

히스토그램 손실함수와 순차적 작업을 이용한 CCTV 영상 화질 향상

정민교 최종인 *정제창

한양대학교

minkyo96@gmail.com, jichoi2014@naver.com, *jjeong@hanyang.ac.kr

CCTV Image Quality Enhancement using Histogram Loss and Sequential Task

Minkyoo Jeong Jongin Choi *Jechang Jeong

Hanyang University

요 약

본 논문에서는 CCTV 영상 화질을 향상하고 해상도를 높이기 위해 딥 러닝(Deep Learning)을 이용하여 잡음 제거(Denoising) 와 초해상도(Super-resolution) 작업을 수행한다. 데이터 증강(Data Augmentation)을 통한 초해상도 성능 향상을 위해서 잡음 제거 네트워크의 출력 영상을 초해상도 네트워크의 입력으로 사용하는 순차적 작업을 사용한다. 또한 딥 러닝을 이용한 영상처리에서 발생하는 평균 밝기 오차 문제를 해결하기 위한 손실함수(Loss Function)와 두 가지 이상의 순차적인 딥 러닝 작업에서 발생하는 문제점을 극복하기 위한 손실함수를 제안한다. 제안하는 손실함수는 네트워크의 출력 영상과 타겟 영상의 밝기 오차를 줄이는 것이 가능하고, 순차적 작업에서 보다 정확한 모델 성능 판단이 가능하다.

1. 서론

CCTV 의 범죄 예방과 추적 효과로 인해 CCTV 의 설치 및 증가 대수는 2018 년 이후 꾸준히 증가하고 있다. 최근 CCTV 에 대한 중요성의 부각으로 고화질 CCTV 가 일부 설치되고 있지만 큰 비용이 발생하고, 기존에 설치 되어있는 CCTV 는 추적 효과가 떨어진다. CCTV 영상의 경우 24 시간 촬영하는 특성으로 인해 저해상도로 저장되는 경우가 많고, 한정된 저장 자원 때문에 물체의 움직임이 존재할 때 프레임을 조절하는 방식을 사용 중이다. 저화질 CCTV 영상의 화질을 개선하고 저장 자원의 문제를 해결하기 위해 본 논문에서는 후처리 알고리즘을 통해 영상의 화질을 개선하고 해상도를 높이는 방법을 제안한다.

CCTV 영상의 화질을 개선하기 위해 잡음 제거

(Denoising)와 초해상도(Super-resolution) 기법을 순차적으로 사용한다. 잡음 제거를 하는 이유는 날씨, 조도, 미세먼지 등의 환경적 요소로 인해 CCTV 영상에 잡음이 많이 존재하기 때문이다. 잡음 제거를 통해 화질을 개선할 수 있고 객체의 식별률을 높일 수 있다. 범죄 추적의 특성상 객체의 선명도가 중요하고, 같은 크기의 화면에서 영상을 표현할 때 해상도가 높을수록 선명하다. 초해상도 작업을 통해 고해상도의 선명한 영상을 얻을 수 있다.

영상 잡음 제거와 초해상도 성능을 향상하기 위해 두 가지 손실함수(Loss Function)를 제안한다. 첫 번째는 히스토그램(Histogram) 손실함수이다. 딥 러닝(Deep Learning)을 이용한 영상처리를 수행하면 히스토그램의 오차가 발생한다. 기존 L1 손실함수에 영상의 히스토그램으로부터 얻은 L1 손실함수를 결합하여 히스토그램 손실함수를 구한다. 이를 통해 타겟(Target) 영상과 출력 영상의 히스토그램 오차를 최소화하여 영상의 밝기

