

프롬프트 기반 영상처리에 대한 결과 및 한계 분석: InstructPix2Pix를 중심으로

시종욱*, 김성영**

Analysis of Results and Limitations in Prompt-based Image Processing: Focusing on InstructPix2Pix

Jongwook Si*, Sungyoung Kim**

이 연구는 국립금오공과대학교 RISE 사업의 지원으로 수행되었음

요 약

최근 Text-to-Image 형태의 생성형 AI가 높은 인기를 얻으며, 텍스트 프롬프트만으로도 현실적인 이미지를 생성하는 기술이 발전하고 있다. 그러나 기존 연구들은 텍스트를 기반으로 실제 영상처리 작업을 수행하는 능력에 대한 연구는 상대적으로 부족하다. 본 논문에서는 원본 영상에 대하여 지시가 가능한 InstructPix2Pix 모델을 활용하여 프롬프트 기반으로 영상처리의 가능성을 탐색한다. 이를 위해 색공간 변환, 필터링, 밝기 및 명암 조정, 히스토그램 기법과 같은 대표적인 영상처리 기법을 프롬프트를 통해 모델에 입력하고, 생성된 이미지와 실제 영상처리 결과를 비교 분석하였다. 실험 결과, InstructPix2Pix는 일부 필터링 기법이나 색공간 변환의 경우 실제 영상처리 기법과 유사한 결과를 생성할 수 있었으나, 전문적인 영상처리를 수행하는 것은 한계가 있음을 확인하였다.

Abstract

Recently, Text-to-Image generative AI models have gained significant popularity, enabling the generation of highly realistic images solely based on textual prompts. However, existing studies primarily focus on image generation rather than the capability of these models to perform actual image processing tasks based on textual instructions. This paper explores the feasibility of prompt-driven image processing using the InstructPix2Pix model, which allows instruction-based modifications on an input image. To evaluate this capability, we apply fundamental image processing techniques, including color space transformation, filtering, brightness and contrast adjustment, and histogram-based methods, through textual prompts. The generated outputs are systematically compared against the results obtained from conventional image processing algorithms. Experimental results demonstrate that while InstructPix2Pix can generate comparable results for certain filtering techniques and color space transformations, its effectiveness remains limited in performing complex and precise image processing operations.

Keywords

text prompt, image processing, InstructPix2Pix, generative AI, analysis

* 국립금오공과대학교 컴퓨터·AI융합공학과 박사과정
- ORCID: <http://orcid.org/0000-0003-2092-2769>
** 국립금오공과대학교 컴퓨터공학과 교수(교신저자)
- ORCID: <http://orcid.org/0000-0002-7722-6759>

• Received: Jun. 18, 2025, Revised: Jul. 14, 2025, Accepted: Jul. 17, 2025
• Corresponding Author: Sungyoung Kim
Dept. of Computer Engineering, Kumoh National Institute of Technology,
61 Daehak-ro (yangho-dong), Gumi, Gyeongbuk, [39177] Korea
Tel.: +82-54-478-7530, Email: sykim@kumoh.ac.kr

I. 서론

영상처리는 디지털 이미지에서 유용한 정보를 추출하고, 분석하거나 변환하는 핵심 기술로, 다양한 산업과 연구 분야에서 폭넓게 활용되고 있다 [1][2]. 영상처리 기술은 다양한 응용 도메인에서 경계 검출, 특징점 매칭, 색공간 변환, 필터링 등을 수행하며, 특정 목적에 맞게 영상을 조정하는 역할을 한다[3]. 전통적인 영상처리 기법은 수학적 모델과 신호처리 기법을 기반으로 픽셀 단위 연산을 수행하며, 특정 연산을 적용하여 원하는 효과를 얻는다. 예를 들어, 고주파 강조 필터는 선명화를 수행하여 영상의 경계를 강조하고, 가우시안 블러는 저주파 성분을 강조하여 영상의 노이즈를 제거하거나 부드러운 효과를 제공한다. 기존 영상처리 기법은 주로 OpenCV 등의 라이브러리를 이용하여 수학적 연산을 기반으로 필터링, 밝기 및 명암 조절, 히스토그램 정규화, 색공간 변환, 주파수 변환 등의 방법을 수행한다. 이러한 기법들은 대부분 픽셀 단위 및 행렬 연산을 기반으로 하며, 명확한 수학적 원리를 따른다.

