Motion Magnification RW ([5],[7])

💡 발표하시는 논문인 **Multi Domain Learning for Motion Magnification에서 언급된 부분**:

The majority of approaches have a **narrow range of applicability** [5], [7], so their scope of work is limited.

DeepMag: Source Specific Motion Magnification Using Gradient Ascent [5]

url: https://arxiv.org/pdf/1808.03338.pdf

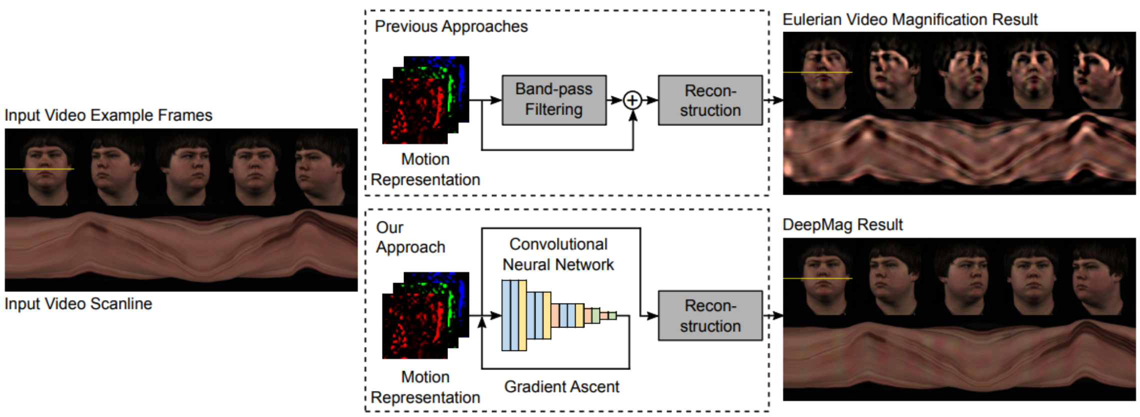


Figure 1: We present a novel end-to-end deep neural framework for video magnification. Our method allows measurement, magnification and synthesis of subtle color and motion changes from a specific source even in the presence of large motions. We demonstrate this via pulse and respiration manipulation in 2D videos. Our approach produces magnified videos with substantially fewer artifacts when compared to the state-of-the-art.

결론: 당시(2018) SOTA보다 잘 된다!

Algorithm

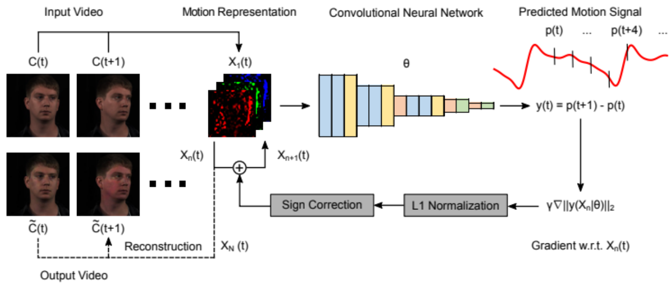


Figure 2: The architecture of DeepMag. The CNN model predicts the motion signal of interest based on a motion representation computed from consecutive video frames. Magnification of the motion signal in video can be achieved by amplifying the L2 norm of its first-order derivative and then propagating the changes back to the motion representation using gradient ascent.

We propose to use a deep convolutional neural network (CNN) to model the relationship between the motion representation and the motion of interest

Input of CNN:

$X_1(t)$ **:** any change happening between two consecutive frames C(t) and C(t+1)

Output of CNN:

The first-order derivative y(t) of the target motion signal p(t)

As the CNN has established the relationship between the input motion representation $X_1(t)$ and the target motion signal p(t), magnification of p(t) in $X_1(t)$ can be achieved by amplifying the L2 norm of its first-order derivative y(t) and then propagating the changes back to $X_1(t)$ using gradient ascent.

pseudo-code

ALGORITHM 1: DeepMag video magnification

Require: $C(t), t = 1, 2, \ldots, T$ is a series of video frames, $\mathcal{M}$ is a motion representation estimator, θ is the pre-trained CNN weights for predicting a target motion signal y, γ is the step size, and N is the number of iterations

for $t = 1$ to $T - 1$ **do**
  Compute motion representation: $X_1(t) \leftarrow \mathcal{M}(C(t), C(t+1))$
  for $n = 1$ to $N - 1$ **do**
    Compute gradient: $G_n(t) \leftarrow \nabla\|y(X_n(t)|\theta, t)\|_2$
    L1 normalization: $G_n(t) \leftarrow G_n(t)/\|G_n(t)\|_1$
    Sign correction: $G_n(t) \leftarrow G_n(t) \odot sgn(G_n(t) \odot X_n(t))$
    Gradient ascent: $X_{n+1}(t) \leftarrow X_n(t) + \gamma G_n(t)$
  end for
end for
$\widetilde{C}(1) = C(1)$
for $t = 1$ to $T - 1$ **do**
  Reconstruct magnified frame $\widetilde{C}(t+1) \leftarrow \mathcal{M}^{-1}(\widetilde{C}(t), X_N(t))$
end for
return $\widetilde{C}(t), t = 1, 2, \ldots, T$

