Acknowledgement

This work was conducted as part of the research project MAGE (ANR-22-CE23-0010) funded by the *Agence Nationale de la Recherche*. This work was granted access to the HPC resources of IDRIS under the allocation AD011014327R2 made by GENCI.

References

[1] Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*, 2023. 2, 3, 4

[2] Michael S Albergo, Mark Goldstein, Nicholas M Boffi, Rajesh Ranganath, and Eric Vanden-Eijnden. Stochastic interpolants with data-dependent couplings. *arXiv preprint arXiv:2310.03725*, 2023. 3, 5

[3] Ollin Boer Bohan. Tiny AutoEncoder for Stable Diffusion. https://github.com/madebyollin/taesd. 6

[4] Derrick Bonafilia, Beth Tellman, Tyler Anderson, and Erica Issenberg. Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 5, 15

[5] Valentin De Bortoli, Iryna Korshunova, Andriy Mnih, and Arnaud Doucet. Schrodinger bridge flow for unpaired data translation. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 3, 4, 6, 12, 16

[6] L. Bruzzone and D.F. Prieto. Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2):456–460, 2001. 2

[7] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. pages 801–818. 5

[8] Seun-An Choe, Ah-Hyung Shin, Keon-Hee Park, Jinwoo Choi, and Gyeong-Moon Park. Open-set domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 23943–23953, June 2024. 2

[9] Kai Norman Clasen, Leonard Hackel, Tom Burgert, Gencer Sumbul, Begüm Demir, and Volker Markl. reBEN: Refined BigEarthNet Dataset for Remote Sensing Image Analysis. 5, 15

[10] Bharath Bhushan Damodaran, Benjamin Kellenberger, Rémi Flamary, Devis Tuia, and Nicolas Courty. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 447–463, 2018. 2

[11] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709, 2021. 2, 3

[12] Wenkai Dong, Song Xue, Xiaoyue Duan, and Shumin Han. Prompt tuning inversion for text-driven image editing using diffusion models. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7396–7406, 2023. 3

[13] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first international conference on machine learning*, 2024. 2, 3, 6, 12

[14] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12873–12883, 2021. 4

[15] Kuiliang Gao, Anzhu Yu, Xiong You, Chunping Qiu, and Bing Liu. Prototype and context-enhanced learning for unsupervised domain adaptation semantic segmentation of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023. 2

[16] Obsa Gilo, Jimson Mathew, Samrat Sohel Mondal, and Rakesh Kumar Sanodiya. Rdaot: Robust unsupervised deep sub-domain adaptation through optimal transport for image classification. *IEEE Access*, 11:102243–102260, 2023. 2

[17] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-or. Prompt-to-prompt image editing with cross-attention control. In *The Eleventh International Conference on Learning Representations*, 2023. 3

[18] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6629–6640, Red Hook, NY, USA, 2017. Curran Associates Inc. 7

[19] Ronny Hänsch, Jacob Arndt, Dalton Lunga, Matthew Gibb, Tyler Pedelose, Arnold Boedihardjo, Desiree Petrie, and Todd M. Bacastow. Spacenet 8 - the detection of flooded roads and buildings. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1471–1479, 2022. 5, 14

[20] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, 2017. 2, 3, 6, 7, 15, 16, 18

[21] Sadeep Jayasumana, Srikumar Ramalingam, Andreas Veit, Daniel Glasner, Ayan Chakrabarti, and Sanjiv Kumar. Rethinking fid: Towards a better evaluation metric for image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9307–9315, 2024. 7

[22] Xingshuo Jing, Kun Qian, Tudor Jianu, and Shan Luo. Unsupervised Adversarial Domain Adaptation for Sim-to-Real Transfer of Tactile Images. 72:1–11. 2

[23] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, 26(11):3365–3385, 2019. 2

[24] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative

models. *Advances in neural information processing systems*, 35:26565–26577, 2022. 12

[25] Benjamin Kellenberger, Onur Tasar, Bharath Bhushan Damodaran, Nicolas Courty, and Devis Tuia. *Deep Domain Adaptation in Earth Observation*, chapter 7, pages 90–104. John Wiley & Sons, Ltd, 2021. 1, 2

[26] Beomsu Kim, Yu-Guan Hsieh, Michal Klein, marco cuturi, Jong Chul Ye, Bahjat Kawar, and James Thornton. Simple reflow: Improved techniques for fast flow models. In *The Thirteenth International Conference on Learning Representations*, 2025. 3, 6, 12

[27] Beomsu Kim, Gihyun Kwon, Kwanyoung Kim, and Jong Chul Ye. Unpaired image-to-image translation via neural schrödinger bridge. In *ICLR*, 2024. 2, 3, 6, 7, 18

[28] Seon-Hoon Kim and Dae-won Chung. Conditional brownian bridge diffusion model for vhr sar to optical image translation. *arXiv preprint arXiv:2408.07947*, 2024. 3

[29] Theodoros Kouzelis, Ioannis Kakogeorgiou, Spyros Gidaris, and Nikos Komodakis. Eq-vae: Equivariance regularized latent space for improved generative image modeling. *arXiv preprint arXiv:2502.09509*, 2025. 4

[30] Geun-Ho Kwak and No-Wook Park. Assessing the potential of multi-temporal conditional generative adversarial networks in sar-to-optical image translation for early-stage crop monitoring. *Remote Sensing*, 16:1199, 03 2024. 2

[31] Trung Le, Tuan Nguyen, Nhat Ho, Hung Bui, and Dinh Phung. Lamda: Label matching deep domain adaptation. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 6043–6054. PMLR, 18–24 Jul 2021. 2

[32] Sangyun Lee, Zinan Lin, and Giulia Fanti. Improving the training of rectified flows. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 3

[33] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2022. 2, 3

[34] Qiang Liu. Rectified flow: A marginal preserving approach to optimal transport. *arXiv preprint arXiv:2209.14577*, 2022. 3, 4, 12

