

## 자기 지도 적응을 통한 동영상 초해상도 기법

유진수 김태현

한양대학교

{jinsuyoo,taehyunkim}@hanyang.ac.kr

## Video Super-Resolution via Self-Supervised Adaptation

Yoo, Jinsu Kim, Tae Hyun

Hanyang University

## 요약

최근 많은 단일 영상 초해상도 네트워크에서 입력 저 화질 영상 자체의 내부 정보를 테스트 타임에 이용하여 파라미터를 업데이트하는 방법을 통해 높은 성능 향상을 이루어냈다. 본 원고에서는, 해당 방법에서 더 나아가 동영상 초해상도네트워크의 파라미터를 테스트 타임의 저 화질 영상만을 가지고 업데이트 하는 기법을 소개한다. 첫째로, 동영상 내에 일반적으로 존재하는 반복되는 패치의 특성을 분석하고, 다음으로 기존의 복원된 동영상을 관찰하여 자기 지도 적응의 가능성을 보인다. 마지막으로, 폭넓은 실험을 통해 제안하는 기법을 검증한다.

## 1. 서론

초해상도는 주어진 저해상도 영상을 고해상도 영상으로 복원하는 기법을 말한다. 이는 의료, 항공, 가전제품 등의 분야에 핵심적으로 적용되는 기술이기 때문에 지속적인 주목을 받고 있다. 저화질 영상을 복원하는 데 있어서 해결해야 하는 문제는 일대다 속성에서 비롯된다. 한 가지의 저화질 영상은 무수히 많은 고화질 영상과 대응될 수 있기 때문에 이 가운데 최적의 정답을 찾기 위해 다양한 연구가 시도되었다 [1, 2]. 한편, 최근 전통적인 예시 기반 초해상도 기법에도 또한 딥러닝이 적용되기 시작하였다 [3, 4]. 이는 초기화된 얇은 네트워크를 저화질 영상이 주어질 때마다 테스트 타임에 새롭게 학습시킴으로써 이루어진다. 학습 시 사용되는 참값 및 입력 데이터는 각각 주어진 저화질 영상과 이를 한 번 더 저화질로 만든 “초저화질” 영상으로, 참값이 없는 환경이기 때문에 자기 지도 학습을 따른다. 해당 기법은 주어진 저화질 영상이 생성된 조건에 따라 유연하게 학습이 될 수 있는 장점에 힘입어 일반적인 딥러닝 기반 초해상도 기법보다 더욱 실제계에 응용될 수 있는 여지를 보인다. 하지만, 이러한 내부 데이터셋 기반 딥러닝 연구는 대부분 단일 영상에 대해 제한적으로 연구되었다. 본 연구에서는, 단일 영상에서 탁월한 성능을 보인 위의 내부 데이터셋 기반 딥 네트워크 학습 방법을 동영상으로 응용한다. 먼저, 우리는 동영상에는 단일 영상 내의 반복되는 패치 뿐만 아니라 서로 다른 프레임에서의 여러 스케일 간 반복되는 패치가 더욱 풍부하게 존재함을 소개하고, 이를 이용해 동영상 초해상도 딥 네트워크를 테스트 저화질 영상에 대해 적용시킬 수 있음을 입증한다. 또한, 기존의 자기 지도 적응이 상대적으로 작은 스케일 팩터 (x2)에서만 잘 동작하는 점을 지적하며, 큰 스케일 팩터 (x4)에 대해서 우월한 성능을 보이는 새로운 가상 데이터 페어 생성 방법을 제안하는 알고리즘과 함께 소개한다. 그 후, 다양한 실험을 통해 제안하는 기법이 일관적으로 기존의 딥 네트워크의 성능을 높임을 보인다.

