encoder)

(b) t-SNE of embeddings after training our learned projection

Text Prompt가 Projection head를 통해 Projection되기 전과 후의 t-SNE 시각화 결과

Image Restoration for Composite Degradation

Experiments, Conclusion

1) 축소한 데이터셋을 기반으로 성능 평가, 해당 데이터셋에서는 SOTA에 버금가는 성능 달성
2) 확장한 원본 데이터셋에서는 조금 부족한 성능을 보임

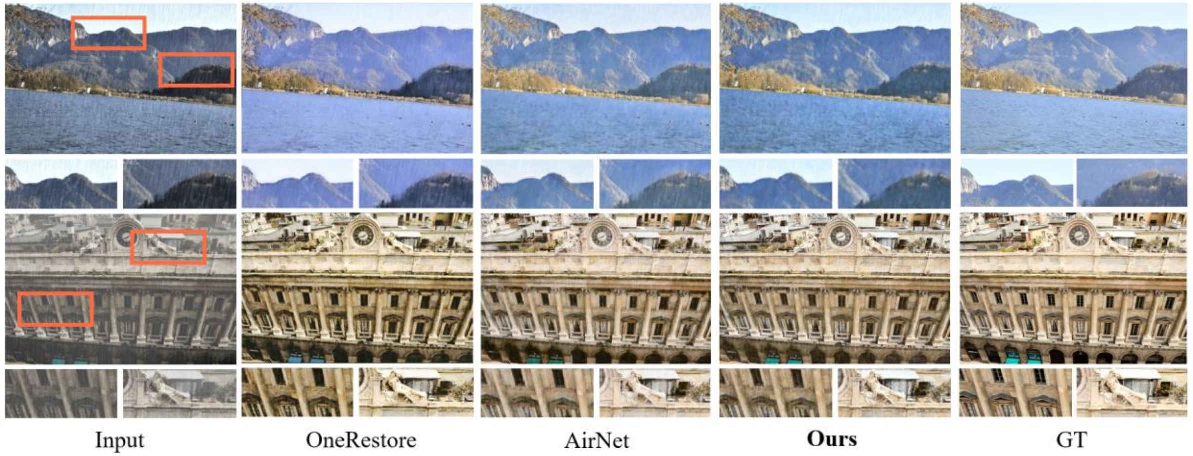
Metrics	CDD-11 (before)	AirNet	OneRestore	*Ours*
PSNR ↑	15.52	22.15	21.43	**22.28**
SSIM ↑	0.5881	0.8499	0.8285	**0.8501**

Comparison of quantitative results on CDD-11 dataset.

Method	Venue	Rain	Haze	Low	Rain+Haze	Rain+Low	Haze+Low	Rain+Haze+Low
AirNet	CVPR' 22	25.63/0.9124	22.23/**0.9309**	22.47/0.8406	20.99/0.8946	21.84/0.8080	20.72/**0.7945**	**20.24/0.7856**
OneRestore	ECCV' 24	25.82/0.9200	**22.52**/0.8061	22.60/**0.9310**	21.13/0.9012	19.77/0.7511	19.12/0.7477	19.05/0.7424
Ours	-	**26.30/0.9244**	22.37/0.9239	**23.16**/0.8466	**21.40/0.9036**	**22.03/0.8102**	**20.82**/0.7772	19.83/0.7647

Performance comparisons by each type of degradations on CDD-11 dataset.