[35] Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, and qiang liu. Instaflow: One step is enough for high-quality diffusion-based text-to-image generation. In *The Twelfth International Conference on Learning Representations*, 2024. 3

[36] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: a fast ode solver for diffusion probabilistic model sampling in around 10 steps. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. 12

[37] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-solver++: Fast solver for guided sampling of diffusion probabilistic models, 2023. 12

[38] Björn Lütjens, Brandon Leshchinskiy, Océane Boulais, Farrukh Chishtie, Natalia Díaz-Rodríguez, Margaux Masson-Forsythe, Ana Mata-Payerro, Christian Requena-Mesa, Aruna Sankaranarayanan, Aaron Piña, Yarin Gal, Chedy Raïssi, Alexander Lavin, and Dava Newman. Generating physically-consistent satellite imagery for climate visualizations. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–1, 2024. 3

[39] Nanye Ma, Mark Goldstein, Michael S Albergo, Nicholas M Boffi, Eric Vanden-Eijnden, and Saining Xie. Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers. *arXiv preprint arXiv:2401.08740*, 2024. 3

[40] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802, 2017. 15

[41] Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 4296–4304, 2024. 3

[42] Zak Murez, Soheil Kolouri, David Kriegman, Ravi Ramamoorthi, and Kyungnam Kim. Image to Image Translation for Domain Adaptation. pages 4500–4509. IEEE Computer Society. 2

[43] Stefano Peluchetti. Non-denoising forward-time diffusions. *arXiv preprint arXiv:2312.14589*, 2023. 2, 3

[44] Fabio Pizzati, Raoul de Charette, Michela Zaccaria, and Pietro Cerri. Domain bridge for unpaired image-to-image translation and unsupervised domain adaptation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2990–2998, 2020. 2

[45] Yuanyuan Qing, Jiang Zhu, Hongchuan Feng, Weixian Liu, and Bihan Wen. Two-way generation of high-resolution eo and sar images via dual distortion-adaptive gans. *Remote Sensing*, 15(7), 2023. 3

[46] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 4

[47] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain Adaptation for Image Dehazing. pages 2808–2817. 2

[48] Jacob Shermeyer, Daniel Hogan, Jason Brown, Adam Van Etten, Nicholas Weir, Fabio Pacifici, Ronny Hansch, Alexei Bastidas, Scott Soenen, Todd Bacastow, et al. Spacenet 6: Multi-sensor all weather mapping dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 196–197, 2020. 5, 14

[49] Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[50] Vignesh Ram Somnath, Matteo Pariset, Ya-Ping Hsieh, Maria Rodriguez Martinez, Andreas Krause, and Charlotte Bunne. Aligned diffusion schrödinger bridges. In *The 39th Conference on Uncertainty in Artificial Intelligence*, 2023. 5

[51] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv:2010.02502*, October 2020. 3

[52] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. 3, 6

[53] Reihaneh Teimouri, Marta Kersten-Oertel, and Yiming Xiao. CT-Based Brain Ventricle Segmentation via Diffusion Schrödinger Bridge without target domain ground truths. In Marius George Linguraru, Qi Dou, Aasa Feragen, Stamatia Giannarou, Ben Glocker, Karim Lekadir, and Julia A. Schnabel, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, pages 135–144. Springer Nature Switzerland. 2

[54] Alexander Tong, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Kilian FATRAS, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023. 4, 8, 12

[55] Devis Tuia, Claudio Persello, and Lorenzo Bruzzone. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):41–57, 2016. 2

[56] Florence Tupin, J. Inglada, and J. M. Nicolas. *Remote Sensing Imagery*. ISTE - Wiley, 2014. 2

[57] Jiangshan Wang, Yue Ma, Jiayi Guo, Yicheng Xiao, Gao Huang, and Xiu Li. COVE: Unleashing the diffusion feature correspondence for consistent video editing. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 3

[58] Jinyu Wang, Haitao Yang, Yu He, Fengjie Zheng, Zhengjun Liu, and Hang Chen. An unpaired sar-to-optical image translation method based on schrödinger bridge network and multi-scale feature fusion. *Scientific Reports*, 14, 11 2024. 2

[59] Lei Wang, Xin Xu, Yue Yu, Rui Yang, Rong Gui, Zhaozhuo Xu, and Fangling Pu. Sar-to-optical image translation using supervised cycle-consistent adversarial networks. *IEEE Access*, 7:129136–129149, 2019. 2

[60] Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Tally Portnoi, Itamar Chinn, Jacob Silterra, Tommi Jaakkola, et al. Boltz-1: Democratizing biomolecular interaction modeling. biorxiv. 2024. 2, 3

[61] Sidi Wu, Chenn Yizi, Samuel Mermet, Lorenz Hurni, Konrad Schindler, Nicolas Gonthier, and Loic Landrieu. StegoGAN: Leveraging steganography for non-bijective image-to-image translation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2, 3, 6, 7, 18

[62] Haifeng Xia, Handong Zhao, and Zhengming Ding. Adaptive adversarial network for source-free domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9010–9019, October 2021. 2

[63] Xinpeng Xie, Jiawei Chen, Yuexiang Li, Linlin Shen, Kai Ma, and Yefeng Zheng. Self-Supervised CycleGAN for Object-Preserving Image-to-Image Domain Adaptation. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 498–513. Springer International Publishing. 2

[64] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4084–4094, 2020. 2

[65] Hannuo Zhang, Huihui Li, Jiarui Lin, Yujie Zhang, Jianghua Fan, and Hang Liu. Seg-cyclegan : Sar-to-optical image translation guided by a downstream task, 08 2024. 3

[66] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 7

[67] Sicheng Zhao, Bo Li, Xiangyu Yue, Yang Gu, Pengfei Xu, Runbo Hu, Hua Chai, and Kurt Keutzer. Multi-source domain adaptation for semantic segmentation. *Advances in neural information processing systems*, 32, 2019. 2

[68] Linqi Zhou, Aaron Lou, Samar Khanna, and Stefano Ermon. Denoising diffusion bridge models. *arXiv preprint arXiv:2309.16948*, 2023. 2, 3

