HST-GAN: Historical Style Transfer GAN for Generating Historical Text Images

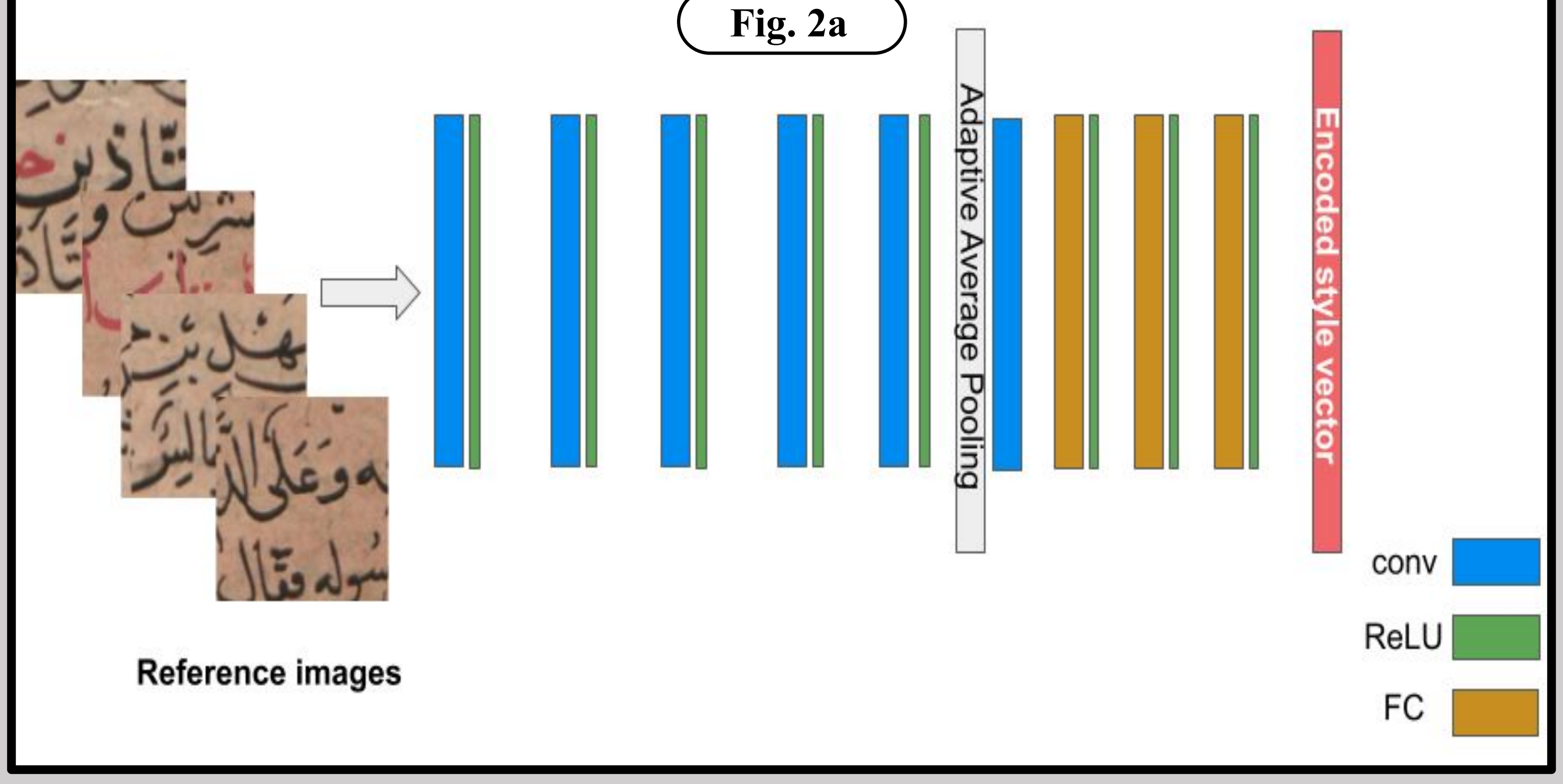
Boraq Madi, Reem Alaasam, Ahmad Droby and Jihad El-Sana
The Department of Computer Science, Ben-Gurion University of the Negev

Introduction:

Manual restoration of historical manuscripts is time-consuming, requires highly trained experts, and may spoil the document's original writing style. In recent years, Generative Adversarial Networks(GAN) and their variations have led to impressive progress in image manipulation and generation. Therefore, Our work presents Historical Style Transfer Generative Adversarial Networks (HST-GAN) for generating historical text images. Our model consists of three blocks: Encoder, Generator, and Discriminator. The Encoder encodes the style to be generated, and the Generator applies an encoded style, S to an input text image, I , and generates a new image with the content of I and the style S . Multiple loss functions are applied to ensure the generation of quality images. We evaluated our model against three challenging historical handwritten datasets of two different languages and compare the performance of HST-GAN with the state of art approaches.