축소한 CDD-11 데이터셋에서의 PSNR, SSIM Score

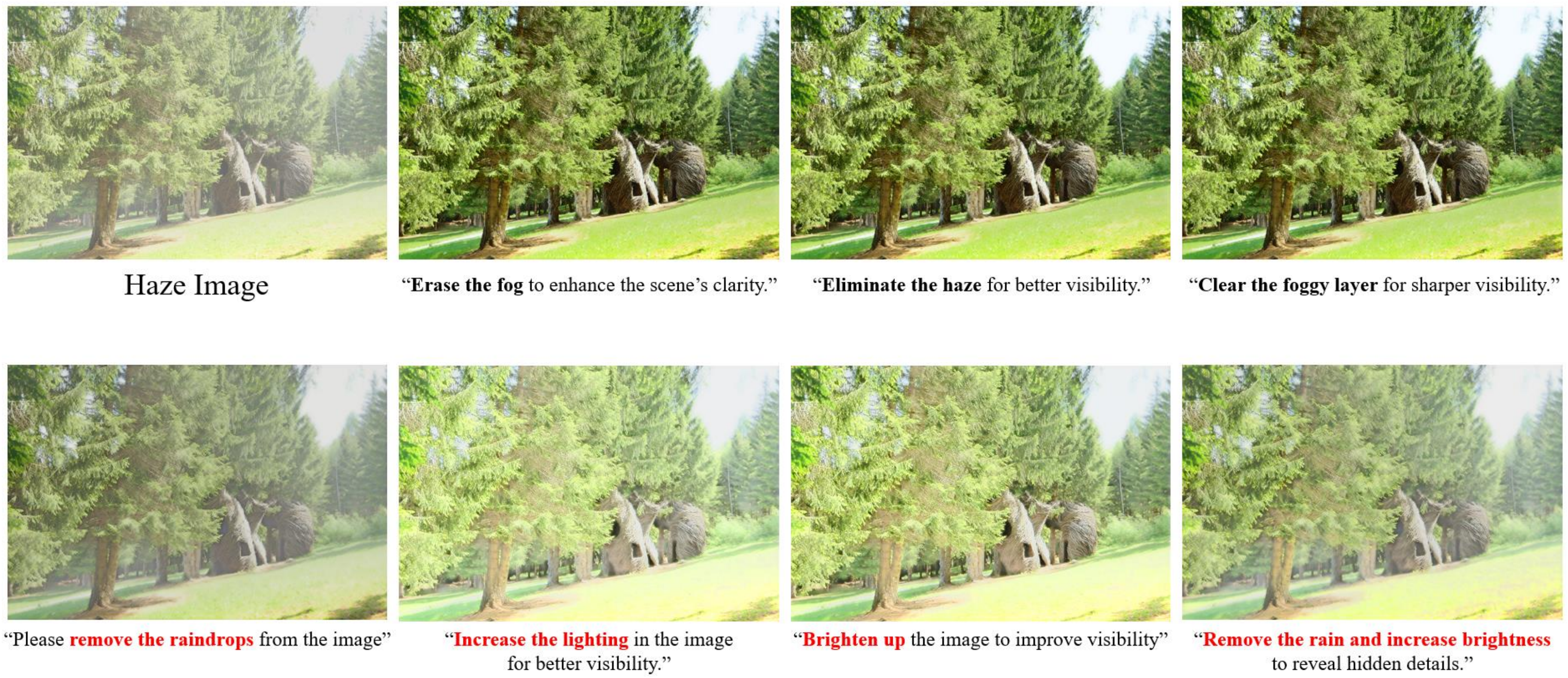


선행 연구와의 복원능력 시각적 비교

Image Restoration for Composite Degradation

Experiments, Conclusion

1) 축소한 데이터셋을 기반으로 성능 평가, 해당 데이터셋에서는 SOTA에 버금가는 성능 달성
2) 확장한 원본 데이터셋에서는 조금 부족한 성능을 보임



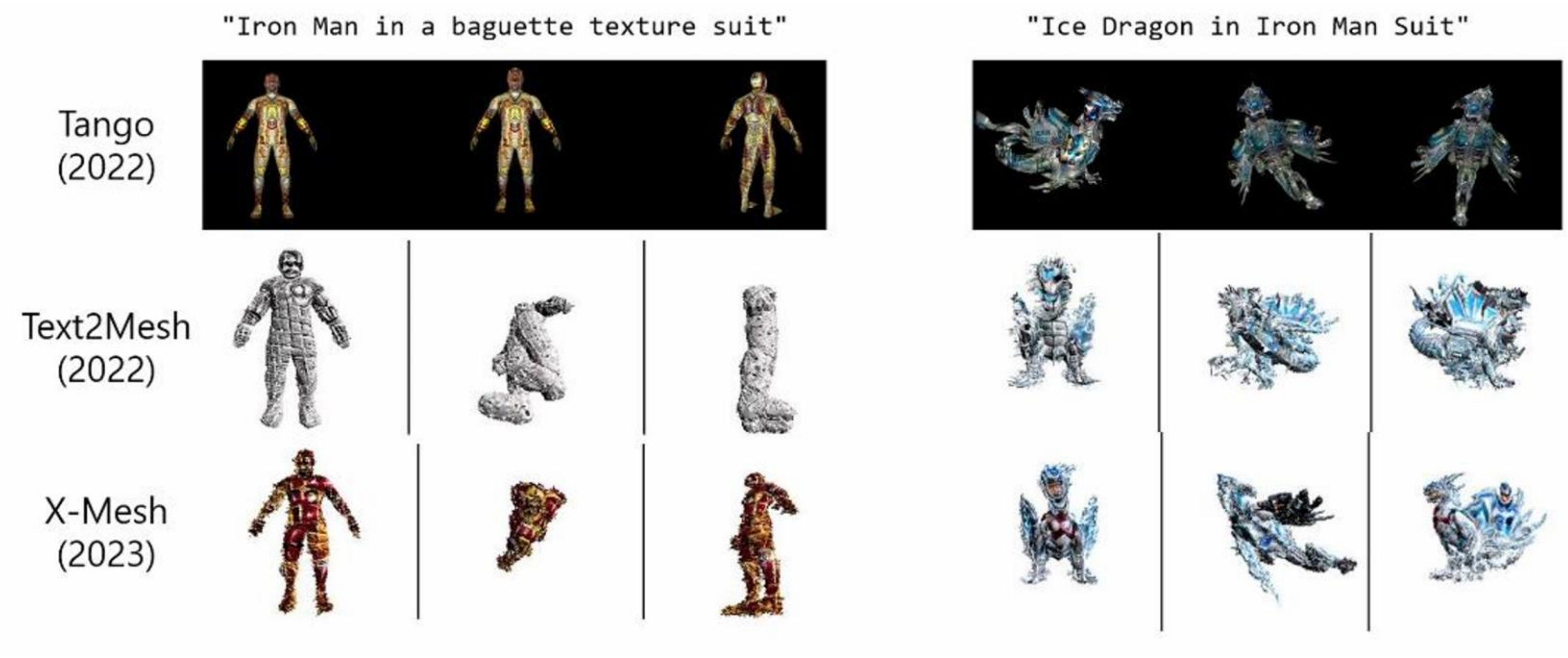
다양한 Text Prompt에 따른 복원 시도 결과 / 원본 이미지에 없는 손상을 제거하려고 시도할 때의 case 분석