[69] 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. 2, 3, 6, 7, 18

A. Details and ablations

A.1. Couplings

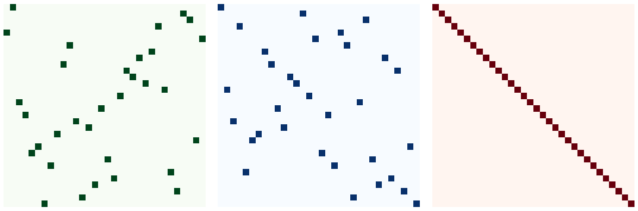

Figure 6. Comparison between the pairing matrices generated with the different couplings for a batch on SpaceNet 8, from left to right: independent coupling $p(x_0)p(x_1)$, OT-coupling $\pi(x_0, x_1)$, data-dependent coupling $p(x_1 \mid x_0)p(x_0)$.

The choice of the coupling has been of prime importance to improve generation capabilities for flow matching models [5, 34, 54]. Figure 6 shows the pairing matrices M obtained with each coupling *i.e.* $M_{ij} = 1$ iff latents x_0^i and x_1^j are paired. The training batches are built by stacking strongly or weakly aligned x_0 and x_1 images in order. Because the data-dependent coupling matches x_0^i with x_1^i, its pairing matrix is diagonal. We observe that the optimal transport-based coupling (left) is poorly aligned with the data-dependent coupling (center), suggesting that semantic information matching cannot be solely recovered through optimal transport.

In addition, we provide visual ablation results in Fig. 7, which illustrate the necessity to use data-dependent couplings to train FlowEO.

A.2. VAE finetuning

A.2.1 Implementation details

We use a distilled version of the VAE from StableDiffusion 3 [13] to speed up training and inference. The encoder is trained to reconstruct the latents produced by the original encoder to preserve the latent space structure of the full model. As shown in the main paper, our experiments show that the reconstructions $\mathcal{D}(\mathcal{E}(x))$ of Sentinel-2 images are of poor quality because the range and distribution of multispectral images deviates from the pretraining dataset used for Stable Diffusion. For the reBEN and Sen1Floods11 datasets that use Sentinel-2 as source data, we finetune the decoder of the distilled VAE on each dataset for 5000 iterations with a learning rate of 10^{-4}, 250 warmup steps, and cosine decay learning rate scheduler. The decoder remains frozen when training the flow. The remaining datasets use the original pretrained decoder.

		SpaceNet 8 Post-flood → Pre-flood			
		mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
RGB	Base	**44.65**	**48.79**	**60.32**	**45.50**
	Finetuned	44.33	48.71	81.75	51.64
		SpaceNet 6 SAR → RGB			
		mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
RGB	Base	**65.07**	**72.33**	**94.02**	**39.96**
	Finetuned	64.63	72.17	111.66	42.77
		Sen1Floods11 SAR → Optical			
		mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
S2	Base	51.45	57.63	24.33	29.22
	Finetuned	**54.92**	**69.04**	**12.96**	**29.21**
		ReBEN SAR → Optical			
		AP^{M}	$\mathrm{F1}^{\mathrm{M}}$	FID↓	LPIPS ↓
S2	Base	27.02	15.97	168.85	16.88
	Finetuned	**32.14**	**25.72**	**75.80**	**15.51**

Table 5. Impact of VAE fine-tuning on domain adaptation performance and transferred image quality. Fine-tuning is beneficial for Sentinel-2 imagery but not for classical RGB images.

A.2.2 Impact of VAE fine-tuning

Reconstruction SD VAE reconstruction error is higher on non-RGB imagery, VAE finetuning improves reconstruction RMSEs 237.04 *vs.* 357.91 and 0.058 *vs.* 3.760 on respectively reBEN S2 and SpaceNet-6 SAR. This is unnecessary for RGB and can be slightly detrimental. S2 images are normalized from [0;10000] to [-1;+1] via band-wise min-max normalization.

Generation We report in Tab. 5 metrics for flow models trained with and without a fine-tuned VAE decoder. We observe that fine-tuning the VAE decoder prior to learning the flow matching has a positive impact when the final domain differs from usual RGB imagery. Indeed, fine-tuning the decoder is beneficial for Sen11Floods11 and ReBEN, for which the images are transferred in the Sentinel-2 color bands. Because Sentinel-2 imagery uses the [0, 10 000] range instead of the usual [0, 255], the pretrained decoder is less effective, which reflects in image quality. Yet, on SpaceNet 6 and 8, which use both standard RGB images, there is no advantage of fine-tuning the decoder. It is even detrimental, as we hypothesize that the decoder overfits to the small training set, compared to the original dataset used for StableDiffusion.

A.3. Sampling schedule

The choice of time discretization and inference-time sampling strategy plays a crucial role in improving the performance of diffusion models [24, 36, 37]. Recently, [26] introduced a sigmoid time-scheduler tailored for flow matching models (see Eq. (4)). This scheduler is parametrized by κ

