StegoGAN: Leveraging Steganography for Non-Bijective Image-to-Image Translation

Supplementary Material

In this appendix, we first detail the construction of the three datasets used in the experiments (Section 1), then provide additional implementation details (Section 2), ablation experiments (Section 3), and qualitative results (Section 4).

1. Details on Dataset Construction

1.1. GoogleMaps

While this paper focuses on non-bijective translation, the official GoogleMaps dataset was introduced in CycleGAN [13] for bijective translation between serial photos and maps. Consequently, we propose a protocol to create controllable non-bijectivity in this dataset.

We select the "highway" class for its prevalence and its distinctiveness on maps: they are always represented by the same orange hue. This allows us to easily detect highways on maps by thresholding in color space: a pixel is a highway if all color channels are closer than 20 units from $(240, 160, 30)$. In total, 356 image/map pairs of the train set of the GoogleMaps contain highways, and 740 do not.

For the training set, the source domain is always defined as 548 aerial images that do not contain highways. We define different versions of the target domain for the test set by fixing the ratio of maps that contains highways, from 0% to 60%, for a fixed total of 548 images. The test set is composed of 899 pairs of aligned aerial photos and maps *that do not contain the highways class* from the test set of the GoogleMaps dataset.

1.2. Brats MRI

We adapt the protocol of Cohen *et al*. [3] from the Brats2013 datasets [11] to the more recent, larger, and more diverse Brats2018 dataset [2]. We consider two MRI modalities: native (T1) and Fluid Attenuated Inversion Recovery (FLAIR). We selected transverse slices from the 60° to 100° range in the caudocranial direction [1] for both modalities of scans.

We label each scan as tumorous if more than 1% of its pixels are labelled as such, and as healthy if it contains no tumor pixels. We only use high-grade gliomas (HGG) instead of low-grade gliomas (LGG) as the are more easily observable [10]. In total, we obtain 5035 pathological pairs and 1135 healthy pairs. The train set is composed of a source domain of 800 T1 images of healthy brains, while the target domain set is composed of FLAIR scans of which 480 (60%) are tumorous and 320 healthy. The test set is composed of 335 aligned scans of healthy brains in both modalities.

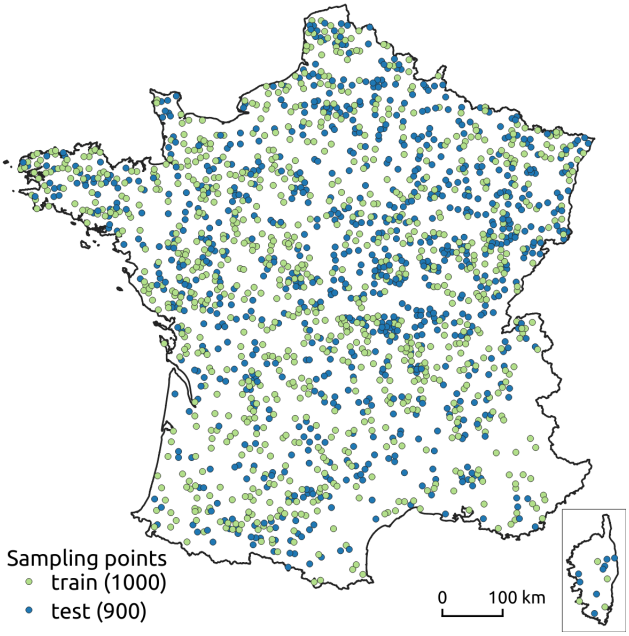

Figure A-1. **Spatial Distribution of Samples in PlanIGN.**

1.3. PlanIGN

We construct the PlanIGN dataset from two open-access sources available on the French governmental geoportal: aerial orthophotos and Plan IGN cartographic product, both projected in RGF93-Lambert-93. As the maps are derived directly from the orthophotos, we ensure the precise spatial alignment between both modalities.

Sampling. We consider aligned image/map pairs of resolution 256×256 at a scale of 1:12500 and a graphics resolution of 96 dpi, corresponding to a ground sampling distance of 3.3m per pixel. We randomly select samples across the French territory with a 3km buffer between images. We removed images that were blurry, with significant radiometric aberrations, over sensitive areas, or for which the roads were significantly occluded. In total, we sample 1900 such pairs, whose spatial distribution is shown in Figure A-1, and whose semantic distribution is given in Figure A-2.

Processing. We apply the following processing to the maps to make them easier to translate:

- We remove underground objects and small paths.