Supplementary #3

3D Mesh Stylization with Multiple Prompting

3D Mesh Stylization with Multiple Prompting

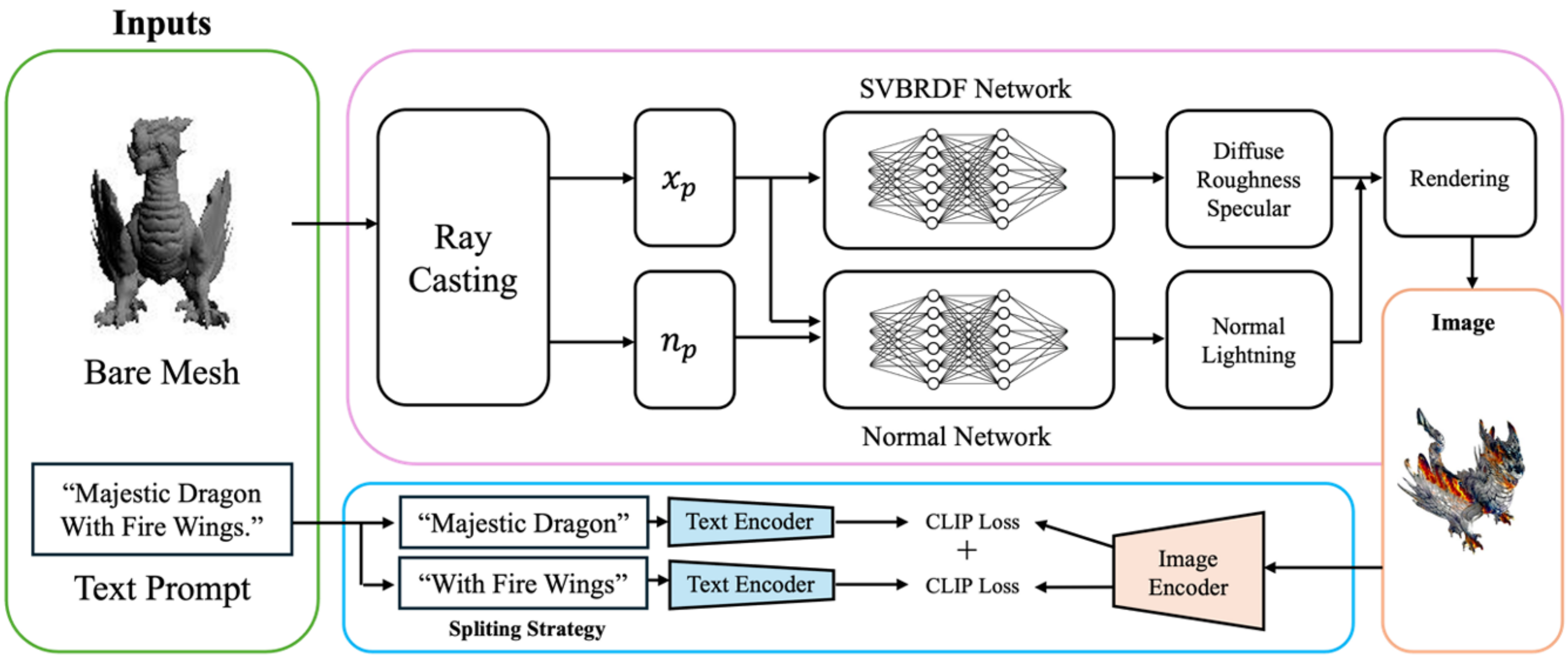
Problem: CLIP 기반의 3D Mesh Stylization 선행연구들이 여러 지시사항을 잘 반영하지 못함



Tango (NeurIPS 2022)의 경우 지시사항을 반영하기는 하지만 얼굴이 머리의 뒤에 오는 등 이상한 결과가 발생
Text2Mesh (CVPR 2022), X-Mesh (ICCV 2023)의 경우 여러 지시사항을 포함하는 Prompt가 입력되면 3D 객체가 과도하게 깨져버리는 현상 발생

3D Mesh Stylization with Multiple Prompting

Method: Tango 아키텍쳐를 baseline으로 사용해 Text Prompt를 Split하여 loss를 계산하는 방법론을 적용



여러 지시사항이 포함된 Text Prompt를 여러 단일 지시사항으로 분리하여 개별적인 Loss 계산을 수행

3D Mesh Stylization with Multiple Prompting

Conclusion

1) User Study를 통한 시각적 비교 결과에서는 가장 좋은 점수를 획득
2) CLIP Score, CLIP R-Precision 평가에서는 다소 아쉬운 성능을 보임. **선행 연구들은 학습 과정에서 전체 Prompt 자체와 생성 결과를 CLIP Loss로 직접적으로 비교하고 평가에서도 전체 Prompt를 사용해 평가 방식에 과적합**된 양상을 보인다고 판단 됨
3) 앞선 CLIP Score, CLIP R-Precision 등의 평가방식을 대체하기 위한 **새로운 평가 방식 도입의 필요성**