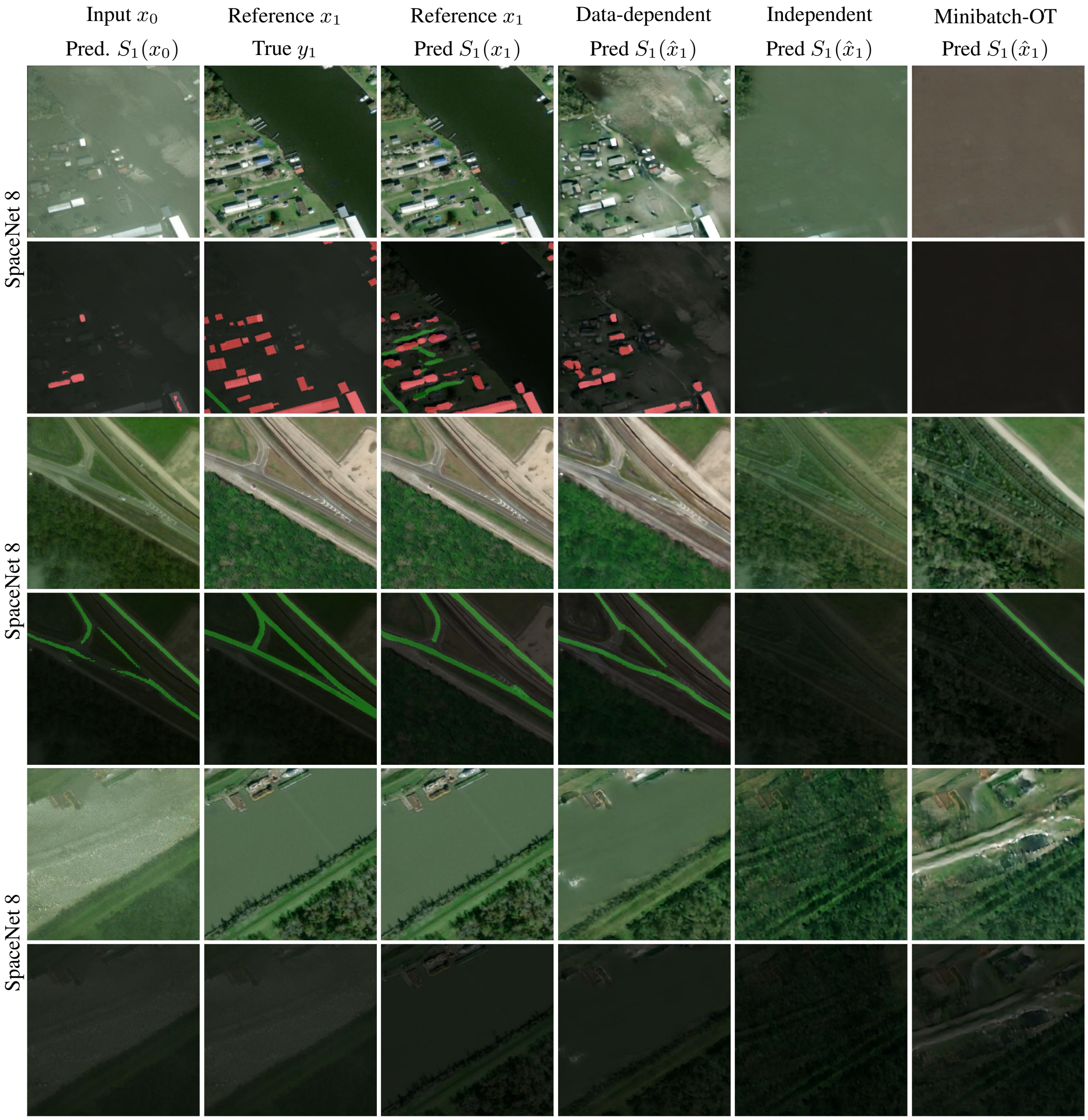

Figure 7. Impact of the training coupling $p(x_0, x_1)$ on preserving semantic information during image translation. FlowEO employs *data-dependent coupling* $p(x_1|x_0)p(x_0)$, which outperforms both *minibatch-OT* coupling $\pi(x_0, x_1)$ and *independent* coupling $p(x_0, x_1)$.

which controls the distribution of sampling steps across time. Higher values of κ concentrate computational effort near the endpoints ($t \approx 0$ and $t \approx 1$), whereas $\kappa \to 0$ corresponds to the linear time schedule (see Figure 8).

$$\left\{ t_i = \frac{\text{sig}\Big(\kappa\Big(\frac{i}{N} - 0.5\Big)\Big) - \text{sig}\Big(-\frac{\kappa}{2}\Big)}{\text{sig}\Big(\frac{\kappa}{2}\Big) - \text{sig}\Big(-\frac{\kappa}{2}\Big)} : i = 0, ..., N \right\} \tag{4}$$

Despite originally designed for generative modeling with flow matching models, *i.e*. mapping a Gaussian prior distribution to the data distribution, this time scheduling is well-motivated in our setting where increasing the number of sampling steps near the data distributions p_0 and p_1 is beneficial. Tab. 6 presents a comparison between sigmoid and linear time discretization, demonstrating consistent improvements in segmentation metrics across all datasets and for all numbers of inference steps. Image quality metrics exhibit only marginal improvements and, in some cases—such as on

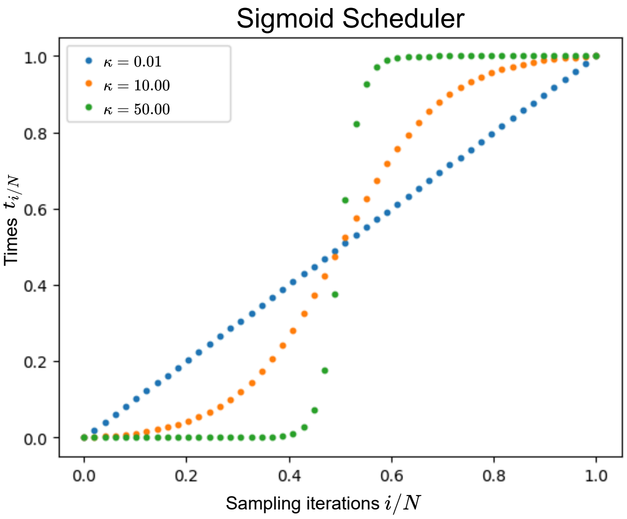

Figure 8. Sigmoid time discretization, allocating more sampling steps near the endpoints ($t \approx 0$ and $t \approx 1$).

the Sen1Floods11 dataset—even show slight deterioration. Nevertheless, the performance gains in segmentation metrics from using a sigmoid rather than a linear schedule diminish as the number of inference steps increases. Also observe that more sampling steps might not be beneficial for domain adaptation. On the two datasets used for validation, 25 sampling steps tends to perform on-par or better than 50 and 100 steps. We attribute this to slightly better preservations of semantics with a low number of steps, which reduce small but accumulating errors in the Euler integration. In practice, we set $\kappa = 10$ and use 50 sampling steps for all experiments.