최근 생성형 AI(Generative AI) 기술의 발전으로 인해 영상처리 방식이 기존의 고전적인 방법에서 딥러닝 기반 접근법으로 변화하였다[4-5]. 대표적인 딥러닝 기반 생성형 AI 모델로는 Stable Diffusion, DALL-E, Midjourney 등이 있으며, 이러한 모델은 사용자가 제공하는 텍스트 프롬프트를 기반으로 새로운 이미지를 생성하는 능력을 갖추고 있다[6-7]. 이러한 모델은 특정한 객체, 스타일, 조명 조건을 명시하는 텍스트 입력을 통해 원하는 형상의 이미지를 생성할 수 있으며, 최근 연구에서는 이러한 생성형 AI 모델이 프롬프트를 반영한 고품질의 이미지 생성과 스타일을 변경하는 방향으로 발전하고 있다[8]. 그러나 기존 영상에 대한 편집 시 프롬프트만으로 구성되기에 의미만 유지되고 전체적인 형상은 모두 변형되게 된다. 이를 해결한 InstructPix2Pix[9]은 사용자의 명령을 해석하여 기존 이미지에 대해 특정한 변환을 수행하는 기능을 갖고 있다. 또한, 텍스트 기반 지시를 통해 기존 영상을 편집할 수 있는 기능을 제공한다.

InstructPix2Pix[9]는 딥러닝 모델이 학습한 데이터셋을 기반으로 사용자의 프롬프트를 해석하여 이미지를 수정한다. 기존 영상처리 기법과 유사한 결과를 생성할 수 있는지 평가할 필요성이 존재하며, 생성된 결과물이 실제 영상처리 기법을 대체할 수 있을 만큼 신뢰성이 있는지 검증하는 것이 본 연구의 핵심 목적이다. 본 연구에서는 InstructPix2Pix[9]를 활용하여 프롬프트 기반 영상처리를 수행하고, 이를 전통적인 영상처리 기법과 비교하여 AI 기반 영상처리의 가능성과 한계를 분석한다. 분석 결과, 색공간 변환의 경우 일부 알고리즘에 대하여 성능이 상대적으로 높았으나, 필터링 등의 세부적인 내용이 필요한 알고리즘의 경우 성능이 매우 낮게 나타남을 확인할 수 있다.

2절에서는 기존 영상처리 알고리즘과 제안하는 방법을 나타낸다. 3절에서는 생성 결과와 이에 따른 분석을 진행하며, 4절에서는 결론과 향후 연구에 대해서 소개한다.

II. InstructPix2Pix 모델 분석

InstructPix2Pix[9]는 텍스트 기반 프롬프트와 원본 이미지를 함께 입력받아 이미지의 편집 및 변환 작업을 수행할 수 있는 생성형 이미지 편집 모델로, 기존 Pix2Pix의 조건부 이미지 변환 구조를 기반으로 하면서도 Stable Diffusion의 이미지 생성 방식을 결합한 것이 특징이다. 이 모델은 사용자가 입력한 텍스트 프롬프트를 텍스트 인코더를 통해 임베딩하여 지시 내용을 벡터 형태로 변환한 뒤, 이를 이미지 생성 네트워크의 조건으로 주입하여 이미지의 Latent Representation에 변화를 주며 프롬프트가 지시한 변환이 적용된 이미지를 생성한다. 생성 과정에서 U-Net 기반의 구조와 Skip Connection이 활용되어 원본 이미지의 구조적 정보를 보존하며, 텍스트 프롬프트에 따라 색상, 형태, 스타일 등이 변환되도록 작동한다. 또한 Stable Diffusion 기반의 Reverse Process를 통해 노이즈가 추가된 이미지로부터 점진적으로 원하는 이미지를 생성하기 때문에, 텍스트 조건을 반영하면서도 이미지의 세부 구조를 유지하는 것이 가능하다.

이러한 구조로 인해 InstructPix2Pix는 기존의 단순 텍스트 기반 이미지 생성과 달리 원본 이미지를 기반으로 사용자가 원하는 편집이나 영상처리 효과를 직관적으로 수행할 수 있다. 그러나 생성 방식이 실제 영상처리 알고리즘이 수행하는 수학적 연산을 그대로 적용하는 것이 아니라, 학습 과정에서 해당 작업의 시각적 패턴을 학습해 이를 유사하게 재현하는 방식으로 작동하기 때문에, 로컬 연산 기반의 정밀 작업에서는 한계가 발생한다.