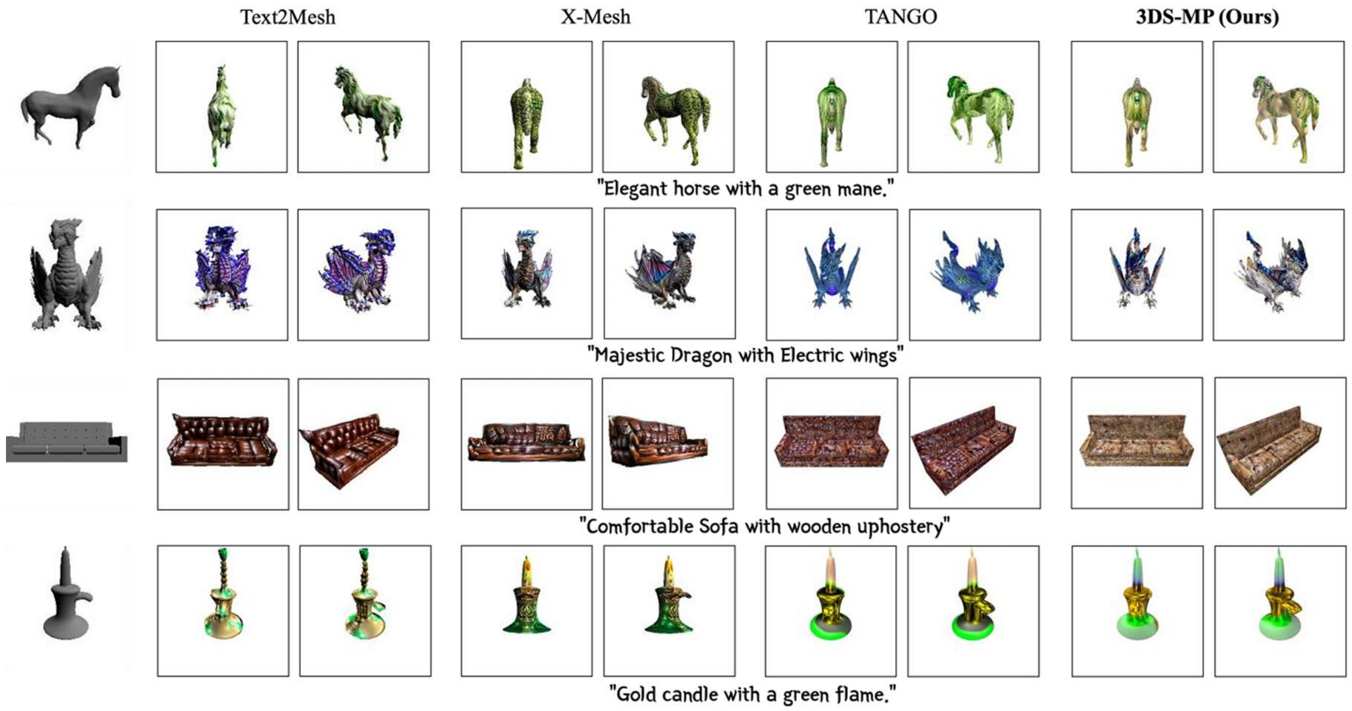
	VIT-B/32@336PX			VIT-L/14		
	CLIP Score	CLIP R-Precision		CLIP Score	CLIP R-Precision	
		Top-1 ACC (%)	Top-3 ACC (%)		Top-1 ACC (%)	Top-3 ACC (%)
TEXT2MESH[2]	0.303	**88%**	92%	0.249	**90%**	94%
TANGO[3]	**0.308**	76%	94%	**0.264**	80%	96%
X-MESH[4]	0.299	**88%**	**98%**	0.242	80%	**100%**
OURS: 3DS-MP	0.301	70%	90%	0.258	70%	90%

선행 연구와의 CLIP Score, CLIP R-Precision 비교

	Q1	Q2	Q3
TEXT2MESH[2]	3.4	3.2	3.1
TANGO[3]	3.6	3.7	3.8
X-MESH[4]	3.3	3.3	3.3
OURS: 3DS-MP	**3.8**	**3.8**	**3.9**

선행 연구와의 User Study 점수 비교

Q1: "생성된 결과가 Text prompt를 충실히 반영하는가?"
Q2: "생성된 결과의 해상도 및 품질이 좋은가?"
Q3: "기존 Mesh 형태를 해치지 않고 잘 유지하는가?"



선행 연구와의 생성능력 시각적 비교

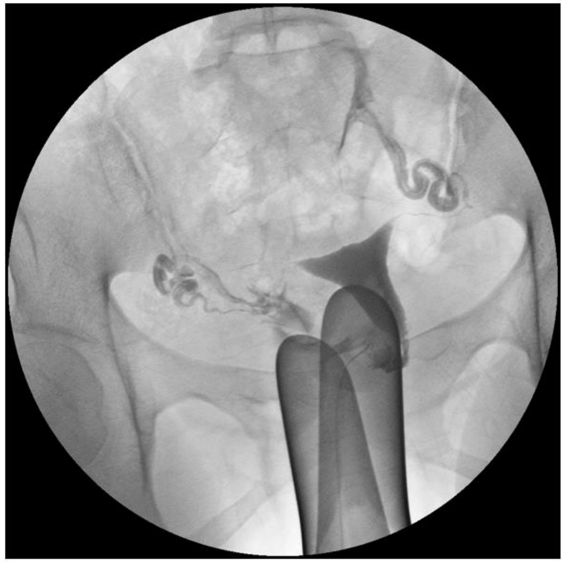
Supplementary #4

Diagnosis for Tubal Patency Using Contrastive Learning

Diagnosis for Tubal Patency Using Contrastive Learning

Problem: 자궁난관조영술 영상(X-ray)에서의 난관 유착 및 자궁 내 질병 진단 모델 개발
적은 데이터양 및 클래스의 심한 불균형으로 인한 Challenge 존재

난관 유착(Occlusion)