A.4. Compute time and memory footprint

We report memory and times in Tab. 7. We agree that inference time is an issue, as flow matching is slower than GANs. This is why we use a lighter distilled version of SD3's VAE (0.24s *vs.* 2.11s for encoding-decoding). Despite relying on ODE integration, FlowEO transfers a batch of 256 images in 7.79s on a single A100 with 50 NFE (≈30 ms/image).

B. Dataset details

For all datasets, we define three distinct splits: train, validation, and test. The training set is used to train both domain adaptation methods and predictive models. To reflect real-world scenarios – where retraining a generative model on new data batches is impractical – we restrict the training of image translation models to the training set. The validation set is used for hyperparameter tuning and model selection based on performance metrics, while the final reported metrics are computed on the test set.

Sen1Floods11 SAR → Optical				
	mIoU ↑	mAcc ↑	FID↓	LPIPS ↓
25 Sampling Steps				
Linear	54.60	72.22	**13.99**	**28.91**
Sigmoid $\kappa = 10$	**55.05**	**72.50**	14.38	29.02
50 Sampling Steps				
Linear	54.26	71.79	**13.06**	**28.86**
Sigmoid $\kappa = 10$	**54.46**	**71.94**	13.46	28.90
100 Sampling Steps				
Linear	54.10	71.59	**12.87**	**28.85**
Sigmoid $\kappa = 10$	**54.19**	**71.66**	12.95	28.86
SpaceNet 6 SAR → RGB				
	mIoU ↑	mAcc ↑	FID↓	LPIPS ↓
25 Sampling Steps				
Linear	64.23	71.68	117.30	**42.78**
Sigmoid $\kappa = 10$	**64.46**	**71.93**	**113.64**	42.96
50 Sampling Steps				
Linear	63.98	71.46	119.68	**42.89**
Sigmoid $\kappa = 10$	**64.07**	**71.57**	**118.06**	42.98
100 Sampling Steps				
Linear	63.79	71.28	121.28	**42.98**
Sigmoid $\kappa = 10$	**63.83**	**71.34**	**120.38**	43.03

Table 6. Sigmoid schedule vs linear schedule (preliminary results, FlowEO performances with only 100 000 training steps).

Model	Train Mem. (GB)	Inference Mem. (GB)	Inference Time (s)
Pix2pix	29.44 (64)	14.59 (256)	0.09 (256)
CycleGAN	30.75GB (12)	14.56 (256)	0.06 (256)
StegoGAN	31.67GB (8)	24.05 (256)	1.94 (256)
UNSB	34.00GB (12)	0.398 (1)	0.11 (1)
FlowEO	30.42GB (256)	22.21 (256)	7.79 (256)

Table 7. Memory footprints and inference times on A100 40GB. Batch sizes are indicated in brackets: measure (batch size). UNSB official implementation only supports inference batch size of 1.

SpaceNet 6 [48] is a multimodal dataset including optical imagery (RGB bands) and SAR data (we select VV/HH/VH polarizations) at a resolution of 2 m/px. From initial tiles, we crop 256 × 256 images and apply an overlap of 50% to create the training set. The segmentation masks have two different classes: background and building. We use three different splits: training (≈ 50000 samples), validation (≈ 1800), and test (≈ 1800) sets. For the optical data, we use bands [4, 3, 2], while for the SAR data, we utilize VV, HH, and VH polarizations.

SpaceNet 8 [19] is a segmentation dataset that contains pre and post-flood RGB images from Maxar for two different locations: Germany and Louisiana. The segmentation masks include three different classes: background, building, and roads. Original tiles are downsampled with a factor 2 and then cropped 256 × 256 images with an overlap of 70% to produce the training data. The final numbers of samples of

Datasets	SpaceNet 8				SpaceNet 8 Germany				SpaceNet 8 Louisiana			
	Post-flood → Pre-flood				Post-flood → Pre-flood				Post-flood → Pre-flood			
	mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓	mIoU ↑	Acc ↑	FID ↓	LPIPS ↓	mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
No adaptation	40.05	42.40	75.62	63.66	37.09	39.08	89.54	63.27	36.51	38.85	96.60	63.80
Upper bound	63.10	72.09	00.00	00.00	55.27	66.77	00.00	00.00	66.91	75.97	00.00	00.00
CycleGAN data-dependent	40.70	43.35	**54.31**	55.70	39.35	41.79	**62.80**	59.46	42.39	45.14	**52.80**	52.92
CycleGAN independent	40.64	43.26	**52.85**	55.17	40.34	43.54	88.04	62.01	41.94	44.80	58.70	53.82
FlowEO	**44.65**	**48.79**	60.32	**45.50**	**41.27**	**45.29**	82.74	**53.63**	**47.19**	**52.30**	59.65	**41.95**

Table 8. Quantitative results on domain adaptation for weakly aligned datasets. We report both segmentation (mIoU, mAcc) and image quality metrics (FID, LPIPS) for SpaceNet 8 and its geographic subsets. CycleGAN benefits from the data-dependent coupling on SpaceNet 8 and Louisiana, despite being suited for unaligned data-translation.

