



Estimation of High Framerate Digital Subtraction Angiography Sequences at Low Radiation Dose

Nazim Haouchine, Parikshit Juvekar, Xin Xiong, Jie Luo, Tina Kapur, Rose Du, Alexandra Golby, Sarah Frisken

► To cite this version:

Nazim Haouchine, Parikshit Juvekar, Xin Xiong, Jie Luo, Tina Kapur, et al.. Estimation of High Framerate Digital Subtraction Angiography Sequences at Low Radiation Dose. MICCAI 2021 - Medical Image Computing and Computer Assisted Interventions, Sep 2021, Strasbourg, France. hal-03675006

HAL Id: hal-03675006

<https://inria.hal.science/hal-03675006>

Submitted on 22 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation of High Framerate Digital Subtraction Angiography Sequences at Low Radiation Dose

Nazim Haouchine^{1,2}, Parikshit Juvekar^{1,2}, Xin Xiong^{2,3}, Jie Luo^{1,2}, Tina Kapur^{1,2}, Rose Du^{1,2}, Alexandra Golby^{1,2}, and Sarah Frisken^{1,2}

¹ Harvard Medical School, Boston, MA, USA

² Brigham and Women's Hospital, Boston, MA, USA

³ Columbia University, New York City, NY, USA

Abstract. Digital Subtraction Angiography (DSA) provides high resolution image sequences of blood flow through arteries and veins and is considered the gold standard for visualizing cerebrovascular anatomy for neurovascular interventions. However, acquisition frame rates are typically limited to 1-3 fps to reduce radiation exposure, and thus DSA sequences often suffer from stroboscopic effects. We present the first approach that permits generating high frame rate DSA sequences from low frame rate acquisitions eliminating these artifacts without increasing the patient's exposure to radiation. Our approach synthesizes new intermediate frames using a phase-aware Convolutional Neural Network. This network accounts for the non-linear blood flow progression due to vessel geometry and initial velocity of the contrast agent. Our approach outperforms existing methods and was tested on several low frame rate DSA sequences of the human brain resulting in sequences of up to 17 fps with smooth and continuous contrast flow, free of flickering artifacts.

Keywords: Biomedical Image Synthesis · Digital Subtraction Angiography · Video Interpolation · Convolutional Neural Networks

1 Introduction and Related Work

Cerebrovascular diseases disrupt the circulation of blood in the brain and include aneurysms, developmental venous angiomas and arteriovenous malformations (AVMs). These diseases may cause blood vessels to rupture that can result in brain hemorrhage with complications including severe headache, seizures, hydrocephalus, brain damage and death. Surgical removal, which is curative, is a preferred treatment [8]. However, it is not without risks. The main risk is post-surgical deficit due to hemorrhage [10]. These risks depend on the malformation, clot size, proximity to the eloquent cortex, the presence of diffuse feeding arteries and deep venous drainage (in the case of AVMs). Risks can be mitigated with meticulous surgical planning and prior understanding of blood flow circulation by careful study of preoperative imaging, which often includes digital subtraction

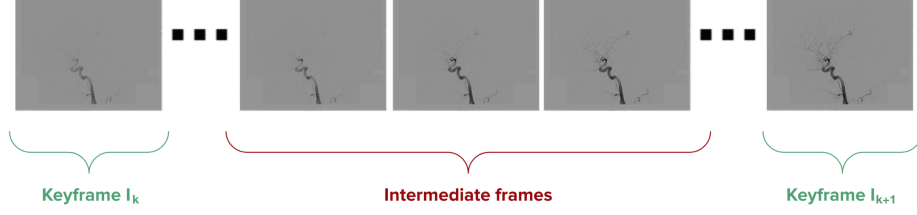


Fig. 1: Our method produces new intermediate images $\{\hat{\mathbf{I}}_{(k,k+1)}\}_N$ between each pair of input images \mathbf{I}_k and \mathbf{I}_{k+1} from a low frame rate DSA sequence to generate a new high frame rate DSA sequence.

angiography (DSA), computed tomography angiography or magnetic resonance angiography.

