

# 동영상 자동 채색화 성능 향상을 위한 깊은 신경망 손실함수 개선

이혜진, 김병조, 김수진, 이재구\*  
국민대학교 소프트웨어학부  
\*jaekoo@kookmin.ac.kr

## Modify Loss function of Deep Neural Network to Improve Video Colorization Performance

Hyejin Lee, Byeongjo Kim, Sujin Kim, Jaekoo Lee\*  
College of Computer Science, Kookmin Univ.

### 요 약

본 논문은 깊은 신경망 기반의 최신 동영상 자동 채색화 기술[1]을 보완하여 채색 신경망의 채색 성능을 올릴 수 있는 방법을 제안한다. 제안된 동영상 자동 채색화 기술은 채색 신경망과 정제 신경망으로 이루어져 있다. 우리는 채색 신경망의 학습 과정에서 손실함수를 수정하여 채색 신경망을 동영상과 관련없이 독립적으로 분리하였다. 또한, 채색 신경망에 사용되는 K-최근접 이웃 알고리즘의 파라미터의 경향성을 측정하여 성능을 높이는 하이퍼 파라미터를 제시한다. 위의 제안된 방법으로 학습한 결과, 기존의 방법보다 2.8% 더 선명하게 채색할 수 있었다. 향후 연구에서는 기존보다 성능이 좋은 채색 신경망을 연구하여 성능을 높이는 연구가 필요하다고 판단된다.

### I. 서 론

최근 깊은 신경망을 이용한 채색 기법으로 과거의 흑백 영상에 색깔을 입힘으로써 생동감을 주고자 하는 시도가 계속되고 있다. 색이 희미하거나 없었던 영상과 동영상을 깊은 신경망 기반의 채색 작업을 통해 컬러 영상이나 동영상으로 바꿔줌으로써 더 생동감 있게 만들 수 있다.

깊은 신경망을 이용한 채색화 기법은 3 가지로 나눌 수 있다. 첫 번째는 자동 채색화(Automatic Colorization)이다. 자동 채색화는 영상이 들어왔을 때 자동으로 흑백 이미지의 영역에 적합한 색을 입힌다. 두 번째는 참조 채색화이다. 참조 채색화(Reference Colorization)는 색을 참조할 이미지와 흑백 이미지를 입력으로 넣어주고, 참조 이미지의 색에 기반을 두어 흑백 이미지를 채색한다. 세 번째는 사용자 지정 채색화(User-Guided Colorization)이다. 이미지의 영역별로 사용자가 원하는 색깔을 선택하면 그 색깔에 기반을 두어 채색을 한다.

본 논문에서 사용하는 기법은 자동 채색화 기법이다. 동영상 자동 채색화는 참조 이미지와 사용자 지정 색깔이 없어, 신경망에서 시공간 일관성과 물체의 다양한 색을 고려할 수 있도록 학습이 되어야 한다. 시공간 일관성은 프레임들 사이에 같은 물체는 같은 색이어야 하며, 한 프레임 안에서 한 물체에 대한 픽셀들은 비슷한 색깔이어야 한다는 것이다. 본 연구는 최신 동영상 자

동 채색화 기법을 보완하여 시공간 일관성과 물체의 다양한 색깔을 모두 고려하였다. 그 결과, 우리는 기존 기술보다 더 자연스럽고 선명한 채색 결과를 얻을 수 있었다.

우리는 동영상 자동 채색화 최신 논문인 [1]에서 제안된 방법을 기반으로 실험을 진행하였다. [1]에서 제안된 동영상 자동 채색화 기법은 그림 1 과 같이 두 개의 깊은 신경망으로 이루어져 있다. 첫 번째 신경망은 채색 신경망이며 흑백 프레임을 적합한 색으로 채색하는 역할을 한다. 두 번째 신경망은 정제 신경망으로, 채색된 t 프레임을 Optical Flow 를 이용하여 t+1 프레임로 정방향 워프시킨 프레임과, Optical Flow 를 이용하여 계산한 Confidence Map 을 입력으로 넣는다. 이 신경망은 연속되는 프레임에 어색함이 없게 시간적인 일관성을 고려하여 t+1 프레임을 정제해주는 역할을 한다. 이 방법은 시공간적 연결성을 고려하기 위해 학습 과정에서 시공간의 일관성을 위한 손실함수를 사용한다. 시공간을 고려한 손실함수는 채색 신경망, 정제 신경망에 적용된다.

