



Digital Subtraction Angiography Generation with Deep Decoupling Network

Ruibo Liu¹, Li Wang¹, He Ma¹, Wei Qian¹, Guobiao Liang², Guangxin Chu², Hai Jin^{2*}

¹ College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110167, China

² Department of Neurosurgery, The General Hospital of Northern Theater Command, Shenyang 110167, China

Corresponding Author Email: kingsea300809@163.com

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410445>

ABSTRACT

Received: 10 June 2024
Revised: 17 July 2024
Accepted: 29 July 2024
Available online: 31 August 2024

Keywords:

cerebrovascular disease, disentangled representation learning, decoupling training strategy, digital subtraction angiography

Cerebrovascular disease constitutes a set of disorders within the brain, and its rates of disability and mortality are extremely high. Digital subtraction angiography serves as a crucial clinical image data for the diagnosis and treatment of cerebrovascular diseases, and the quality of it will influence the complexity of diagnosis and interventional therapy. Nevertheless, artifacts frequently occur in the images, potentially obscuring or blocking the vascular structure. In this paper, it was proposed that a novel deep learning framework could utilize the training strategy of information decoupling to generate superior quality DSA images. We proposed the deep decoupling network, a feature decoupling framework. Through a decoupling training strategy founded on disentangled representation learning, it is capable of maximizing the disparity among diverse structures. The results indicated that our method obviously superior to existing methods in all metrics with the SSIM of 93.57% and the PSNR of 24.18dB. Our framework has the capacity to generate DSA images with few artifacts and create DSA images of superior quality characterized by complete and distinct vascular structures.

1. INTRODUCTION

Cerebrovascular diseases, such as arteriovenous fistula, intracranial aneurysm, intracranial arterial stenosis, and malformation, etc., affect the brain [1-4]. These diseases have a high prevalence among middle-aged and elderly people, along with a high rate of disability and fatality. Digital subtraction angiography (DSA) is frequently utilized in the clinical diagnosis and interventional treatment of cerebrovascular diseases [5-7]. Compared to CTA and MRA, DSA is regarded as the gold standard for cerebrovascular diseases due to its high resolution and clear vascular structure [8-10]. In clinical practice, as illustrated in Figure 1, DSA images are procured by performing a dual subtraction of live images (with the contrast agent) and mask images (without the contrast agent) by virtue of the same detector to nullify the impact of tissues unrelated to the vascular structure. Given the protracted inter-frame time interval during the acquisition process between acquisition processes, the advent of artifacts such as motion artifacts, aliasing artifacts, and beam hardening artifacts becomes inevitable. Due to the protracted time interval during the acquisition process, the occurrence of artifacts becomes unavoidable. Furthermore, even the slightest movement of the patient can give rise to discernible artifacts in the DSA images. The existence of these artifacts significantly reduces image quality and blurs out the small vascular structures that are otherwise difficult to notice. This makes it difficult to diagnosis the types of diseases and determine the location of lesions, which indirectly reduces the success rate of interventional therapy. Consequently, reducing

or even eliminating the artifacts in DSA images has always been regarded as an essential yet challenging undertaking.

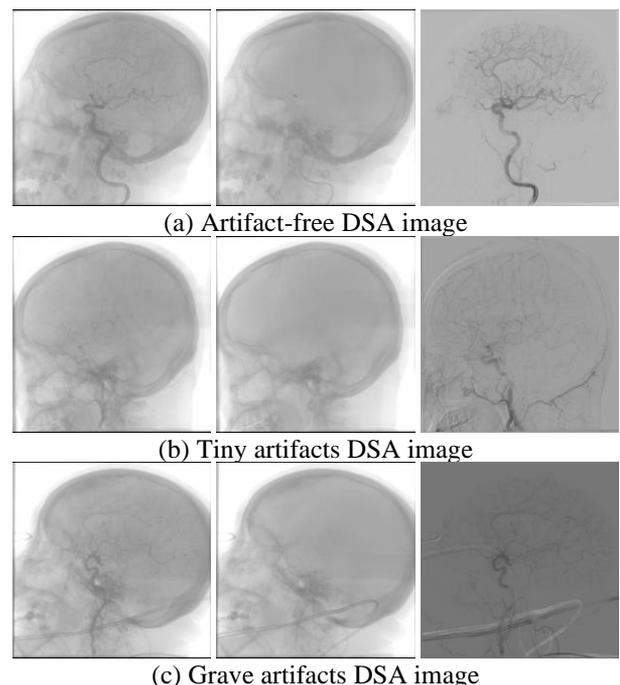


Figure 1. Illustrations of cerebrovascular DSA images are presented. The images are respectively the live image, the mask image, and the subtraction image from left to right