DSA is considered the "gold standard" imaging method for evaluating cerebral malformations. It provides high resolution (0.3mm pixels), dynamic imaging of blood flow through the brain during arterial filling and venous drainage [9]. To visualize blood flow and vascular malformations, a contrast agent/dye is injected into a large artery progressing through the smaller arteries, the capillaries, and then out through the veins (and back to the heart). DSA acquisition protocols involve several parameters, including radiation dose and frame rate [6]. Because higher frame rates acquisitions expose patients to more radiation, an adequate trade-off between acquisition frame rate and radiation dose must be made to reduce patients exposure to radiation[11]. Therefore, current best practice routinely limits the frame rate to 1-3 fps, with frame rates of up to 7.5 fps in some cases [6]. Unfortunately, low frame rate DSA produces stroboscopic (or flicker) effects [11] that make it more challenging for clinicians to interpret the complex dynamic cerebrovascular blood flow, especially in the presence of complex pathology such as arteriovenous malformations.

To eliminate flicker effects, a higher frame rate is required. Because this would require an unsafe radiation dose using conventional methods, we propose a new approach that synthesizes intermediate images to increase the frame rate of low frame rate DSA sequences without increasing radiation dose. Synthesizing images has been studied for the purpose of video interpolation [2] and slow-motion generation [3], for registration of medical images [5] and for graphical animations and rendering [7]. Our solution follows these approaches while preserving coherent blood hemodynamics in intermediate images and does not alter images from the original sequence. Because the progression of contrast agents in the neurovascular structure is non-linear, our method first decomposes the sequence into three phases: arterial, capillary and venous. We extract from these phases an estimate of the volume of the contrast agent that, combined with the original images, will be fed to a Convolutional Neural Network (CNN) to generate new images. The optimized loss function is designed to focus on the regions with high entropy to preserve details of the predicted images. The final sequence is assem-

bled by iteratively interleaving the intermediate images with the original images (see Figure 1). We believe that our method will reduce the challenges of interpreting DSA, particularly in the presence of complex cerebrovascular malformations, thus helping to improve preoperative planning and surgical monitoring with the goal of reducing surgical complication rates.

2 Methods

As illustrated in Figure 1, given two input images \mathbf{I}_k and \mathbf{I}_{k+1} , our goal is to predict the intermediate image $\hat{\mathbf{I}}_{(k,k+1)}$. Inspired by recent work in video frame interpolation, we rely on CNNs to predict the intermediate image. Because we want to control the contribution of the input images \mathbf{I}_k and \mathbf{I}_{k+1} on $\hat{\mathbf{I}}_{(k,k+1)}$, we take into account the non-linear progression of the contrast through the vascular network. We estimate the contrast agent volume for each image by decomposing the DSA sequence into arterial, capillary and venous phases. In addition, we constrain the model to learn features in regions of the image with rich vessel information. The final composition includes the input and intermediate images and can be repeated iteratively, resulting in a high frame rate DSA sequence.

2.1 Phase Decomposition using Independent Component Analysis

In order to obtain an estimate of the contrast progression per image, we start by decomposing the DSA sequence into arterial, capillary and venous phases as illustrated in Figure 2. Because blood flow propagation behaves differently during these three phases, this decomposition permits us to adapt the interpolation mechanism to each phase. We perform the decomposition using Independent Component Analysis (ICA) [4]. ICA is a statistical method that extracts sub-components of a 1D-signal under the assumption that these subcomponents are independent. The DSA images composing a sequence are first stacked and vectorized to produce a unique 1D-signal. Using ICA on this signal (with three classes) will generate three separate signals corresponding to the arterial, capillary and venous phases. These signals are transformed back to 2D to obtain three distinct images which encode signal information, as well as noise and outliers. To clearly separate meaningful information from the noise and outliers, we use image-base binary thresholding on the histogram distribution. The three thresholded images are used as binary masks to estimate time-density curves (TDCs) [4]. Using the TDCs, we can estimate contrast agent volume that runs through the vascular network between two consecutive images. This amount is normalized between 0 and 1 w.r.t the phase’s peak volume. Assuming a sequence $\{\mathbf{I}_i\}_{i=1}^N$ of N DSA images, we can now define the function $v_{(i,j)} = p(\{\mathbf{I}_i\}_{i=1}^N, i, j)$ that estimates for any consecutive images \mathbf{I}_i and \mathbf{I}_j , the contrast volume $v_{i,j}$, with $j \in (1, N)$ and $j > i$. This volume will be used to control the contribution of the input images to the intermediate image.