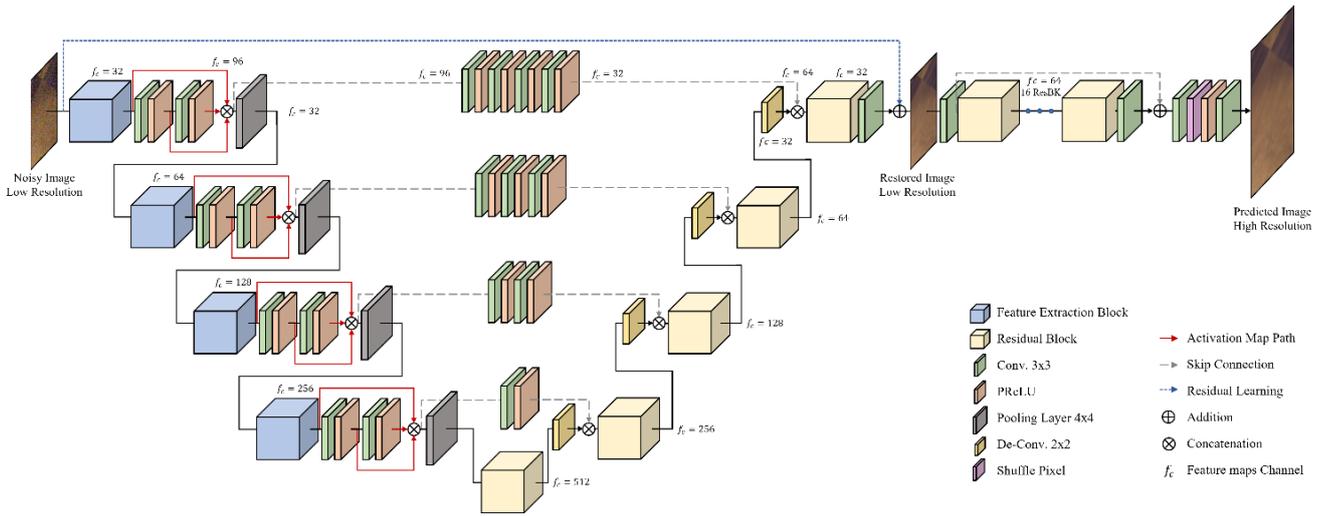


그림 1. 잡음 제거와 초해상화를 위한 네트워크 구조

변화를 억제한다. 두 번째는 순차적 작업(Sequential Task) 손실함수로 두 가지 딥 러닝 네트워크를 효율적으로 학습하기 위한 손실함수이다. 잡음 제거와 초해상도 네트워크를 각각 학습하는 것보다 잡음 제거 네트워크의 출력으로 초해상도 네트워크를 학습할 경우 초해상도 네트워크의 입력은 잡음 제거 네트워크를 통해 데이터 증강(Data Augmentation) 효과를 얻는다. 기존 손실함수로 순차적 작업을 학습할 경우 두 가지 작업의 학습 방향을 제어하지 못하기 때문에 잡음 제거와 초해상도 네트워크의 출력을 통해 L1 손실함수를 각각 계산하고 두 손실함수의 차를 통해 초해상도 네트워크의 학습 방향을 결정한다.

본 논문의 구성은 다음과 같다. 2 장에서는 사용한 잡음 제거와 초해상도 네트워크 구조를 설명하고, 히스토그램 손실함수와 순차적 작업 손실함수에 대해 설명한다. 3 장에서는 기존 방법과 제안한 손실함수를 이용한 결과를 비교하고, 4 장에서 결론을 맺는다.

2. 제안하는 방법

CCTV 에 존재하는 가우시안(Gaussian) 잡음과 ISO 잡음을 제거하기 위한 잡음 제거와 고해상도 영상을 얻기 위한 초해상화가 딥 러닝 네트워크를 통해 수행된다. 초해상화가 잡음 제거보다 먼저 수행될 경우 원 영상대비 4 배 크기의 해상도에서 잡음제거를 수행해야하기 때문에 연산량이 증가하고 초해상도 작업에서 잡음의 분포가 변경되기 때문에 잡음 제거 작업을 먼저 수행한다.

