Stabilize, Decompose, and Denoise: Self-Supervised Fluoroscopy Denoising

Ruizhou Liu[1][0000−0002−2940−9630], Qiang Ma[1][0000−0003−0791−1731], Zhiwei Cheng[1][0000−0002−1980−811X], Yuanyuan Lyu[1][0000−0001−9049−7812], Jianji Wang[2][0000−0001−5373−6247], and S. Kevin Zhou[3][0000−0002−6881−4444]

[1] Z2Sky Technologies Inc.
[2] Affiliated Hospital of Guizhou Medical University
[3] University of Science and Technology of China
s.kevin.zhou@gmail.com

Abstract. Fluoroscopy is an imaging technique that uses X-ray to obtain a real-time 2D video of the interior of a 3D object, helping surgeons to observe pathological structures and tissue functions especially during intervention. However, it suffers from heavy noise that mainly arises from the clinical use of a low dose X-ray, thereby necessitating the technology of fluoroscopy denoising. Such denoising is challenged by the relative motion between the object being imaged and the X-ray imaging system. We tackle this challenge by proposing a self-supervised, three-stage framework that exploits the domain knowledge of fluoroscopy imaging. (i) Stabilize: we first construct a dynamic panorama based on optical flow calculation to stabilize the non-stationary background induced by the motion of the X-ray detector. (ii) Decompose: we then propose a novel mask-based Robust Principle Component Analysis (RPCA) decomposition method to separate a video with detector motion into a low-rank background and a sparse foreground. Such a decomposition accommodates the reading habit of experts. (iii) Denoise: we finally denoise the background and foreground separately by a self-supervised learning strategy and fuse the denoised parts into the final output via a bilateral, spatiotemporal filter. To assess the effectiveness of our work, we curate a dedicated fluoroscopy dataset of 27 videos (1,568 frames) and corresponding ground truth. Our experiments demonstrate that it achieves significant improvements in terms of denoising and enhancement effects when compared with standard approaches. Finally, expert rating confirms this efficacy.

Keywords: Fluoroscopy Denoising · Image Decomposition · Self-Supervised Learning

1 Introduction

Fluoroscopy is a medical imaging technique that uses X-ray to monitor the interior structure of human body in real-time. It helps surgeons to observe pathological structures and tissue functions especially during intervention, without destroying the external epidermis. While it is desirable to reduce a radiation dose for less harm in clinical practice, the use of low dose results in heavy noise in raw fluoroscopy, thereby necessitating the technology of fluoroscopy denoising. But, fluoroscopy denoising is challenging due to the relative motion of the object being imaging with respect to the imaging system as well as a lack of ground truth clean data.

Video denoising, one of the most fundamental tasks in computer vision, is a procedure that eliminates measurement noise from corrupted video data and recovers the original clean information. In mathematical terms, a corrupted video y can be represented as $y = x + n$, where x is clean video data and n is measurement noise. Conventionally, not only edge-preserving adaptive filters[10], non-local means[6,13,15,33] denoising methods, but Robust Principle Components Analysis (RPCA) [7,19,25,26,36,39], a prominent method on foreground-background problem, is often used on video denoising task, which uses a low-rank subspace model to estimate the background and a spatially sparse model to estimate the foreground. Some models [23,24] employing Total Variation (TV) on RPCA to separate foreground and background, like TVRPCA [8]. While Inc-PCP[39] can iteratively align the estimated background component. However, for non-static background video, a RPCA-based [17] model was proposed, but the approach considers only the common view of the video. In addition, when processing such video data with the Inc-PCP method, the decomposed background information is lost due to the moving background.

Recently, deep learning based denoising approaches become prominent. Fully supervised methods usually train a neural network with *corrupted/clean data pairs* [12,43,48]. However, they require collecting paired data and also have the likelihood of learning an identity mapping instead of statistic feature of noise. Self-supervised models aim to avoid this issue by either inferring real data by the local perceived corrupted data, or predicting clean images based on characteristics of noise [3,18,28,31,38]. Meanwhile, most approaches are designed for natural images, while denoising algorithms for medical imaging data are less common.