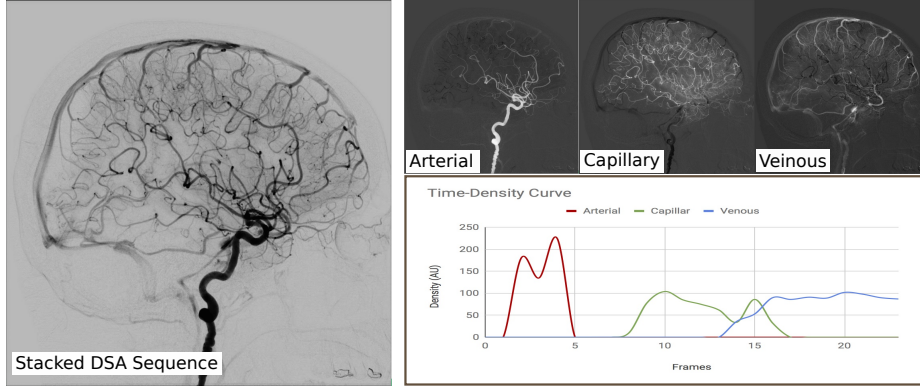


Fig. 2: Decomposition of DSA sequence into arterial, capillary and venous phases using ICA. After stacking and vectorizing all images into a 1D-signal, we can separate the signal to three distinct components and obtain a time-density curve.

2.2 Training and Optimization

Given the training set $\mathcal{T}^e = \{(\mathbf{I}_j, \mathbf{L}_j)\}_j$ composed of DSA images \mathbf{I}_j and their corresponding binary labels \mathbf{L}_j , with $j \in (1, N_{\mathcal{T}^e})$, we first start by training a region-of-interest extractor $e(\mathbf{I}; \theta_r)$ with θ_r being the learned parameters for network e . Because a binary segmentation may discard vessels with small diameters and information about the vessels/background boundaries, we use the extractor to generate a per-pixel entropy map \mathbf{M} that will associate with each pixel of an image \mathbf{I} the probability of that pixel being part of the vessels or the background (See Figure 3). The higher the probability the richer the information around the pixel. We used a patch-based method [1] that optimizes a mean-square error loss function over the parameters θ_r of the network e .

Then given the training set $\mathcal{T}^g = \{(\mathbf{X}_j, \mathbf{Y}_j)\}_j$, we train a generator network $g(\mathbf{X}; \mathbf{M}; \theta_g)$ over the parameters θ_g to interpolate new images. \mathbf{X}_j is composed of a pair of images \mathbf{I}_k and \mathbf{I}_{k+2} and the contrast agent volume $v_{(k,k+2)}$ with $k \in (1, N_{\mathcal{T}^g})$. \mathbf{Y}_j corresponds to the output \mathbf{I}_{k+1} that consists of a skipped image that represents the true intermediate image. Finally, $\mathcal{M} = \{\mathbf{M}_k\}_{k=1}^{N_{\mathcal{T}^g}}$ is the set of entropy maps. Note that \mathbf{M} is not considered an input to train g but will be used in the optimization loss function.