Evaluation

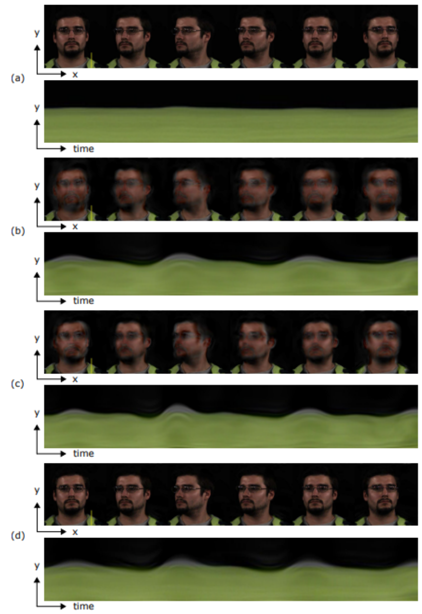


Figure 7: Scan line comparisons of motion magnification methods for a Task B video: a) original video, b) phase-based Eulerian video magnification, c) video acceleration magnification, d) Our method. The yellow line shows the source of the scan line in the frames. The section of video shown was 15 seconds in duration. Our method produces comparable magnification of the respiration motion and significantly fewer artifacts and blurring.

→ 다른 방법들에 비해 제안된 방법이 잡음(artifact)와 블러링이 훨씬 덜하다.

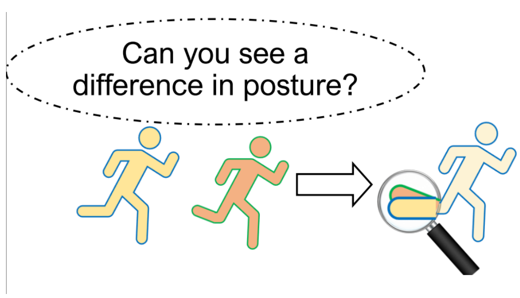
(제가 생각한 한계점: experiments를 facial dataset에만 적용했기에 다른 분야에선 어떤 성능을 보이는지 알기 어려운 것 같습니다)

Unsupervised Magnification of Posture Deviations Across Subject [7]

url:
https://openaccess.thecvf.com/content_CVPR_2020/papers/Dorkenwald_Unsupervised_Magnification_of_Posture_Deviations_Across_Subjects_C

(이해에 도움을 많이 받은 reference: https://compvis.github.io/magnify-posture-deviations/images/poster.pdf)

연구의 전체적 목표: 밑 그림에서 보이는 것처럼 두 posture 사이의 차이점을 부각하여 알려준다!



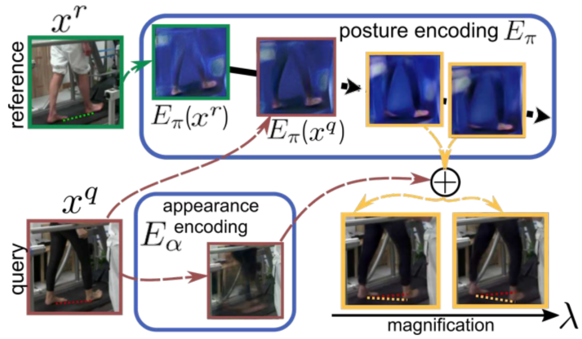


Figure 1. Magnification of Posture Deviations Across Subjects. We visually emphasize subtle posture deviations between a query x^q and reference frame x^r by magnifying their differences in the posture encoding. x^q walks with its legs apart, highlighted by the red line in comparison to the green line of x^r. We first disentangle posture from appearance (the blue boxes show visualization of E_π, E_α). Then, we extrapolate in the posture encoding the distance of x^r and x^q in the direction of x^q. The magnified images (bottom right) are generated by combining the appearance encoding of x^q and the magnified posture encoding using different magnification intensities λ. The generated images allow a user to easier see differences.

→ 우선 posture과 appearance사이를 disentangle 한 후 extrapolation을 활용하여 magnification을 통한 새로운 posture을 생성한다.

Part 1: Disentanglement of Magnification

목표: Magnifying only posture differences

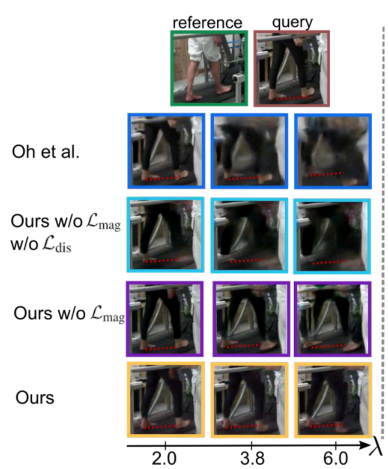
→ Unsupervised disentanglement of posture π and appearance α

→ Use one autoencoder with two encoders (E_π and E_α)

Part 2: Learning to Magnify

Produce realistic and valid magnifications via extrapolation.

Evaluation



→ Magnification parameter (λ)가 높아져도 다른 연구에 비해 훨씬 robustness를 보이는 것이 이 연구의 contribution이다. (= magnification intensity에 robust한 결과를 보여준다.)

Qualitative Evaluation

comment: 발표준비하시는 논문(**Multi Domain Learning for Motion Magnification**)에서 한계점으로 지적하는 부분은 posture에만 적용한 알고리즘이라 ‘narrow range of applicability’라고 지적했다고 생각합니다 :)