Deblurring near infrared fluorescence images with CycleGAN

Zhuoran Ma
Department of Chemistry
Stanford University
zma2@stanford.edu

Haotian Du
Department of Chemistry
Stanford University
dht14@stanford.edu

Abstract
Compared to other imaging modalities, fluorescence imaging provides benefits of real-time image acquisition with high spatial resolution via non-radioactive optical system. However, the resolution has been limited mainly due to photon scattering, absorption and autofluorescence of biological tissues. In this project, we used CycleGAN to transform a scattering limited traditional near infrared window (NIR-I) fluorescence image to a high-resolved long wavelength (NIR-IIb) image.

1 Introduction
In vivo imaging is an emerging and promising technique to profile the internal status of biological systems. Among various in-vivo imaging modalities, fluorescence imaging provides benefits of real-time image acquisition with high spatial resolution via non-radioactive optical system. However, the resolution has been limited mainly due to photon scattering, absorption and autofluorescence of biological tissues. Imaging in NIR-II window has received much interest for its advances of improved penetration depth and higher spatial resolution, owing to reduced photon scattering, minimized tissue absorption and negligible autofluorescence from endogenous molecules. Recent research has achieved imaging in NIR-IIb window (1500nm-1700nm) in which signal to noise ratio has been improved dramatically. However, compared to biocompatible probes emitting in traditional near infrared window (NIR-I, 700-900nm) like ICG which is FDA approved, NIR-IIb fluorophores involving rare-earth nanoparticles and lead/cadmium-based quantum dots always suffer from safety concerns. By training a neural network to transform NIR-I images to NIR-IIb images, we will be able to generate high spatial resolution fluorescence images from scattering-limited NIR-I images, enabling the possibility to conduct NIR-IIb imaging using FDA-approved biocompatible NIR-I probes.

Our inputs are real NIR-I images taken with camera. Our outputs are NIR-IIb images generated by cycleGAN neural network.

2 Related work
Image-to-image translation is a computer vision problem, where a mapping is learned to transfer images from one domain to another. The results of image-to-image translation have been largely improved since the introduction of generative adversarial networks (GAN). For example, pix2pix uses conditional GAN to transfer paired data of images in a fully supervised way. However, collecting paired data can be time-consuming and even not feasible. Unpaired image-to-image translation methods have also been proposed, such as CycleGAN, DualGAN and StarGAN. These frameworks introduced a cycle consistency constrain to preserve key features between the input image and the generated image. These methods have been successfully applied to different tasks, including season transfer, collection style transfer and object transfiguration.

Image-to-image translation techniques have also been used to synthesize medical images. For
example, pix2pix \cite{2017arXiv170306836P} and CycleGAN \cite{2017arXiv170306836P} were used to transfer medical images from one modality to another. GAN has also been used to transfer diffraction-limited fluorescence microscopic images to super-resolved ones \cite{2017arXiv170306836P}. Although traditional image processing methods have been used to reduce the effect of scattering for NIR-I images \cite{2017arXiv170306836P}, deep learning based methods have never been tried.

### 3 Dataset and Features

Both NIR-I and NIR-IIb fluorescence images were collected in the lab using an InGaAs camera \cite{2017arXiv170306836P}. The raw data were 16-bit grayscale images with a resolution of $640 \times 512$ pixels. The images were rescaled to $256 \times 256$ pixels to reduce training time. For neural style transfer, the images were transformed to 3-channel RGB images. Intensity of the images were normalized to the range $[-1, 1]$ to help training. There are 693 NIR-I and 869 NIR-IIb images in total. Because the amount of data is relatively small, we split the data to training/validation set randomly with a ratio of $80\%/20\%$ (554/694 NIR-I/NIR-IIb images in training set and 139/175 NIR-I/NIR-IIb images in validation set). We also performed random horizontal flip for data augmentation. Four sample images are shown in Fig. 1.

#### Figure 1: (a) NIR-I images for mice ventral, lateral, brain and hindlimb. (b) NIR-IIb images for mice ventral, lateral, brain and hindlimb.

### 4 Methods

#### 4.1 Neural style transfer

We used neural style transfer \cite{2017arXiv170306836P} as our baseline method for image-to-image translation. Neural style transfer is a image-to-image translation algorithm which combine the content of one image with the style of another. A NIR-I and a NIR-IIb image were selected as the content image and the style image, respectively, to create a "NIR-IIb-styled" NIR-I image. The images were used as inputs for a pre-trained VGG19 network \cite{2017arXiv170306836P}, and the activations of different layers were used to calculate the style loss, content loss and total loss, which are defined as follows:

$$J_{\text{style}} = \sum_{l=1}^{n} w_l \frac{1}{(2n_C n_H n_W)^2} ||\text{Gram}(S(l)) - \text{Gram}(G(l))||_2^2$$

$$J_{\text{content}} = \frac{1}{4n_C n_H n_W} ||A(C(l)) - A(G(l))||_2^2$$

$$J_{\text{total}} = \alpha J_{\text{style}} + \beta J_{\text{content}}$$

where $S$, $C$ and $G$ are the style image, content image and generated image, respectively. The matrix $\text{Gram}$ and $A$ are the gram matrix and the activation, respectively. The layers $\text{conv1}_1, \text{conv2}_1, \text{conv3}_1, \text{conv4}_1$ and $\text{conv5}_1$ of the original VGG network were used to calculate the style loss, each with a weight of 0.2. The layer $\text{conv4}_2$ was used to calculate the content loss. The total loss was a weighted sum of the style loss and the content loss.