또한 InstructPix2Pix는 입력되는 프롬프트의 표현 방식과 구체성에 따라 생성 이미지의 품질과 처리 결과가 달라진다는 특징을 가진다. 동일한 변환을 지시하더라도 표현 방식이 다르면 결과물의 색조, 변환 강도, 처리 범위 등이 달라질 수 있어 사용자의 언어적 표현이 결과에 직접적으로 영향을 미치며, 복잡한 배경이나 다양한 색상이 혼합된 이미지일수록 지시된 작업을 일관되게 적용하지 못하고 일부 영역만 변환하거나 과도하게 변환하는 경우도 발생한다[10]. 이러한 한계에도 불구하고 프롬프트 기반으로 직관적이고 빠르게 이미지 편집 및 변환 작업을 수행할 수 있다는 점에서, 영상처리 워크플로우의 대안적 방법으로 활용 가능성을 보여주고 있다.

III. InstructPix2Pix를 사용한 영상처리

본 연구에서는 InstructPix2Pix 모델을 중심으로 생성형 AI가 다양한 영상처리 기법을 정확하게 수행할 수 있는지 평가하기 위해 표 1과 같이 텍스트 프롬프트를 설계하였다. 본 연구에서는 색공간 변환, 필터링, 밝기 및 명암 조정, 히스토그램 기법 등의 대표적인 영상처리 기법을 채택한다. 특히 밝기 및 명암 조정과 관련하여 변화율(%) 기반의 조정 방식을 적용한다. 제안된 프롬프트는 기존 영상처리 기법을 수행하는 방식과 비교하기 위해 구성하였다. 일부 기법은 단순한 기술 명칭만 포함된 프롬프트로 제공하며, 일부 기법은 필터 크기 및 파라미터 값을 포함한 형태로 나타낸다.

이러한 프롬프트 설계 방식은 InstructPix2Pix가 명확한 수학적 연산 없이 프롬프트의 의미를 올바

르게 해석하고, 영상처리 기법을 수행할 수 있는지 평가하는 데 중요한 기준이 된다. 본 논문에서 사용된 프롬프트는 크게 두 가지 유형으로 나뉘며, 첫 번째는 기술 명칭만 포함된 프롬프트이고, 두 번째는 필터 크기나 파라미터 값을 함께 포함한 프롬프트이다. 이러한 두 가지 프롬프트 유형을 설계한 목적은 InstructPix2Pix가 단순한 영상처리 기법을 얼마나 정확하게 수행할 수 있는지를 평가하고, 복잡한 연산이 필요한 기법에서 정확한 수치를 해석할 수 있는지를 검증하는 데 있다. 기술 명칭만 포함된 프롬프트는 색공간 변환과 같은 기본적인 영상처리 기법을 적용할 때 사용한다. 예를 들어, RGB에서 HSV나 YCbCr로 변환하는 작업이나 그레이스케일 변환은 비교적 명확한 정의를 가지므로, 올바른 결과를 생성할 가능성이 높다.

또한, 히스토그램 변환 기법의 경우, 히스토그램 평활화와 히스토그램 정규화는 모두 명확한 연산으로, 이러한 영상처리 기법을 단순한 명칭만으로 적용할 수 있는지를 평가하기 위해 별도의 파라미터 없이 제공하였다. 이러한 유형의 프롬프트는 InstructPix2Pix가 특정 영상처리 기법의 개념을 올바르게 이해하고 있는지 평가하는 데 초점을 둔다. 밝기 및 명암 조정과 같은 영상 조정 기법에는 기술 명칭에 단순한 값들이 포함되는 프롬프트를 사용한다. 밝기를 50% 증가/감소시키는 경우, 혹은 명암을 20% 증가/감소시키는 경우에는 단순히 증가 또는 감소의 정도를 명확하게 기술하는 것만으로도 모델이 올바르게 해석하는지 확인할 수 있다. 감마 보정의 경우에도 특정한 감마 값($\gamma=2.2$)을 제공하며, 이를 통해 모델이 픽셀 값의 비선형 변환을 수행할 수 있는지 평가하고자 한다.

반면, 필터 크기나 파라미터 값과 같이 복잡한 내용을 포함한 프롬프트는 특정한 연산 방식이 요구되는 영상처리 기법을 평가하기 위해 설계하였다. 그러나 필터링 기법에서는 커널 크기와 같은 필수적인 연산 요소를 포함하여 모델이 이를 정확히 반영할 수 있는지를 검증한다. 선명화 필터의 경우에는 특정한 커널을 제공하였으며, 이를 통해 InstructPix2Pix가 선명화 연산을 수행할 때 정확한 커널을 적용하는지를 확인하고자 한다.