Although artificial intelligence has demonstrated excellent performance in clinical tasks like lesion detection, segmentation, and classification [11, 12], relatively few studies have focused on DSA image generation. At present, two principal research orientations currently exist in this task. The first one pertains to the registration of the two images through motion alignment [13, 14]. And the second one, referred to as virtual DSA, endeavors to generate DSA images from single live images by Generative Adversarial Nets (GANs) [15-17]. In GANs, the generator is employed to obtain DSA images from the single live images, and the discriminator distinguishes the global truth and the generated results. Despite certain accomplishments of these algorithms in DSA image generation, two major issues persist to be addressed. Firstly, mask images exhibit a high correlation with artifacts, entailing their minimal utilization. Secondly, the small vessels might become blurred or disappear in the absence of the assistance of the mask images. Thus, the actual mask image simultaneously serves as both an element that disturbs the generated image and a protective measure against the vanishing of the vascular structure. A positive settlement is indispensable to handle this equivocal issue.

In this paper, we present the deep decoupling network (DDN), which consists of the vessel network (V) and the mask network (M). V is responsible for extracting the vascular structure, and M is for extracting bone tissue to assist V in obtaining accurate features. Moreover, a conjoint loss function is established to restrict the framework, intending to gain the greatest differences from the extracted features. With the process of the decoupling of the vascular structure from other structures, they can be differentiated from live images. Alternatively, the two generators extract more comprehensive vascular structures and less artifacts. The key contributions of this work can be recapitulated as follows:

- DDN can generate DSA images without actual mask images. The C-arm detector only need to work once, which can reduce the radiation dose they receive.
- The specific tubular vessel structure can be extracted more effectively with the help of the axial residual block.
- A learnable sampling approach is put forward to effectively prevent the lack of crucial features. Meanwhile, the artifacts are got rid of and the integrity of the vascular structure is ensured.

2. RELATED WORK

The temporal subtraction algorithm is the typical one currently employed in clinical practice for subtraction. It subtracts the live images from the mask images in order to obtain unobscured vascular structures and eliminate bone tissues. Nevertheless, the algorithm might generate a considerable amount of artifacts. In this part, several methods for generating DSA images based on deep learning will be introduced. Subsequently, the disentangled representation learning will be presented.

2.1 Motion alignment method

The differences between the live images and the corresponding mask images were initially detected by the motion alignment algorithms. Afterwards, the mask images were deformed in accordance with the relevant live images. Eventually, in accordance with the principle of DSA imaging,

the live images and mask images were subjected. The results obtained through these methods depended on the results of the mask image generation. In consequence, the outcome was not satisfactory.

2.2 Virtual DSA method

The virtual DSA algorithms were a series of methods that generate DSA images from the single live images. Initially, a straightforward GAN method [15] was proposed to mitigate artifacts. Nevertheless, issues emerged like the problem of the disappearance of the small vascular structures. Subsequently, a series of methods were put forward to generate DSA images with GANs [16, 17]. The dataset is divided depending on the severity of artifacts before training by these approaches to reduce artifacts. But these algorithms could bring about unanticipated problems, for instance, the loss of vascular structures and blurry vascular structures.

2.3 Disentangled representation learning

Disentangled representation learning imitates the processes of human's cognition and is at present mostly employed in some simple scene generation tasks [18, 19]. High-dimensional abstract representations within the latent space lead to the generation of target features. DR-GAN [20] achieved pose-invariant face recognition task based on the disentangled representation learning. With the assistance of this model, one could frontalize or rotate a face with any pose. CausalVAE [21] combined disentangled representation learning with variational autoencoder (VAE) to obtain the latent representation. Disentangled representation was employed by DR-MTCDR [22] to learn the user-level domain-shared and domain-specific information on several financial datasets for multi-target cross-domain recommendation.