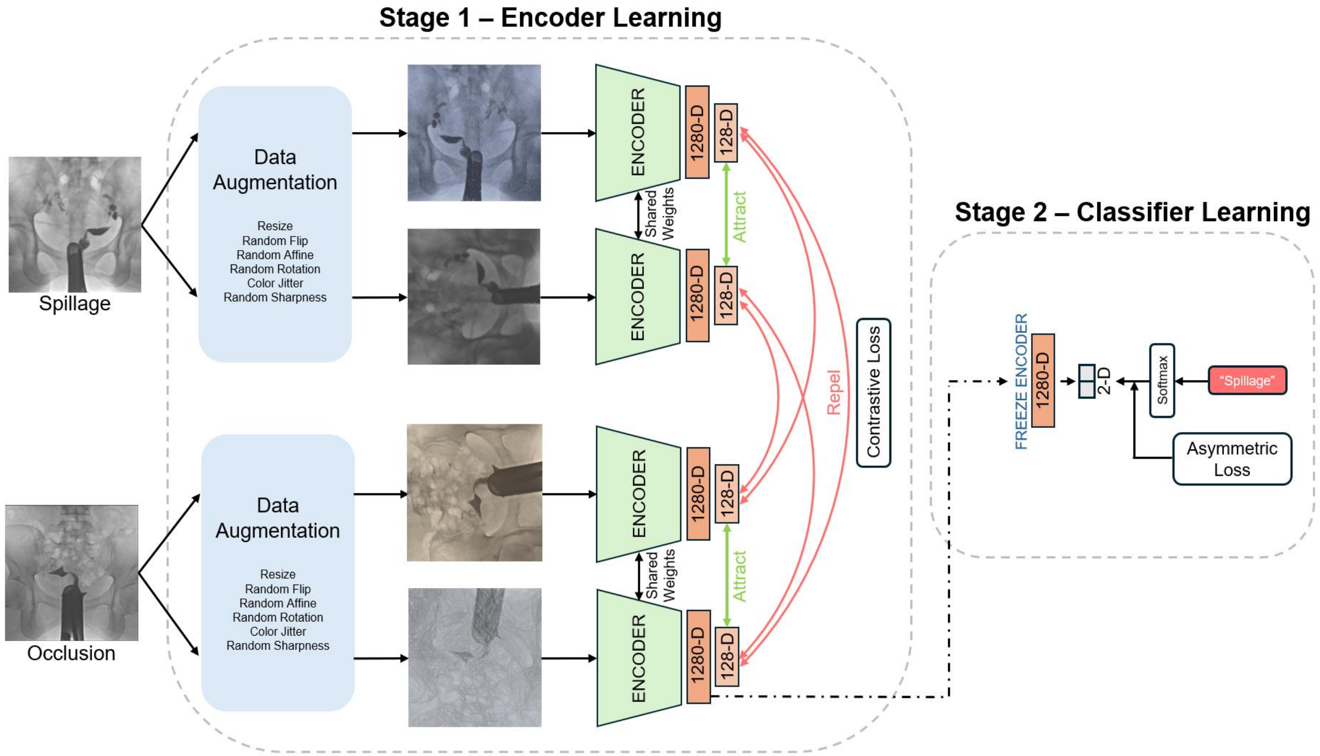


정상 난관(Spillage)

자궁난관조영술 X-ray 영상 기반 질병 진단 모델은 선행 연구가 존재하지 않음
전공의들의 자궁난관조영술 판독능력 향상 및 업무 효율화를 위해 자궁난관조영술을 인공지능으로 진단하는 것이 필요함

Diagnosis for Tubal Patency Using Contrastive Learning

Methods: Supervised Contrastive Learning + Asymmetric Loss



적은 데이터에서도 효과적으로 클래스간 분별을 수행하기 위해 Supervised Contrastive Learning 적용
심한 클래스 불균형을 해소하기 위해 Cross Entropy 대신 Asymmetric Loss 적용

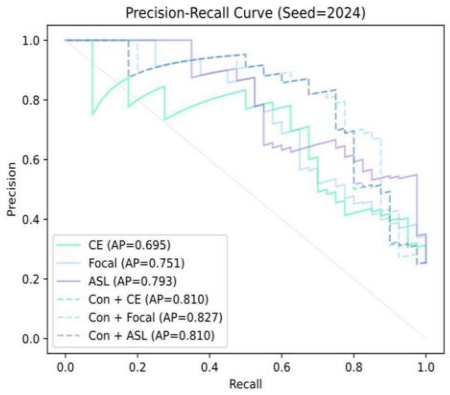
Diagnosis for Tubal Patency Using Contrastive Learning

Experiments, Conclusion

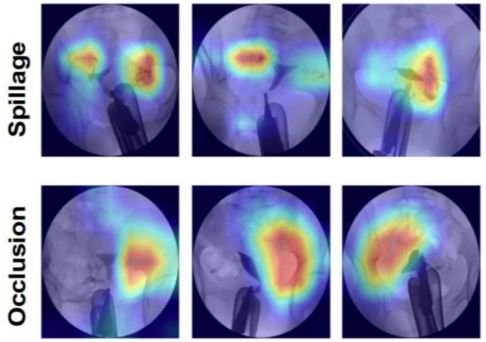
1) 난관이 개통되어 양쪽 난관 모두 정상적인 조영제 유출이 보이는 케이스는 Grad-CAM에서 양쪽 난관 모두 잘 집중하는 양상을 보이며, 난관 폐색이 있을 경우 폐색된 난관 쪽에 집중하는 양상을 보임
2) Supervised Contrastive Loss와 Asymmetric Loss를 결합한 방법은 **Accuracy와 F1-Score에서 가장 높은 성능을 달성했으며, Precision과 Recall에서 두 번째로 높은 성능**을 달성
3) 제안하는 접근 방식이 자궁난관조영술 영상에서 난관의 정상적인 개통 여부를 판별하는 것에 효과적인 것으로 확인할 수 있었으며, **불균형한 의료 데이터를 사용할 때 성능을 향상시킬 수 있을 것으로 기대**