Given an intermediate image \mathbf{I}_k and our predicted intermediate image $\hat{\mathbf{I}}_k$, the optimization loss function over the parameters θ_g of the network g is as follows:

$$\begin{aligned}
 \mathcal{L}(\theta_g; \mathcal{M}; \mathcal{T}^g) &= \mathcal{L}_{\text{reconstruction}} + \mathcal{L}_{\text{perceptual}} \\
 &= \frac{\alpha}{N_{\mathcal{T}^g}} \sum_{k=1}^{N_{\mathcal{T}^g}} \mathbf{M}_k \sqrt{\|\hat{\mathbf{I}}_k - \mathbf{I}_k\|^2 + \epsilon^2} + \frac{\beta}{N_{\mathcal{T}^g}} \sum_{k=1}^{N_{\mathcal{T}^g}} \|\phi(\hat{\mathbf{I}}_k) - \phi(\mathbf{I}_k)\|_2
 \end{aligned} \tag{1}$$

where α and β are meta-parameters to control the interaction of the loss function components. The *reconstruction loss* models how good the generation of the image is. We opted for a Charbonnier loss function and confirmed the observations in [17] that it performs better than an ℓ_1 loss function. We penalize this loss using the entropy maps \mathcal{M} to enforce the network to focus on rich vessel information. Moreover, using the reconstruction error alone might produce blur in the predictions. We thus use a *perceptual loss* [14] to make interpolated images sharper and preserve their details, where ϕ denote the `conv4_3` features of an ImageNet pretrained VGG16 model [19].

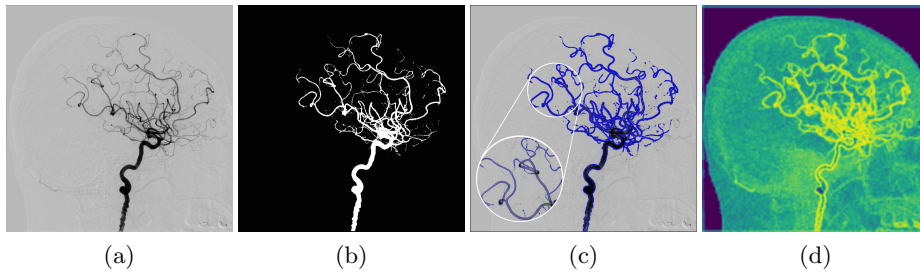


Fig. 3: Segmentation and entropy maps generation: (a) input DSA image, (b) binary segmentation, (c) extracted contours and (d) entropy map. Although binary segmentation gives accurate geometry, it may discard useful information leading to discrepancies. We use entropy maps instead of binary images to preserve small vessels and information about the vessels/background boundaries.

2.3 Network Details

We adopt a U-Net architecture [18] for both e and g networks. The network e architecture is straightforward and built upon a ResNet34 pretrained model [13] and a `Softmax` final activation function to generate the entropy map. The network g consists of 6 hierarchies in the encoder, composed of two convolutional and one ReLU layers. Each hierarchy except the last one is followed by an average pooling layer with a stride of 2 to decrease the spatial dimension. There are 6 hierarchies in the decoder part. A bilinear upsampling layer is used at the beginning of each hierarchy, doubling the spatial dimension, followed by two convolutional and ReLU layers. The input consists of a stacked pair of images, while the contrast volume scalar is pointwise added as a 2D features array at the bottom of the network where features are the most dense and spatial dimensions are the least.

For the parameters of the loss function in Eq. 1, we empirically chose, using a validation set, $\alpha = 1$ and $\beta = 0.01$. We optimize the loss using gradient descent [15] for 80 epochs. using mini-batches of size 16 and with a learning rate of 0.001.

2.4 Final Composition

To increase the framerate of the DSA sequence $\{\mathbf{I}_i\}_{i=1}^N$ we infer an intermediate image for each pair of successive images following:

$$\hat{\mathbf{I}}_{(i,i+1)} \leftarrow g(\{\mathbf{I}_i, \mathbf{I}_{i+1}, p(\{\mathbf{I}_i\}_{i=1}^N, i, i+1)\}; e(\mathbf{I}_{i+1}; \hat{\theta}_r); \hat{\theta}_e) \quad (2)$$