에지 검출을 위한 소벨 필터는 방향성과 커널 크기를 명시적으로 제공하였으며, 수평 또는 수직 방향의 경계를 올바르게 감지하는지를 평가하는 데 초점을 맞춘다. 언샤프 마스크와 같은 영상 선명화 기법에서는 가우시안 블러 강도, 선명화 강도, 마스크 적용 비율 등의 수치를 제공하여 모델이 세부적인 필터링 파라미터를 올바르게 반영하는지를 검증하고자 한다.

IV. 결과 분석 및 비교

InstructPix2Pix 모델을 활용하여 표 1의 프롬프트를 적용한 결과를 추론 후, 원본 영상과 비교하여 PSNR(Peak Signal-to-Noise Ratio) 및 SSIM(Structural Similarity Index) 값을 측정하였으며 결과는 표 2와 같다. PSNR은 영상의 신호 대 잡음비를 평가하는 지표로, 값이 클수록 원본과 변환된 영상의 차이가 적음을 의미한다. 반면, SSIM 값은 영상의 구조적 유사도를 평가하며, 1에 가까울수록 원본과 구조적으로 유사한 영상을 생성했음을 의미한다. 그림 1은 #1-#6, 그림 2는 #7-#13, 그림 3은 #14-#20까지의 변환 결과를 나타낸다. 각 그림에서 1열은 InstructPix2Pix의 결

과를, 2열은 OpenCV 기반의 영상처리 알고리즘을 사용한 결과를 나타내며, 상위에 있는 그림은 원본 사진을 나타낸다.

표 2. 결과 분석을 위한 정량적 평가
Table 2. Quantitative evaluation for result analysis

		PSNR	SSIM
Color space transformation	#1	15.74	0.649
	#2	9.51	0.645
	#3	6.95	0.274
	#4	11.82	0.625
	#5	13.14	0.659
	#6	18.95	0.711
Brightness & contrast adjustment	#7	5.68	0.422
	#8	13.27	0.270
	#9	13.57	0.610
	#10	12.95	0.582
Histogram transformation	#11	16.14	0.614
	#12	17.34	0.595
Filtering	#13	17.82	0.693
	#14	8.59	0.163
	#15	5.87	0.027
	#16	10.11	0.338
	#17	17.45	0.595
	#18	17.15	0.416
	#19	15.81	0.349
	#20	13.87	0.371

표 1. 채택한 영상처리 알고리즘을 적용하기 위한 프롬프트
Table 1. Prompts for applying the selected image processing algorithms

	Prompts contents	
Color space transformation	#1	Generate an image with the red (R) channel emphasized.
	#2	Generate an image with the green (G) channel emphasized.
	#3	Generate an image with the blue (B) channel emphasized.
	#4	Generate an image converted from the RGB color space to the HSV color space.
	#5	Generate an image converted from the RGB color space to the YCbCr color space.
	#6	Generate an image with a grayscale filter applied.
Brightness & contrast adjustment	#7	Generate an image with brightness increased by 50%.
	#8	Generate an image with brightness decreased by 50%.
	#9	Generate an image with gamma correction (gamma=2.2) applied.
	#10	Generate an image with contrast increased by 20%.
	#11	Generate an image with contrast decreased by 20%.
Histogram transformation	#12	Generate an image with histogram equalization.
	#13	Generate an image with histogram normalization.
Filtering	#14	Generate an image with a sharpening filter applied (kernel = [[0, -1, 0], [-1, 5, -1], [0, -1, 0]]).
	#15	Generate an image with a Sobel edge detection filter applied (kernel size=3×3, direction=X-axis)
	#16	Generate an image with unsharp masking applied (sigma=1.5, strength=1.2, mask ratio=0.7).
	#17	Generate an image with a Gaussian blur applied (sigma=2, kernel size 5×5).
	#18	Generate an image with a median filter applied (kernel size=3×3).
	#19	Generate an image with a bilateral filter applied (sigma_space=10, sigma_color=25, filter size=9×9).
	#20	Generate an image with a high-boost filter (A=2.0, Gaussian blur kernel=5×5) applied.