Methods	Accuracy (%)	Precision	Recall	F1	AUPRC
CE	84.388 ± 2.223	0.813 ± 0.107	0.508 ± 0.029	0.623 ± 0.037	0.761 ± 0.100
Focal	85.564 ± 1.593	0.750 ± 0.023	0.650 ± 0.087	0.695 ± 0.049	0.758 ± 0.025
ASL	85.564 ± 0.967	**0.832 ± 0.105**	0.558 ± 0.063	0.663 ± 0.020	**0.827 ± 0.048**
SupCon+CE	86.709 ± 2.282	0.792 ± 0.112	**0.675 ± 0.090**	0.720 ± 0.032	0.766 ± 0.038
SupCon+Focal	87.342 ± 1.675	0.820 ± 0.008	0.641 ± 0.095	0.717 ± 0.050	0.776 ± 0.045
SupCon+ASL	**87.764 ± 0.731**	0.827 ± 0.057	0.650 ± 0.066	**0.724 ± 0.030**	0.769 ± 0.036

난관 폐색 여부에 따른 Binary Classifcation의 수치적 결과



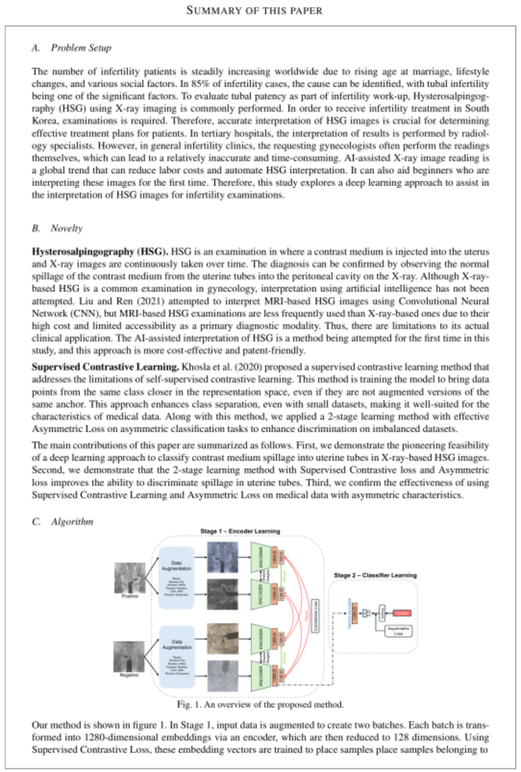
난관 폐색 분류 PR Curve



난관 폐색 분류 GradCAM

Diagnosis for Tubal Patency Using Contrastive Learning

본 연구는 한국인공지능학회 2024 하계학술대회 포스터 세션에서 발표됨



SUMMARY OF THIS PAPER

A. Problem Setup

The number of infertility patients is steadily increasing worldwide due to rising age at marriage, lifestyle changes, and various social factors. In 85% of infertility cases, the cause can be identified, with tubal infertility being one of the significant factors. To evaluate tubal patency as part of infertility work-up, Hysterosalpingography (HSG) using X-ray imaging is commonly performed. In order to receive infertility treatment in South Korea, examinations is required. Therefore, accurate interpretation of HSG images is crucial for determining effective treatment plans for patients. In tertiary hospitals, the interpretation of results is performed by radiology specialists. However, in general infertility clinics, the requesting gynecologists often perform the readings themselves, which can lead to a relatively inaccurate and time-consuming. AI-assisted X-ray image reading is a global trend that can reduce labor costs and automate HSG interpretation. It can also aid beginners who are interpreting these images for the first time. Therefore, this study explores a deep learning approach to assist in the interpretation of HSG images for infertility examinations.

B. Novelty

Hysterosalpingography (HSG). HSG is an examination in where a contrast medium is injected into the uterus and X-ray images are continuously taken over time. The diagnosis can be confirmed by observing the normal spillage of the contrast medium from the uterine tubes into the peritoneal cavity on the X-ray. Although X-ray-based HSG is a common examination in gynecology, interpretation using artificial intelligence has not been attempted. Liu and Ren (2021) attempted to interpret MRI-based HSG images using Convolutional Neural Network (CNN), but MRI-based HSG examinations are less frequently used than X-ray-based ones due to their high cost and limited accessibility as a primary diagnostic modality. Thus, there are limitations to its actual clinical application. The AI-assisted interpretation of HSG is a method being attempted for the first time in this study, and this approach is more cost-effective and patent-friendly.