where $(\hat{\theta}_r, \hat{\theta}_e)$ are the network parameters found during the training. Using the entropy to enforce learning features on the regions with rich vessel information has the drawback of mis-interpolating the background and creates a visually disturbing effect. Thus, we add a blending step to build the final image $\hat{\mathbf{I}}_{(i,i+1)}^f$ that linearly interpolates the pixels with low entropy as follows:

$$\hat{\mathbf{I}}_{(i,i+1)}^f = \begin{cases} \delta \odot \mathbf{I}_i + (1 - \delta) \odot \mathbf{I}_{i+1}, & \text{if } e(\hat{\mathbf{I}}_{(i,i+1)}; \hat{\theta}_r)(x) \leq \eta \\ \hat{\mathbf{I}}_{(i,i+1)}, & \text{otherwise} \end{cases} \quad (3)$$

where \odot is an element-wise multiplication function and $x \in (1, w \times h)$ represents a pixel of the entropy map of size $w \times h$. The parameter δ that controls the interpolation is set empirically to $\delta = 0.6$ while η , a threshold parameter to discard regions with high entropy is set to $\eta = 0.1$.

3 Results

Dataset: Our dataset is composed of 32 DSA sequences for a total of 3216 DSA images (after geometric transformations augmentation). These images were randomly split and 25% of the images were kept for validation. All images were acquired on human subjects using a standard clinical protocol with a biplane General Electric imaging system. We used both frontal and lateral image sequences. The sequences were acquired at 1-3 fps and captured the full cycle of contrast inflow and washout after injection. The dataset \mathcal{T}^e is composed of the the totality of images with their corresponding labeled images while the dataset \mathcal{T}^g is composed of pair of images with a leave-one-image-out strategy, skipping every other image to be used as output for the model.

	woP-MSE	P-MSE	P-eMSE	P-Ch	P-eCh (full model)
PSNR (db)	38.31	38.80	39.90	40.10	40.40
IE	11.93	10.90	8.76	8.60	8.00

Table 1: Effectiveness of different components of our model.

Ablation study on the impact of different components design: We first perform an ablation study to analyze the contribution of each component of our approach. We test the impact of having a phase-constrained model, the impact of using a Charbonnier loss instead of a mean-square loss (MSE) and finally the impact of using the entropy maps. To this end, we train five variants of our model: **woP-MSE** which is a traditional U-Net without phase constraints and optimized using an MSE loss function, **P-MSE** and **P-eMSE** which are phase-constrained models using MSE loss function with and without the entropy maps respectively and finally, **P-Ch** and **P-eCh** (full model) which are phase-constrained models using Charbonnier loss function with and without the entropy maps respectively. To quantify the accuracy of our method we use the Peak Signal-to-Noise Ratio (PSNR) as a measure of corruption and noise and the interpolation error (IE), which is defined as the root-mean-squared difference between the ground-truth image and the interpolated image, as a measure of accuracy.

We can observe from Table 1 that removing the phase knowledge from the model harms performance, particularly the interpolation error, while using the Charbonnier loss slightly improves the results w.r.t MSE loss. We also verify that adding the entropy maps improves the predictions, which validate our hypothesis to enforce a learning on regions with rich vessel information.

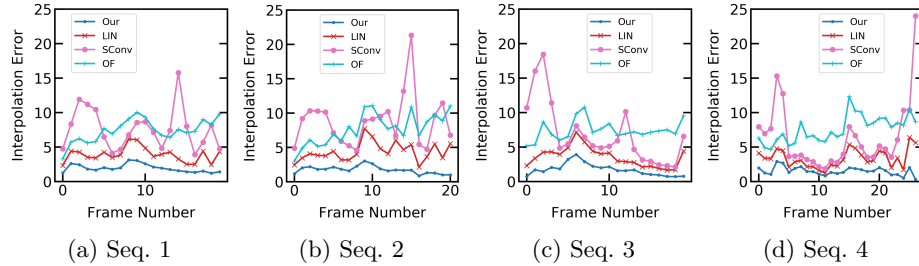


Fig. 4: Comparison with state-of-the-art methods.