In this paper, we propose a novel self-supervised, three-stage fluoroscopy denoising framework for processing a fluoroscopy video with a non-static background. In the first stage, **Stabilize**. In order to preserve temporal consistency and stabilize the given fluoroscopy imaging data, which might still undergo a global motion, a dynamic panorama is constructed for frame registration. In the second stage, **Decompose**. We proposed a simple but novel mask-based RPCA decomposition method to separate the given imaging video with detector motion into a low-rank background and a sparse foreground. In the third stage, **Denoise.** The background and foreground will be denoised, respectively, with a self-supervised denoising strategy, and the foreground is denoised again with spatiotemporal bilateral filtering. Finally, these parts are fused together to obtain final result. We improve the denoising performance of our framework by 5∼6 dB on a dedicated fluoroscopy imaging video dataset comparing with other

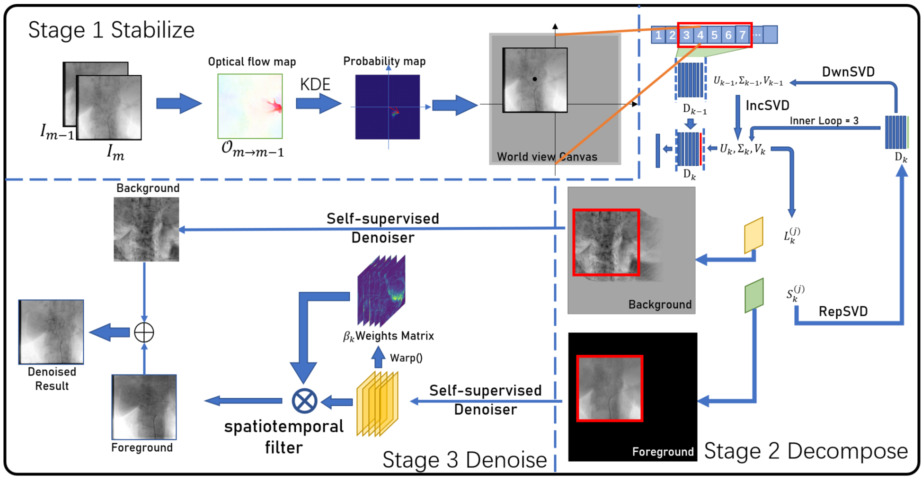

Fig. 1. Overview of the Stabilize, Decompose, and Denoise framework for fluoroscopy denoising.

approaches. Expert evaluations confirm the efficacy of our framework on clinical images too.

2 Method

We design a self-supervised framework for reducing the noise in fluoroscopy, tackling problem of background shift in RPCA decomposition, and supporting fast data processing. As in Fig. 1, our framework consists of three stages — Stabilize, Decompose and Denoise, whose details are elaborated below.

2.1 Stage 1 - Stabilize

In a fluoroscopy video, the motion of X-ray detector results in a shift in the background, which makes temporal denoising challenging. Inspired by [36], we construct a panorama on a world view canvas for each input video frame to preserve the temporal consistency of input video by compensating the translation between frames.

For a given video stream $Y := \{Y_0, Y_1, ..., Y_{T-1}\} \in \mathbb{R}^{M\times N\times T}$, we first construct a world view canvas. We use the first frame Y_0 as the reference frame of the entire video and place it at the center of the world view canvas. When any given frame Y_m is fed into this stage, we calculate an optical flow field $\mathcal{O}_{m\to m-1} \in \mathbb{R}^{M\times N\times 2}$ between Y_{m-1} and Y_m with optical flow estimator $\mathcal{G}_{flow}$, which is implemented by PWC-Net [41].

Though it is likely that the video foreground possesses more pixels than background, the motion patterns of foreground pixels are random, while those of background pixels are consistent. Thus, the distribution of the optical flow field $\mathcal{O}_{m\to m-1}$ is estimated using the kernel density estimator (KDE), then the value of $[u_{max}, v_{max}]$ with a maximum probability density value is selected as the background offset between the current and previous frames. Finally, we can obtain a panoramic current frame Y_m^{pano} by compensating the translation transformation. Each frame is placed on their appropriate position and temporal consistency of video is preserved.