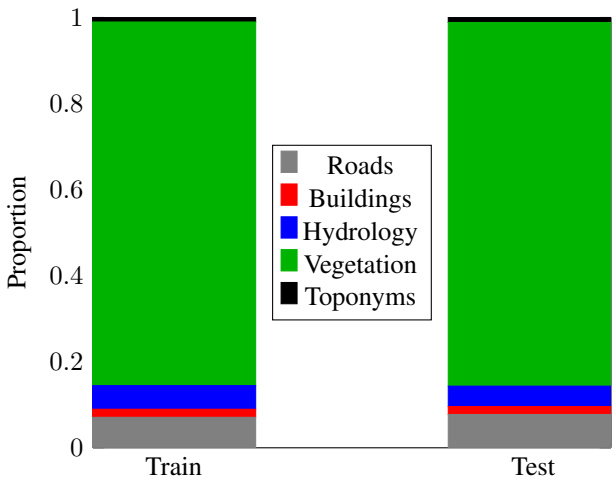

Figure A-2. **Semantic Distribution in PlanIGN.**

- We homogenize the palette: we use the same color for all roads except highways, and for buildings and hydrology.
- We optionally add the toponyms to the maps. When we do, we also compute a toponym mask by applying a 4-pixel dilation to the binary difference between the map with and without toponyms.

Dataset. The training set is composed of 1000 orthophotos for the source domain and 1000 maps with toponyms for the target domain. The test set is composed of 900 aligned pairs of orthophotos and maps without toponyms.

2. Additional Implementation Details

We follow the same architecture and hyperparameters as CycleGAN, including the ResNet-based generator [5] with 9 residual blocks, PatchGAN discriminator [6], and weights in th loss. We train our model for 200 epochs with a learning rate of 0.002 and the ADAM optimizer [8].

The hyperparameters for each dataset are given in Table A-1. Most methods use similar hyperparameters with two exceptions:

- Due to the high heterogeneity and noisiness of scans across different MRI machines, we use a larger batch size of 12.
- We observed better unmatchability masks with shallower encoders for PlanIGN, whereas it was the contrary for other datasets. Section 4.4 has pointed out that shallower encoder seems more influenced by the variation in appearance. Despite that class like hydrology exits both in the source and target domain, variations in colors or occlusions by vegetation occurring often in the aerial images challenge the model to establish correct correspondences. This should also be empirically regarded as mismatch, which shallower encoder performs better to capture.

Table A-1. **Hyperparameters.** We report the value for different parameters across the datasets used in the experiments.

Dataset	λ_{reg}	Encoder_Depth	Batch_Size	λ_{match}
GoogleMaps	0.3	8	1	1
PlanIGN	0.25	1	1	1
Brats	0.3	8	12	1

Table A-2. **Impact of λ_{reg}.** We report the performance on the **GoogleMaps** dataset of our method for different values of regularization strength. Values between 0.3 and 0.5 give good results, while the performance rapidly decreases above 0.6.

$\mathcal{L}_{\text{reg}}$	RMSE↓	Acc(σ_1)↑	Acc(σ_2)↑	pFPR(‰)↓	iFPR↓	FID↓	KID↓
0.1	26.1	11.0	60.4	0.1	7.7	254.4	25.8
0.2	23.3	36.5	64.0	**0.0**	**0.0**	141.2	17.4
0.3	**22.7**	41.7	**67.1**	**0.0**	**0.0**	**77.3**	**6.8**
0.4	23.5	**41.8**	65.7	**0.0**	0.2	84.8	8.3
0.5	24.4	39.5	64.6	**0.0**	1.1	86.1	8.1
0.6	25.2	40.0	64.8	19.1	16.8	103.5	11.3
0.7	25.7	36.8	62.7	18.8	22.2	110.1	11.3

3. Additional Ablation Study

The hyperparameter λ_{reg} is crucial as it enforces the sparsity of the unmatchability masks. We report its impact in Table A-2. Too small values may lead to a too-liberal use of the unmatchability masks, resulting in a loss of details in the clean generation. Too large values will prevent our model from using the unmatchability masks altogether.

4. Additional Qualitative Results

We provide additional results for **GoogleMaps** in Figure A-3, **PlanIGN** in Figure A-4 and **Brats MRI** in Figure A-5.

Application to Natural Images. We apply our method to natural image datasets; see Fig. A-6. StegoGAN performs well in this setting and generates faithful yet realistic images. Compared to CycleGAN, the clean translation $y_{\text{gen}}^{\text{clean}}$ produces fewer unmatchable features like snow or color shifts (Summer $\mapsto$ Winter example), or internal structures of fruits (Apple $\mapsto$ Orange example). However, since such features often contribute to the realism and visual appeal of the translated images. Therefore, our method is better suited for domains that value reliability over aesthetics, such as medical images or cartography.