Comparison with state-of-the-art methods: We then compare our approach with state-of-the-art methods including neural and non-neural approaches. In addition to our approach (Our), we include a simple bilinear interpolation method (Lin), an optical-flow based interpolation (OF) [12] and a CNN-based method (SConv) [16] for slow motion video generation. We conduct the comparison on four 7.5 fps DSA sequences that are not used during the training. For each triplet of images we leave the middle image out to serve as ground truth. We report the interpolation error in Figure 4 that shows that our model achieves the best performance on all sequences, on all images. We can notice that our method is only slightly impacted by the contrast agent progression over time, as opposed to the other methods. The performance of our model validates the

generalization ability of our approach. Furthermore, in addition to the quantitative measurements, Figure 5 shows the visual differences between our and others methods and highlights its efficiency.

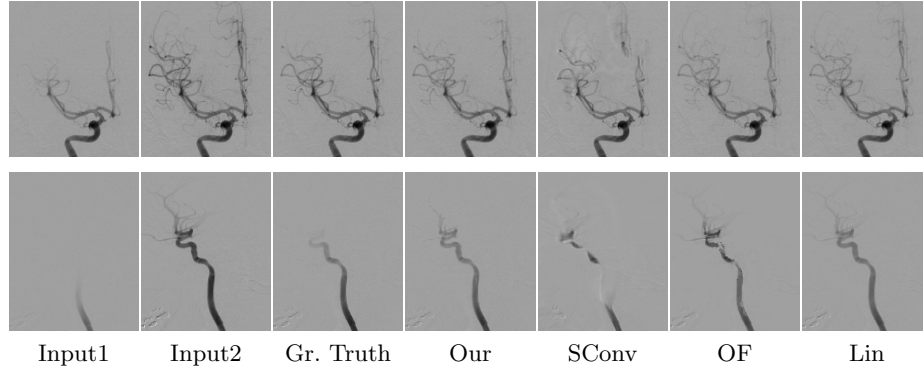


Fig. 5: Samples from the DSA sequences used in our experiments.

Iterative interpolation: Finally, using our solution we successfully produce high frame rate sequences of up to 17 fps from 3-fps DSA sequences by iteratively interpolating intermediate images from previously predicted images. Figure 6 shows examples from our dataset including arterial, capillary and venous phases with and without the presence of AVMs.

4 Conclusion

We have presented a solution to generate high frame rate DSA sequences at low radiation dose from low frame rate DSA sequences. Using our method, we can increase the framerate of DSA sequences to obtain a continuous blood and contrast flow progression through the cerebral vasculature. The presented approach is clinically practical and can be used with commercially available systems to help clinicians understand the complex dynamic cerebrovascular blood flow, especially in the presence of complex malformations. Our solution is applicable to different organs and procedures, although our experiments involved neurovascular imaging, there is no actual technical limitation for the use of our method in the diagnosis of pulmonary embolisms, renal artery stenosis or any treatment of arterial and venous occlusions.

Our current method is limited to single-frame interpolation and could be extended to produce variable-length multi-image video, where multiple intermediate images are interpolated simultaneously. In addition, future work will investigate ways to estimate 3D high frame rate DSA with the ultimate goal of improving the understanding and diagnosis of arteriovenous malformations.