2.2 Stage 2 - Decompose

For a given frame Y_i, it can be decomposed into sparse foreground S_i and low-rank background L_i according to RPCA assumption. Then, given a mask $\mathcal{M}_i \in \mathbb{R}^{MN\times 1}$ of S_i, in which ones elements in the $\mathcal{M}_i$ indicats non-zero elements of S_i, so we have $\|\mathcal{M}_i\|_F \ll \|1-\mathcal{M}_i\|_F$, and the $\|\cdot\|_F$ is Frobenius norm. Since the component L_i complies with the assumption of RPCA, it does not change significantly along the temporal axis, which means that most of the noise energy has been accumulated on the foreground. Therefore, if we denoise the background L_i to obtain $\hat{L}_i$ and denoise the foreground S_i to obtain $\hat{S}_i$, then noise energy of $\hat{Y}_i = \hat{L}_i + \hat{S}_i$ must be no more than that of $\hat{X}_i$, which is the result obtained by directly denoising the corrupted image Y_i. Assuming that X_i is a clean version of Y_i, we can prove (in supplementary materials) that

$$\|(\mathbf{1}-\mathcal{M}_i)\odot(X_i-\hat{L}_i)\|_F^2 \le \|(\mathbf{1}-\mathcal{M}_i)\odot(X_i-\hat{X}_i)\|_F^2, \tag{1}$$

where the $\odot$ is element-wise multiplication. Denoting the errors in the denoised results $\hat{Y}_i$ and $\hat{X}_i$ by $\epsilon(X_i,\hat{Y}_i)$ and $\epsilon(X_i,\hat{X}_i)$, respectively, using (1), we can further prove that $\epsilon(X_i,\hat{Y}_i) \le \epsilon(X_i,\hat{X}_i)$ (refer to supplementary materials for more details), which means that **video decomposition results in better denoising performance**.

However, since two consecutive frames after translation compensation do not perfectly overlap with each other in the canvas, we need to deal with this issue. We use $\mathcal{P}_M$ and $\mathcal{P}_{\bar{M}}$ to indicate non-overlapped area and overlapped area between Y_{i-1} and Y_i, respectively. Generally speaking, the non-overlapped area introduces new information, which makes $\mathcal{P}_M(S_i)=0$ and $\mathcal{P}_M(L_i)=\mathcal{P}_M(Y_i)$. Therefore, we have

$$\begin{cases}\mathcal{P}_{\bar{M}}(\|(\mathbf{1}-\mathcal{M}_i)\odot(X_i-\hat{L}_i)\|_F^2) \le \mathcal{P}_{\bar{M}}(\|(\mathbf{1}-\mathcal{M}_i)\odot(X_i-\hat{X}_i)\|_F^2),\\ \mathcal{P}_M(\|X_i-\hat{L}_i\|_F^2) = \mathcal{P}_M(\|X_i-\hat{X}_i\|_F^2).\end{cases} \tag{2}$$

With the help of (2), the same statement $\epsilon(X_i,\hat{Y}_i) \le \epsilon(X_i,\hat{X}_i)$ still holds (again refer to supplementary materials).

Because the conventional RPCA decomposition methods cannot tackle aforementioned problem, such as Inc-PCP[39], we proposed mask-based RPCA decomposition method as decomposition module in the framework. It inherits four operations from Inc-PCP, PartialSVD, IncSVD, RepSVD and DwnSVD. Through these operations, the U matrix reserving previous background information and two weight matrices Σ and V in the algorithm are maintained and updated. However, because the positions of two consecutive panoramic frames on world view are not the same, we need to fill the non-overlapped areas on Y_i and U before decomposition using the below equations:.

$$\begin{cases}\mathcal{P}_{M_Y}(Y_i) = \mathcal{P}_{M_Y}(U)\mathrm{diag}(\Sigma)(\mathcal{P}_{\bar{M}}(U)\mathrm{diag}(\Sigma))^{+}\mathcal{P}_{\bar{M}}(Y_i);\\ \mathcal{P}_{M_U}(U) = \mathcal{P}_{M_U}(Y_i)\mathcal{P}_{\bar{M}}(Y_i)^{+}(\mathcal{P}_{\bar{M}}(U)\mathrm{diag}(\Sigma))\mathrm{diag}(\Sigma)^{+},\end{cases} \tag{3}$$

where $\mathcal{P}_{M_U}$ indicates unknown area on U relatively to known area on Y_i, and $\mathcal{P}_{M_Y}$ refers as unknown area on Y_i corresponding to known area on U. The $(\cdot)^+$ is pseudo inverse matrix. The first equation of (3) is used to complete non-overlapped area on Y_i, and the second equation of (3) is used to fill non-overlapped blank area on U. The aforementioned Inc-PCP operations are then used to decompose Y_i into the foreground S_i and background L_i and update the U matrix, Σ matrix and V matrix.