Currently, disentangled representation learning still has no superior definitions or metrics. Moreover, its performance is unsatisfactory in complicated situations. The initial definition of disentangled representations [23] was proposed under the assumption that a feasible group decoupling has been determined. Nevertheless, in numerous circumstances, several intricate correlations among the groups are required to be decoupled.

3. METHOD

Taking advantage of the theory underlying DSA imaging, DSA images can be generated through a disentangled approach. Firstly, we will present the overall framework and detailed modules of DDN. Secondly, the task-specific loss function will be put forward.

3.1 Deep decoupling framework

Specifically, it should be noted that the live image x and the mask image m are respectively generated as a result of X-rays penetrating the body before and after the contrast agent is added. The DSA image then can be defined as:

$$I_v = I_x - I_m \quad (1)$$

where, v represents the DSA image, I represents the intensity of \cdot .

Due to the prolonged scan time and the slight movement, the artifact (donated by ε) is frequently present. This implies that v is composed of the predicted vascular structure \hat{v} and ε , and m is regarded as made up of \hat{m} and ε , and \hat{m} is the impartial appraisal of m in x . Accordingly, the intensity of vascular structure can be defined as:

$$I_{\hat{v}} = I_x - I_{\hat{m}} = I_x - (I_m - \varepsilon) \quad (2)$$

Upon conducting a comparison between (1) and (2), it becomes evident that ε is the additional element of v and the deficient portion for m . This further constitutes a kind of pivotal related attributes between v and m . Consequently, minimizing ε is directed towards disentangling the live images. With the intent of maximizing the disparity in outputs, the most appropriate features are extracted and filtered by these networks. Being both adversarial and mutually reinforcing, this constitutes a process for the networks. The overall framework, comprising two networks and adversarial loss functions with regression loss, which is shown in Figure 2.

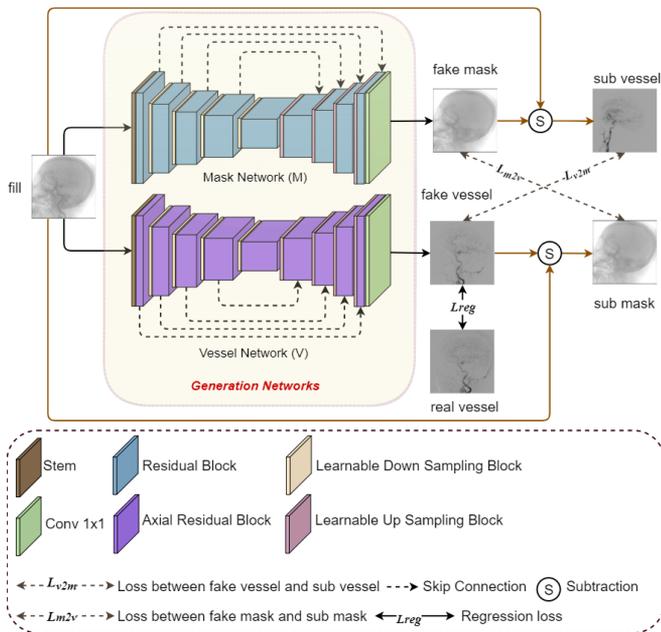


Figure 2. The framework of DDN

3.2 Network details

Both the two networks are constructed on an architecture similar to U-net. The stem block is a convolution with a 7-sized kernel for both networks. V initially utilizes an axial residual block (ARB) for feature extraction. Since the special vascular structure, the conventional kernel encounters difficulties in extracting effective features. Although the dynamic snake convolution [24] is sensitive to the tubular structures, this algorithm can be misled by the noise. We find that, in contrast to ordinary convolutions, ARB is more sensitive to tubular features and have lower computational burdens. Hence, ARB is defined as:

$$F_{out} = (W_x C_x(F_{in}) \parallel W_y C_y(F_{in})) + W_s C_s(F_{in}) \quad (3)$$

Here, C represents convolution, \parallel indicates the concatenation operation, and W donates learned convolution weights. The features then are concatenated along the channel

dimension. The longer size of the kernels in the axial convolutions is 5 to acquire larger receptive field. Eventually, the pixel wise convolution is utilized to adjust the number of channels. This structure constitutes a residual block to prevent the loss of long-term features.

