DETAIL FUSION GAN: HIGH-QUALITY TRANSLATION FOR UNPAIRED IMAGES WITH GAN-BASED DATA AUGMENTATION

Ling Li[1], Yaochen Li[1*], Chuan Wu[1], Hang Dong[2] and Peilin Jiang[1]

[1] School of Software Engineering, Xi'an Jiaotong University, [2] Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University,

* Y. Li is the corresponding author yaochenli@mail. xjtu.edu.cn

Motivation

Image-to-image translation, a task to learn the mapping relation between two different domains, is a rapid-growing research field in deep learning.

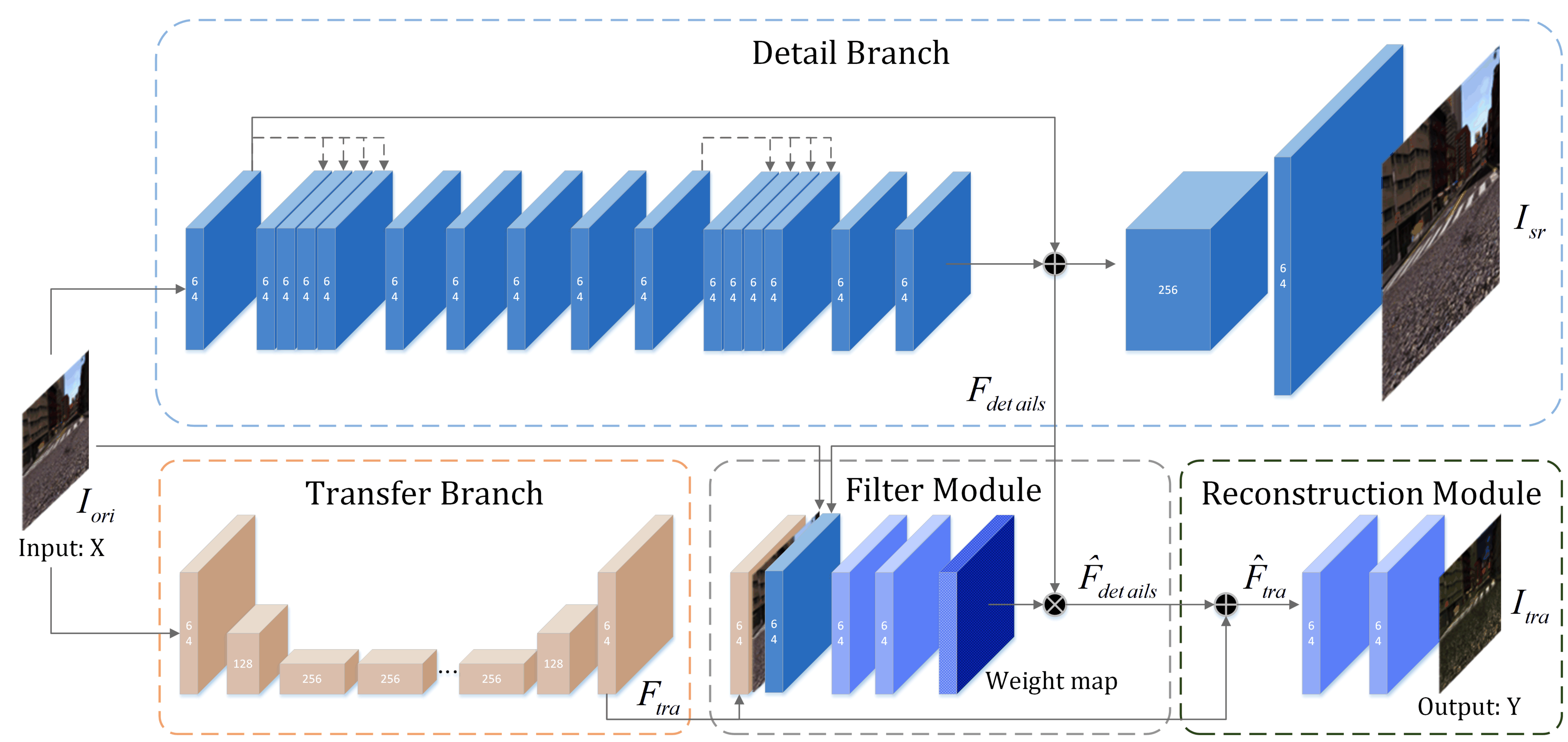
Although existing Generative Adversarial Network (GAN)-based methods have achieved decent results in this field, there are still some limitations in generating high-quality images for practical applications.

Contributions

The main contributions of our work are summarized as follows:

1. To the best of our knowledge, this is the first GAN-based method to introduce super-resolution loss as a guidance to improve the quality of translated results.
2. We divide the proposed network into two branches. One branch is to obtain feature maps with more details, and another branch is to obtain style translation feature maps. And we propose a filter module, which can effectively suppress the style information in the detail feature maps to avoid affecting the style translation feature maps.
3. Compared with the state-of-the-art GAN-based methods, our model generates better results and converges faster.