2.3 Stage 3 - Denoise

At this stage, a single-frame self-supervised denoising network $\mathcal{F}_\theta(\cdot)$ is deployed to denoise the background L_i and foreground S_i decomposed from a corrupted frame Y_i, and the denoised results are denoted by $\hat{L}_i$ and $\hat{S}_i$, respectively. In this work, the Self2Self [38] denoising network is used, and it is worth noting that the Denoise stage is a general stage for self-supervised denoising, which means the denoising model can be substituted with other self-supervised denoising approaches. In the training phase of the Self2Self network, we first generate a lot of training samples through Bernoulli sampling $\boldsymbol{x}^i \sim \mathcal{X}$, donated as $\{\hat{\boldsymbol{x}}^i_m\}^M_{m=1}$,where $\hat{\boldsymbol{x}}^i_m := \boldsymbol{b}_m \odot \boldsymbol{x}^i$. Next, let $\overline{\boldsymbol{x}}^i_m := (\mathbf{1} - \boldsymbol{b}_m) \odot \boldsymbol{x}^i$ and $\boldsymbol{b}_m$ is a down sampling mask. The network is then trained by minimizing the following loss function

$$\mathcal{L}(\theta) = \mathbb{E}_{\boldsymbol{x} \sim \boldsymbol{\mathcal{X}}} \left[\sum_{m=1}^{M} \| \mathcal{F}_\theta(\hat{\boldsymbol{x}}_m) - \overline{\boldsymbol{x}}_m \|^2_{\boldsymbol{b_m}} \right]. \tag{4}$$

But in the testing stage, for reducing prediction time, a U-Net [40] as a student module $\mathcal{D}_\omega(\cdot)$ is added into the framework, learning the features of Self2Self network in a fully-supervised manner. The loss function $\mathcal{L}_{student}(\omega)$ is given as

$$\mathcal{L}_{student}(\omega) = \mathbb{E}_{\boldsymbol{x} \sim \boldsymbol{\mathcal{X}}} \left[\sum_{m=1}^{M} \| \mathcal{D}_\omega(\boldsymbol{x}) - \mathcal{F}_\theta(\boldsymbol{x}) \|^2_F \right]. \tag{5}$$

Finally, an optical flow based spatiotemporal bilateral filter on foregrounds is adopted for multi-frame denoising. Given $2K + 1$ consecutive foregrounds $\hat{S}_{t-K}, ..., \hat{S}_t, ..., \hat{S}_{t+K}$, for each $k \in \{-K, ..., K\}$, the optical flow $\mathcal{O}_{t+k \to t} \in \mathbb{R}^{M \times N \times 2}$ is calculated. Then we use a bilateral filter to average the warped $2K + 1$ frames for denoising $\overline{S}_t$,

$$\overline{S}_t = \sum_{k=-K}^{K} \beta_k \cdot \mathcal{W}(\hat{S}_{t+k}, \mathcal{O}_{t+k \to t}), \tag{6}$$

$$\beta_k \propto exp\{-(\mathcal{W}(\hat{S}_{t+k}, \mathcal{O}_{t+k \to t}) - \hat{S}_t)/\rho\}, \sum_k \beta_k = 1, \tag{7}$$

where $\mathcal{W}(\cdot)$ is a warping function and ρ is a parameter controlling the smoothness, with a larger ρ value producing a smoother denoising result.

Table 1. Denoising performances of various self-supervised denoising methods. In each cell, the PSNR and SSIM values are presented. The italics shows our framework adopted other denoiser in Denoise stage. The **bold** shows the best scores.

Gaussian	Method					
	Noise2Void	*Ours w/ N2V*	Noise2Self	*Ours w/ N2S*	Self2Self	*Ours w/ S2S*
0.001	19.27/.919	25.13/**.960**	30.69/.935	34.78/.948	34.39/.941	**36.47**/.941
0.003	15.74/.856	21.98/**.920**	23.99/.854	29.80/.895	29.96/.870	**34.18**/.864
0.005	14.25/.820	19.29/.890	21.01/.801	26.86/.854	27.72/.819	**32.78/.938**

3 Experiment

3.1 Setup Details

Fluoroscopy dataset. We collect 27 clean fluoroscopy videos (1,568 frames in total) with a high X-ray dose as ground truth, including 20 static-background videos and 7 non-static background videos. For evaluation, we add Gaussian noise with different variances of 0.001, 0.003, and 0.005 into clean images to simulate corrupted data with different noise levels.