## 2. 동영상에서의 반복되는 패치

예시 기반 초해상도 기법은 단일 영상 내에서 동일한 패치가 크고 작은 크기로 반복되는 특성에 기반한다 [1]. 이처럼 패치가 반복되는 현상은 동영상에서 카메라 및 객체의 움직임에 의해 더욱 빈번히 일어난다. 동일한 내용의 패치에 대해 큰 패치는 작은 패치보다 더욱 세밀한 정보를 담고 있고, 이러한 대응되는 패치 쌍이 많아진다면 예시 기반 초해상도 방법의 유용성 또한 증가한다. 이러한 특성은 기존의 전통적인 초해상도 기법 뿐 만 아니라, 최근 딥 네트워크를 이용한 자기 지도 학습 기반 초해상도 기법에서도 응용되었으며, 참값이 주어지지 않는 상황에서도 딥 네트워크의 성능을 더욱 증가시킬 수 있는 여지를 보여준다.

## 3. 가상 데이터 페어 생성

기존의 예시 기반 초해상도 방법은 가상의 고해상도 참값을 주어진 저해상도 영상 자체에서 탐색한다. 이는 스케일 팩터가 증가함에 따라 어려워지고, 이를 완화시키기 위한 전통적인 방법으로는 원하는 스케일 팩터가 될 때까지 계층적으로 영상을 복원시키는 것인데, 이에 따라 연산 시간이 증가하기 때문에 테스트 타임에 작업을 수행해야 하는 예시 기반 방법의 특성의 내재적인 큰 한계점이 된다.

본 연구에서는 이러한 문제점을 해결하기 위해 새로운 가상 데이터 페어 생성 방법을 제안한다. 먼저, 그림 1 (a) 의 두 패치  $A_{gt}$  와  $b_{gt}$  는 참값 동영상 내의 인접한 프레임에서 반복되는 서로 다른 크기의 패치이다. 큰 패치의 크기를 작은 패치와 같게 줄인다면, 큰 패치와 작은 패치가 지니고 있는 정보 또한 비슷한 점을 볼 수 있다. 그림 1 (b) 의 두 패치  $A$  와  $b$  는 딥 네트워크를 통해 복원된 영상에서의 그림 1 (a) 와 동일한 부분의 패치를 나타낸다. 흥미롭게도, 큰 패치의 크기를 줄인 패치  $a$  가 여전히 작은 패치보다 더욱 선명한 고주파 정보를 포함한다.

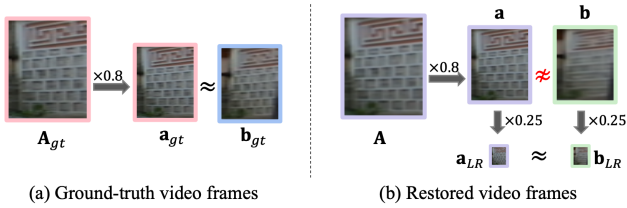


그림 1. (a) 참값 영상에서의 서로 다른 크기의 반복되는 패치. (b) 복원된 영상에서의 서로 다른 크기의 반복되는 패치 및 이의 크기를 줄인 결과.

또한, 같은 크기의 두 패치를 다시 한 번 원하는 만큼의 스케일 팩터를 통해 줄인다면, 두 패치의 구별점이 완화되는 것을 볼 수 있다 ( $a_{LR}$  과  $b_{LR}$ ). 그렇다면, 학습 데이터의 입력값, 참값을 각각  $a$  와  $a_{LR}$  로 설정하여 딥 네트워크를 학습시키면 결과적으로  $b_{LR}$  에 대한 복원값은 기존의 복원값  $b$  보다 더욱 화질이 좋은  $a$  의 방향으로 최적화 될 것이다. 본 연구에서는 위의 기대를 바탕으로, 기존의 학습된 딥 네트워크를 통해 복원된 동영상에서 임의의 패치  $A$  를 얻고, 이를 다운샘플링 하는 과정을 통해 가상의 데이터셋을 생성하였다. 이는 테스트 시의 저화질 영상만을 가지고 이루어지기 때문에, 자기 지도 학습을 따른다.

#### 4. 실험 결과

Method	Vid4 (PSNR/SSIM/tOF)	REDS4 (PSNR/SSIM/tOF)
TOFlow	25.86/0.7625/0.2116	29.32/0.8193/1.7378
TOFlow+ Adaptation	26.08/0.7731/0.1954	29.73/0.8303/1.3957
EDVR	27.35/0.8264/0.1362	32.45/0.8937/0.8024
EDVR + Adaptation	27.53/0.8305/0.1277	32.69/0.8957/0.6356

표 1. 제안하는 알고리즘의 정량적 실험 결과.

