Ruibo Liu¹, Zhenzhou Li¹, Xiaoyan Shen¹, Ronghui Tian¹, Ligang Chen¹, Xinyu Yang¹, Wei Qian¹, Guobiao Liang¹, Guangxin Chu¹, Hai Jin¹, and He Ma¹

¹Affiliation not available

March 07, 2024
GDN: Generative Decoupling Network for Digital Subtraction Angiography Generation

Ruibo Liu, Zhenzhou Li, Xiaoyan Shen, Ronghui Tian, Ligang Chen, Xinyu Yang, Wei Qian, Guobiao Liang, Guangxin Chu, Hai Jin and He Ma

Abstract—Objective: Digital subtraction angiography (DSA) is significantly important for cerebrovascular disease diagnosis and treatment. However, artifacts and noise are inevitable and reduce image quality. These problems could make clinical diagnosis difficult. In this paper, we introduce a novel deep learning architecture, exploiting the information decoupling training strategy to generate high-quality DSA images. Methods: We propose the generative decoupling network, a feature decoupling convolutional network, which maximizes the difference between different structures throughout a decoupling training strategy. In this network, an axial residual block and a learnable sampling method are proposed to enhance the strength of feature extraction. Results: The results showed that our proposed method significantly outperforms the existing methods in the DSA generation task. Furthermore, we quantified the method using the metrics of SSIM, PSNR, VSI, FID and FSIM, with the results of 93.57%, 24.18dB, 98.04%, 351.59, and 89.95%, respectively. Conclusion: Our method can produce high-quality DSA images with little or even no artifact and noise. Significance: The proposed method can effectively reduce artifacts and noise, and generate high quality DSA images with complete and clear vascular structures.

Index Terms—Cerebrovascular diseases, decoupling training strategy, digital subtraction angiography, generative decoupling network, interactive loss function.

I. INTRODUCTION

The high resolution of Digital Subtraction Angiography (DSA) makes it unbeatable in the field of cerebrovascular imaging [1], even though CT and MRI images are commonly used for clinical diagnosis [2], [3]. DSA is therefore treated as the gold standard for cerebrovascular diseases. Nevertheless, the problems in cerebral vessels can be visually observed by DSA. Furthermore, it offers imaging information for relevant interventional surgery that is superior to CTA and MRA [4]–[6]. In clinical, DSA images are acquired by subtracting live images (images contain the contrast agent) and mask images (images without the contrast agent) by using the C-arm detector twice to eliminate the effect of bone and soft tissue structures, as shown in Fig. 1. Due to the extended time required for acquisition between a mask image and the live image, artifacts are inevitable. Additionally, the patient’s light movement can result in visible artifacts in the DSA images. Additionally, despite being shot on the same equipment for both mask images and live images, noise is unavoidable as separate shots. Due to the presence of artifacts and noise, the image quality is significantly reduced, as well as blurring out the small vascular structure that is not easily noticeable. This makes it more challenging for physicians to determine the cause of the disease. As a result, the reduction or even elimination of artifacts and noise from DSA images has consistently been considered an essential but challenging task.

Although artificial intelligence has been shown to perform well on clinical tasks such as lesion detection, segmentation, and classification, there are relatively few research focus on DSA image generation [7], [8]. There currently exist two main research directions in the DSA image generation task. The first step is to register the mask images and live images with the motion alignment [9], [10]. The second-fold is named as virtual DSA, with the aim of generating DSA images from the live images without mask images by using Generative Adversarial Nets(GANs). GANs [11]–[13] use the generative network to acquire subtraction images from live images. A discriminator network is used to classify the real images and generated images. Despite the fact that these algorithms have made some achievements for DSA image generation, two main problems still need to be solved. First, the mask images are highly correlated with the artifacts. This problem requires us to minimize the use of mask images. Second, some small vascular structures may blur or disappear without the help of mask images. Consequently, the real mask image is a disturbing factor for the generated image but can prevent the disappearance of the vascular structure at the same time. A successful resolution is necessary to address this ambiguous issue.