Datasets	Sen1Floods1				SpaceNet 6				ReBEN					
	SAR → Optical				SAR → RGB				SAR → Optical					
	mIoU	mAcc	FID	LPIPS	mIoU	mAcc	FID	LPIPS	AP^{μ}	AP^{M}	F1^{μ}	F1^{M}	FID	LPIPS
No adaptation	06.22	49.72	297.22	84.84	31.94	41.01	275.05	79.48	17.46	17.43	02.31	01.31	339.36	85.99
Upper bound	55.14	71.28	00.00	00.00	84.94	90.74	00.00	00.00	79.26	65.28	74.28	62.84	00.00	00.00
CycleGAN data-dependent	42.12	48.47	20.97	36.35	50.01	55.85	132.75	50.72	26.09	19.79	26.93	15.75	81.54	19.67
CycleGAN independent	44.23	51.04	393.88	97.35	51.02	57.51	110.90	49.89	24.01	19.88	28.13	19.77	78.63	24.08
FlowEO	**54.92**	**69.04**	**12.96**	**29.21**	**65.07**	**72.33**	94.02	**39.96**	37.16	**32.14**	36.04	25.72	75.80	**15.51**

Table 9. Quantitative results on domain adaptation for strongly aligned datasets. We report both segmentation (mIoU, mAcc) or classification (AP/F1) and image quality metrics (FID, LPIPS). On SAR-to-optical translation datasets, CycleGAN trained with independent coupling (i.e., unaligned training) yields marginally superior performance on downstream task metrics compared to data-dependent coupling. Nonetheless, the coupling strategy does not alter its relative ranking with respect to FlowEO.

each split are 5688/88/88 for Germany and 17173/244/244 for Louisiana. The full SpaceNet 8 dataset is obtained by merging the two subsets for each split.

Sen1Floods11 [4] provides SAR data (Sentinel-1) and optical imagery (Sentinel-2) alongside water/non-water pixel-level annotations at a resolution of 10 m/px. Random cropping of 256 × 256 images is computed for training images, and deterministic cropping without overlap is provided for validation and test sets. It results in a total of 64 512 patches for training. To match the number of SAR bands with the optical ones we duplicate the VH band, and then we use bands $[4, 3, 2]$ for optical data and VV/HH/VH polarization for SAR data.

BigEarthNet2 (reBEN) [9] is a multi-sensor dataset including Sentinel-1 and Sentinel-2 imagery. We used 237 871 training patches with the multiclass annotations for both classification and domain-adaptation models training, 122 342 for validation, and 119 825 for testing following the original paper's splits. To match the number of SAR bands with the optical ones we duplicate the VH band, and then we use bands $[4, 3, 2]$ for optical data and VV/HH/VH polarization for SAR data. We resize the original 120 × 120 patches with bilinear interpolation to match the 256 × 256 used for the other datasets.

C. Hyperparameters

Pix2Pix We train two Pix2Pix models, one translating images from p_0 to p_1 and vice versa. We use the reference PyTorch implementation available [1] and train the models with the data-dependent coupling. We train the models with a batch size of 1 for 200 000 training steps with a learning rate of 2×10^{-4} and learning rate linear decay. Following the reference implementation, we use the *LSGAN* [40] adversarial loss. We deviate from the default hyperparameters for λ_{L1}, which we decrease from 100 to 10 to fix blurry image generation issues on ours datasets. The generator is a 9-blocks ResNet and we use the PatchGAN discriminator [20] with instance normalization.

CycleGAN The implementation of CycleGAN follows the same hyperparameters set as the Pix2Pix mentioned above. We train the models with a batch size of 1 for 200 000 training steps with a learning rate of 2×10^{-4} and learning rate linear decay. We keep $\lambda_{\text{L1}} = 100$ since it does not negatively impact the training or the generated images' quality. We used the same network architectures as for Pix2Pix.

StegoGAN While the StegoGAN models use two generators, translating respectively from domain $\mathcal{X}_0$ to $\mathcal{X}_1$ and vice versa, the training process is asymmetrical. Thus, we trained two different models for each dataset, using the of-

[1] https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

ficial implementation[2]. We use *LSGAN* adversarial loss, instance normalization, and train the model for 200 000 iterations with a learning rate of 2×10^{-4}. We select the set of loss weightings used for the GoogleMismatch dataset in the original paper: $\lambda_A = 10$, $\lambda_B = 10$,$\lambda_A = 10$, $\lambda_{\text{id}} = 0.5$, $\lambda_{\text{cycle}} = 0.5$ and $\lambda_{\text{reg}} = 0.3$ for the mask regularization loss. Note that this last value is similar for all remote sensing datasets used in StegoGAN: $\lambda_{\text{cycle}} = 0.5$ for GoogleMismatch and $\lambda_{\text{cycle}} = 0.3$ for PlanIGN. The generator is a 9-blocks-Resnet and we use the PatchGAN discriminator [20] with instance normalization.

UNSB Schrödinger bridges map two arbitrary distributions with forward and backward stochastic processes. Nevertheless, UNSB leverages an adversarial loss on p_1 making the training asymmetrical. Thus we train two different models, translating respectively from domain $\mathcal{X}_0$ to $\mathcal{X}_1$ and vice versa. We use the official implementation [3] and train the models for 200 000 iterations with a learning rate of 2×10^{-4}. We use the proposed set of hyperparameters: $\lambda_{\text{GAN}} = 1$, $\lambda_{\text{NCE}} = 1$, $\lambda_{\text{SB}} = 1$. We use the same architectures as the other methods, namely 9-blocks-Resnet and PatchGAN discriminator with instance normalization. We use 5 sampling steps at inference, following original paper guidelines.

Diffusion Bridges Diffusion bridges establish mappings between arbitrary distributions via forward and backward stochastic processes. We adopt the formulation of [5] and train the models for 200 000 iterations using an x_1-prediction objective, with a batch size of 32 and a learning rate of 2×10^{-4}. The UNet backbone follows the same design as FlowEO, but is adapted to operate directly on image inputs rather than latent representations. Inference is performed with 50 sampling steps, consistent with FlowEO.

D. CycleGAN with unaligned training

CycleGAN is a data-to-data translation framework originally designed to handle unaligned datasets through its cyclical loss. However, in the context of pre- and post-disaster datasets, we observe that CycleGAN benefits from the availability of co-registered pairs (data-dependent coupling improves segmentation metrics) (Table 8). For SAR-to-optical translation, the use of unpaired datasets can offer certain advantages, though the performance gains are marginal and do not alter its relative ranking compared to our method (Table 9).