Clinical dataset. We collect 60 groups of real corrupted samples by sampling low dose X-ray videos. Each group consists of 5 images, one of which is the original noisy image, and the remaining four are denoised results obtained by feeding the original noisy image into our method, Noise2Self, Noise2Void and Self2Self, respectively. In addition, for each group, the five images are permuted randomly.

Implementation and training stage. We adopt the Self2Self as denoiser, which is trained on X-ray Coronary Angiograms dataset (XCA) of 22 videos and set the probability of Bernoulli sampling as 0.3. When we train student network, the Adam optimizer is used and the learning rate is set as 10^{-4}, and it is trained with 100 epochs with a batch size of 8. In the Stabilize stage, the size of world view canvas is 2048×2048 and frame size $M \times N$ is 1024×1024. The Inc-PCP algorithm is implemented on GPU except the part of filling blank area, with the rank $r = 1$, and the windows size is 30. In spatiotemporal bilateral filter, we set $\rho = 0.02$. All experiments are implemented on an NVIDIA GeForce RTX 2080Ti GPU and using PyTorch.

Metrics. We use peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) to evaluate denoising performance by comparing the denoised images and its corresponding clean images. To evaluate the extend of blurriness for a given image, image entropy (IE) is used for quantification.

Expert rating. A proficient radiologist with about 20 years of reading experience is invited to rate the denoised image quality for our Clinical dataset. The criteria include perceived noisiness and completeness of micro-structures like small vessels. The radiologist is asked to rate all the images from each group presented in a random order, from 1 (bad) to 5 (good).

3.2 Results and Discussion

Restoration. We select some state-of-the-art self-supervised deep models, including Noise2Void[28], Noise2Self[3], Self2Self[38] for comparison. The comparison performances are demonstrated in Table 1. It is evident that the PSNR

Table 2. Denoising performances of our framework with and without the Stabilize stage. In each cell, the PSNR and SSIM values are presented.

Method	Gaussian			FPS
	0.001	0.003	0.005	
Full three-stage	**36.466/.941**	**34.183/.864**	**32.783/.938**	0.72
w/o Stabilize	36.223/.939	33.229/.852	30.239/.933	4.20

and SSIM scores of our framework (Ours+S2S, the last column) among these self-supervised denoising methods are the best, contributing an improvement of about 5∼6 dB compared with other self-supervised methods. Especially the comparison between our results and the results of Self2Self (the second to last column) clearly demonstrates the effectiveness of the Stabilize and Decompose stages. Similarly, in terms of SSIM, ours framework records the best performance. It is noted that even when the noise level increases, our framework is robust too. Fig. 2 visualizes the processed results of these methods.

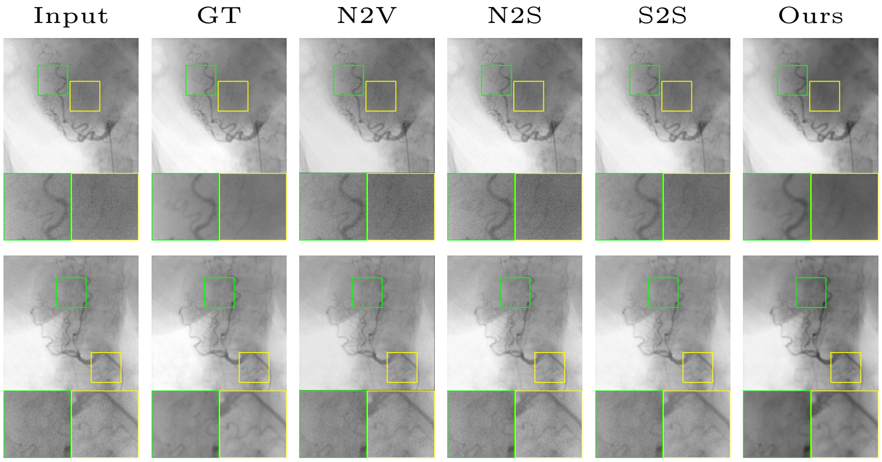

Fig. 2. Left: Visual comparison of denoising results. The images are with Gaussian noise ($\sigma = 0.003$).