각 네트워크는 방대한 CCTV 데이터를 처리하기 위해 적은 파라미터(Parameter)를 사용하는 구조가 요구된다. 잡음 제거 네트워크는 AMCNet [1]의 수축 경로(Contraction Path)에 위치한 어텐션 블록(Attention Block)을 제거하여 사용하고, 초해상도 네트워크는 EDSR [2]의 구조를 사용한다. 두 네트워크 배치는 그림 1 과 같이 잡음 제거 네트워크의 출력이 초해상도 네트워크의 입력으로 사용되도록 구성한다.

2.1 히스토그램 손실함수

제안하는 히스토그램 손실함수는 딥 러닝 네트워크의 출력 영상은 타겟 영상과 평균 밝기의 차이가 존재한다. 히스토그램으로 분석한 결과 이러한 차이는 히스토그램 분포의 이동을 야기한다. 히스토그램의 오차는 출력 영상과 타겟 영상의 유사도를 측정하는 또 하나의 지표이다. 기존 손실함수는 같은 위치의 픽셀(Pixel)에 대해 오차를 계산한다. 이러한 공간영역의 오차만을 사용하는 것은 히스토그램 오차를 발생시킬 수 있다. 본 논문에서는 히스토그램 영역에서 오차를 계산하여 영상 간 유사도를 측정할 수 있는 손실함수로 사용한다. 이러한 손실함수를 통해 두 영상간 히스토그램 오차를 줄인다. 제안하는 히스토그램 손실함수 $L_{Histogram}$ 는 공간 영역의 손실함수 L_{Pixel} 와 정규화된 히스토그램 오차 E_{Hist} 의 가중합으로 계산된다.

$$L_{Histogram} = L_{Pixel} + \lambda \frac{E_{Hist}}{h * w} \quad (1)$$

$$E_{Hist} = \sum_0^{255} \|y_{Hist} - f(x)_{Hist}\| \quad (2)$$

λ 는 가중치를 결정하는 변수이고, h 와 w 는 영상의 높이와 너비이다. 히스토그램 오차 E_{Hist} 는 원 영상 히스토그램 y_{Hist} 와 출력 영상 히스토그램 $f(x)_{Hist}$ 의 절대 오차 합이다. 색 밝기 오차 정보가 포함된 히스토그램 손실함수를 이용하여 밝기 오차를 줄일 수 있다.

2.2 순차적 작업 손실함수

두 가지 딥러닝 모델을 동시에 학습할 경우 잡음 제거 네트워크가 초해상도 네트워크의 입력 영상을 변형시키고 데이터 증강 효과를 얻을 수 있다. 순차적으로 진행되는 두 가지 작업을 동시에 학습할 때, 손실함수 한 개로 최적화하는 경우 손실함수가 학습 방향을 결정하지 못한다. 이는 두 네트워크가 종단 간 학습(End to End Learning)으로 구성되어 있지 않기 때문에 손실함수가 특정 네트워크의 결과에 의해 편향될 수 있다. 반대로 두 가지 작업에 대한 손실함수를 각각 사용하는 경우에는 초해상도 작업의 입력으로 잡음이 제거된 영상이 사용되기 때문에 학습이 진행됨에 따라 입력과 타겟 영상의 손실함수 값이 작아진다. 따라서 초해상도의 학습 방향성이 잘못되더라도 잡음 제거 네트워크의 출력에 따라 손실함수 값이 작아질 수 있다.

앞서 언급한 문제들을 해결하기 위해 논문에서는 순차적 작업 손실함수를 제안한다. 순차적 작업 손실함수는 잡음 제거 네트워크의 손실함수 L_{DN} 와 초해상도 네트워크의 손실함수 L_{SR} 로 구성되며 초해상도 손실함수 L_{SR} 은 전체 네트워크의 손실함수 L_{Total} 을 이용하여 다음과 같이 계산된다.

$$L_{SR} = L_{Total} - L_{DN} \tag{3}$$

3. 실험 결과 및 분석

실험 결과 및 분석의 구성은 다음과 같다. 3.1 절에서 실험 환경에 대해 설명한다. 3.2 절에서는 제안하는 손실함수와 네트워크의 성능을 평가한다.

