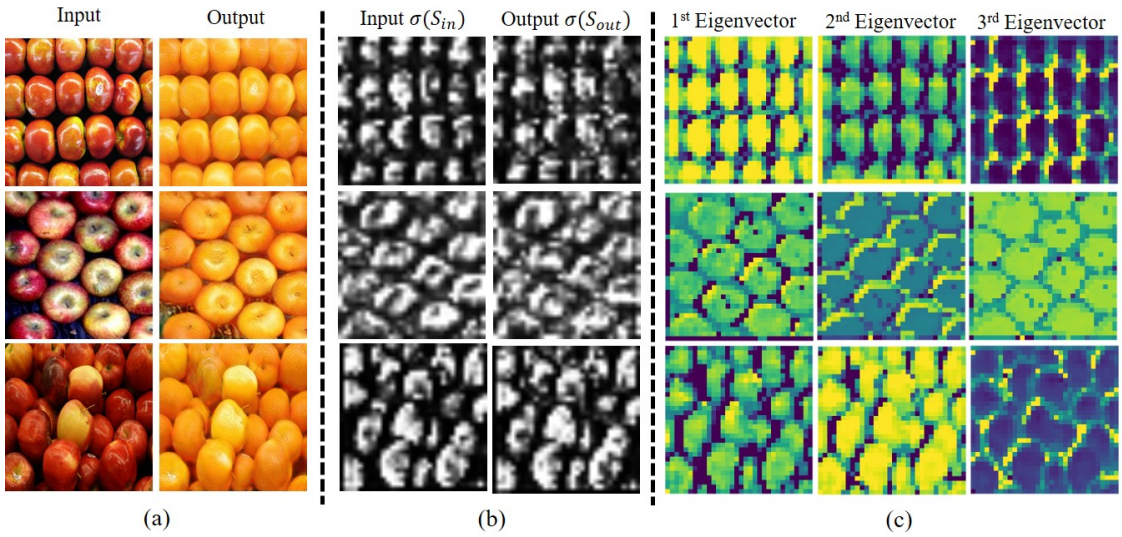

Figure 9: Analysis of the proposed method: (a) Input and the output images. (b) Visualization of $\sigma(S_{in}), \sigma(S_{out})$. The vector p allocates higher weights for the object parts which are task-relevant. Similar appearance refers the correspondence between input and output. (c) Eigenvectors of the Laplacian matrix of A, which are coherent to the semantics of the image.

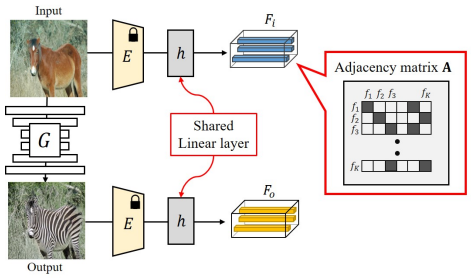

Figure 10: The adjacency matrix A is constructed from F_i which is the output of learnable h. Here, h is updated by the gradient from the F_o similar to CUT (**?**).

	Settings				H→Z	
	# of Hop (**n**)	Thresh (**t**)	# of Pool	Down sample	FID↓	KID↓
GNN Ablation (**n, t**)	1	0.1	1	1/4	37.9	0.438
	3	0.1	1	1/4	39.9	0.374
	2	0.0	1	1/4	34.5	0.551
	2	0.4	1	1/4	36.8	0.293
	2	0.6	1	1/4	38.3	0.332
Pooling Ablation	2	0.1	0	-	37.6	0.432
	2	0.1	2	1/4	35.0	0.625
	2	0.1	1	1/8	37.7	0.340
Proposed	**2**	**0.1**	**1**	**1/4**	**34.5**	**0.271**

Table 2: Quantitative results of ablation studies. Our setting shows the best performance in both of FID and KID×100.

graph has fewer nodes which leads to fewer negative pairs for the contrastive learning. Additionally, we provide the results with varying downsampling rate. For the downsampling of 1/8, the pooled graph consists of fewer nodes, which leads to similar problem with the excessive pooling layers. This again confirms that a sufficiently dense graph after the pooling can capture the semantically meaningful hiearchy. We provide additional ablation study for the graph construction in the supplementary material.

Conclusion

In conclusion, we proposed a novel patch-wise graph representation matching method for image translation task. For structural consistency between input and output images, we proposed to match the constructed graphs between input and outputs. In this part, we used the same adjacency matrix for input and output images for graph consistency. To further leverage the topological information in an hierarchical manner, we applied graph pooling on initial graphs. Our experimental results showed state-of-the-art performance, which again confirms that graph-based patch representation have obvious advantage over baseline methods.

Acknowledgements

This research was supported by National Research foundation of Korea(NRF) (**RS-2023-00262527**)

Ethical Impacts

Regarding on the social impact, the realistic fake images generated by the proposed method may produce a social disinformation, as most of image generation methods shares. Also, the model has potential risk of violating copyright as the model learns the mapping function from input to target distribution. Magnam non mollitia, doloremque autem temporibus non fuga in dolorem nam doloribus praesentium laudantium.Expedita recusandae aliquam unde voluptatum ratione commodi provident, ad eveniet nulla doloremque

harum assumenda suscipit, eum fugiat adipisci necessitatibus.