Ablation study. First of all, the detector motion in fluoroscope yields background shifting, which leads to serious blurriness in the background obtained in video decomposition stage. To quantify this artifact, we calculate the image entropy (IE) for the background to verify the effectiveness of the Stabilize stage for preserving temporal consistency of video. The higher IE is, the better. We plot the IE curves of one fluroscopy video with and without the Stabilize stage as shown in Fig. 3. It is clear that the existence of the Stabilize stage preserves more information on the background. Table 2 indicates that preserving more background information can improve the denoising performance. In addition, in order to verify the generalization of our framework, we replace the denoiser in the Denoise stage with other self-supervised models, like Noise2Void, Noise2Self, the result is shown in Table 1. It is clear that our framework boosts a significant performance for all denoising network.

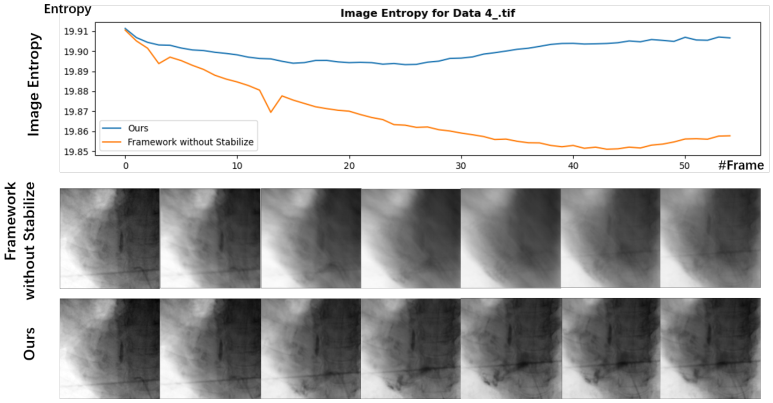

Fig. 3. Top: the image entropy curves of backgrounds generated by our framework with and without the Stabilize stage, respectively. Bottom: the visualization of backgrounds for the two results.

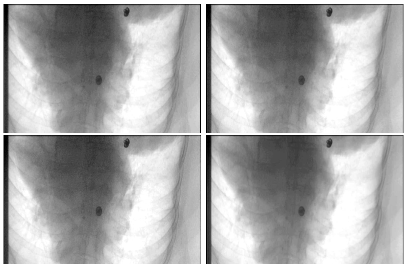

Fig. 4. Visualization of Clinical dataset.

Table 3. Ratings results on Clinical dataset

	Clinical Dataset	
	Ranking	P-Value
Raw	1.600±0.693	<0.001
Noise2Void	3.300±0.962	<0.001
Noise2Self	1.867±0.769	<0.001
Self2Self	3.350±0.840	<0.001
Ours	**4.883±0.584**	**nan**

Clinical study. Table 3 summarizes the average ratings and P-values for comparison between our model and other competing methods. The performance of our framework is significantly better than Noise2Self[3], Noise2Void[28], Self-2Self[38] on our clinical dataset. Fig. 4 shows one group of denoising results of real corrupted data. The top-left, top-right, bottom-left, and bottom-right are denoised by Noise2Void[28], Noise2Self[3], Self2Self[38] and our framework, respectively. It is obvious that the denoised image generated by our framework possesses less noise and better preserves the micro-structures in the image.

4 Conclusion

We propose a three-stages self-supervised denoising framework, consisting of the Stabilize, Decompose, and Denoise stages. In the Stabilize stage, we firstly estimate the optical flow map and then the background offset for each frame to build a panorama. Then in the second Decompose stage, a mask-based RPCA decomposition method is proposed for separating foreground and background. Finally, we invoke a self-supervised denoising method to denoise foreground and background, respectively, and fuse them together with a bilateral temporal-spatial filter as final denoised result. In experiments, visual comparisons and qualitative evaluations demonstrate that our framework yields better image quality than competing methods and exhibits a great potential of boosting self-supervised learning denoising method on Fluoroscopy dataset and Clinical dataset. In the future, we plan to employ deep learning model to estimate affine parameters so that our framework can work on more complex circumstance.

