

이미지 비율 보정을 위한 딥러닝 기반 아웃페인팅 응용에 대한 연구

장수현, 최윤영, 강지헌

덕성여자대학교

tngus7238@duksung.ac.kr, cyy0509@duksung.ac.kr, jhkang@duksung.ac.kr

A Study on Image Aspect Ratio Correction Using Deep Learning - Based Outpainting

SuHyeon Jang, YunYeong Choi, Jiheon Kang

School of Software, Duksung Women's University

요약

본 논문은 이미지 비율 보정을 위한 딥러닝 기반 아웃페인팅(outpainting) 응용 연구를 수행하였다. 제안하는 방법은 입력 이미지의 핵심 정보를 유지하면서 다양한 목표 비율에 따라 외곽 영역을 생성할 수 있도록 아웃페인팅 모델을 설계하고, 비율 조건에 따른 생성 특성을 분석하는 데 목적이 있다. 이를 통해 기존의 고정된 비율 중심 이미지 확장 방식의 한계를 극복하고, 데이터 표현 구조의 유연성을 확보하고자 한다. 더 나아가 본 연구는 제안한 아웃페인팅 기법을 MFCC(Mel-Frequency Cepstral Coefficient) 기반 이미지 표현의 음성 신호에 적용함으로써, 이미지 데이터에 국한되지 않은 신호 처리 영역으로의 확장 가능성을 탐색한다. 실험을 통해 제안 기법의 유효성과 다양한 데이터 형태에 대한 적용 가능성을 검증한다.

I. 서론

딥러닝 기반 이미지 생성 기술의 발전과 함께 아웃페인팅(outpainting)은 입력 이미지의 외곽 영역을 자연스럽게 확장하는 방법으로 주목받고 있다. QueryOTR(Query-based Outpainting Transformer)는 쿼리 기반 구조를 활용한 대표적인 아웃페인팅 모델로, 우수한 이미지 생성 성능을 보여왔다. 그러나 기존 QueryOTR는 정사각형(1:1) 비율의 이미지에만 아웃페인팅이 적용되는 구조적 한계를 지니고 있어, 다양한 이미지 비율을 요구하는 데이터 환경에 대한 적용이 제한적이다.

본 논문은 이러한 한계를 개선하여 QueryOTR 기반 아웃페인팅을 다양한 종횡비(aspect ratio)에 적용하는 것을 목표로 한다. 이를 위해 Scenery6000 데이터셋과 Place365 데이터셋의 mountain 카테고리에서 수집한 이미지를 활용하여 총 11,040장의 유효 이미지로 구성된 데이터셋을 구축하였으며, 이를 학습용 8,832장과 검증용 2,208장으로 분할하였다. 제안하는 방법은 기존 1:1 비율에 한정된 아웃페인팅을 16:9, 9:16, 4:5 비율로 확장함으로써, 이미지 비율 보정 및 구조적 확장 측면에서 보다 유연한 이미지 생성이 가능하도록 한다.

본 연구를 통해 다양한 비율 조건에서의 아웃페인팅 가능성을 검증하고, QueryOTR 모델의 활용 범위를 확장하고자 한다.

II. 본론

본 연구는 기존 1:1 비율을 가정한 이미지 아웃페인팅 모델(QueryOTR 기반)을 수정하여, 다양한 종횡비(1:1, 16:9, 9:16, 4:5)에 대해 비율 보정을 수행하는 딥러닝 기반 아웃페인팅 응용 가능성을 탐구한다. 제안 방식의 핵심은 원본 이미지의 가시 영역을 최대한 보존한 상태에서 보이지 않는 영역을 의미적으로 일관되게 재생성하는 데 있다. 모든 실험에서 입력 이미지의 중앙 영역은 동일하게 유지하였으며 모델은 관측되지 않은 외곽 영역을 완전히 새로운 시각 정보로 생성하도록 설정하였다. 이는 단순한 패딩이나 보간이 아닌, 손실된 시각 정보를 재구성하는 생성 문제로 정의

된다.



Figure 1. Aspect-Ratio-based Outpainting Results

Figure 1은 서로 다른 비율에 대해 아웃페인팅된 결과를 시각적으로 비교한 예시이다. 모든 비율에서 원본 영역과 생성 영역의 경계가 명확히 드러나지 않으며, 이는 모델이 단순히 경계를 기준으로 확장하는 것이 아니라 전역적인 문맥을 고려한 생성 전략을 학습했음을 의미한다.