3.1 실험 환경

실험을 위해 GPU는 NVIDIA GeForce RTX 3090을 사용했고, CPU는 Intel Core i9-10920X를 사용했다. 프레임워크는 PyTorch를 사용하였고, DIV2K와 SIDD [3]를 학습 데이터셋으로 사용했다. 학습 과정에서 네트워크 입력으로는 256×256 크기의 패치를 사용했고, 학습률을 초깃값 $1e-4$ 부터

표 1. 순차적 작업 손실함수에 따른 초해상화 성능 비교

Loss Function	Independently Learning	DN L1 + SR L1	Sequential task L1
PSNR [dB]	32.39	31.12	32.74
SSIM	0.924	0.921	0.925

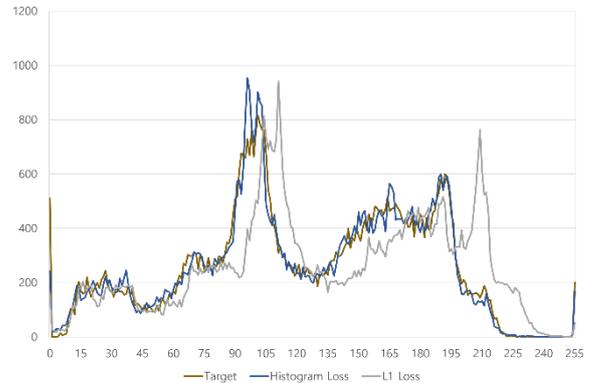


그림 2. 손실함수에 따른 히스토그램 비교

$1e-7$ 까지 코사인 파형에 따라 감소시키며, 총 250 에폭(Epoch) 동안 진행되었다. 옵티마이저(Optimizer)는 아담(Adam)을 사용했다. 잡음 제거 네트워크 학습에는 10에서 20 사이의 분산을 갖는 가우시안 잡음을 랜덤하게 추가했다. 초해상도 네트워크 학습은 높이와 너비를 두 배 크기로 확장하도록 학습했다. 손실 함수는 히스토그램 손실함수가 포함된 순차적 손실함수를 사용했다. 네트워크와 손실함수의 성능 평가를 위해 Div2k 테스트 데이터 셋과 이상행동 CCTV 데이터 셋 [4]을 이용하여 객관적 화질 평가 지표인 PSNR 과 주관적 화질 비교를 통해 평가를 진행하였다.

3.2 손실함수 성능 비교

그림 2는 원본 영상과 히스토그램 손실 함수, L1 손실함수를 사용한 출력 영상의 히스토그램을 각각 갈색, 남색, 회색으로 나타낸 그래프이다. 그림에서 히스토그램 손실 함수를 사용한 네트워크는 원본 영상의 히스토그램에 더 근접하게 표현된다는 것을 알 수 있다. 또한 히스토그램 오차를 수치상으로 확인할 때 히스토그램 손실 함수를 사용한 네트워크는 기존 L1 손실 함수를 사용할 때보다 오차가 적다. 이를 통해 히스토그램 로스를 사용하면 밝기 오차를 줄일 수 있다는 것을 알 수 있다.