References

1. Amiot, C., Girard, C., Chanussot, J., Pescatore, J., Desvignes, M.: Spatio-temporal multiscale denoising of fluoroscopic sequence. IEEE Transactions on Medical Imaging **35**(6), 1565–1574 (2016). https://doi.org/10.1109/TMI.2016.2520092
2. Arias, P., Morel, J.M.: Video denoising via empirical bayesian estimation of space-time patches. Journal of Mathematical Imaging and Vision **60**(1), 70–93 (2018)
3. Batson, J., Royer, L.: Noise2self: Blind denoising by self-supervision. In: International Conference on Machine Learning. pp. 524–533. PMLR (2019)
4. Brand, M.: Incremental singular value decomposition of uncertain data with missing values. In: European Conference on Computer Vision. pp. 707–720. Springer (2002)
5. Buades, A., Coll, B., Morel, J..: A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 2, pp. 60–65 vol. 2 (2005). https://doi.org/10.1109/CVPR.2005.38
6. Buades, A., Coll, B., Morel, J.M.: Non-Local Means Denoising. Image Processing On Line **1**, 208–212 (2011)
7. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? Journal of the ACM (JACM) **58**(3), 1–37 (2011)
8. Cao, X., Yang, L., Guo, X.: Total variation regularized rpca for irregularly moving object detection under dynamic background. IEEE Transactions on Cybernetics **46**(4), 1014–1027 (2016). https://doi.org/10.1109/TCYB.2015.2419737
9. Cerciello, T., Bifulco, P., Cesarelli, M., Paura, L., Romano, M., Pasquariello, G., Allen, R.: Noise reduction in fluoroscopic image sequences for joint kinematics analysis. In: Bamidis, P.D., Pallikarakis, N. (eds.) XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010. pp. 323–326. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)
10. Cerciello, T., Romano, M., Bifulco, P., Cesarelli, M., Allen, R.: Advanced template matching method for estimation of intervertebral kinematics of lumbar spine. Medical engineering and physics **33**, 1293–302 (07 2011). https://doi.org/10.1016/j.medengphy.2011.06.009
11. Chen, Y.C.: A tutorial on kernel density estimation and recent advances. Biostatistics & Epidemiology **1**(1), 161–187 (2017)
12. Claus, M., van Gemert, J.: Vidенn: Deep blind video denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–0 (2019)
13. Dabov, K., Foi, A., Egiazarian, K.: Video denoising by sparse 3d transform-domain collaborative filtering. In: 2007 15th European Signal Processing Conference. pp. 145–149 (2007)
14. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on image processing **16**(8), 2080–2095 (2007)
15. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on Image Processing **16**(8), 2080–2095 (2007). https://doi.org/10.1109/TIP.2007.901238
16. Dewil, V., Anger, J., Davy, A., Ehret, T., Facciolo, G., Arias, P.: Self-supervised training for blind multi-frame video denoising. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2724–2734 (2021)
17. Ebadi, S.E., Guerra-Ones, V., Izquierdo, E.: Approximated robust principal component analysis for improved general scene background subtraction. ArXiv **abs/1603.05875** (2016)

18. Ehret, T., Davy, A., Morel, J.M., Facciolo, G., Arias, P.: Model-blind video denoising via frame-to-frame training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11369–11378 (2019)
19. Feng, J., Xu, H., Yan, S.: Online robust pca via stochastic optimization. In: Advances in neural information processing systems. pp. 404–412. Citeseer (2013)
20. Foi, A.: Clipped noisy images: Heteroskedastic modeling and practical denoising. Signal Processing **89**(12), 2609–2629 (2009), special Section: Visual Information Analysis for Security
21. Gilboa, G., Sochen, N., Zeevi, Y.: Variational denoising of partly textured images by spatially varying constraints. IEEE Transactions on Image Processing **15**(8), 2281–2289 (2006). https://doi.org/10.1109/TIP.2006.875247
22. Grissom II, A., He, H., Boyd-Graber, J., Morgan, J., Daum e, H.: Don't until the final verb wait: Reinforcement learning for simultaneous machine translation. In: Empirical Methods in Natural Language Processing (2014)
23. Guyon, C., Bouwmans, T., hadi Zahzah, E.: Foreground detection via robust low rank matrix decomposition including spatio-temporal constraint. In: ACCV Workshops (2012)
24. Guyon, C., Bouwmans, T., ZAHZAH, E.h.: Foreground detection via robust low rank matrix factorization including spatial constraint with iterative reweighted regression (11 2012)
25. Han, S., Cho, E.S., Park, I., Shin, K., Yoon, Y.G.: Efficient neural network approximation of robust pca for automated analysis of calcium imaging data. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 595–604. Springer International Publishing, Cham (2021)
26. He, J., Balzano, L., Szlam, A.: Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1568–1575. IEEE (2012)
27. Kostadin, D., Alessandro, F., Karen, E.: Video denoising by sparse 3d transform-domain collaborative filtering. In: European signal processing conference. vol. 149 (2007)
28. Krull, A., Buchholz, T.O., Jug, F.: Noise2void-learning denoising from single noisy images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2129–2137 (2019)
29. Lai, W.S., Huang, J.B., Wang, O., Shechtman, E., Yumer, E., Yang, M.H.: Learning blind video temporal consistency. In: Proceedings of the European conference on computer vision (ECCV). pp. 170–185 (2018)
30. Laine, S., Karras, T., Lehtinen, J., Aila, T.: High-quality self-supervised deep image denoising. arXiv preprint arXiv:1901.10277 (2019)
31. Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T.: Noise2noise: Learning image restoration without clean data. arXiv preprint arXiv:1803.04189 (2018)
32. Li, J., Monroe, W., Ritter, A., Jurafsky, D., Galley, M., Gao, J.: Deep reinforcement learning for dialogue generation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 1192–1202. Association for Computational Linguistics, Austin, Texas (Nov 2016). https://doi.org/10.18653/v1/D16-1127, https://aclanthology.org/D16-1127
33. Maggioni, M., Boracchi, G., Foi, A., Egiazarian, K.: Video denoising using separable 4d nonlocal spatiotemporal transforms. Proceedings of SPIE