Thereafter, down-sampling blocks and up-sampling blocks in a learnable manner are implemented to enhance the presentation of the output features. With respect to the down-sampling blocks, a dilated convolution is incorporated that can gain larger receptive field. Subsequently, the depth-wise convolution is used. Furthermore, an average pooling is used to rescale the size of the features and a pointwise convolution is followed. For the up sampling block, a bilinear up sampling block is used, and other structures are the same as the down-sampling block. Furthermore, the output features of this structure are richer in semantics. In the end, a pixel-wise convolution with a Tanh activation function is integrated to project the output results.

M is designed to acquire the virtual mask images. Therefore, we used Res-UNet with learnable down-sampling and up-sampling blocks. GELU and BatchNorm are used respectively following by the convolutions. In addition, the hyperparameters are the same in both V and M to keep the same level of the semantic information.

3.3 Loss function

V aims to extract the features of vascular structure, but for M, the aim is to obtain features of other structures. That means the distance of \hat{v} and \hat{m} needs to be maximizing:

$$\mathbb{E}_{v2m} = \text{minimize} (\sqrt{\hat{v} - (x - \hat{m})^2}) \quad (4)$$

In other words, the objective function is the same as minimizing the Euler distance between \hat{v} and $x - \hat{m}$. As far as we know, minimizing the Euler distance is the function of the mean square error (MSE) loss. Therefore, the objective function is defined as:

$$\mathcal{L}_{v2m} = \frac{1}{n} \sum_{i=1}^n (\hat{v} - (x - \hat{m}))^2 \quad (5)$$

Similarly, the objective function of M is defined as:

$$\mathcal{L}_{m2v} = \frac{1}{n} \sum_{i=1}^n (\hat{m} - (x - \hat{v}))^2 \quad (6)$$

Additionally, the regression loss is used to supervise \hat{v} keep away from unanticipated features. Based on (2), the expectation and standard deviation of v is the same as that of \hat{v} . L_l is used as the loss of the expectation, and MSE is used to approximate the standard deviation. Thus, the regression loss is defined as:

$$\mathcal{L}_{v2\hat{v}} = L_1(\hat{v}, v) + MSE(\hat{v}, v) \quad (7)$$

We use the smooth L_l function as the loss function, which can avoid zero-point distortions. The smooth L_l function is defined as:

$$L_1(\hat{v}, v) = \begin{cases} 0.5(\hat{v} - v)^2, & \text{if } |\hat{v} - v| < 1 \\ |\hat{v} - v| - 0.5, & \text{otherwise} \end{cases} \quad (8)$$

At the beginning, \hat{v} and \hat{m} are far from the expected values. v is predominant, guiding \hat{v} to approach v . As the training process the factor of $\mathcal{L}_{v2\hat{v}}$ needs to be lower to avoid the unexpected objects. Therefore, the total loss function of V is defined as:

$$\mathcal{L}_v = (1 - \lambda)\mathcal{L}_{v2m} + \lambda\mathcal{L}_{v2\hat{v}} \quad (9)$$

λ stands for the weight factor of $\mathcal{L}_{v2\hat{v}}$, which decreases from 0.999 to 0.0001 along with the training process, and \mathcal{L}_{m2v} is the loss function of M.

3.4 Training strategy

Since the two networks are required to facilitate each other mutually, they are trained in a manner similar to GANs. V generates the DSA images and use the backpropagation method to update the optimization function. Subsequently, M generates the mask images and uses the same progress as V but keep no gradient broadcast. In this way, these two network can be trained like GANs.

4. EXPERIMENTS

4.1 Data collection

In this paper, the data is from the Department of Neurosurgery, the General Hospital of Northern Theater Command in Shenyang, China. We collected totally 342 sequences from 312 patients, consisting of 45486 frames in total, acquired by SIEMENS AXIOM Artis equipment and the patient information was removed. These images had size of 960×960 , 960×1024 pixels. We further divided these images into two dataset, artifact-free dataset (41496 images) and artifact dataset (3990 images). We divided the artifact-free dataset by a ratio of 7:1:2 for the training set, validation set and the test set. The artifact dataset was used as the additional test set.

4.2 Preprocessing

Because the images had varying pixel sizes, we first padded the images to a uniform size and then resized all images to 256×256 . To facilitate the convergence of the network, we used the image inversion operation firstly. The Contrast Limited Adaptive Histogram Equalization (CLAHE) was used for enhancing the contrast.

