

# Lab 5. Perceptual Losses for Real-Time Style Transfer and Super-Resolution Summary

Summarized by: 202355514 강지원

Justin Johnson, Alexandre Alahi, Li Fei-Fei  
 1 Department of Computer Science, Stanford University

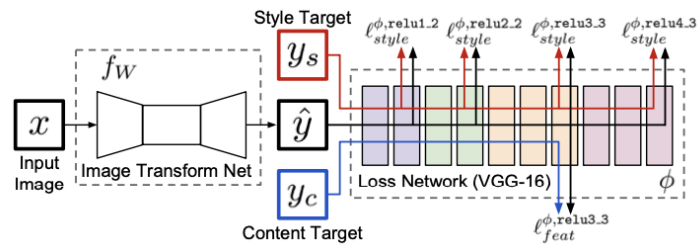


Figure 1: Perceptual Loss를 활용한 이미지 변환 네트워크 구조

본 논문에서는 사전 학습된 CNN의 고수준 특징을 이용한 perceptual loss를 통해 이미지 변환 작업을 수행하는 feed-forward 네트워크를 제안하며, 스타일 변환과 초해상도(super-resolution)에서 기존과 차별화된 성능을 보이고자 한다.

기존의 픽셀 손실 기반 학습은 효율적이지만 시각적 유사성을 반영하지 못하는 한계를 지니며, 이에 본 논문은 이미지의 의미적 특징을 고려한 perceptual loss가 더 적합함을 보여준다. 제안된 방법은 스타일 변환(style transfer)과 초해상도(super-resolution) 작업에서 기존보다 훨씬 빠르면서도 시각적으로 우수한 결과를 실시간으로 생성함을 입증한다.

제안된 방법은 입력 이미지  $x$ 를 출력 이미지  $\hat{y}$ 로 변환하는 feed-forward 네트워크를 학습하며, 사전 학습된  $\phi$ 을 손실 네트워크로 사용하여 특징 재구성 손실(feature reconstruction loss)과 스타일 재구성 손실(style reconstruction loss)을 계산한다. 전체 손실은 다음 수식으로 정의된다.

$$W^* = \arg \min_W E_{x, \{y_i\}} \left[ \sum_{i=1}^k \lambda_i \ell_i(f_W(x), y_i) \right]$$

이를 통해 네트워크는 픽셀 단위 손실 대신 시각적(perceptual) 기준에 기반한 학습을 수행하게 된다.

논문에서 제안한 이미지 변환 네트워크는 풀링 대신 스트라이드 합성곱과 residual block을 이용하여 효율적이면서도 깊은 구조를 형성한다. fractionally-strided convolution을 통해 업샘플링 과정을 학습 가능하게 하고, 다운 샘플링을 통해 연산의 효율성과 넓은 수용 영역을 동시에 확보하였다. 또한 Residual Connection을 적용해 입력과 유사한 구조의 출력을 안정적으로 학습할 수 있도록 설계하였다.

이미지 간의 고수준 시각적, 의미적 차이를 측정하기 위해 두 가지 Perceptual Loss를 정의한다. 특징 재구성 손실(feature reconstruction loss)은 두 이미지가 내용(content) 면에서 유사하도록 하며, 스타일 재구성 손실(Style Reconstruction Loss)은 각 층의 특징 간의 상관관계(Gram matrix)를 이용하여 질감, 색상 등의 스타일 유사도를 유지하도록 한다. 위 손실의 여러 층의 계산을 결합해 최종 스타일의 손실을 도출해낸다. 이를 통해 모델은 픽셀 일치보다 사람이 인식하는 시각적 유사성(perceptual similarity)에 더 민감하게 반응한다.

추가적으로 Perceptual Loss 외에 Pixel Loss와 TV Regularization을 함께 사용하여 정답 이미지와의 일치도와 출력의 부드러움을 보장한다.

제시한 네트워크를 스타일 변환과 초해상도 작업에 적용하여 실험을 수행한 결과, 스타일 변환의 경우 제안한 feed-forward 네트워크는 VGG-16 기반의 Perceptual Loss를 사용해 기존의 최적화 방식과 비슷한 시각적 품질을 보이면서도 약 1,000배 빠른 속도를 달성하였다. 또한 사람과 동물 같은 의미적 객체는 명확히 유지하면서 배경은 스타일에 맞게 변형되어,



Figure 2: 스타일 변환(Style Transfer) 결과



Ground Truth	Bicubic	Ours ( $\ell_{pixel}$ )	SRCNN [11]	Ours ( $\ell_{feat}$ )
This image	31.78 / 0.8577	31.47 / 0.8573	32.99 / 0.8784	29.24 / 0.7841
Set5 mean	28.43 / 0.8114	28.40 / 0.8205	30.48 / 0.8628	27.09 / 0.7680

Figure 3: x4 초해상도(Super-Resolution) 결과 비교

지각적 품질(Perceptual Quality)과 실시간 성능을 동시에 확보했음을 확인하였다.

초해상도 실험에서는 저해상도 이미지를 고해상도로 복원하기 위해 픽셀 손실 대신 특징 재구성 손실을 사용하였다. 제안한 네트워크는 VGG-16의 의미적 특징을 활용하여 4배, 8배 배율 초해상도에서 세밀한 윤곽과 경계 표현을 개선하였으며, 수치 지표는 비록 낮지만 사람이 인식하는 시각적 품질(Perceptual Quality)은 매우 우수하였다.

결론적으로, 본 논문은 Perceptual Loss를 활용한 feed-forward image transformation network를 통해 기존 최적화 기반 방식의 품질을 유지하면서 실시간 처리 속도와 시각적 품질을 동시에 확보하였다. 제안된 방법은 스타일 변환에서 빠르고 유사한 품질, 초해상도에서 세밀한 디테일 복원을 달성했으며, 향후 색상 복원이나 의미 분할 등 다양한 이미지 변환 작업으로 확장될 가능성을 제시한다.

[1] "Perceptual Losses for Real-Time Style Transfer and Super-Resolution." European Conference on Computer Vision (ECCV), 2016. Justin Johnson, Alexandre Alahi, and Li Fei-Fei. <https://arxiv.org/abs/1603.08155>

[2] "스타일 전이(Style Transfer)" <https://ejji.tistory.com/5>