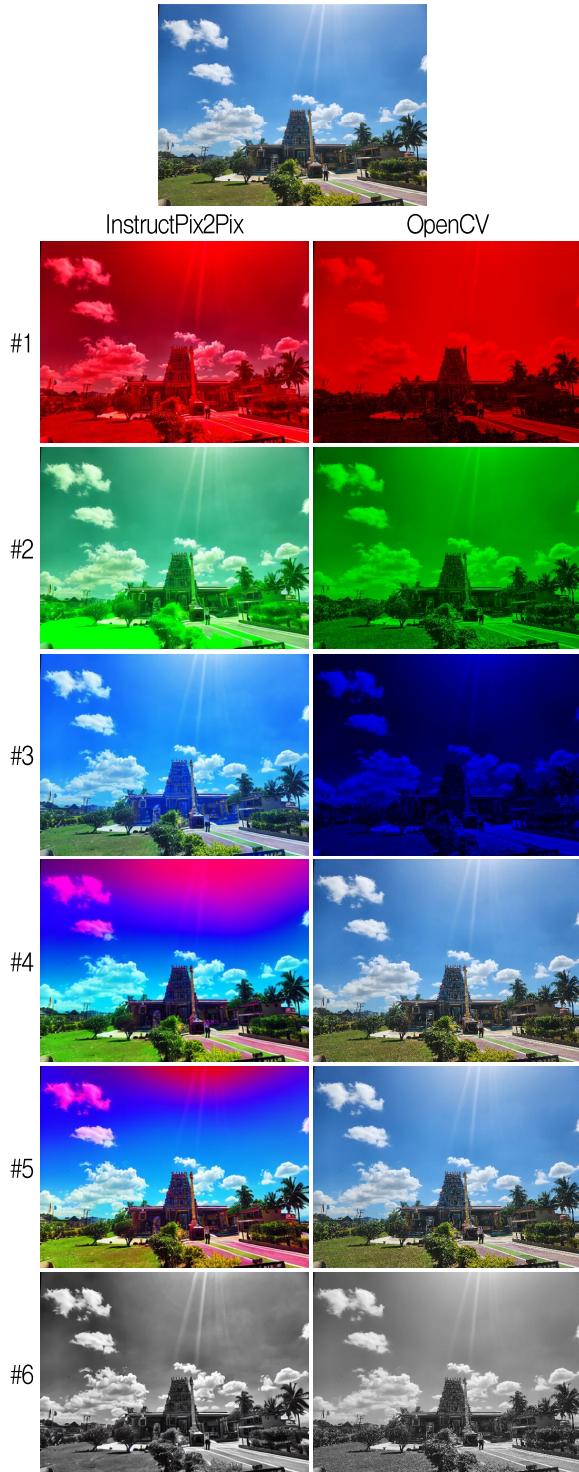


그림 1. 프롬프트에 기반 출력 영상의 비교 (#1-#7)
 Fig. 1. Comparison of output results based on prompt (#1-#7)

#1에서는, R 채널을 강조한 결과 PSNR은 15.74, SSIM은 0.649로 나타났다. InstructPix2Pix가 R 채널로의 변환을 수행하는 과정에서 원본의 구조적 유

사성을 유지하면서도 색상의 강도를 조정했음을 확인할 수 있다. #2에서는, G 채널을 강조한 결과 PSNR은 9.51, SSIM은 0.645로 나타났다. R 채널 강조보다 PSNR이 낮으며, 이는 모델이 G 채널을 강조하는 과정에서 차이가 발생했음을 시사한다.

#3에서는, B 채널을 강조한 결과 PSNR은 6.95, SSIM은 0.274로 기록되었다. SSIM이 매우 낮다는 점에서 원본과의 구조적 차이가 상당히 크며, B 채널을 분리하는 과정에서 원본과의 유사성을 충분히 유지하지 못했음을 알 수 있다. #4와 #5는 색공간 변환 후 비교를 위해 다시 RGB로 변환하여 나타낸다. #4에서 HSV 색공간으로 변환한 결과, PSNR은 11.82, SSIM은 0.625로 나타났다. 원본과의 유사성이 일정 부분 유지되었으나, 색조 및 채도 변환 과정에서 차이가 발생했다. #5에서는, YCbCr 색공간으로 변환한 결과 PSNR은 13.14, SSIM은 0.659로 나타났다. HSV 변환보다 PSNR이 높으며, 명암 대비 및 색상 정보를 비교적 안정적으로 변환했음을 의미한다. #6에서는, 그레이스케일 변환을 수행한 결과 PSNR은 18.95, SSIM은 0.711로 나타나 전체 실험 중 가장 높은 성능을 기록하였다. 이는 명암 정보만을 이용한 변환을 가장 잘 수행했음을 확인할 수 있다.