4.3 Experiment results

In our experiment, we merely trained our model and the models for comparison using the artifact-free dataset, since we discovered that data with artifact would lead to a significant degradation of the comparison models and prevent obtaining normal results.

4.4 Evaluation metrics

Although the artifacts in the real DSA images within the dataset have a minor visual impact, they still exert a considerable influence on numerical calculation. A solitary result indicator cannot fully depict the quality of the generated results. Firstly, it should be noted that the aforementioned

indicators assess the generated results in a multifaceted and comprehensive fashion, encompassing similarity, distortion, and distribution. Therefore, the structure similarity index measure (SSIM), the peak signal-to-noise ratio (PSNR), the visual saliency-induced index (VSI), Fréchet Inception Distance (FID), and the feature similarity index measure (FSIM) are used for evaluation metrics.

In this paper, SSIM pertains to the resemblance of the overall structure in the generated image with respect to the real DSA image. PSNR indicates the integrity of the small vessel area. VSI can evaluate the degree of the sharpness of the vascular. FID is utilized to describe the difference of the total distribution between the generated image and the gold standard. Moreover, the FSIM represents the similarity in detail.

4.5 Results on artifact-free dataset

Firstly, we compared DDN with several state-of-art methods on our artifact-free dataset.

The quantitative results are presented in Table 1. In this task, CycleGAN [24], DiscoGAN [25], and DualGAN [26] fail to produce outcomes effectively. This is attributed to the fact that the style domain is too similar for the live images and the mask images. Notably, although the metric results of DualGAN are not bad, the FID reaches 803.92, indicating that the distribution significantly distorted. The result of RDBGAN [15] shows that the method does not work. UNIT [27] and MUNIT [28] yield disparate results although these two methods are based on disentangled representation learning. Distortion is appeared in the result of UNIT, whereas the results of MUNIT are superior with the FID of 367.70, which means MUNIT has greater potential in this task. UNet [29] has a higher structure similarity with 2.2% than that of Pix2Pix and MUNIT. However, the PSNR, VSI, and FID are respectively 0.39 dB, 0.021, and 237.75 inferior to these methods. The results of Pix2Pix show that it works better than these methods, yet VSI of 0.001 and FSIM of 0.001 lower than MUNIT. For DDN, the results outperform other methods. Compared to the other methods, all metrics are superior.

The qualitative results of DDN were thoroughly demonstrated through a comprehensive comparison with MUNIT, Pix2Pix, and UNet, as illustrated in Figure 3. Evidently, it is manifest that DNN achieves the best results. Notably, this conclusion is in perfect alignment with that derived from the quantitative analysis.

Table 1. DSA image generation performance for different networks on artifact-free dataset

	SSIM	PSNR	VSI	FID	FSIM
CycleGAN	0.401	6.46	0.764	2261.19	0.487
DiscoGAN	0.759	13.90	0.865	1522.72	0.789
DualGAN	0.805	21.70	0.920	803.92	0.813
RDBGAN	0.732	18.40	0.899	865.91	0.795
UNIT	0.752	12.98	0.837	841.13	79.06
MUNIT	0.877	24.01	0.969	367.70	0.873
Pix2Pix	0.882	24.05	0.968	360.96	0.872
UNet	0.904	23.66	0.947	598.71	0.874
DDN	0.936	24.18	0.980	351.59	0.900

4.6 Results on artifact dataset

According to the results of the artifact-free dataset, we compared our method with three models MUNIT, Pix2Pix and

UNet. The quantitative results are listed in Table 2. We did not use FID because the definition is conflict with the existence of artifacts. From the results, our model were robust and kept stable results. The performance was better than other methods. The qualitative results are shown in Figure 4. Compared with other methods, our model kept the clear and intact vascular structure. Compared with the label images, our model removed the artifacts successfully, while other methods remained a small number of artifacts.