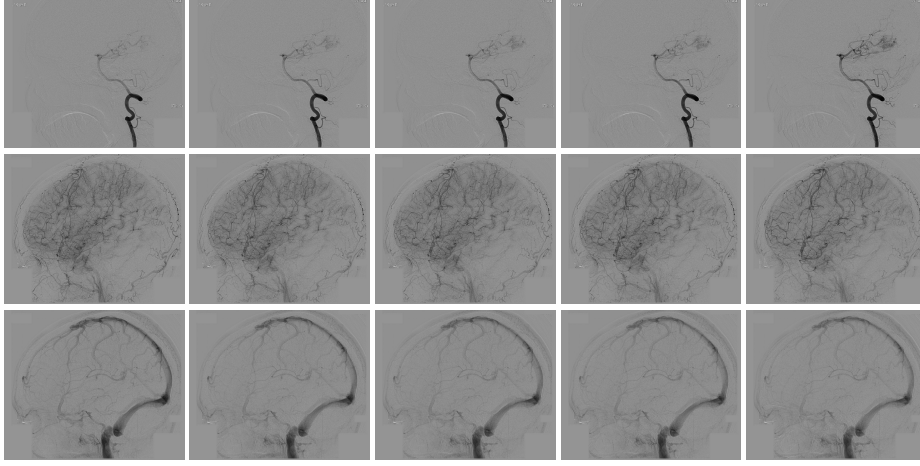


Fig. 6: Samples of high frame rate sequences generated using our method. The first and last column represents the input successive images, and the middle columns represent the estimated intermediate images. Row 1 is the arterial phase (with an AVM), row 2 is the capillary phase and row 3 is the venous phase.

Acknowledgement

The authors were supported by the following funding bodies and grants: NIH: R01 EB027134-01, NIH: R03 EB032050 and BWH Radiology Department Research Pilot Grant Award.

References

1. Meng et al., C.: Multiscale dense convolutional neural network for dsa cerebrovascular segmentation. *Neurocomputing* **373**, 123 – 134 (2020)
2. Herbst et al., E.: Occlusion reasoning for temporal interpolation using optical flow. In: Microsoft Technical report (2009)
3. Jiang et al., H.: Super slo-mo: High quality estimation of multiple intermediate frames for video interpolation. In: *IEEE CVPR*. pp. 9000–9008 (2018). <https://doi.org/10.1109/CVPR.2018.00938>
4. Hong et al., J.S.: Validating the automatic independent component analysis of dsa. *American Journal of Neuroradiology* (2019). <https://doi.org/10.3174/ajnr.A5963>
5. Leng et al., J.: Medical image interpolation based on multi-resolution registration. *Computers and Mathematics with Applications* **66**(1), 1 – 18 (2013)
6. Pearl et al., M.: Practical techniques for reducing radiation exposure during cerebral angiography procedures. *Journal of neurointerventional surgery* **7**, 141–145 (01 2014). <https://doi.org/10.1136/neurintsurg-2013-010982>
7. Narita et al., R.: Optical flow based line drawing frame interpolation using distance transform to support inbetweens. In: *IEEE ICIP*. pp. 4200–4204 (2019)

8. Starke et al., R.: Treatment guidelines for cerebral arteriovenous malformation microsurgery. *Neurosurg.* **23**(4), 376–386 (2009)
9. Chng et al., S.M.: Arteriovenous malformations of the brain and spinal cord. In: *Neurology and Clinical Neuroscience*, pp. 595 – 608. Mosby (2007)
10. Thana et al., T.: Microsurgery for cerebral arteriovenous malformations: postoperative outcomes and predictors of complications in 264 cases. *Neurosurgical Focus FOC* **37**(3), E10 (2014)
11. Balter, S.: Practical techniques for reducing radiation exposure during cerebral angiography procedures. *IAJR Am J Roentgenol.* **3**, 234–236 (2014). <https://doi.org/10.2214/AJR.13.11041>
12. Brox, T., Malik, J.: Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE transactions on pattern analysis and machine intelligence* **33**, 500–13 (03 2011). <https://doi.org/10.1109/TPAMI.2010.143>
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
14. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *European Conference on Computer Vision* (2016)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (2015)
16. Niklaus, S., Mai, L., Liu, F.: Video frame interpolation via adaptive separable convolution. In: *IEEE International Conference on Computer Vision* (2017)
17. Park, J., Ko, K., Lee, C., Kim, C.S.: Bmbc: Bilateral motion estimation with bilateral cost volume for video interpolation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) *Computer Vision – ECCV 2020*. pp. 109–125. Springer International Publishing, Cham (2020)
18. Ronneberger, O., P.Fischer, Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. LNCS, vol. 9351, pp. 234–241 (2015)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations* (2015)