#7에서는, 밝기를 50% 증가시킨 결과 PSNR은 5.68, SSIM은 0.422로 매우 낮은 값을 기록하였다. 그리고 #8에서는, 밝기를 50% 감소시킨 결과 PSNR은 13.27, SSIM은 0.270으로 측정되었다. PSNR은 밝기 증가보다 높았지만, SSIM이 낮아 구조적 유사성이 유지되지 못했음을 확인할 수 있다. 실제 밝기의 증가 및 감소는 발생한 것으로 보이지만, 정량적으로 제시한 수치만큼은 변화하지 못한 것으로 확인된다. #9에서는, 감마 보정을 수행한 결과 PSNR은 13.57, SSIM은 0.610으로 나타났다. 모델이 감마 보정을 비교적 안정적으로 수행했음을 시사한다. #10에서는, 명암을 20% 증가시킨 결과 PSNR은 12.95, SSIM은 0.582로 나타났다. 명암 증가 과정에서 색상 왜곡이 발생했으며, 구조적 유사성이 감소한 것으로 보인다. #11에서는, 명암을 20% 감소시킨 결과 PSNR은 16.14, SSIM은 0.614로 측정되었다. 명암 감소의 경우 상대적으로 구조적 특성이 유지되면서도 정보 손실이 적었음을 확인할 수 있다.

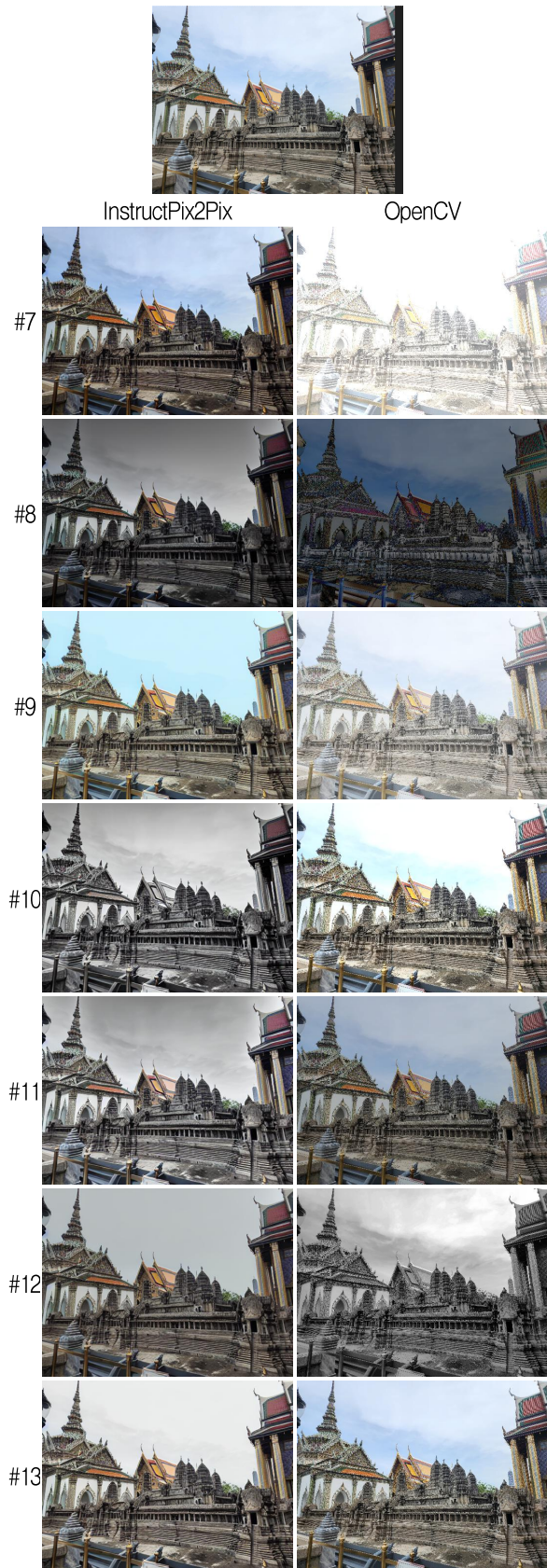


그림 2. 프롬프트 기반 출력 영상 비교 (#8-#13)
Fig. 2. Comparison of output results based on prompt (#8-#13)



그림 3. 프롬프트 기반 출력 영상 비교 (#14-#20)
Fig. 3. Comparison of output results based on prompt (#14-#20)