In this paper, we propose the generative decoupling network (GDN) consisting of the vessel network and the mask network. The vessel network extracts the vascular structure...
while the mask network extracts the other structures to help the vessel network acquire the correct features. Furthermore, a symmetric loss function is constructed to constrain these two models to maximize the difference from extracted features. By decoupling the vascular structure and other structures, they are distinguished from live images in a contrasting and reinforcing manner. Alternatively, these two generators can create training structures that are antagonistic and mutually reinforce, leading to more complete vascular structures and fewer artifacts. The main contributions of this work are summarized as follows:

- GDN generates DSA images from the live images without collecting real mask images. The patients could be scanned by the C-arm detector only once to reduce the radiation dose suffered by the patients. In this way, we are able to generate DSA images within the constraints of the generated mask images to achieve mutual reinforcement.
- This particular tubular vascular structure is extracted more efficiently with the axial residual block, which is more sensitive than other convolutional blocks for this type of arrangement.
- An efficient learnable sampling method is proposed to avoid the deficiency of important features. The unnecessary artifacts can be removed while ensuring the integrity of fine vascular structures.

II. RELATED WORK

The typical subtraction algorithm currently used in clinical applications is the temporal subtraction algorithm. This algorithm subtracts live images and mask images to acquire clear vascular structures and remove extraneous tissue structures. However, this algorithm produces a large number of noise and artifacts. In this section, we will first introduce several DSA image generation methods based on deep learning. Decoupling representation learning then will be introduced as a preliminary.

A. Motion Alignment

The motion alignment algorithms [9], [10] detected differences between live and mask images firstly. Subsequently, the mask images were distorted on the basis of the corresponding live images. Finally, live images and mask images were subtracted based on the principle of DSA imaging. The quality of the results produced by these algorithms depended on the quality of the mask image generation. As a result, artifacts could be reduced but not satisfactory. Moreover, the noise was not resolved.

B. Virtual DSA

The virtual DSA algorithms were used to generate DSA images from live images without the corresponding mask images. At first, a simple GAN method [11] was proposed to reduce noise and artifacts. However, the problems with the disappearance of vascular structures occurred. After that, a series of methods [12], [13] based on GANs were proposed to generate DSA images. These methods can reduce artifacts by dividing the dataset according to the presence or absence of artifacts before training. However, these algorithms may cause unexpected problems such as the disappearance of vascular structures and blurred vascular structures.

C. Disentangled Representation Learning

Disentangled representation learning mimicked human cognitive processes and was currently mainly applied to some simple scene generation tasks [14], [15]. Target features were generated by extracting high-dimensional abstract representations in the latent space. Existing disentangled learning methods such as β-VAE [16], factor-VAE [17] and InfoGAN [18] are combined with other existing deep learning methods. Since disentangled Representation Learning is still in its infancy, there are no better definitions and metrics. Additionally, it performs poorly in complex scenarios. The initial definition [19] of disentangled representations was proposed based on the condition that a feasible group decomposition has been found. However, in many cases, there will be complex correlations between groups that need to be decomposed.

III. METHODOLOGY

Benefiting from the principle of DSA imaging, we can generate DSA images in a decoupling way. First, we will introduce the principle of DSA imaging. The GDN will be proposed secondly. In the end, a distinct decoupling training framework will be established by the special loss function and its specific training strategy.

A. Principle of DSA Imaging

Specifically, the live image \( x \) is generated by the intensity of X-rays passing through the body, and the X-ray intensity signal follows an exponentially decreasing law:

\[
I_{T_1} = I_0 e^{-(\mu_B d_B + \mu_T d_T)},
\]

where \( I_0 \) and \( I_{T_1} \) represents the intensity of emitted X-rays and the intensity of incident X-rays. Furthermore, \( \mu_B \) and \( \mu_T \) represents linear absorption coefficients for bone and soft tissues. Besides, \( d_B \) and \( d_T \) represents the thicknesses of bone and soft tissues. Accordingly, after adding the contrast agent to the vascular:

\[
I_{T_2} = I_0 e^{-(\mu_B d_B + \mu_T (d_T - d_1) + \mu_I d_1)},
\]

where \( \mu_I \) and \( d_1 \) represents the linear absorption coefficient and thickness of the contrast agent, respectively. From this, the intensity of DSA image \( v \) can be defined as:

\[
I_v = \ln I_{T_2} - \ln I_{T_1} = (\mu_T - \mu_I) d_1,
\]

where \( I_{T_2} \) is the intensity of the live image \( x \), and \( I_{T_1} \) is the intensity of the mask image \( m \). Hereby, the DSA image is the subtraction of the intensity of the soft tissues and the intensity of the contrast so as to subtract the mask image from the live image:

\[
I_v = I_x - I_m.
\]
B. Generative Decoupling Framework

Due to the long acquisition time and movement of patients, noise and artifacts $\varepsilon$ are common in the DSA images. That means that $v$ is made up of the vascular structure $\hat{v}$ and $\varepsilon$, and $m$ can be identified as consisting of $\hat{m}$ and $\varepsilon$, where $\hat{m}$ is the unbiased estimate of $m$ in $x$. Thus, the intensity of the vascular structure can be defined as:

$$I_v = I_x - I_{\hat{m}} = I_x - (I_m - \varepsilon)$$  \hspace{1cm} (5)

Compared (4) with (5), $\varepsilon$ can not only be treated as the extra part of $v$, but also the missing part of $m$. It is also one of the key relevant features for $v$ and $m$. Therefore, the process of reducing $\varepsilon$ is decoupling the live images. These two networks select the features in $x$ and filter them to obtain the most suitable features for themselves to maximize the difference in outputs. It is a process that is both adversarial and mutually reinforcing for the networks. The total framework is shown in Fig. 2 that consists of two models, a pair of adversarial loss functions, and a regression loss function.

C. Network Details

The vessel net and mask net are based on U-net like architecture. At the start of both networks, a convolutional kernel size of 7 and stride of 1 is used as the stem block. The vessel net first uses an axial residual block (ARB) to extract features. Because the vascular structure is a tubular structure, the normal convolution kernel is difficult to extract effective features. Dynamic snake convolution [20] can effectively extract tubular structures. By calculating the offset of each axis, the method alters the direction of convolution feature extraction and extracts features in a specific direction. However, it is computationally complex and sensitive to noise. We find that axial convolutions are more sensitive to tubular features and are less computationally expensive than ordinary convolutions. Hence ARB is proposed. The ARB is defined as:

$$F_{out} = (W_xC_x(F_{in})||W_yC_y(F_{in}) + W(s)C_s(F_{in}),$$  \hspace{1cm} (6)

where $C$ donates convolution, $\|$ represents concatenation of channel dimensions and $W$ is the weight of the convolution. We use two axial convolutions to extract features in different directions and concatenate them in the channel dimension. Furthermore, the kernel size of the axial convolution is set to (5,1) and (1,5) to get a larger receptive field. Finally, a convolution with a kernel size of 1 is used to vary the number of channels of the input feature map and form a residual-shaped block to avoid long-term feature loss.

After that, a learnable down-sampling block and a symmetrical learnable up-sampling block are proposed to make the sampling process learnable. For the learnable down sampling block, a dilated convolution with kernel size of 3 and dilation rate of 2 is added to enlarge the receptive field. After that, a depthwise convolution is followed to make a difference in the channel dimension. Besides, an average pooling and a pointwise convolution are used to down sampling and reassemble the features. The learnable up-sampling block just changes the average pooling in the learnable down-sampling block to a bilinear up-sampling to make the features symmetrical. Common down-sampling methods such as pooling and convolution with step size of 2 are concerned only with the region where the kernel is located. The addition of dilated convolution can make the down-sampling operation take into account the location of the convolution kernel and the relationship with the neighborhood features. In addition, it can improve the representation of the sampling features. Moreover, the combination of depthwise convolution, sampling, and pointwise convolution can obtain a richer representation of features and avoid the loss of important features.