Aspect Ratio	Epoch	Inception Score	Original Region Preservation	Boundary Region FID	CLIP Similarity
1:1	100	3.449	0.011	16.490	0.884
16:9	100	3.477	0.013	19.061	0.936
9:16	100	4.553	0.014	15.898	0.935
4:5	100	4.069	0.031	48.368	0.770

Table 1. Quantitative Evaluation Across Aspect Ratios

Table 1은 각 비율별 아웃페인팅 결과에 대해 정량적 지표를 비교한 결과를 나타낸다. 평가 지표로는 생성 이미지의 다양성과 품질을 평가하기 위한 IS(Inception Score), 원본 이미지 영역이 얼마나 잘 유지되었는지를 측정하는 Original Region Preservation, 생성 이미지 분포와 실제 이미지 분포 간의 차이를 평가하는 FID, 그리고 생성 결과와 입력 이미지 간의 유사도를 측정하는 CLIP Similarity를 사용하였다.

본 연구에서는 아웃페인팅을 단순히 이미지의 크기를 늘리는 문제로 보지 않고, 손실된 정보를 재생성하는 문제로 해석한다. 원본 이미지에서 잘려 나가거나 관측되지 않은 영역은 정보 관점에서 보면 부분적인 신호 손

실에 해당하며, 아웃페인팅은 이를 복원하는 과정으로 볼 수 있다. 이러한 관점은 신호 처리 분야의 주파수 해석과 자연스럽게 연결된다. 이미지에서 중앙 영역은 주로 저주파 성분(전체 구조, 레이아웃)을 안정적으로 포함하고 있는 반면, 경계 바깥의 세부 패턴은 고주파 성분과 연관된다. 제안 방식은 저주파 문맥을 조건으로 고주파 정보를 생성하는 구조를 갖는다. 이는 음성 신호 처리에서 MFCC(Mel-Frequency Cepstral Coefficients)가 수행하는 역할과 유사한 해석이 가능하다. MFCC는 시간 영역의 음성 신호를 주파수 영역으로 변환하여, 인간 청각에 중요한 저주파 대역의 특징을 중심으로 정보를 요약한다. 마찬가지로, 본 연구의 아웃페인팅 모델은 가시 영역(저주파적 구조)을 기반으로 비가시 영역(고주파적 세부 정보)을 생성한다는 점에서, 시각 신호의 주파수적 보완 문제로 해석될 수 있다.

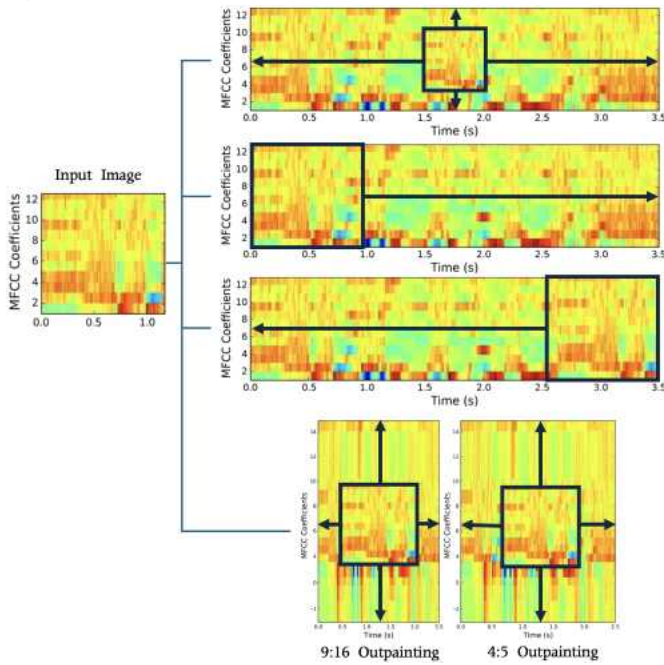


Figure 2. MFCC generation and expansion based on input signal (adapted from [2])

Figure 2는 MFCC 표현을 기반으로 한 아웃페인팅 개념을 시각적으로 나타낸 예시이다. 정사각형 형태의 MFCC 구간을 입력으로 사용하여 제한된 시간 및 주파수 정보만이 관측된 상황을 가정하며, 입력 영역을 기준으로 관측되지 않은 시간-주파수 영역을 확장적으로 생성하는 과정을 도식화하였다.

그림에서 각 영역은 서로 다른 출력 비율에 대응되는 아웃페인팅 결과를 의미하며, 이를 통해 출력 비율에 따라 시간 방향과 주파수 방향 확장 특성이 어떻게 달라지는지를 나타낸다. 특히, 세로 방향 확장이 우세한 경우는 제한된 주파수 성분으로부터 고주파 영역을 확장하는 주파수 방향 확장을, 가로 방향 확장이 우세한 경우는 제한된 시간 구간을 기반으로 미래 또는 과거 시간 영역을 예측·확장하는 시간 방향 확장을 각각 의미한다. 이러한 구성은 제한적으로 관측된 MFCC 입력만으로도 시간 및 주파수 방향 확장이 가능함을 개념적으로 보여준다.