- The International Society for Optical Engineering **7870** (02 2011). https://doi.org/10.1117/12.872569

34. Maggioni, M., Boracchi, G., Foi, A., Egiazarian, K.: Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms. IEEE Transactions on image processing **21**(9), 3952–3966 (2012)
35. Matviychuk, Y., Mailhé, B., Chen, X., Wang, Q., Kiraly, A., Strobel, N., Nadar, M.: Learning a multiscale patch-based representation for image denoising in x-ray fluoroscopy. In: 2016 IEEE International Conference on Image Processing (ICIP). pp. 2330–2334 (2016). https://doi.org/10.1109/ICIP.2016.7532775
36. Moore, B.E., Gao, C., Nadakuditi, R.R.: Panoramic robust pca for foreground–background separation on noisy, free-motion camera video. IEEE Transactions on Computational Imaging **5**(2), 195–211 (2019)
37. Paulus, R., Xiong, C., Socher, R.: A deep reinforced model for abstractive summarization. In: International Conference on Learning Representations (2018), `https://openreview.net/forum?id=HkAClQgA-`
38. Quan, Y., Chen, M., Pang, T., Ji, H.: Self2self with dropout: Learning self-supervised denoising from single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1890–1898 (2020)
39. Rodriguez, P., Wohlberg, B.: Incremental principal component pursuit for video background modeling. Journal of Mathematical Imaging and Vision **55**(1), 1–18 (2016)
40. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
41. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8934–8943 (2018)
42. Tassano, M., Delon, J., Veit, T.: Fastdvdnet: Towards real-time deep video denoising without flow estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1354–1363 (2020)
43. Wang, C., Zhou, S.K., Cheng, Z.: First image then video: A two-stage network for spatiotemporal video denoising. arXiv preprint arXiv:2001.00346 (2020)
44. Wang, N., Yao, T., Wang, J., Yeung, D.Y.: A probabilistic approach to robust matrix factorization. In: European Conference on Computer Vision. pp. 126–139. Springer (2012)
45. Xu, J., Ithapu, V.K., Mukherjee, L., Rehg, J.M., Singh, V.: Gosus: Grassmannian online subspace updates with structured-sparsity. In: Proceedings of the IEEE international conference on computer vision. pp. 3376–3383 (2013)
46. Yu, C., Liu, J., Nemati, S.: Reinforcement learning in healthcare: A survey. ArXiv **abs/1908.08796** (2019)
47. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing **26**(7), 3142–3155 (2017)
48. Zhang, K., Zuo, W., Zhang, L.: Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. IEEE Transactions on Image Processing **27**(9), 4608–4622 (2018)
49. Zhou, T., Tao, D.: Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In: Proceedings of the 28th International Conference on Machine Learning, ICML 2011 (2011)

50. Zhou, T., Tao, D.: Shifted subspaces tracking on sparse outlier for motion segmentation. In: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence. p. 1946–1952. IJCAI '13, AAAI Press (2013)
51. Zhou, X., Yang, C., Yu, W.: Moving object detection by detecting contiguous outliers in the low-rank representation. IEEE transactions on pattern analysis and machine intelligence **35**(3), 597–610 (2012)