Afterwards, skip connections are used to transmit the features from the encoder blocks to the corresponding blocks in the decoder. Finally, a Tanh activation function and a convolution with kernel size of 1 are combined to obtain the output features.

The mask net aims to extract the other structures except the vascular structure. U-net is then used and replaces downsampling and upsampling with the sampling method proposed so as to keep the extracted feature space consistent with the vessel net. The residual block is used as the backbone. The residual block contains two convolutions with a kernel size of 3 and a convolution with a kernel size of 1 is used as the shortcut.

In ARB and residual block, each convolution is followed by GELU and Batch Norm as the activation function and the normalization function, respectively. Moreover, in both the vessel net and the mask net, the number of encoder and decoder layers is 4. The bottleneck layers of vessel net and mask net are set as ARB and residual block respectively. Besides that, the number of the channels of the same layer of the vessel net and the mask net is the same. The number of the channel in the stem block is 32. After that, each layer of the encoder has twice as many channels as the previous layer. In particular, the number of channels in the bottleneck layer is set to 512. The number of channels in each layer in the decoder is then aligned with the corresponding layer in the encoder.

D. Loss Function

For the vessel net, the vascular structure is expected to be extracted, while for the mask net, other structures are expected. The vessel net needs to maximize the distance of $\hat{v}$ and $\hat{m}$:

$$\mathbb{E}_{v2m} = \text{Distance}_{\max}(\hat{v}, \hat{m})$$
$$= \text{Distance}_{\min}(\hat{v}, x - \hat{m})$$
$$\text{minimize} \sqrt{\hat{v}^2 - (x - \hat{m})^2}, \hspace{1cm} (7)$$

that is, the objective function is equal to minimize the Euler distance between $\hat{v}$ and $x - \hat{m}$. To best our knowledge, mean square error (MSE) loss is a function to minimize the Euler distance, thus the objective function can be defined as:

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

\hspace{1cm} (8)
For the mask net, analogously, the objective function is defined as:

\[ L_{m2v} = \frac{1}{n} \sum_{i=1}^{n} (\hat{m} - (x - \hat{v}))^2. \]  

(9)

In addition, a regression loss function is added to monitor \( \hat{y} \) for unexpected features. Because \( v = \hat{v} + \varepsilon \), \( v \) and \( \hat{v} \) have the same expectation and standard deviation. The expectation loss can be expressed in terms of \( L_1 \), while the standard deviation can be approximated by \( \text{MSE} \). Hence, the regression loss is defined as:

\[ L_{v2\hat{v}} = L_1(\hat{v}, v) + \text{MSE}(\hat{v}, v), \]  

(10)

where \( L_1 \) is the smooth L1 loss, which can prevent from zero-point irregularities. The definition of the smooth L1 loss is:

\[ \text{smooth}_{L_1}(\hat{v}, v) = \begin{cases} 
0.5(\hat{v} - v)^2, & \text{if } |\hat{v} - v| < 1 \\
|\hat{v} - v| - 0.5, & \text{otherwise}. 
\end{cases} \]  

(11)

At the start of the training process, \( \hat{y} \) and \( \hat{m} \) are far from the expectation, the supervisory role of \( y \) is dominant to lead \( \hat{y} \) to close to \( y \). As the training process goes on, \( \hat{y} \) is closed to the vascular structure gradually. If the factor of \( y \) is still high, the noise in \( y \) can have a unexpected influence to \( \hat{y} \). Hence, the weight of \( L_{v2\hat{v}} \) needs to decrease gradually to reduce the impact on \( y \). Therefore, the final loss function for vessel net is defined as:

\[ L_v = (1 - \lambda)L_{v2\hat{v}} + \lambda L_{v2\hat{v}}, \]  

(12)

where \( \lambda \) represents the weight of \( L_{v2\hat{v}} \), which decreases with the training process from 0.999 to 0.0001. and the final loss function for mask net is \( L_{m2v} \).