E. Additional quantitative results on reBEN

We include in Table 10 a detailed comparison of Pix2pix and FlowEO on the ReBEN SAR-to-Optical domain adaptation dataset. It reveals that Pix2pix exhibits a pronounced bias toward forest classes (*Coniferous forest* and *Mixed forest* classes), which are disproportionately represented relative to other categories. This class imbalance inflates micro-averaged metrics, thereby explaining the discrepancy in ranking between FlowEO and Pix2pix under micro- versus macro-averaging.

F. Additional qualitative results

F.1. Qualitative classification results on reBEN

We provide here qualitative domain adaptation results for reBEN, with transferred images for baselines and FlowEO and predicted labels shown in Figure 9. As for the segmentation tasks, this underlines both the visual quality of the generated images by FlowEO and the accuracy of the predictions by the pre-trained classification model on the adapted images. In addition to the generated optical images, we show the top-3 predicted classes, *i.e*. the 3 classes with the highest probabilities predicted by the classification model C_1^*.

F.2. Additional image generation results

We provide in Figure 10 additional image generation results for a more exhaustive assessment of our image translation approach. We can observe that FlowEO tends to better capture the color range of the reference images, avoid hallucinations, and better reconstruct the scene geometry. In particular, note that FlowEO is robust to changes between the source and target images, *e.g*. clouds and boats that have moved. Interestingly, this shows the potential of flow matching for inverse problems in Earth observation, such as cloud removal.

[2] https://github.com/sian-wusidi/StegoGAN

[3] https://github.com/cyclomon/UNSB

	Pix2Pix	FlowEO	Pix2Pix	FlowEO	#test samples	Proportions
	AP		F1			
Macro metric M	27.88	32.14	25.79	25.72		
Micro metric μ	41.09	37.16	43.93	36.04		
Industrial or commercial units	13.79	25.43	22.47	34.09	2018	0.0058
Arable land	64.25	73.77	62.05	69.89	50052	0.1446
Permanent crops	6.69	11.42	05.02	12.19	5710	0.0165
Pastures	35.01	42.38	22.84	36.22	26722	0.0772
Complex cultivation patterns	24.70	30.58	08.06	36.28	22078	0.0638
Land principally occupied by agriculture, with significant areas of natural vegetation	31.46	35.99	33.35	30.75	29846	0.0862
Agro-forestry areas	22.62	44.25	05.55	18.56	9942	0.0287
Broad-leaved forest	32.76	41.63	22.76	20.68	36377	0.1051
Coniferous forest	54.65	54.95	57.82	30.66	39043	0.1128
Mixed forest	52.64	49.57	58.93	29.07	44284	0.1280
Natural grassland and sparsely vegetated areas	01.57	02.30	00.08	02.32	2211	0.0064
Moors, heathland and sclerophyllous vegetation	03.74	05.31	02.39	02.70	3759	0.0109
Transitional woodland, shrub	43.34	44.00	45.68	29.54	40523	0.1171
Beaches, dunes, sands	00.92	00.75	03.88	02.29	152	0.0004
Inland wetlands	05.26	04.98	08.96	09.24	4519	0.0131
Coastal wetlands	00.09	00.09	00.28	00.11	117	0.0003
Inland waters	33.79	34.17	26.53	25.78	16846	0.0487
Marine waters	69.16	68.78	66.72	55.48	11854	0.0343

Table 10. Performance comparison of Pix2pix and FlowEO on the ReBEN SAR-to-Optical domain adaptation dataset. Pix2pix shows a strong bias toward forest classes, which are overrepresented relative to other categories. The high performance on these dominant classes inflates micro-averaged metrics, accounting for the difference in ranking between FlowEO and Pix2pix under micro- versus macro-averaging.