표 1은 순차적 작업 손실함수 사용 유무에 따른 초해상도 작업의 성능을 객관적 화질 지표인 PSNR 과 SSIM 을 통해 비교한 결과이다. 두 작업을 독립적으로 학습한 것보다 제안한 순차적 작업 손실함수를 이용하여 학습한 것이 높은 성능을

표 2. 가우시안 잡음이 포함된 Div2k 데이터 셋에서의 잡음 제거와 초해상화의 객관적 성능

Task	Noisy ($\sigma = 15$)	Denoised	Denoised + Bicubic	Denoised + Super-resolution
PSNR [dB]	24.46	37.83	30.98	32.74
SSIM	0.838	0.966	0.901	0.925

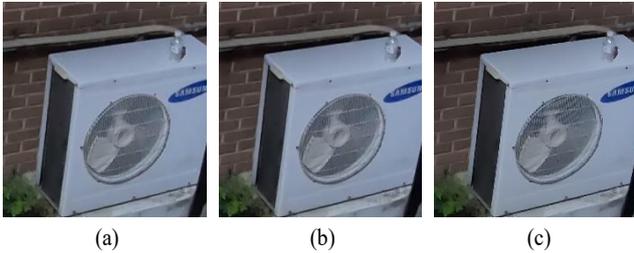


그림 3. 잡음 제거와 초해상화의 주관적 화질 향상
(a) 원본, (b) 잡음 제거된 영상, (c) 초해상화 영상

보이고, 초해상도 작업의 손실 값을 초해상도 네트워크의 출력과 타겟 영상으로부터 직접 구하는 경우 독립적으로 학습한 것보다 낮은 성능을 보인다.

3.3 CCTV 영상 향상 성능

그림 3은 제안한 손실함수와 네트워크를 평가하기 위해 실제 CCTV 영상 데이터 셋인 이상행동 CCTV 데이터 셋 영상을 부분 확대한 영상이다. 그림 3(a)의 벽돌, 물병, 로고 등에 존재하는 잡음이 잡음 제거한 영상인 그림 3(b)에서는 제거된다. 그림(b)의 초해상화 영상인 그림 3(c)는 실외기 팬 부분의 격자 형태가 뚜렷해 진 것을 확인할 수 있다.

표 2 는 네트워크의 절제 연구 결과이다. 잡음 제거 네트워크의 성능은 PSNR 기준 37.83dB 이고, SSIM 은 0.966 이다. 잡음 제거 네트워크의 출력 영상을 입력하여 얻은 초해상화 네트워크의 출력 영상은 Bicubic 을 이용하여 초해상화한 영상보다 PSNR 수치가 1.76dB 이 높고, SSIM 값은 0.024 이 높다.

4. 결론

본 논문에서는 히스토그램, 순차적 작업 손실함수를 적용하여 CCTV 영상의 잡음을 제거하고, 초해상화하는 딥 러닝

네트워크를 제안한다. 히스토그램 손실 함수는 잡음 제거와 초해상화 작업에서 영상의 색 밝기 오차를 이용하여 영상의 밝기 변화를 억제한다. 순차적 작업 손실함수는 순차적 작업을 가능하게 하여 데이터 증강을 통해 초해상도 성능을 향상한다. 그 결과 영상의 해상도가 4 배 확장되고, 영상 내의 사람의 손, 얼굴 등에서 물체의 디테일과 질감을 보존하면서 잡음이 효과적으로 제거된다. 제안하는 기법은 딥 러닝을 통해 동영상 처리하는 과정 중 발생하는 플리커 현상을 억제하는 기법으로 적용될 수 있고, 또한 지능형 CCTV 등에 적용하여 객체 검출의 성능의 향상 가능성을 제공한다.

감사의 글

이 연구는 2022 년도 산업통상자원부 및 한국산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임('20013726')

참고문헌

- [1] M. Jeong, K. Lee and J. Jeong, "Activation Map Concatenation Network for Denoising" In *the Image Processing and Image Understanding(IPIU) Workshops*, Feb 2022.
- [2] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. "Enhanced deep residual networks for single image super-resolution," In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.
- [3] A. Abdelhamed et al., "A High-Quality Denoising Dataset for Smartphone Cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1692-1700, 2018.
- [4] 이상행동 CCTV 영상 데이터 셋, <https://aihub.or.kr/aidata/139> (accessed May. 1, 2022).