Network Architecture



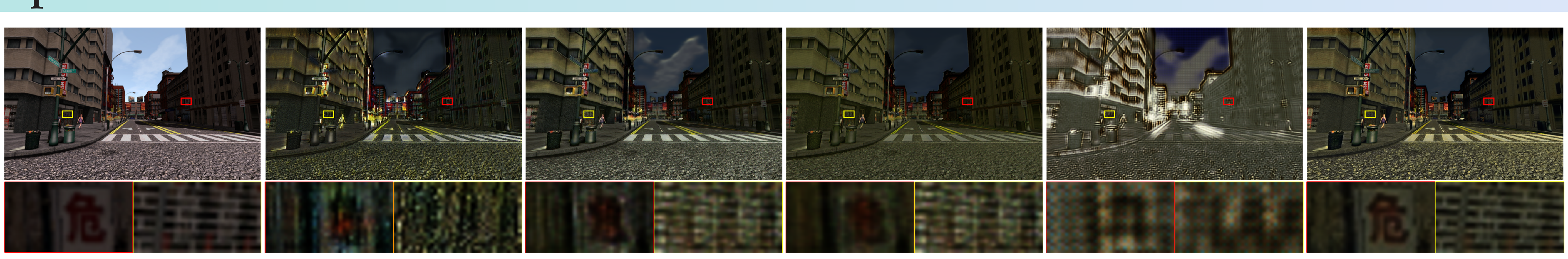
Detail Branch. The detail branch aims to extract the detail feature F_{detail} from the original image I_{ori}. To transfer the details, we feed the feature map F_{detail} before the upscaling layer to the filter module for further processing.

Transfer Branch. The backbone of transfer branch is modified based on the generator of CycleGAN by removing the last convolution layer. It encodes the images and transfer the style information to generate transferred feature F_{tra}.

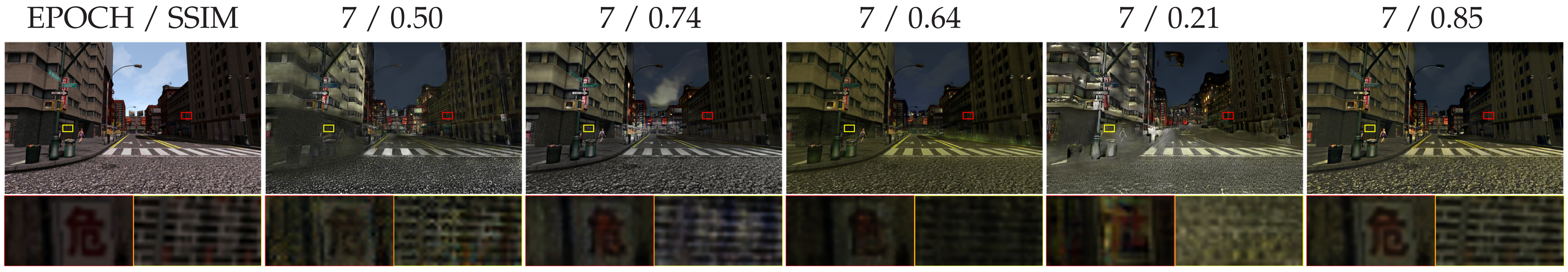
Filter Module. Since the detail feature F_{detail} from detail branch contains strong style attribute of the original image, the direct fusion with transferred feature F_{tra} will affect the translated result. Therefore, we propose a filter module to suppress the style information. The style-free detail feature $\hat{F}_{detail}$ will be fed into the reconstruction module for future fusing.

Reconstruction Module. We obtain detail-enhanced feature $\hat{F}_{tra}$ by fusing $\hat{F}_{detail}$ and F_{tra}, and then it generates the final result I_{tra} through the reconstruction module.

Experiments and Results



(a) Input EPOCH / SSIM | (b) CycleGAN 7 / 0.50 | (c) UNIT 7 / 0.74 | (d) MUNIT 7 / 0.64 | (e) DRIT 7 / 0.21 | (f) DFGAN(ours) 7 / 0.85



(a) Input EPOCH / SSIM | (b) CycleGAN 49 / 0.63 | (c) UNIT 30 / 0.79 | (d) MUNIT 35 / 0.74 | (e) DRIT 1199 / 0.36 | (f) DFGAN(ours) 7 / 0.85

We evaluate the proposed DFGAN on day-to-night translation task. For fair comparisons, we compare the results of all the evaluated methods in the same iteration. To better evaluate the detail recovering ability of each method, we zoom in two local patches of each image.

We also show the best result of each method which can preserve most structural information in the original image (i.e., with highest SSIM score).

Method	Input		Output		acc. of FCN		AP of YOLO		FID
	I_{ori}	I_{hr}	I_{tra}	I_{sr}	Day	Night	Day	Night	
Daytime Only	-	-	-	-	0.7946	0.4798	0.7999	0.5840	-
CycleGAN	256×256	-	256×256	-	0.8223	0.7560	0.8326	0.7294	62.53
UNIT	256×256	-	256×256	-	0.7983	0.8097	0.8462	0.7132	52.03
MUNIT	256×256	-	256×256	-	0.8393	0.7931	0.8411	0.6873	53.47
DRIT	256×256	-	256×256	-	0.8389	0.7972	0.8502	0.7408	65.62
DFGAN-s	128×128	256×256	128×128	256×256	**0.8421**	**0.8162**	**0.8624**	**0.7501**	**48.87**
DFGAN	256×256	512×512	256×256	512×512	**0.8518**	**0.8227**	**0.8801**	**0.7553**	**47.76**

Considering that in the previous experiment, our detail branch is trained by 512×512 images, which may cause our model to gain potential advantages. In order to make a fair comparison, we use 256×256 images to train the detail branch, 128×128 images to train the transfer branch, and test the performance of this model (DFGAN-s).

We use Fréchet Inception Distance (FID) to measure the distribution difference between the generated nighttime images by the evaluation models and the images in the SYNTHIA-NIGHT dataset. To evaluate these methods' capability of data augmentation, we combine the daytime images and the translated nighttime images by each method as the training dataset to train the FCN and YOLO. Both DFGAN and DFGAN-s obtain lower FID than other models, which means that our model can generate more realistic nighttime images. Since the FCN and YOLO trained on the dataset augmented by our model achieve higher accuracy and precision, the proposed DFGAN is more proper to augment the datasets for semantic segmentation and detection.