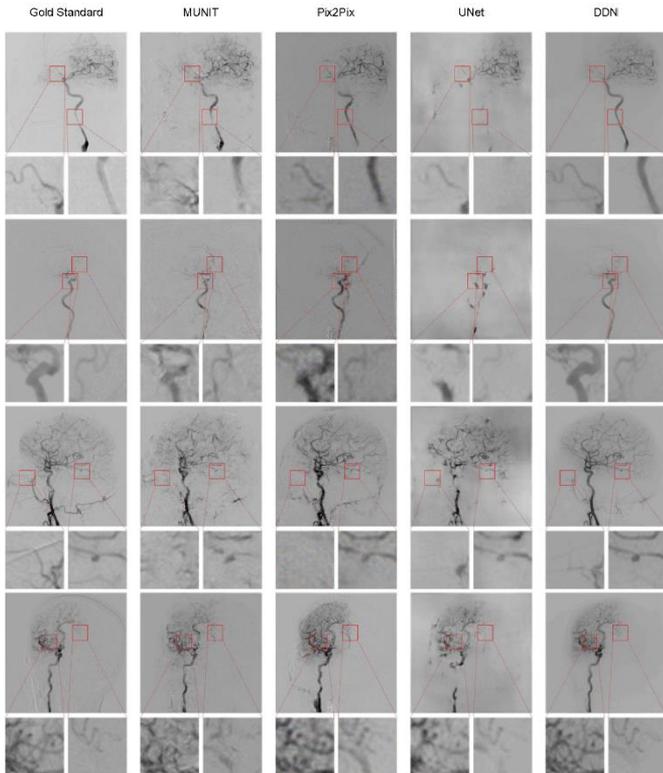


Figure 3. Generation results on the artifact-free dataset (the images from left to right are gold standard, MUNIT, Pix2Pix, UNet and DDN, respectively.)

Table 2. Generation results on artifact dataset

	SSIM	PSNR	VSI	FSIM
MUNIT	0.869	23.93	0.970	0.874
Pix2Pix	0.871	23.90	0.969	0.872
UNet	0.900	23.42	0.971	0.875
DDN	0.921	23.93	0.979	0.892

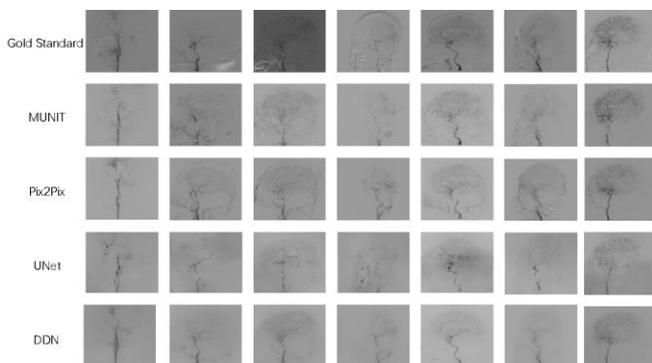


Figure 4. Generation comparisons of MUNIT, Pix2Pix, UNet and DDN on artifact dataset

4.7 Ablation studies

A comprehensive ablation study has been presented to validate the efficacy of the modules integrated within our approach. Specifically, we scrupulously evaluate the importance of ARB (denoted as A), the sampling blocks (denoted as B), and the decoupling strategy (denoted as C). The corresponding results has been illustrated in Table 3.

The results is evident that each of A, B, and C can independently improve the results. A and B have a more pronounced impact on enhancing structural completeness with SSIM is improved by 0.001 and 0.022. C significantly enhances the authenticity in details; VSI is improved by 0.021. The combination of A and B effectively enhances not only the overall fidelity but also the structure similarity in details, PSNR is improved by 1.4 dB and FSIM is improved by 0.03. The results of Baseline + A + C is similar to Baseline + B + C. DDN demonstrates a obviously enhancement through the combined utilization of A, B, and C. Specifically, SSIM increases by 0.032, VSI increases by 0.033, and FID decreases by 247.12.

Table 3. The ablation experiment results

	SSIM	PSNR	VSI	FID	FSIM
Baseline	0.904	23.66	0.947	598.71	0.874
Baseline+A	0.905	23.74	0.954	477.85	0.884
Baseline+B	0.926	24.62	0.959	396.08	0.897
Baseline+C	0.909	23.65	0.968	598.18	0.868
Baseline+A+B	0.929	25.06	0.957	432.35	0.904
Baseline+A+C	0.926	24.40	0.956	592.35	0.893
Baseline+B+C	0.927	24.77	0.958	482.30	0.899
DDN	0.936	24.18	0.980	351.59	0.900

5. CONCLUSION

As mention above, we have delved into a novel approach for DSA generation, named as DDN. This innovative methodology demonstrates the proficiency to mitigate artifacts effectively.