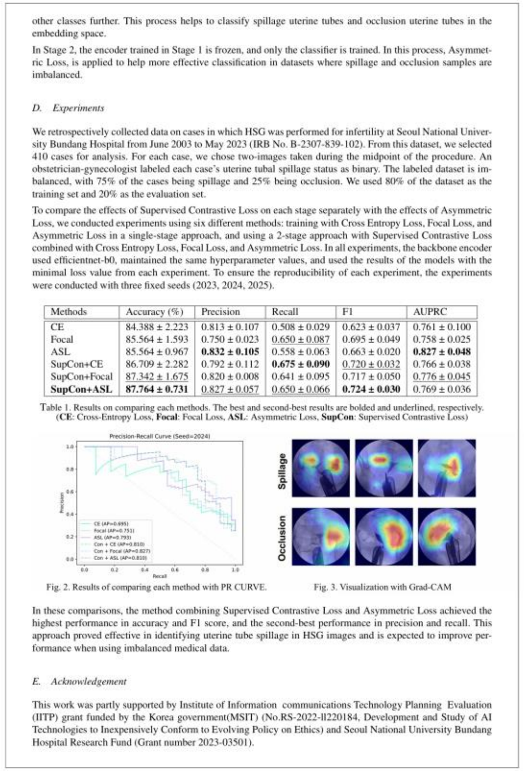
Supervised Contrastive Learning. Khosla et al. (2020) proposed a supervised contrastive learning method that addresses the limitations of self-supervised contrastive learning. This method is training the model to bring data points from the same class closer in the representation space, even if they are not augmented versions of the same anchor. This approach enhances class separation, even with small datasets, making it well-suited for the characteristics of medical data. Along with this method, we applied a 2-stage learning method with effective Asymmetric Loss on asymmetric classification tasks to enhance discrimination on imbalanced datasets.

The main contributions of this paper are summarized as follows. First, we demonstrate the pioneering feasibility of a deep learning approach to classify contrast medium spillage into uterine tubes in X-ray-based HSG images. Second, we demonstrate that the 2-stage learning method with Supervised Contrastive loss and Asymmetric loss improves the ability to discriminate spillage in uterine tubes. Third, we confirm the effectiveness of using Supervised Contrastive Learning and Asymmetric Loss on medical data with asymmetric characteristics.

C. Algorithm

Fig. 1. An overview of the proposed method.

Our method is shown in figure 1. In Stage 1, input data is augmented to create two batches. Each batch is transformed into 1280-dimensional embeddings via an encoder, which are then reduced to 128 dimensions. Using Supervised Contrastive Loss, these embedding vectors are trained to place samples place samples belonging to



other classes further. This process helps to classify spillage uterine tubes and occlusion uterine tubes in the embedding space.

In Stage 2, the encoder trained in Stage 1 is frozen, and only the classifier is trained. In this process, Asymmetric Loss, is applied to help more effective classification in datasets where spillage and occlusion samples are imbalanced.

D. Experiments

We retrospectively collected data on cases in which HSG was performed for infertility at Seoul National University Bundang Hospital from June 2003 to May 2023 (IRB No. B-2307-839-102). From this dataset, we selected 410 cases for analysis. For each case, we chose two-images taken during the midpoint of the procedure. An obstetrician-gynecologist labeled each case's uterine tubal spillage status as binary. The labeled dataset is imbalanced, with 75% of the cases being spillage and 25% being occlusion. We used 80% of the dataset as the training set and 20% as the evaluation set.

To compare the effects of Supervised Contrastive Loss on each stage separately with the effects of Asymmetric Loss, we conducted experiments using six different methods: training with Cross Entropy Loss, Focal Loss, and Asymmetric Loss in a single-stage approach, and using a 2-stage approach with Supervised Contrastive Loss combined with Cross Entropy Loss, Focal Loss, and Asymmetric Loss. In all experiments, the backbone encoder used efficientnet-b0, maintained the same hyperparameter values, and used the results of the models with the minimal loss value from each experiment. To ensure the reproducibility of each experiment, the experiments were conducted with three fixed seeds (2023, 2024, 2025).

Methods	Accuracy (%)	Precision	Recall	F1	AUPRC
CE	84.388 ± 2.223	0.813 ± 0.107	0.508 ± 0.029	0.623 ± 0.037	0.761 ± 0.100
Focal	85.564 ± 1.593	0.750 ± 0.023	0.650 ± 0.087	0.695 ± 0.049	0.758 ± 0.025
ASL	85.564 ± 0.967	**0.832 ± 0.105**	0.558 ± 0.063	0.663 ± 0.020	**0.827 ± 0.048**
SupCon+CE	86.709 ± 2.282	0.792 ± 0.112	**0.675 ± 0.090**	0.720 ± 0.032	0.766 ± 0.038
SupCon+Focal	87.342 ± 1.675	0.820 ± 0.008	0.641 ± 0.095	0.717 ± 0.050	0.776 ± 0.045
SupCon+ASL	**87.764 ± 0.731**	0.827 ± 0.057	0.650 ± 0.066	**0.724 ± 0.030**	0.769 ± 0.036

Table 1. Results on comparing each methods. The best and second-best results are bolded and underlined, respectively. (**CE**: Cross-Entropy Loss, **Focal**: Focal Loss, **ASL**: Asymmetric Loss, **SupCon**: Supervised Contrastive Loss)