#12에서는, 히스토그램 평활화를 수행한 결과 PSNR은 17.34, SSIM은 0.595로 나타났다. InstructPix2Pix가 명암 대비를 조정하는 과정에서 원본과의 유사성을 일정 부분 유지하였지만, 색상과 구조적 면에서 차이가 발생하였음을 의미한다. #13에서는, 히스토그램 정규화를 수행한 결과 PSNR은 17.82, SSIM은 0.693으로 측정되었다. 전체 실험 중 두 번째로 높은 성능을 기록하였으며, 명암 대비를 조정하는 작업에서 InstructPix2Pix가 비교적 안정적인 결과를 생성했음을 알 수 있다. 하지만, 실제 영상처리 알고리즘의 결과와 비교하였을 때, 하늘 부분에서 색상의 차이가 큰 것을 확인할 수 있다.

#14에서는, 선명화 필터를 적용한 결과 PSNR은 8.59, SSIM은 0.163으로 나타났다. 모델이 선명화라는 키워드를 해석하지 못했음을 확인할 수 있다. #15에서는, 소벨 엣지 검출을 수행한 결과 PSNR은 5.87, SSIM은 0.027로 전체 실험 중 가장 낮은 성능을 기록하였다. 이는 InstructPix2Pix가 엣지 검출을 정확하게 수행하지 못했음을 의미하며, 전통적인 영상처리 기법과 상당한 차이가 발생했음을 알 수 있다. #16에서는, 언샤프 마스킹을 적용한 결과 PSNR은 10.11, SSIM은 0.338로 측정되었다. 모델이 이 기법을 효과적으로 수행하지 못했으며, 단순한 대비 조정과 같은 방식으로 처리했을 가능성이 높다. #17에서는, 가우시안 블러를 적용한 결과 PSNR은 17.45, SSIM은 0.595로 나타났다. InstructPix2Pix가 블러 효과가 적용되었음을 확인할 수 있으나, 파라미터에 대한 고려가 배제된 것으로 보인다.

#18에서는, 중간값 필터를 적용한 결과 PSNR은 17.15, SSIM은 0.416으로 측정되었다. 노이즈 제거 효과를 반영하는 과정에서 원본 대비 일부 차이가 발생했으나, 필터링 작업이 비교적 안정적으로 수행되었음을 의미한다. #19에서는, 양방향 필터를 적용한 결과 PSNR은 15.81, SSIM은 0.349로 나타났다. 경계를 유지하면서 블러링을 적용하는 과정에서 구조적 유사성이 다소 감소했으며, 색상 차이가 발생하였다. #20에서는, 하이부스트 공간 필터를 적용한 결과 PSNR은 13.87, SSIM은 0.371로 측정되었다. 구조적 유사성이 낮아졌으며, 모델이 고주파 강조 필터를 정확하게 반영하지 못했음을 시사한다.

V. 결론 및 향후 과제

본 논문에서는 InstructPix2Pix 모델을 중심으로 프롬프트 기반 영상처리의 가능성을 평가하고 정량적으로 분석하였다. 실험 결과, 그레이스케일 변환이 전체 실험에서 가장 높은 성능을 기록하며 가장 안정적인 변환을 수행했음이 확인되었다. 반면, 소벨 엣지 검출과 선명화 필터와 같은 로컬 연산 기반의 영상처리 기법은 낮은 성능을 기록하였다. 이는 픽셀 단위의 수학적 연산을 직접 수행하는 것이 아니라, 학습된 패턴을 기반으로 영상을 변환하는 방식으로 동작하기 때문이다.

이러한 한계를 극복하기 위해서는 모델이 영상처리 알고리즘의 실제 연산 과정을 더 잘 반영할 수 있도록 학습 데이터를 재구성하여 파인튜닝하는 방식으로 개선할 필요가 있다. 구체적으로, 소벨 엣지 검출, 샤프닝, 블러링, 그리고 그레이스케일 변환 등 다양한 영상처리 알고리즘이 적용된 이미지와 원본 이미지의 쌍을 대량으로 구축하여 모델이 각 알고리즘의 효과를 학습하도록 유도할 수 있다. 또한, 로컬 연산의 특성을 반영하기 위해 국소 영역별로 연산 패턴을 정확히 학습하도록 하고, 그라디언트 손실을 추가해 경계선 보존 능력을 높이는 방법도 고려할 수 있다.

