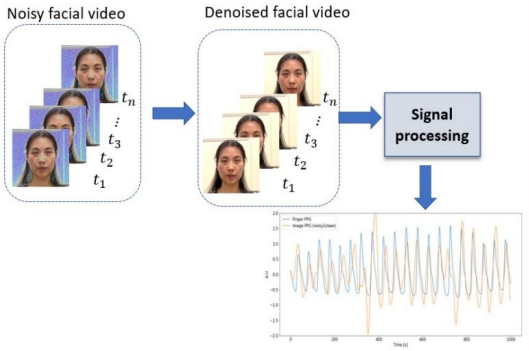
Deep Video Denoising for Facial Signal Processing

Xiaoyan Li
University of Toronto

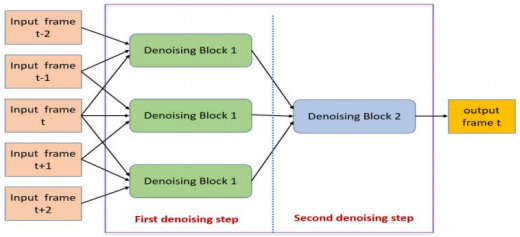
Motivation

- Spatial-temporal denoising. Temporal coherency.
- Video contains much more information.
- FastDVDnet (supervised) and UDVD (unsupervised).
- Apply two advanced video denoising methods to facial videos.
- Ground-truth video is not available.
- Image photoplethysmography (iPPG) reals the change of blood flow.
- Clean video could reflect blood flow better.

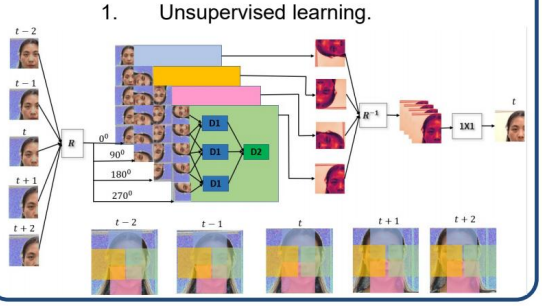


Related Work

- **FastDVDnet**
  1. Supervised learning
  2. Fast , a wide range of noise level.



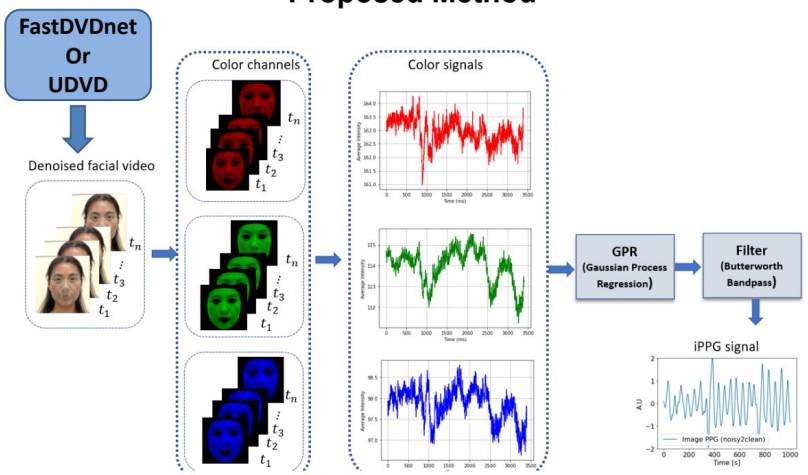
- **UDVD**
  1. Unsupervised learning.



References

[1] M. Tassano, J. Delon, and T. Veit, “Fastdvdnet: Towards real-time deep video denoising without flow estimation,” in CVPR, 2020.
[2] D. Y. Sheth, S. Mohan, J. Vincent, R. Manzorro, P. A. Crozier, M. M. Khapra, E. P. Simoncelli, and C. Fernandez-Granda, “Unsupervised deep video denoising,” in ICCV, 2021.
[3] R. H. Goudarzi, S. Somayyeh Mousavi, and M. Charmi, “Usingimaging photoplethysmography (ippg) signal for blood pressure estimation,” in MVIP, 2020.

Proposed Method



- **UDVD**

$$\mathcal{L}(\mu_x, \Sigma_x) = \frac{1}{2}[(y-\mu_x)^T(\Sigma_x+\sigma^2 I)^{-1}(y-\mu_x)] + \frac{1}{2}\log|\Sigma_x+\sigma^2 I|$$

Where $\mu_x \in R^3$ and $\sigma_x \in R^3$ denote the mean vector and covariance matrix; y is observed noisy pixel

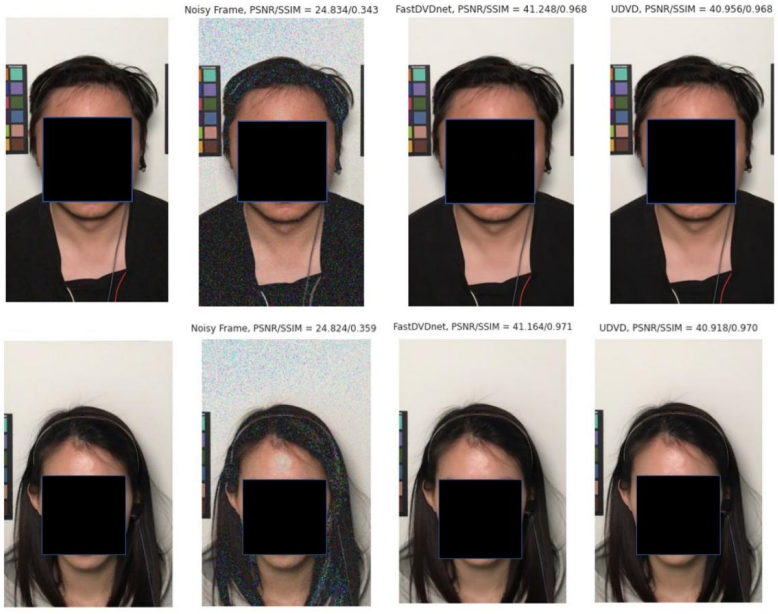
- **FastDVDnet**

$$\mathcal{L}(\theta) = \frac{1}{2m_t}\sum_{j=1}^{m_t} \|\hat{\boldsymbol{P}}_t^j - \boldsymbol{P}_t^j\|^2, \quad (2)$$

where $\hat{\boldsymbol{P}}_t^j = \mathcal{F}((S_t^j, M^j); \theta)$ is the output of the network, and θ is the set of all learnable parameters.

Experimental Results

- **Results of frame denoising**



- **Results of iPPG extraction**