제안하는 알고리즘을 검증하기 위해 본 연구에서는 최적화 된 두 베이스라인 딥 네트워크 모델 (TOFlow, EDVR) 과 두 가지 데이터셋을 사용하였다. 각 실험 결과의 성능은 세 종류의 평가지표 (PSNR, SSIM, tOF) 로 평가되었으며, 이 중 tOF 는 인접한 프레임 간의 광학 흐름의 성능을 비교한 것으로, 동영상의 시간적 일관성에 대한 지표이다. 가상 데이터 패어를 생성하기 위한 다운스케일 팩터의 범위는 0.8 ~ 0.95 사이로 설정하였으며, 적응 이터레이션은 Vid4 와 REDS4 데이터셋에 대해 각각 1K, 3K 번으로 설정하였다. 모든 실험은 4배의 스케일 팩터에 대해 진행되었다.



그림 2. 제안하는 알고리즘의 정성적 실험 결과.

표 1 은 실험의 정량적 결과를 보여준다. 모든 모델, 데이터셋에서 제안하는 알고리즘이 세 가지 평가지표에 대해 성능을 일관적으로 높여준다. 특히, 상대적으로 프레임 수가 많고 모션이 큰 REDS4 데이터셋에



EDVR (top) / EDVR + Adaptation (bottom)

그림 3. 제안하는 알고리즘의 시간적 일관성을 보여주는 시각화 결과.

대해서는 Vid4 에서보다 더 높은 성능 향상을 보이는데, 이는 적응 과정에서 이용될 수 있는 반복되는 패치가 더욱 풍부하기 때문이다.

그림 2 는 제안하는 알고리즘의 정성적 결과를 보여준다. 기존의 수렴된 네트워크를 통해 복원된 영상에서의 참값과 거리가 먼 패턴들이 자기 지도 적응을 통해 참값과 가까워지는 것을 볼 수 있다.

그림 3 은 제안하는 알고리즘이 동영상의 시간적 일관성을 높일 수 있음을 보여주는 시각화 결과이다. 이는 동영상에서 고정된 선을 추적한 것으로, 결과에서 노이즈가 보인다면 영상에서 해당 부분이 시간 축으로 일관되지 못하다고 해석된다. 결과에서 보이듯이 기존의 학습된 네트워크에 의해 복원된 동영상은 시간적 비일관성을 보이는 데 반해 이를 제안하는 알고리즘을 통해 적응시킨 결과는 노이즈가 완화되어 시간적으로 일관된 영상을 생성하였다.

#### 5. 결론

본 연구에서는 기존에 활발히 연구되었던 단일 영상에서의 딥 네트워크를 통한 초해상도 기법을 동영상 복원으로 확장하였다. 이를 위해 먼저 서로 다른 프레임 간에 반복되는 패치들을 소개하고, 참값과 복원된 영상에서의 이러한 패치들의 차이점을 관찰하였다. 그 후, 관찰을 바탕으로 새로운 가상 데이터셋 생성 방법을 소개하였다. 특히, 제안하는 방법은 높은 스케일 팩터에 효과적이었다. 마지막으로, 다양한 실험을 통해 제안하는 알고리즘의 우월성을 입증하였다.

#### 감사의 글

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-01373, 인공지능대학원지원(한양대학교))

#### 참고 문헌

- [1] Glasner et al. "Super-resolution from a single image." Proceedings of the *IEEE International Conference on Computer Vision*. 2009.
- [2] Dong et al. "Learning a deep convolutional network for image super-resolution." Proceedings of the *European Conference on Computer Vision*. 2014.
- [3] Shocher et al. "'zero-shot' super-resolution using deep internal learning." Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
- [4] Park et al. "Fast adaptation to super-resolution networks via meta-learning." Proceedings of the *European Conference on Computer Vision*. 2020.