#### 4.2 CycleGAN

The goal of CycleGAN \cite{2017arXiv170306836P} is to learn mapping functions between two domains $A$ (NIR-I images) and $B$ (NIR-IIb images) given unpaired data. A generator $G_A : A \rightarrow B$ is learned, such that the
distribution of images from $G_A(A)$ is indistinguishable from the distribution of images from B. Similarly, a generator $G_B : B \rightarrow A$ is learned to transform images from domain B to domain A. Two discriminators $D_A$ and $D_B$ are used to distinguish between real images and generated images in domain A and B, respectively. In order to guarantee the input image and the output image are paired in a meaningful way, a cycle consistency loss is used to make sure $G_B(G_A(x)) \approx x$ for an image $x$ in domain A, and vice versa. During training time, all the generators and discriminators were optimized. During test time, we only used $G_A$ to transfer a NIR-I image to a NIR-IIb one.

4.2.1 Model architecture

The architecture of the generator and the discriminator are shown in Fig.2(a) and (b), respectively. The generator contains an encoding network, a transformation network and a decoding network. The encoding part consists of 3 convolutional layers, the transformation part is a repeated application of 6 or 9 residual blocks, and the decoding part contains 2 deconvolutional layers followed by 1 convolutional layer. Each convolutional layer is followed by an instance normalization layer and a ReLU activation layer, except for the last convolutional layer, which uses tanh as the activation function.

The discriminator consists of 5 convolutional layers, each followed by an instance normalization layer and a ReLU activation layer, except for the last convolutional layer, which uses tanh as the activation function.

The adversarial loss $L$ is applied to both generators. The cycle consistency loss $L_{cyc}$ is introduced to make sure the reconstructed images $G_B(G_A(x))$ and $G_A(G_B(y))$ look similar to the original image $x$ and $y$, respectively. The full loss function is a weighted sum of the adversarial loss and the cycle consistency loss. In each iteration of the training, the generators $G_A$ and $G_B$ were first updated to minimize $L(G_A, G_B, D_A, D_B)$, and then the discriminators $D_A$ and $D_B$ were updated to minimize $-L_{adv}(G_A, D_B) + L_{adv}(G_B, D_A)$. We aim to solve the following min-max problem:

$$G_A^*, G_B^* = \arg\min_{G_A, G_B} \max_{D_A, D_B} L(G_A, G_B, D_A, D_B)$$

5 Experiments/Results/Discussion

5.1 Training details

For neural style transfer, we chose the hyperparameters $\alpha = 10$ and $\beta = 40$. The loss function was optimized using an Adam optimizer with a learning rate of 0.01 for 5,000 epochs. The NIR-IIb image

\[ \text{Figure 2: The architecture of the generator (a) and the discriminator (b)} \]

4.2.2 Loss function

The loss function of the network is defined as follows:

$$L_{adv}(G_A, D_A) = \mathbb{E}_{x \sim A}[||D_B(G_A(x)) - 1||^2_2] + \mathbb{E}_{y \sim B}[||D_A(y)||^2_2]$$

$$L_{adv}(G_B, D_B) = \mathbb{E}_{y \sim B}[||D_A(G_B(y)) - 1||^2_2] + \mathbb{E}_{x \sim A}[||D_B(x)||^2_2]$$

$$L_{cyc}(G_A, G_B) = \mathbb{E}_{x \sim A}[||G_B(G_A(x)) - x||_1] + \mathbb{E}_{y \sim B}[||G_A(G_B(y)) - y||_1]$$

$$L(G_A, G_B, D_A, D_B) = L_{adv}(G_A, D_B) + L_{adv}(G_B, D_A) + \lambda L_{cyc}(G_A, G_B)$$

The adversarial loss $L_{adv}$ is applied to both generators. The cycle consistency loss $L_{cyc}$ is introduced to make sure the reconstructed images $G_B(G_A(x))$ and $G_A(G_B(y))$ look similar to the original image $x$ and $y$, respectively. The full loss function is a weighted sum of the adversarial loss and the cycle consistency loss.
shown in Fig. 1 (b, ventral) was selected as the style image. The generated image was initialized to be the same as the content image. The generated RGB image was transferred to a grayscale image as the final output.

For CycleGAN, we used an Adam optimizer with the hyperparameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The learning rate was kept at 0.0002 in the first 100 epochs, and decayed to 0 linearly in the next 100 epochs. A mini-batch size of 1 was used during training. We chose these hyperparameters according to the original CycleGAN paper [7]. Two hyperparameters that we tuned are the number of Resnet blocks (6 or 9) and weights for the cycle consistency loss ($\lambda = 1$ or 10). These hyperparameters were selected by training on the training set, and then comparing the results using the validation set. The losses of the generators and the discriminators during training are shown in Fig. 3 (Resnet6, $\lambda = 1$). The loss values converged during training. It should be noted that the loss itself cannot represent the performance of CycleGAN.