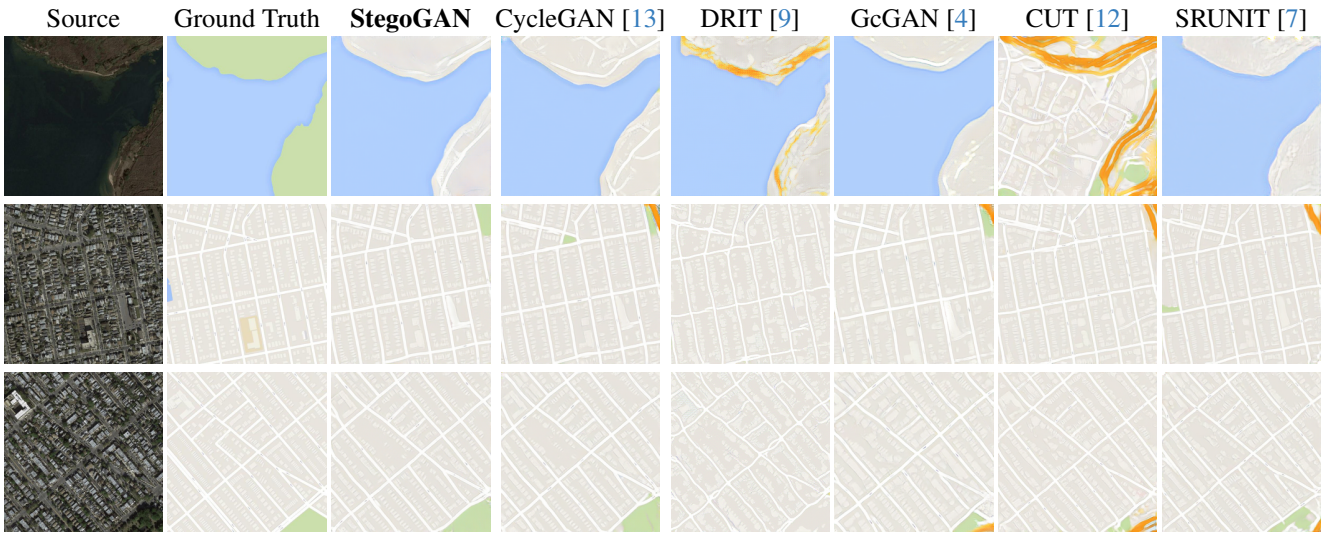

Figure A-3. Additional qualitative comparison on **Google Photo→Map**.

Figure A-4. Additional qualitative comparison on **PlanIGN**.

References

[1] Simon Andermatt, Antal Horváth, Simon Pezold, and Philippe Cattin. Pathology segmentation using distributional differences to images of healthy origin. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in*