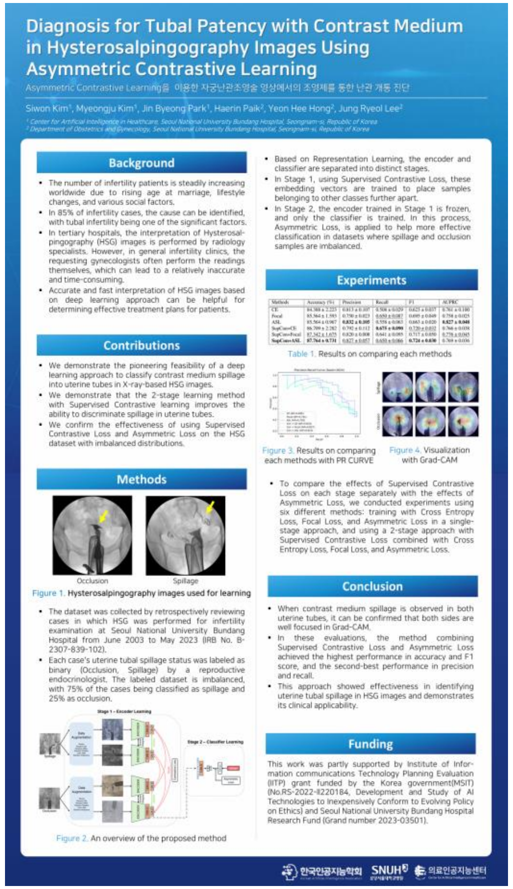
Fig. 2. Results of comparing each method with PR CURVE.

Fig. 3. Visualization with Grad-CAM

In these comparisons, the method combining Supervised Contrastive Loss and Asymmetric Loss achieved the highest performance in accuracy and F1 score, and the second-best performance in precision and recall. This approach proved effective in identifying uterine tube spillage in HSG images and is expected to improve performance when using imbalanced medical data.

E. Acknowledgement

This work was partly supported by Institute of Information communications Technology Planning Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2022-II220184, Development and Study of AI Technologies to Inexpensively Conform to Evolving Policy on Ethics) and Seoul National University Bundang Hospital Research Fund (Grant number 2023-03501).



Diagnosis for Tubal Patency with Contrast Medium in Hysterosalpingography Images Using Asymmetric Contrastive Learning

Asymmetric Contrastive Learning을 이용한 자궁난관조영술 영상에서의 조영제를 통한 난관 개통 진단

Siwon Kim[1], Myeongju Kim[1], Jin Byeong Park[1], Haerin Paik[2], Yeon Hee Hong[2], Jung Ryeol Lee[2]
[1] Center for Artificial Intelligence in Healthcare, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea
[2] Department of Obstetrics and Gynecology, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea

Background

- The number of infertility patients is steadily increasing worldwide due to rising age at marriage, lifestyle changes, and various social factors.
- In 85% of infertility cases, the cause can be identified, with tubal infertility being one of the significant factors.
- In tertiary hospitals, the interpretation of Hysterosalpingography (HSG) images is performed by radiology specialists. However, in general infertility clinics, the requesting gynecologists often perform the readings themselves, which can lead to a relatively inaccurate and time-consuming.
- Accurate and fast interpretation of HSG images based on deep learning approach can be helpful for determining effective treatment plans for patients.

Contributions

- We demonstrate the pioneering feasibility of a deep learning approach to classify contrast medium spillage into uterine tubes in X-ray-based HSG images.
- We demonstrate that the 2-stage learning method with Supervised Contrastive learning improves the ability to discriminate spillage in uterine tubes.
- We confirm the effectiveness of using Supervised Contrastive Loss and Asymmetric Loss on the HSG dataset with imbalanced distributions.

Methods

Figure 1. Hysterosalpingography images used for learning

- The dataset was collected by retrospectively reviewing cases in which HSG was performed for infertility examination at Seoul National University Bundang Hospital from June 2003 to May 2023 (IRB No. B-2307-839-102).
- Each case's uterine tubal spillage status was labeled as binary (Occlusion, Spillage) by a reproductive endocrinologist. The labeled dataset is imbalanced, with 75% of the cases being classified as spillage and 25% as occlusion.

Figure 2. An overview of the proposed method

- Based on Representation Learning, the encoder and classifier are separated into distinct stages.
- In Stage 1, using Supervised Contrastive Loss, these embedding vectors are trained to place samples belonging to other classes further apart.
- In Stage 2, the encoder trained in Stage 1 is frozen, and only the classifier is trained. In this process, Asymmetric Loss, is applied to help more effective classification in datasets where spillage and occlusion samples are imbalanced.

Experiments

Table 1. Results on comparing each methods

Figure 3. Results on comparing each methods with PR CURVE

Figure 4. Visualization with Grad-CAM

- To compare the effects of Supervised Contrastive Loss on each stage separately with the effects of Asymmetric Loss, we conducted experiments using six different methods: training with Cross Entropy Loss, Focal Loss, and Asymmetric Loss in a single-stage approach, and using a 2-stage approach with Supervised Contrastive Loss combined with Cross Entropy Loss, Focal Loss, and Asymmetric Loss.

Conclusion

- When contrast medium spillage is observed in both uterine tubes, it can be confirmed that both sides are well focused in Grad-CAM.
- In these evaluations, the method combining Supervised Contrastive Loss and Asymmetric Loss achieved the highest performance in accuracy and F1 score, and the second-best performance in precision and recall.
- This approach showed effectiveness in identifying uterine tubal spillage in HSG images and demonstrates its clinical applicability.

Funding

This work was partly supported by Institute of Information communications Technology Planning Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2022-II220184, Development and Study of AI Technologies to Inexpensively Conform to Evolving Policy on Ethics) and Seoul National University Bundang Hospital Research Fund (Grand number 2023-03501).

한국인공지능학회 SNUH 의료인공지능센터