Figure 3: Loss of the generators (a) and discriminators (b) during training

5.2 Results

Fig. 4 shows some random examples of the input NIR-I images from the validation set, the NIR-IIb images generated by CycleGAN with different hyperparameters and the NIR-IIb images generated by neural style transfer. The NIR-I images look blurry due to light scattering. Compared with the original image, the image generated by neural style transfer has higher resolution. However, some features are not correctly translated. For example, in the third image, the bright region which corresponds to the spleen of the mouse turns dark after image translation. Also, the region surrounding the mouse is bright after the application of neural style transfer, which is an artifact. All the CycleGAN results look similar. The images transferred using CycleGAN have lower background signal, higher resolution and better contrast. These images also have less artificial features.

Figure 4: Random examples of the results generated by neural style transfer and CycleGAN
5.3 Discussion

To evaluate the results more quantitatively, we calculated the Fréchet Inception Distance (FID) \[18\] between the generated images and the real NIR-IIb images. Features of the generated images and the real images were extracted by a pre-trained Inception-v3 model\[19\], and viewed as multivariate Gaussian random variables. FID was calculated as follows:

\[
FID(r, g) = ||\mu_r - \mu_g|| + Tr(C_r + C_g - 2(C_r C_g)^{\frac{1}{2}})
\]

where \((\mu_r, C_r)\) and \((\mu_g, C_g)\) are the mean and covariance matrix of the real and generated images, respectively. A lower FID indicates a better image quality. FID of different models are shown in Table 1. All the CycleGAN models show similar FID on the validation set, which is much smaller than the FID by neural style transfer. Compared to larger networks, a smaller network can reduce the training and testing time. A smaller weight for the cycle consistency loss may help generate images that are more similar to the real NIR-IIb images while preserve the feature of the original NIR-I image. The best hyperparameters in our current settings are 6 Resnet blocks with \(\lambda = 1\). The FID of the training set is smaller compared to that of the validation set in all models, which indicates the generator \(G_A\) overfits the training set. It should be noted that FID only evaluate the performance of the generator \(G_A\), which may not fully evaluate the performance of the whole CycleGAN model.

<table>
<thead>
<tr>
<th>Model</th>
<th>training set</th>
<th>validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resnet6 (\lambda = 1)</td>
<td>42.71</td>
<td>59.59</td>
</tr>
<tr>
<td>Resnet6 (\lambda = 10)</td>
<td>39.64</td>
<td>64.43</td>
</tr>
<tr>
<td>Resnet9 (\lambda = 1)</td>
<td>41.41</td>
<td>65.42</td>
</tr>
<tr>
<td>Resnet9 (\lambda = 10)</td>
<td>39.48</td>
<td>64.01</td>
</tr>
<tr>
<td>Neural style transfer</td>
<td>330.44</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5 shows some failed examples of our model (Resnet6, \(\lambda = 1\)). In these examples, part of the mouse seems to be missing after the image translation. These results suggest that the CycleGAN model may treat part of the animal as background signal if the fluorescence intensity is not very high, which can lead to artifacts.

Figure 5: Some failed examples using CycleGAN

6 Conclusion/Future Work

In summary, we applied CycleGAN to translate a scattering limited NIR-I fluorescence image to a corresponding NIR-IIb image. By tuning hyperparameters of the model, we got the best performance and efficiency with 6 Resnet blocks, and weight of the cycle consistency loss \(\lambda = 1\). The model outperformed the baseline model (neural style transfer). To the best of our knowledge, this is the first time that deep learning based method has been used to transform a low-resolution NIR fluorescence image to a high-resolution one.

To further improve the performance of our model, we will try the following:
(1) Collect more data for training. This may help reduce the variance of the generator.
(2) Do some data pre-processing. The images were collected by us under different imaging conditions. We can try adjusting the brightness and contrast of the images before training, so that they have similar mean and variance.
(3) Try to collect paired data of NIR-I and NIR-IIb images. With paired dataset, we’ll be able to use fully supervised image-to-image translation algorithms, such as pix2pix\[5\], which can outperform the CycleGAN model\[18\].
7 Contributions

Zhuoran Ma and Haotian Du conceived the project. Haotian Du collected and labeled the data with the help of Zhuoran Ma. Zhuoran Ma implemented the algorithms and trained the networks with the help of Haotian Du. Zhuoran Ma and Haotian Du wrote the manuscript. All the data were collected in Prof. Hongjie Dai’s lab and supported by the National Institutes of Health through grant DP1-NS-105737. The data were originally collected for other projects supervised by Prof. Dai, but this project is an independent project, not supervised by Prof. Dai or other Dai lab members.

Github link: https://github.com/zhuoranzma/Fluorescence-image-translation-with-CycleGAN

References