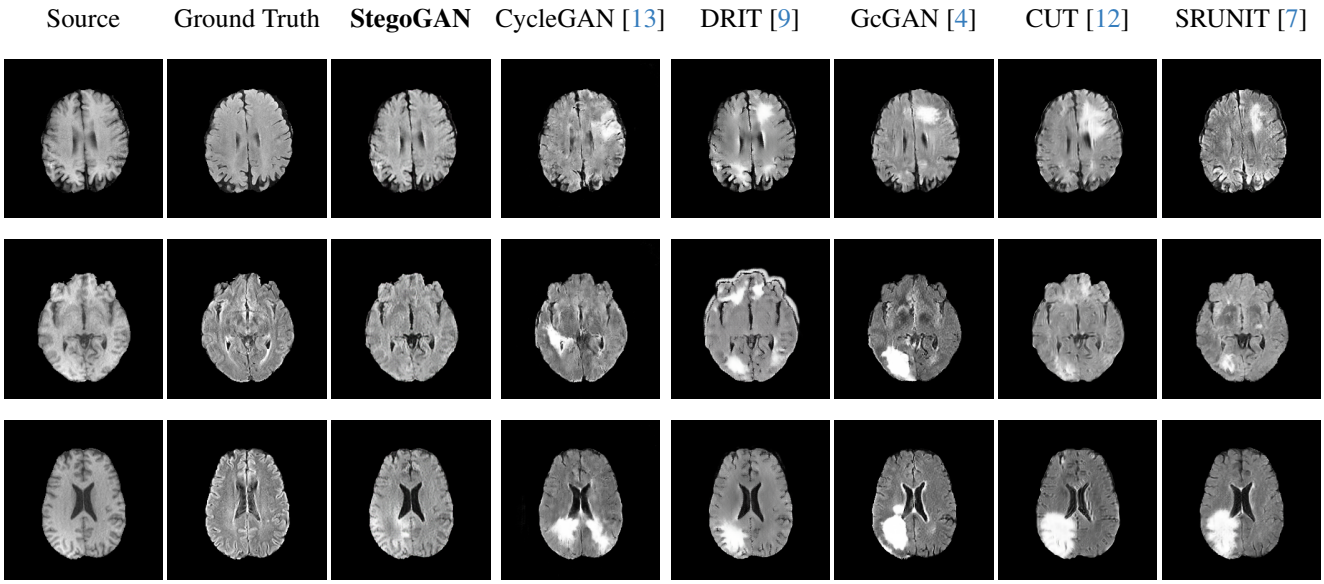

Figure A-5. Additional qualitative comparison on **Brats**.

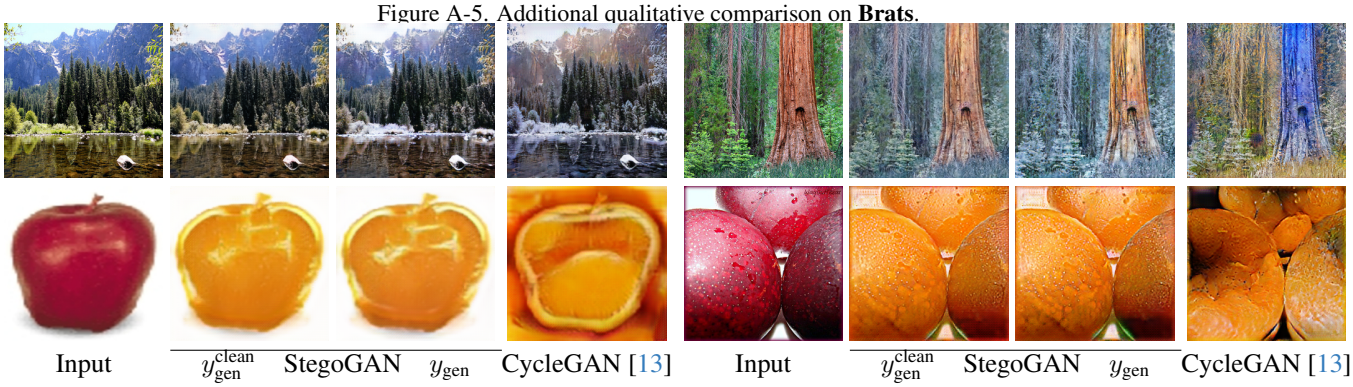

Figure A-6. **Natural Images Translation.** We apply our model to the **Summer** $\mapsto$ **Winter Yosemite** (top row) and **Apple** $\mapsto$ **Orange** datasets (bottom row).

Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4, pages 228–238. Springer, 2019. 1

[2] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1): 1–13, 2017. 1

[3] Joseph Paul Cohen, Margaux Luck, and Sina Honari. Distribution matching losses can hallucinate features in medical image translation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, pages 529–536. Springer, 2018. 1

[4] Huan Fu, Mingming Gong, Chaohui Wang, Kayhan Batmanghelich, Kun Zhang, and Dacheng Tao. Geometry-Consistent Generative Adversarial Networks for One-Sided Unsupervised Domain Mapping. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3, 4

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[6] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1125–1134, 2017. 2

[7] Zhiwei Jia, Bodi Yuan, Kangkang Wang, Hong Wu, David Clifford, Zhiqiang Yuan, and Hao Su. Semantically robust unpaired image translation for data with unmatched semantics statistics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14273–14283, 2021. 3, 4

[8] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 2

[9] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Kumar Singh, and Ming-Hsuan Yang. Diverse image-to-image

translation via disentangled representations. In *European Conference on Computer Vision*, 2018. 3, 4

[10] Szidónia Lefkovits, László Lefkovits, and László Szilágyi. Hgg and lgg brain tumor segmentation in multi-modal mri using pretrained convolutional neural networks of amazon sagemaker. *Applied Sciences*, 12(7), 2022. 1

[11] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014. 1

[12] Taesung Park, Alexei A. Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *European Conference on Computer Vision*, 2020. 3, 4

[13] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017. 1, 3, 4