MFCC(Mel-Frequency Cepstral Coefficient)는 음성 및 통신 신호의 주파수 특성을 시간-주파수 영역에서 효과적으로 표현하는 대표적인 특징 벡터로, 인접 프레임 간 주파수 에너지 분포가 반복적으로 나타나는 구조적 특성을 가진다. 이러한 반복성은 통신 신호 처리 관점에서 시간적 상관성(temporal correlation)과 주파수 대역 간 연속성을 내포하며, 부분적으로 손실된 신호 구간을 주변 정보로부터 추정할 수 있는 기반을 제공한다.

다.

무선 통신 환경에서는 채널 페이딩, 패킷 손실, 전송 지연 등의 요인으로 인해 신호의 일부 프레임이 관측되지 않거나 불완전하게 수신되는 상황이 빈번히 발생한다. 이러한 신호 단절(outage) 현상은 MFCC 기반 시간-주파수 이미지 상에서 특정 영역이 손실된 상태로 해석될 수 있으며, 이는 이미지 도메인에서의 결손 영역 복원 문제와 구조적으로 유사하다.

본 연구에서는 이러한 관점에서 MFCC를 이미지 형태로 변환한 후, 딥러닝 기반 아웃페인팅 기법을 적용하여 관측되지 않은 시간-주파수 영역을 복원하는 문제를 이미지 생성 관점에서 접근하였다. 특히, 입력 MFCC 이미지의 핵심 주파수 패턴을 유지한 채 외곽 영역을 확장함으로써, 과거 또는 미래 프레임에 해당하는 신호 성분을 추정하는 효과를 기대할 수 있다.

또한, 다양한 이미지 비율(1:1, 16:9, 9:16, 4:5)에 따른 아웃페인팅은 시간 축 또는 주파수 축 방향의 확장으로 해석될 수 있다. 예를 들어, 가로 방향 확장은 시간 축상의 신호 연장을, 세로 방향 확장은 주파수 대역 외 성분의 추정을 의미할 수 있으며 이는 통신 시스템에서의 신호 예측 및 대역 보간 문제와 유사한 해석이 가능하다. 이러한 분석을 통해, 제안한 방법은 단순한 시각적 이미지 확장을 넘어 통신 신호 복원 문제로의 확장 가능성을 지닌다.

III. 결론

본 연구에서는 사용자가 보유한 원본 이미지의 시각적·구조적 특성을 최대한 유지하면서 딥러닝 기반 아웃페인팅을 통해 다양한 비율(1:1, 16:9, 9:16, 4:5)로 이미지를 자동으로 확장할 수 있는 시스템의 초기 개념을 제시한다. 실험 결과, 단일 레퍼런스 모델을 활용하여 입력 이미지의 다양한 비율 조정이 가능하며, 생성된 영역은 원본 이미지와 유사한 형태를 갖도록 자연스럽게 보정됨을 확인하였다.

특히, 아웃페인팅 과정은 단순히 이미지 영역을 채우는 것을 넘어 과거-미래 정보와 저주파-고주파 신호 성분을 재생성하는 의미를 가지며, 이는 통신 신호 처리 관점에서 손실된 신호를 복원하는 과정과 유사하다. MFCC와 같은 주파수 기반 특징과 결합하면 시각적 정보뿐만 아니라 신호적 특징까지 분석하여 다양한 산업 분야에서 활용 가능성을 높일 수 있다.

결과적으로, 본 연구는 원본 이미지를 유지하면서 다양한 비율로 자동 확장하는 아웃페인팅 시스템을 대상으로 MFCC를 활용한 신호 기반 분석 가능성을 탐색한 초기 단계 연구이며, 향후 모델 성능 개선, 주파수 영역 분석 연계, 산업적 활용을 통해 보다 정교한 시각-신호 복합 생성 기술로 발전시킬 수 있는 방향성을 제시한다. 즉, 원본과 유사한 형태의 보여지지 않는 영역을 채운다는 것은 손실된 정보들을 재생성한다는 의미로, 이를 주파수 영역까지 이어서 활용할 수 있음을 보여준다.

참 고 문 헌

- [1] K. Yao, P. Gao, X. Yang, K. Huang, J. Sun, and R. Zhang, "Outpaining by Queries" arXiv preprint arXiv:2207.05312, Jul. 2022.
- [2] Prathena, "The Dummy's Guide to MFCC," Medium, Jul. 22, 2022. [Online]. Available: <https://medium.com/prathena/the-dummys-guide-to-mfcc-aceab2450fd>. [Accessed: 2026-01-08].