Firstly, ARB can extract the vascular structure with notable efficacy. Secondly, sampling blocks are presented, enabling the sampling of features in a learnable manner. This compensates for the drawback where existing sampling methods often tend to overlook significant features. Thirdly, loss functions are used to constraint the results of V and M with the decoupling strategy. Finally, the learning strategy can reduce the artifacts resulting from registration. Experimental outcomes clearly indicate that DDN generates DSA images successful and outperforms other methods.

For clinical applications, this approach holds substantial significance as it can diminish artifacts in DSA images, enhance the contrast of the vascular structure. What’s more, it can assist doctors in diagnosing more quickly, and it can also accelerate the speed of interventional therapy and reduce the risks. Hence, DDN has the potential to become a clinically available method for DSA generation.

ACKNOWLEDGMENT

This work was supported by the Science and technology Planning Project of Liaoning Province, China (Grant No. 2022JH2/101500037).

REFERENCES

- [1] Portegies, M., Koudstaal, P., Ikram, M. (2016). Cerebrovascular disease. *Hand-Book of Clinical Neurology*, 138: 239-261.
- [2] Aho, K., Harmsen, P., Hatano, S., Marquardsen, J., Smirnov, V.E., Strasser, T. (1980). Cerebrovascular disease in the community: Results of a WHO collaborative study. *Bulletin of the World Health Organization*, 58(1): 113.
- [3] Shah, N.S., Xi, K., Kapphahn, K.I., Srinivasan, M., Au, T., Sathye, V., Vishal, V., Zhang, H., Palaniappan, L.P. (2022). Cardiovascular and cerebrovascular disease mortality in Asian American subgroups. *Circulation: Cardiovascular Quality and Outcomes*, 15(5): e008651. <https://doi.org/10.1161/CIRCOUTCOMES.121.008651>
- [4] Jiang, H.J., Huang, X.L., Xian, B., Wang, Y., Zhou, Y.F., Ren, C.X., Pei, J. (2023). Chinese herbal injection for cardio-cerebrovascular disease: Overview and challenges. *Frontiers in Pharmacology*, 14: 1038906. <https://doi.org/10.3389/fphar.2023.1038906>
- [5] Brody, W.R. (1982). Digital subtraction angiography. *IEEE Transactions on Nuclear Science*, 29(3): 1176-1180. <https://doi.org/10.1109/TNS.1982.4336336>
- [6] Jeans, W.D. (1990). The development and use of digital subtraction angiography. *The British Journal of Radiology*, 63(747): 161-168. <https://doi.org/10.1259/0007-1285-63-747-161>
- [7] Bashir, Q., Ishfaq, A., Baig, A.A. (2018). Safety of diagnostic cerebral and spinal digital subtraction angiography in a developing country: A single-center experience. *Interventional Neurology*, 7(1-2): 99-109. <https://doi.org/10.1159/000481785>
- [8] Hou, Y., Ren, L., Cao, C., Zhang, H., Zhao, W., Zhu, J., Xia, S. (2023). The additional value of high-resolution vessel wall imaging in screening suitable chronic internal carotid artery occlusion candidates for endovascular recanalization: Comparison with digital subtraction angiography. *Acta Radiologica*, 64(4): 1702-1711. <https://doi.org/10.1177/02841851221127563>
- [9] Al-rudaini, H.E., Han, P., Liang, H. (2019). Comparison between computed tomography angiography and digital subtraction angiography in critical lower limb ischemia. *Current Medical Imaging*, 15(5): 496-503. <https://doi.org/10.2174/1573405614666181026112532>
- [10] Park, J.E., Jung, S.C., Lee, S.H., Jeon, J.Y., Lee, J.Y., Kim, H.S., Choi, C.G., Kim, S.J., Lee, D.H., Kim, S.O., Kwon, S.U., Kang, D.W., Kim, J.S. (2017). Comparison of 3D magnetic resonance imaging and digital subtraction angiography for intracranial artery stenosis. *European Radiology*, 27: 4737-4746. <https://doi.org/10.1007/s00330-017-4860-6>
- [11] Soori, M., Arezoo, B., Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*, 3: 54-70. <https://doi.org/10.1016/j.cogr.2023.04.001>
- [12] Sarker, I.H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6): 420. <https://doi.org/10.1007/s42979-021-00815-1>