E. Training Strategy

Since the vessel net and the mask net need to promote each other, they are trained like GANs. At the beginning of the training process, the vessel net and the mask net generate their own images. Afterwards, the loss function of the mask net is calculated and backward with no gradient propagation of the vessel net, and the optimizer of the mask net is updated. In addition, the loss function of the vessel net is calculated and backward with no gradient propagation of the mask net so as to correct the result of the vessel net with the help of optimized mask net. Finally, the vessel net optimizer is updated.

IV. EXPERIMENTS

A. Data Collection and Pre-Processing

We obtained access to the Department of Neurosurgery of the PLA Northern Theater Command General Hospital in Shenyang, China, and acquired clinical cerebrovascular DSA image data. Data acquisition was carried out using equipment from SIEMENS AXIOM Artis. Furthermore, all DSA images were gathered after the patient’s information was removed. A total of 22,125 sequences from 6,823 patients were collected and 17,140 sequences from 4,997 patients were discarded due to quality problems. Furthermore, the sequences were divided into two categories based on the severity of the subtraction image artifacts. 312 artifact-free sequences were selected as the dataset. We partition per patient to prevent overlapping images between the data sets. Moreover, a ratio of 7:1:2 was used to split the selected data as the training set, the validation set and the test set. Additionally, each sequence in the training set sample 10 images to decrease the repetition rate. Eventually, 2,180 paired images were sampled for training, 300 paired images for validation and 640 paired images for testing. Furthermore, all images were resized to \( 256 \times 256 \), and the image inversion operation was used to ensure that the intensity of the vascular structure was higher than that of the other structures so that it can be conducive to the convergence of the network. The contrast-limited adaptive histogram equalization (CLAHE) was then used to improve the contrast of the vascular structures.

B. Experimental Configuration

We adapted the Adam optimizer with an initial learning rate of 0.0001. The cosine annealing learning rate scheduler was used. In addition, the mini-batch size was 16. We trained the framework with an epoch number of 1000. We used a NVIDIA 4090 graphics card with 24GB memory on a GPU workstation to train and implement the model. The Pytorch (v2.1.0) and monai (v1.2.0) [21] were used to establish our framework, and the programming language is Python (v3.11.5). The training speed was about 1.62 s/iter, and it took about 600 epochs to reach convergence. The inference speed was about 29.84 images/s with a batch size of 1.

C. Evaluation Metrics

In order to make the evaluation method more objective and avoid some unexpected factors affecting the experimental results, we use a variety of evaluation indicators to evaluate from multiple perspectives. Structure similarity, degree of distortion, distribution similarity, and detail feature similarity are used as the indicators. Firstly, we use structure similarity index measure (SSIM) to measure the similarity of the massive structure. Second, peak signal-to-noise ratio (PSNR) and visual saliency-induced index (VSI) are used to evaluate the overall distortion degree and detail distortion degree, respectively. Thirdly, Fréchet Inception Distance (FID) is used to measure the authenticity of the image distribution. Finally, feature similarity index measure (FSIM) is used to evaluate the detailed similarity of images from the level of detail structure and gradient. In this paper, the similarity in structure is indicative of the overall structure of the vascular structures that is consistent with the real DSA image. The degree of distortion represents the integrity of the small vessels. Distribution similarity is used to describe the distribution distance between the generated image and the real DSA image. The similarity of the details of the feature represents the similarity of the details of the vascular structure, which can evaluate the degree of disappearance of the vascular structure. These indicators are defined by the following equations:

\[ \text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\mu_x^2 + \mu_y^2 + C_1(\sigma_x^2 + \sigma_y^2 + C_2)}, \]  

(13)
where $\mu$ and $\sigma$ are expectation and standard deviation respectively in a local window ($11 \times 11$ in this paper). $C_1$ and $C_2$ are constants.