Acknowledgement

2025년도 한국정보기술학회 하계종합학술대회에서 발표한 논문(텍스트 프롬프트를 이용한 영상처리 기법의 적용 가능성 분석)[11]을 확장한 것임

References

- [1] H. Guan, P. T. Yap, A. Bozoki, and M. Liu, "Federated learning for medical image analysis: A survey", *Pattern Recognition*, pp. 110424. Jul. 2024. <https://doi.org/10.1016/j.patcog.2024.110424>.
- [2] J. Lee, R. Lee, S. Hong, and Y. Kim, "Image Processing System based on Deep Learning for Safety of Heat Treatment Equipment", *Journal of*

The Institute of Internet, Broadcasting and Communication (IIBC), Vol. 20, No. 6, pp.77-83, Dec. 2020. <https://doi.org/10.7236/IIBC.2020.20.6.77>.

[3] J. Si and S. Kim, "Defect Detection of Wood Grain Images with Image Gradient Applied to AE-based generation", IEEE International Conference on Knowledge Innovation and Invention (ICKII), pp. 335-337, Aug. 2023. <https://doi.org/10.1109/ICKII58656.2023.10332785>.

[4] C. Song, M. Jang, Y. Park, and S. Chun, "Audit Inspection Method for Data Quality of Generative AI Information System", Journal of The Institute of Internet, Broadcasting and Communication (IIBC), Vol. 24, No. 6, pp. 1-7, Dec. 2024. <https://doi.org/10.7236/IIBC.2024.24.6.1>.

[5] S. Bengesi, et al., "Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers", IEEE Access. Vol. 12, pp. 69812-69837, Jul. 2024. <https://doi.org/10.1109/ACCESS.2024.3397775>.

[6] J. Si, et al., "Generating and Integrating Diffusion Model-Based Panoramic Views for Virtual Interview Platform", IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAET), Kota Kinabalu, Malaysia, pp. 343-348, Aug. 2024. <https://doi.org/10.1109/IICAET62352.2024.10730450>.

[7] G. Bansal, A. Nawal, V. Chamola, and N. Herencsar, "Revolutionizing visuals: the role of generative AI in modern image generation", ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 20, No. 11, pp. 1-22, Nov. 2024. <https://doi.org/10.1145/3689641>.

[8] J. Si and S. Kim, "Research on Designing Actual Virtual Interviewers based on Analysis of Image Generative AI in Web Platforms", Journal of Korean Institute of Information Technology, Vol. 22, No. 11, 145-154. Nov. 2024. <https://doi.org/10.14801/jkiit.2024.22.11.145>.

[9] T. Brooks, A. Holynski, and A. A. Efros,

"Instructpix2pix: Learning to follow image editing instructions", Proc. Of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, pp. 18392-18402, Jun. 2023.

[10] A. Shagidanov, H. Poghosyan, X. Gong, Z. Wang, S. Navasardyan, and H. Shi, "Grounded-instruct-pix2pix: Improving instruction based image editing with automatic target grounding", IEEE International Conference on Acoustics, Speech and Signal Processing, Seoul, Korea, pp. 6585-6589, Apr. 2024.

[11] J. Si and S. Kim, "Feasibility Analysis of Applying Image Processing Techniques Using Text Prompts", Proceedings of KIIT Conference, Jeju, Korea, pp. 121-121, Jun. 2025.

저자소개

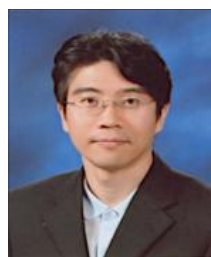
시 종 욱 (Jongwook Si)



2020년 8월 : 국립금오공과대학교
컴퓨터공학과(공학사)
2022년 2월 : 국립금오공과대학교
컴퓨터공학과(공학석사)
2022년 3월 ~ 현재 :
국립금오공과대학교 컴퓨터·AI
융합공학과 대학원 박사과정

2023년 9월 ~ 현재 : 국립금오공과대학교 컴퓨터공학부 강사
관심분야 : 영상처리, 컴퓨터비전, 영상 복원, 이상 감지,
프롬프트엔지니어링, 영상 기반 생성형 AI

김 성 영 (Sungyoung Kim)



1994년 2월 : 부산대학교
컴퓨터공학과(공학사)
1996년 2월 : 부산대학교
컴퓨터공학과(공학석사)
2003년 8월 : 부산대학교
컴퓨터공학과(공학박사)
2004년 3월 ~ 현재 : 국립금오공과

대학교 컴퓨터공학부 컴퓨터공학전공 교수
관심분야 : 영상처리, 컴퓨터비전, 기계학습, 딥러닝,
메타버스