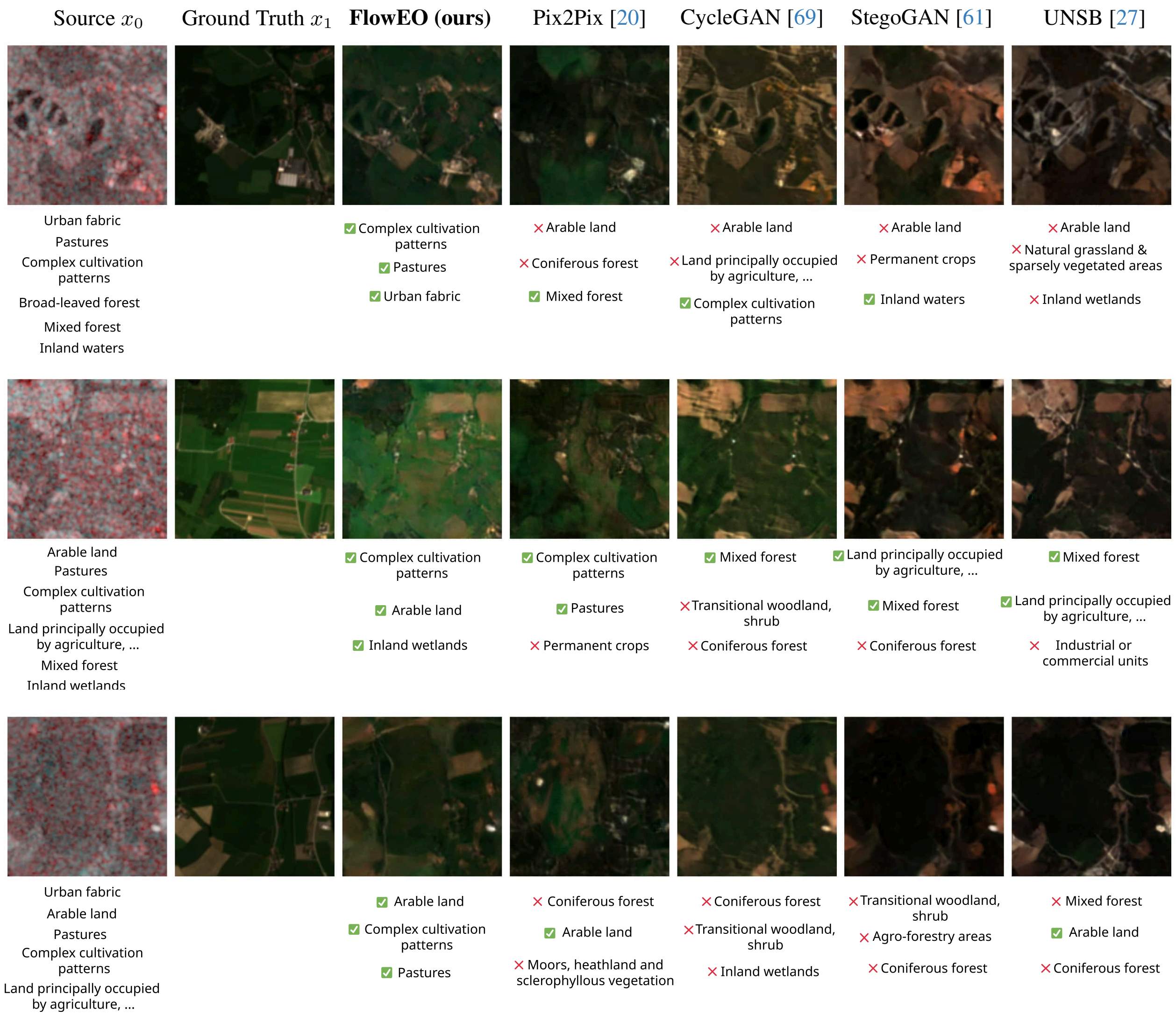

Figure 9. Qualitative comparison of domain adaptation methods on the reBEN dataset, for multiclass classification. The first column represents the source domain image x_0, the second depicts the weakly or strongly aligned x_1, and the others display the images generated by the different methods. Below each image generated, we provide the corresponding top-3 predicted classes by the classification model C_1. For the reference image, we display all the class labels. FlowEO outperforms other methods in both class preservation and image quality.

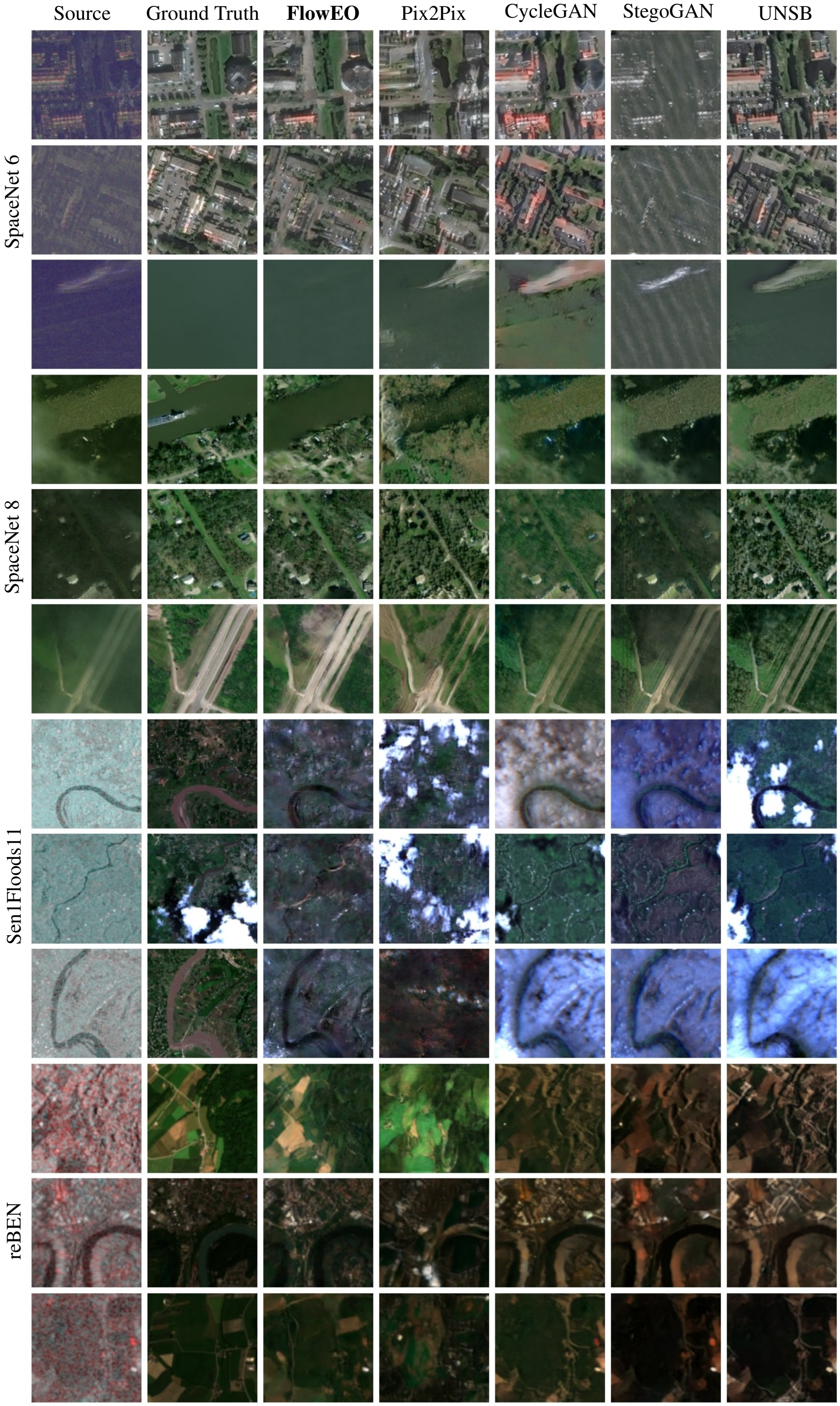

Figure 10. FlowEO generates the highest-quality images while maintaining semantic consistency during the transfer process. In the third row, we observe that our method demonstrates greater robustness to the geometric artifacts present in SAR imagery. Additionally, we note that it successfully learns to map flood-disturbed water states to a more natural appearance (fourth row).