우리는 학습 과정에서 채색 신경망에 사용되는 시간 차 손실함수를 제거함으로써 프레임 간의 연결성과 채색을 독립적으로 분리하였다. 또한, 채색 신경망 학습에 사용되는 K-최근접 이웃 알고리즘의 파라미터 변화를 측정하여 최적의 하이퍼 파라미터를 도출해냈다. 제안된 방법을 적용하여 기존보다 2.8% 더 좋은 성능 결과를 확인하였다.



그림 1. 동영상 자동 채색화 신경망

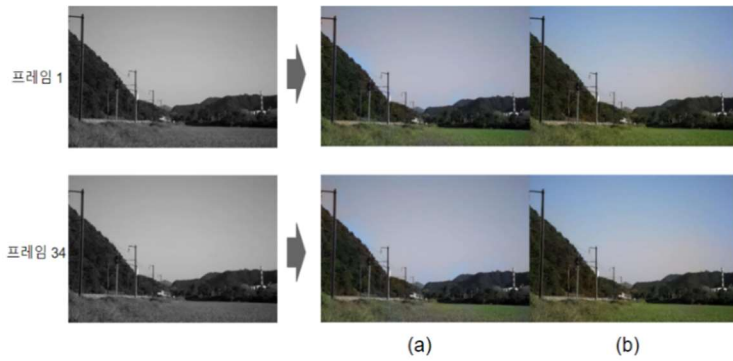


그림 2. 기존 모델(a) 결과와 제안된 모델(b) 결과

## II. 본론

[1]의 동영상 자동 채색화 기술을 보완하여 학습을 진행하였다. 그림 1 을 보면 깊은 신경망은 크게 두 가지로 구성되어 있다. 우리는 첫번째 신경망인 채색 신경망을 학습할 때 시차 손실 함수를 제거하고 학습을 진행하였다. 제거할 채색 신경망의 시차 손실함수  $L_{\text{Temporal}}$  은 식 1 에 나타나 있다.

$$\sum_{t=1}^{n-1} \left\| \left( f(X^t; \theta_f) - \omega_{t+1 \rightarrow t} \left( f(X^{t+1}; \theta_f) \right) \right) \odot M_{t+1 \rightarrow t} \right\|_1 \quad (\text{식 1})$$

이  $L_{\text{Temporal}}$  에서  $\omega_{t+1 \rightarrow t}$  는 Optical Flow 를 이용하여  $t+1$  프레임을  $t$  프레임으로 역방향 워프시키는 연산자이다.  $M_{t+1 \rightarrow t}$  은  $t$  프레임의 픽셀이  $t+1$  프레임에도 보이는지 여부를 가리키는 이진 마스크이다.

채색 신경망은 이 손실함수를 통해  $t$  프레임과  $t+1$  프레임의 연관관계를 고려하여 시차 일관성을 학습한다[1]. 그러나 입력으로는  $t$  프레임만 들어간다. 채색 신경망이 손실함수에 의해 학습된다고 해도 입력으로  $t+1$  프레임을 받지 않아 알 수 없는  $t+1$  프레임을 고려하여  $t$  프레임이 채색되는 것이다. 이러한 학습은 채색에 방해되는 요소라고 가정하여 시차 손실함수를 제거하고 실험을 진행하였다. 실험에 사용된 데이터집합은 DAVIS 데이터집합[3], COCO 데이터집합[4], 직접 수집한 한국 동영상 데이터집합이다. COCO 데이터집합[4]으로 채색 신경망을 사전 학습시킨 뒤, DAVIS 데이터집합[3]과 한국 데이터집합을 학습시켰다. 기존 방식과 성능을 비교하기 위해 기존 방식 또한 동일한 환경으로 학습을 시켰다.