$$PSNR(x, y) = 10 \log_{10}\left(\frac{2^{\text{bits}} - 1}{\sqrt{MSE(x, y)}}\right),$$  \hspace{1cm} (14)$$

$$VSI(x, y) = \frac{\sum_\Omega S(x, y) \cdot V S_m(x, y)}{\sum_\Omega V S_m(x, y)},$$  \hspace{1cm} (15)$$

where $\Omega$ represents the whole pixel space, $S(x, y)$ is the pixel-wise virtual similarity between x and y, and $V S_m(x, y) = \max(V S(x), V S(y))$ is the weight of $S(x, y)$.

$$FID(x, y) = |\mu_x - \mu_y|^2 + T_r(\Sigma_x + \Sigma_y - 2\sqrt{\Sigma_x \Sigma_y}),$$  \hspace{1cm} (16)$$

where $T_r$ is the trace of the covariance matrix, and $\Sigma$ is the covariance matrix.

$$FSIM(x, y) = \frac{\sum_\Omega S_L(x, y) \cdot PC_m(x, y)}{\sum_\Omega PC_m(x, y)},$$  \hspace{1cm} (17)$$

where $S_L(x, y)$ is the similarity of phase and gradient between $x$ and $y$, and $PC_m(x, y)$ is the maximum of the phase congruency between $x$ and $y$.

**D. Comparisons with Existing Methods**

To evaluate the performance of the generation methods, we compare our proposed GDN with several existing methods on our dataset.

The quantitative results are listed in Table I, and the statistical results excluding abnormal values are shown in Fig. 4. The style transfer methods CycleGAN, DualGAN, and DiscoGAN cannot achieve effective results in this task. This occurs because the domain difference between the live images and the corresponding subtraction images is smaller than in the case of the human visualization. In particular, although most results of DualGAN are satisfactory, the FID reaches a high value of 803.92. This means that the distribution of the results differs greatly from that of the gold standard. Similarly, this problem occurs with RDBGAN and UNIT. The result of RDBGAN significantly shows that this method is not valid on our dataset with serious distortion problems. The disentangled representation learning methods UNIT and MUNIT get different results. The UNIT result is distorted, while the MUNIT result is significantly better with the FID of 367.70. For UNet, the structure similarity is higher than Pix2Pix and MUNIT with 2.2%, but the PSNR, VSI and FID are worse than these methods with 0.39dB, 2.15% and 237.75, respectively. Therefore, the degree of the results of UNet is significantly higher than that of the other two methods. The results of Pix2Pix are the best except for GDN, while VSI and FSIM are lower than MUNIT with 0.07% and 0.13%, which means that the degree of structural detail and overall distortion is unsatisfactory. For GDN, the results outperform other methods in all the five metrics. GDN is superior to other methods in both structural integrity and distortion degree. Compared to the best values of other methods in all aspects, SSIM is increased by 3.22%, VSI is increased by 1.17%, FID is decreased by 9.37 and FSIM is increased by 2.54%.

Qualitative results were compared with GDN using MUNIT, Pix2Pix and UNet, which are superior to other methods, as demonstrated in Fig. 3. Obviously, the structural integrity and distortion degree of the results of GDN are significantly better than other methods, and this conclusion is consistent with the conclusion of quantitative analysis.

**E. Ablation studies**

A series of ablation studies are proposed to prove the effectiveness of the modules in our method. We evaluate the effect of ARB (signed with A), learnable sampling block (signed with B) and the decoupling method (signed with C). The results are shown in Table II.

From the results, A, B and C can all be improved when added alone. A and B have a greater effect on improving structural integrity (with the improvement of SSIM by 0.17% and 2.25%, respectively). C can improve the detail fidelity of the results with the improvement of VSI by 2.11%. The combination of A and B can improve the overall fidelity and structural similarity in details (with the improvement of PSNR by 1.4dB and FSIM by 2.97%). The combination of A and C and the combination of B and C have a balanced improvement. The proposed method has an obvious improvement with the help of the combination of A, B and C. The total structure similarity, the detail fidelity and the distribution similarity are improved with SSIM is increased by 3.22%, VSI is increased by 3.39% and FID is decreased by 247.12.