- [13] Whitehead, J.F., Hoffman, C.A., Periyasamy, S., Laeseke, P.F., Speidel, M.A., Wagner, M.G. (2022). A motion compensated approach to quantitative digital subtraction angiography. *Medical Imaging 2022: Physics of Medical Imaging*, 12031: 396-404. <https://doi.org/10.1117/12.2611816>
- [14] Su, R., van der Sluijs, M., Cornelissen, S., van Zwam, W., van der Lugt, A., Niessen, W., Ruijters, D., van Walsum, T., Dalca, A. (2023). *AngioMoCo: Learning-based motion correction in cerebral digital subtraction angiography*. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham: Springer Nature Switzerland, pp. 770-780. https://doi.org/10.1007/978-3-031-43990-2_72
- [15] Gao, Y., Song, Y., Yin, X., Wu, W., Zhang, L., Chen, Y., Shi, W. (2019). Deep learning-based digital subtraction angiography image generation. *International Journal of Computer Assisted Radiology and Surgery*, 14: 1775-1784. <https://doi.org/10.1007/s11548-019-02040-x>
- [16] Ueda, D., Katayama, Y., Yamamoto, A., Ichinose, T., Arima, H., Watanabe, Y., Miki, Y. (2021). Deep learning-based angiogram generation model for cerebral angiography without misregistration artifacts. *Radiology*, 299(3): 675-681. <https://doi.org/10.1148/radiol.2021203692>
- [17] Yonezawa, H., Ueda, D., Yamamoto, A., Kageyama, K., Walston, S.L., Nota, T., Miki, Y. (2022). Maskless 2-dimensional digital subtraction angiography generation model for abdominal vasculature using deep learning. *Journal of Vascular and Interventional Radiology*, 33(7): 845-851. <https://doi.org/10.1016/j.jvir.2022.03.010>
- [18] Bengio, Y., Courville, A., Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8): 1798-1828. <https://doi.org/10.1109/TPAMI.2013.50>
- [19] He, H., Chen, D., Balakrishnan, A., Liang, P. (2018). Decoupling strategy and generation in negotiation dialogues. *arXiv preprint arXiv:1808.09637*. <https://doi.org/10.48550/arXiv.1808.09637>
- [20] Tran, L., Yin, X., Liu, X. (2017). Disentangled representation learning GAN for pose-invariant face recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 1283-1292. <https://doi.org/10.1109/CVPR.2017.141>
- [21] Yang, M., Liu, F., Chen, Z., Shen, X., Hao, J., Wang, J. (2021). Causalvae: Disentangled representation learning via neural structural causal models. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, pp. 9588-9597. <https://doi.org/10.1109/CVPR46437.2021.00947>
- [22] Guo, X., Li, S., Guo, N., Cao, J., Liu, X., Ma, Q., Zhao, Y. (2023). Disentangled representations learning for multi-target cross-domain recommendation. *ACM Transactions on Information Systems*, 41(4): 1-27. <https://doi.org/10.1145/3572835>
- [23] Higgins, I., Amos, D., Pfau, D., Racaniere, S., Matthey, L., Rezende, D., Lerchner, A. (2018). Towards a definition of disentangled representations. *arXiv preprint arXiv:1812.02230*. <https://doi.org/10.48550/arXiv.1812.02230>
- [24] Zhu, J.Y., Park, T., Isola, P., Efros, A.A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 2242-2251. <https://doi.org/10.1109/ICCV.2017.244>

- [25] Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. In International Conference on Machine Learning, pp. 1857-1865.
- [26] Yi, Z., Zhang, H., Tan, P., Gong, M. (2017). Dualgan: Unsupervised dual learning for image-to-image translation. In P 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 2868-2876. <https://doi.org/10.1109/ICCV.2017.310>
- [27] Liu, M.Y., Breuel, T., Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in Neural Information Processing Systems*, 30.
- [28] Huang, X., Liu, M.Y., Belongie, S., Kautz, J. (2018). Multimodal unsupervised image-to-image translation. In Proceedings of the European Conference on Computer Vision (ECCV), pp. 172-189.
- [29] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science, Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28