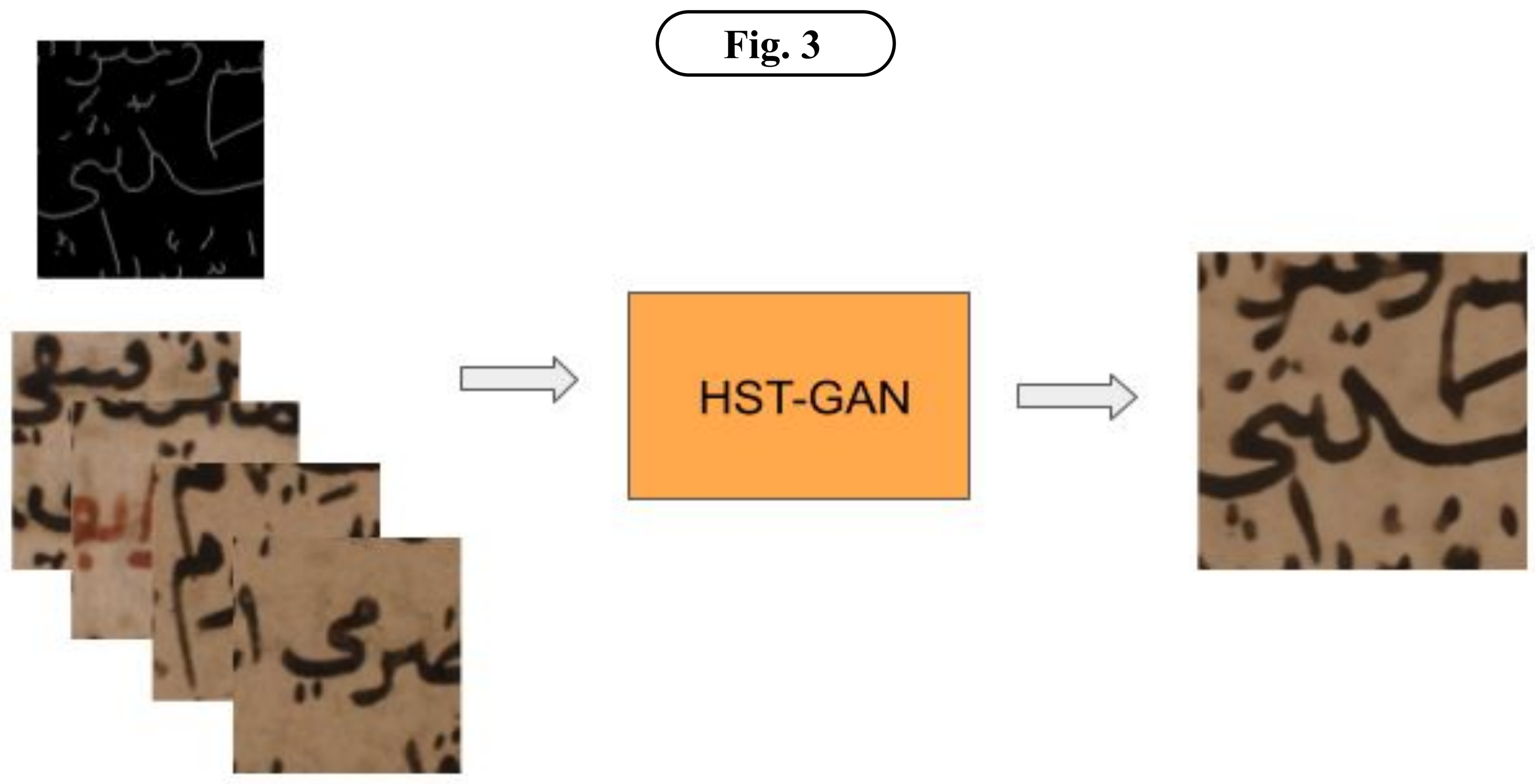
Fig. 1a Fig. 1b Fig. 1c Fig. 1d Fig. 1e

Fig. 2a



GT HST UFET CycleGAN Pix2Pix

Fig. 3



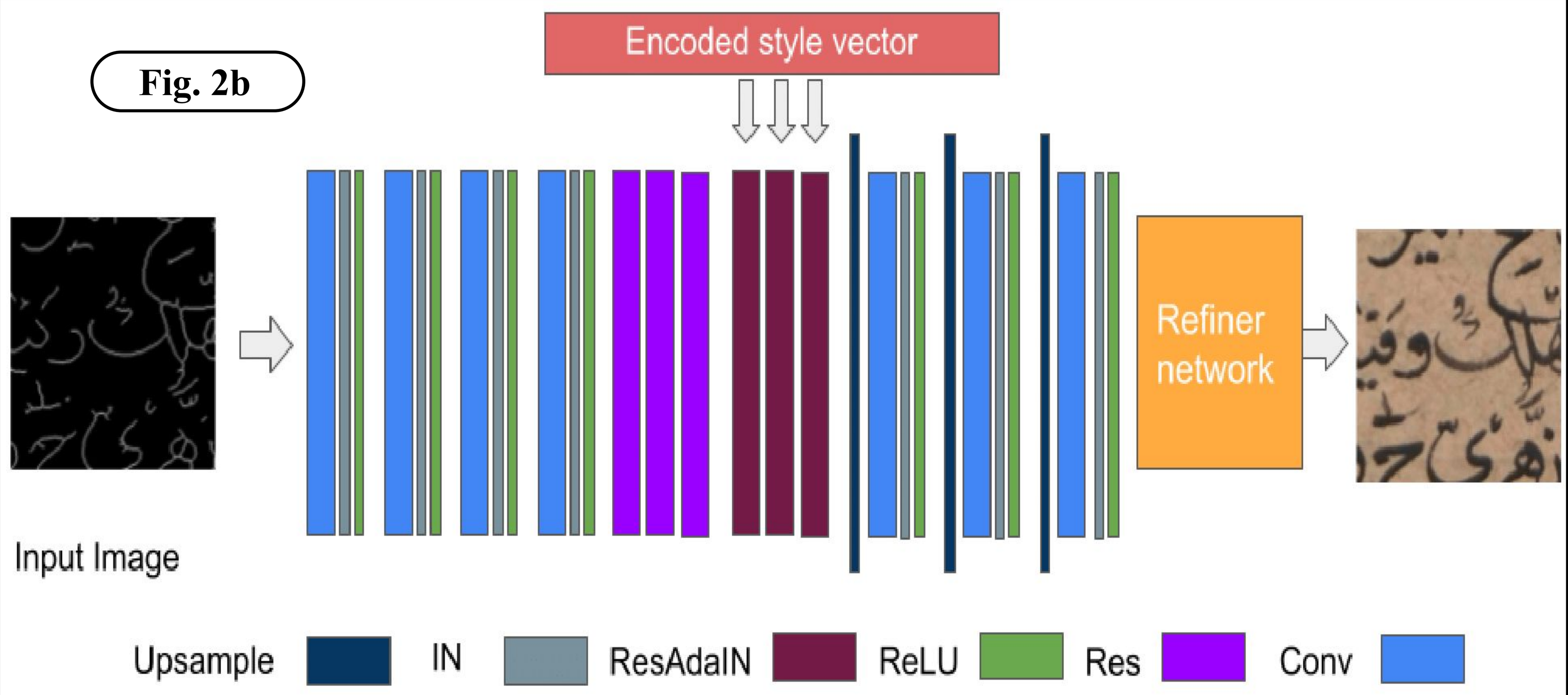
Datasets:

- **DIVA-HisDB:** is a historical Latin dataset that consists of three medieval manuscript, with different styles. We experimented with one manuscript, CSG863, which has a challenging layout. It includes touching components, little overlapping letters, but no punctuation as shown in **Fig.1a**.
- **AHTE-A:** contains bold text with many ink stains but less punctuation and overlapping text compared to AHTE-B. It is more complex than DIVA-HisDB but less challenging than AHTE-B as shown in **Fig.1b**.
- **AHTE-B:** includes complex layouts, crowded diacritics, and many overlapping words and sentences as shown in **Fig.1c**. Although it has sharp text, it includes background texture noise. It is the most complex among the three datasets due to the existence of many small details, such as punctuation.
- **Data Preparation:** The dataset consists of samples created from the three datasets. Each sample is a triplet that includes a patch P(See **Fig.1d**), its skeleton S(P)(See **Fig.1e**), and a set of reference patches, which have styles similar to P.

- **The Generator and Encoder Models**

The Encoder input is a set of reference image and the output is a latent vector, as shown in **Fig. 2a**. The Generator G input is a skeleton text and the output is an image conditioned with the encoded historical style as shown in **Fig. 2b**.

Fig. 2b



Results:

- **HST-GAN** and **UFET-GAN** generate smooth background and remove unnecessary background noise comparing to **CycleGAN** and **Pix2Pix**. For images with crowded diacritics and overlapping words as shown in the third row, HST-GAN produced the best results and **UFET-GAN** comes second, while **CycleGAN** and **Pix2Pix** failed to generate the small details correctly.
- **UFET-GAN** generates images with higher blur level than the other approaches, while **HST-GAN** provides images with sharp edges. Overall, for the three test datasets, DIVA, AHTE-A and AHTE-B, HST-GAN generates images with the best quality.

Data Augmentation Using Style Transfer:

- HST-GAN can project the style of a historical dataset to the content of another dataset(skeleton form) to augment and build new historical dataset as shown in **Fig. 3**.
- We evaluate the quality of augmentation using a trained classifier and human questionnaire consisting of 30 questions.