**V. Conclusion**

In this work, we have investigated a new DSA generation method named GDN. This method can effectively reduce artifacts and noise while generating high-quality DSA images. Firstly, an axial residual block is proposed to extract the vascular structures efficiently. Second, a pair of learnable sampling blocks are proposed to sample features in a learnable way to compensate for the problem that the existing sampling method is easy to miss important features. Moreover, the addition of the dilated convolution and depthwise convolution results in an expanded receptive field that makes computing more lightweight. Thirdly, a symmetric loss function is proposed to decouple the live images into the mask images and the subtraction images. With the help of this loss function, the training process has both antagonism and mutual promotion of two different features. Finally, the learning strategy can avoid the collapse of the model that comes with GANs, and reduce the artifacts and noises caused by registration. Experimental results show that GDN is an effective DSA image generation method that outperforms other methods of all metrics.

For clinical applications, this method can reduce artifacts and noise in DSA images, improve the recognition of vascular structure, and improve the speed and accuracy of clinical diagnosis. Moreover, DSA subtraction is a preprocessing step of DSA reconstruction, and the quality of the DSA images directly affects the accuracy of the reconstruction results. What’s more, the reconstructed images are the main angiography basis for clinical cerebrovascular surgery. It can be seen that the quality of the generated DSA images directly affects...
the accuracy of clinical diagnosis and surgical navigation. Therefore, it is very important for clinical practice to generate DSA images accurately and minimize artifacts and noise. Our method can obviously reduce artifacts and noise and acquire higher quality DSA images than other existing methods. For the proposed learning strategy, different from the limitations of the disentangled representation learning method, it can be applied to many related tasks. For instance, some medical image segmentation tasks that extract regions of interest with specific features in the images can be regarded as this kind of task. In particular, when the difference between the target and the surrounding background is relatively low, this method may be helpful.

### References


### TABLE I

**DSA IMAGE GENERATION PERFORMANCE FOR DIFFERENT NETWORKS.**

<table>
<thead>
<tr>
<th>NETWORK</th>
<th>SSIM</th>
<th>PSNR</th>
<th>VSI</th>
<th>FID</th>
<th>FSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CycleGAN</td>
<td>40.12</td>
<td>6.46</td>
<td>76.37</td>
<td>2261.19</td>
<td>48.72</td>
</tr>
<tr>
<td>Discogan</td>
<td>75.91</td>
<td>13.90</td>
<td>86.51</td>
<td>1522.72</td>
<td>78.91</td>
</tr>
<tr>
<td>DualGAN</td>
<td>80.51</td>
<td>21.70</td>
<td>91.96</td>
<td>803.92</td>
<td>81.28</td>
</tr>
<tr>
<td>RDBGAN</td>
<td>73.22</td>
<td>18.40</td>
<td>89.92</td>
<td>865.91</td>
<td>79.52</td>
</tr>
<tr>
<td>UNIT</td>
<td>75.21</td>
<td>12.98</td>
<td>83.66</td>
<td>841.13</td>
<td>79.06</td>
</tr>
<tr>
<td>MUNIT</td>
<td>87.71</td>
<td>24.01</td>
<td>96.87</td>
<td>367.70</td>
<td>87.29</td>
</tr>
<tr>
<td>Pix2Pix</td>
<td>88.15</td>
<td>24.05</td>
<td>96.80</td>
<td>360.96</td>
<td>87.16</td>
</tr>
<tr>
<td>Unet</td>
<td>90.35</td>
<td>23.66</td>
<td>94.65</td>
<td>598.71</td>
<td>87.41</td>
</tr>
</tbody>
</table>