우리는 이미지의 품질 정도를 평가하는 PSNR 성능지표를 우리의 모델과 기존 모델의 성능을 평가하였다. 테스트 셋으로 50 개의 동영상을 사용하여 측정한 결과, 표 1 과 같이 우리의 모델이 2.8% 더 높은 것을 알 수 있다. 또한, 그림 2 와 같이 우리의 모델의 결과가 더 선명하게 채색이 된 것을 볼 수 있다. 정제 신경망을 거친 최종 결과물은 그림 2 의 출력 이미지에서 볼 수 있듯이 기존 모델(a) 보다 우리의 모델(b)이 더 자연스럽게 선명하게 채

표 1. 기존모델과 제안된 모델의 PSNR

방법	기존 모델	제안된 모델
PSNR	28.93	29.74

표 2. K-최근접 이웃값에 따른 PSNR

K	K=5	K=7	K=8	K=10	K=15
PSNR	28.92	29.23	28.97	31.27	28.88

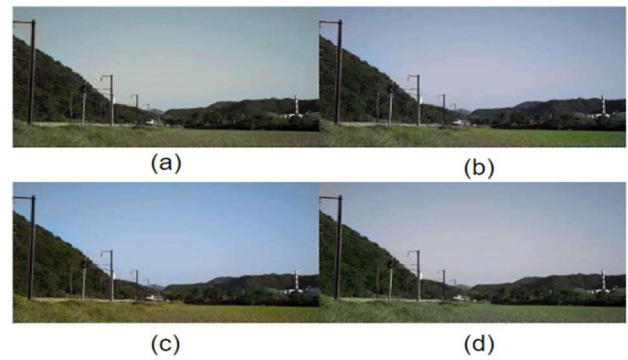


그림 3. K 값에 따른 결과 (왼쪽부터 KNN=5, 7, 10, 15)

색되었다.

우리가 제안한 두 번째 방식은 K-최근접 이웃 알고리즘의 하이퍼 파라미터의 변화를 보고 최적의 하이퍼 파라미터를 찾는 것이다. 실험은 변수를 두지 않기 위해 기존 모델로 진행하였다. K를 5, 7, 8, 10, 15로 변화시키면서 학습을 시킨 결과, 표 2 의 PSNR 결과에서 K가 7 일 때는 기존 파라미터였던 5 보다 1% 더 높고, 10 일 때는 기존보다 8.1%가 더 높은 것을 볼 수 있다. 그림 3 의 (b)와 (c)에서 K가 7 일 때와 10 일 때의 채색 결과가 좋은 것을 확인할 수 있다.

## III. 결론

본 논문에서 제안한 방법으로 학습한 결과, 기존의 방식보다 자동 채색화의 성능이 향상되어 더 선명하게 채색되는 것을 확인할 수 있었다. 채색 신경망은 손실함수를 제거함으로써 이미지 채색화처럼 하나의 프레임 단위로 입력값을 주어 보다 선명하게 채색할 수 있다. 나아가 채색 신경망이 완전한 독립성을 가지면서 기존의 채색 신경망을 보다 성능이 좋은 다른 채색 신경망 모델을 사용할 수 있는 가능성을 보여준다. 성능이 좋은 채색 신경망을 거친 후 정제 신경망을 거친다면 어떤 동영상이든 더 선명하고 자연스러운 채색이 가능하다. 향후 연구에서는 더 좋은 채색 신경망을 연구하여 성능을 높이는 연구가 필요하다고 판단된다.

## ACKNOWLEDGMENT

이 성과는 2020 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원(No. NRF2018R1C1B5086441)과 과학기술정보통신부 및 정보통신기획 평가원의 SW 중심대학지원사업(2016-0-00021)으로 수행된 연구임.

## 참 고 문 헌

- [1] Lei, C., & Chen, Q. (2019). Fully automatic video colorization with self-regularization and diversity. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3753-3761).
- [2] Chen, Q., Li, D., & Tang, C. K. (2013). KNN matting. IEEE transactions on pattern analysis and machine intelligence, 35(9), 2175-2188.
- [3] F. Perazzi, J. Pont-Tuset, L. McWilliams, B. and Van Gool, M. Gross, and A. Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In CVPR, 2016.
- [4] Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., ... & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. arXiv 2014. *arXiv preprint arXiv:1405.0312*.