**GDN**

<table>
<thead>
<tr>
<th>NETWORK</th>
<th>SSIM</th>
<th>PSNR</th>
<th>VSI</th>
<th>FID</th>
<th>FSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDN without A, B and C</td>
<td>90.35</td>
<td>23.66</td>
<td>94.65</td>
<td>598.71</td>
<td>87.41</td>
</tr>
<tr>
<td>GDN without B and C</td>
<td>90.52</td>
<td>23.74</td>
<td>95.36</td>
<td>477.85</td>
<td>88.37</td>
</tr>
<tr>
<td>GDN without A and C</td>
<td>92.60</td>
<td>24.62</td>
<td>95.85</td>
<td>396.08</td>
<td>89.70</td>
</tr>
<tr>
<td>GDN without A and B</td>
<td>90.94</td>
<td>23.65</td>
<td>96.76</td>
<td>598.18</td>
<td>86.79</td>
</tr>
<tr>
<td>GDN without C</td>
<td>92.86</td>
<td>25.06</td>
<td>95.72</td>
<td>432.35</td>
<td><strong>90.38</strong></td>
</tr>
<tr>
<td>GDN without B and C</td>
<td>92.63</td>
<td>24.40</td>
<td>95.59</td>
<td>592.35</td>
<td>89.30</td>
</tr>
<tr>
<td>GDN without A and C</td>
<td>92.63</td>
<td>24.40</td>
<td>95.59</td>
<td>592.35</td>
<td>89.30</td>
</tr>
<tr>
<td>GDN</td>
<td><strong>93.57</strong></td>
<td><strong>24.18</strong></td>
<td><strong>98.04</strong></td>
<td><strong>351.59</strong></td>
<td><strong>89.95</strong></td>
</tr>
</tbody>
</table>

### TABLE II

**The results of the ablation experiments for GDN.**

<table>
<thead>
<tr>
<th>NETWORK</th>
<th>SSIM</th>
<th>PSNR</th>
<th>VSI</th>
<th>FID</th>
<th>FSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDN without A, B and C</td>
<td>90.35</td>
<td>23.66</td>
<td>94.65</td>
<td>598.71</td>
<td>87.41</td>
</tr>
<tr>
<td>GDN without B and C</td>
<td>90.52</td>
<td>23.74</td>
<td>95.36</td>
<td>477.85</td>
<td>88.37</td>
</tr>
<tr>
<td>GDN without A and C</td>
<td>92.60</td>
<td>24.62</td>
<td>95.85</td>
<td>396.08</td>
<td>89.70</td>
</tr>
<tr>
<td>GDN without A and B</td>
<td>90.94</td>
<td>23.65</td>
<td>96.76</td>
<td>598.18</td>
<td>86.79</td>
</tr>
<tr>
<td>GDN without C</td>
<td>92.86</td>
<td>25.06</td>
<td>95.72</td>
<td>432.35</td>
<td><strong>90.38</strong></td>
</tr>
<tr>
<td>GDN without B and C</td>
<td>92.63</td>
<td>24.40</td>
<td>95.59</td>
<td>592.35</td>
<td>89.30</td>
</tr>
<tr>
<td>GDN without A and C</td>
<td>92.63</td>
<td>24.40</td>
<td>95.59</td>
<td>592.35</td>
<td>89.30</td>
</tr>
<tr>
<td>GDN</td>
<td><strong>93.57</strong></td>
<td><strong>24.18</strong></td>
<td><strong>98.04</strong></td>
<td><strong>351.59</strong></td>
<td><strong>89.95</strong></td>
</tr>
</tbody>
</table>
Fig. 1. Examples of cerebrovascular DSA image, the images in each row from left to right are live image, mask image and subtraction image respectively.
Fig. 2. The framework of the proposed GDN-Net. The GDN-Net consists of the mask net and the vessel net. And the loss function between the vessel net and the mask net is proposed to constrain these networks. In addition, the real vessel images are used to be the gold standard.
Fig. 3. Generation comparisons of MUNIT, Pix2Pix, UNet and GDN on our dataset. Subgraphs represent the detail of the same location for each set of images, as marked with rectangular boxes.
Fig. 4. Results of different methods excluding abnormal values. (a) SSIM; (b) PSNR; (c) VSI; (d) FID; (e) FSIM.