FlowEO: Generative Unsupervised Domain Adaptation for Earth Observation

Georges Le Bellier [1]
[1] Cnam, CEDRIC, EA4629
F-75141 Paris, France
georges.le-bellier@lecnam.net

Nicolas Audebert [1,2]
[2] Univ. Gustave Eiffel, ENSG, IGN, LASTIG
F-94160 Saint-Mandé, France
nicolas.audebert@ign.fr

Abstract

The increasing availability of Earth observation data offers unprecedented opportunities for large-scale environmental monitoring and analysis. However, these datasets are inherently heterogeneous, stemming from diverse sensors, geographical regions, acquisition times, and atmospheric conditions. Distribution shifts between training and deployment domains severely limit the generalization of pretrained remote sensing models, making unsupervised domain adaptation (UDA) crucial for real-world applications. We introduce FlowEO, a novel framework that leverages generative models for image-space UDA in Earth observation. We leverage flow matching to learn a semantically preserving mapping that transports from the source to the target image distribution. This allows us to tackle challenging domain adaptation configurations for classification and semantic segmentation of Earth observation images. We conduct extensive experiments across four datasets covering adaptation scenarios such as SAR to optical translation and temporal and semantic shifts caused by natural disasters. Experimental results demonstrate that FlowEO outperforms existing image translation approaches for domain adaptation while achieving on-par or better perceptual image quality, highlighting the potential of flow-matching-based UDA for remote sensing.

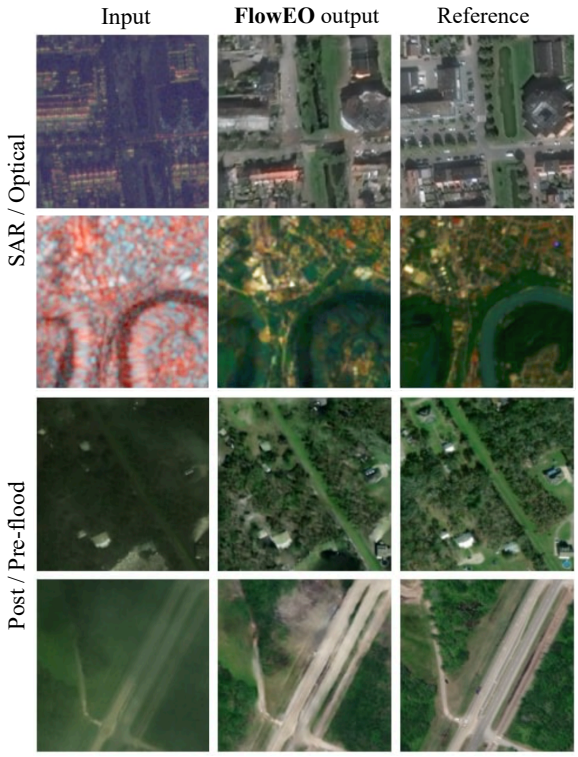

Figure 1. **FlowEO** generates realistic and semantically consistent outputs on various challenging image translation tasks, such as pre- to post-disaster domain adaptation and SAR-to-Optical translation.

1. Introduction

Large amounts of remote sensing images are collected at high frequency to analyze and model the complexity of physical phenomena on Earth. The diversity of the data acquired calls into question the use of pretrained models to process them. Indeed, the phenomena studied on the Earth's surface are non-stationary and subject to great variability due to seasonal variations, human-made changes, and extreme events such as wildfires and floods. This causes drifts in the data distribution, compromising model performance at inference [25]. In addition, sensors with complementary characteristics are used to capture multiple views of the same area and overcome sensor limitations, *e.g.* ground occultation by clouds for optical sensors can be alleviated using radar.

Emergency management of natural disasters requires rapid analysis of the ground-level situation to plan rescue operations and assess environmental consequences. However, the domain shift between post-disaster and ordinary satellite images degrades the performance of off-the-shelf deep models. The urgency of such events makes it impossible to annotate a dataset for supervised training in disaster-affected areas. Domain adaptation is therefore a promising solution to speed up image analysis for disaster management.

Similarly, robust remote sensing pipelines leverage the strengths of all available Earth observation sensors. For ex-

ample, Synthetic Aperture Radar (SAR) provides all-weather day-and-night imaging capabilities, as its wavelength allows it to penetrate clouds and operate independently of illumination conditions. However, due to their speckle noise and sensitivity to terrain geometry, interpreting SAR images is harder for humans than optical images [56]. Therefore, cross-sensor domain adaptation has been well-investigated in Earth observation. SAR-to-Optical translation (S2O) in particular can provide human-interpretable optical images in contexts where only SAR imagery is available [30, 58, 59], *e.g.* to fill in missing optical due to cloud cover. This provides higher frequency images by leveraging co-located multimodal SAR/optical acquisitions. In turn, this enables disaster monitoring and environmental surveillance in scenarios where acquiring cloud-free optical data is challenging.

Major efforts have been made in recent years to overcome distribution drift through domain adaptation [25,44,55]. Due to the low availability of labeled satellite image datasets, *unsupervised* domain adaptation methods have been preferred as they only require labels in the source domain. Unsupervised domain adaptation is mainly studied inside the feature space of a pretrained model [10, 62]. Because of the lower dimensionality of the latent space, this favours classification tasks [16, 31], although some approaches also have been proposed for dense tasks such as segmentation [8, 15, 64]. To overcome this limitation, domain adaptation can be applied in image-space [67]. It facilitates the transfer interpretation and improves explainability while disentangling transfer and downstream tasks. Such image translation approaches are orthogonal to future improvements in classifiers and can be used with any inference model without retraining. To this end, we employ flow matching models [1, 33, 43], a new family of models that have demonstrated high-quality generation across various modalities [13, 60].

FlowEO. We propose FlowEO, a new model that leverages flow matching models for unsupervised domain adaptation in Earth Observation. We introduce a novel domain adaptation method in pixel-space, enabling visual interpretation, and test it extensively on four datasets covering classification and segmentation tasks, demonstrating its effectiveness for dense downstream tasks in challenging scenarios of post-disaster domain adaptation and sensor translation. In summary:

1. We introduce FlowEO, a new generative UDA method, downstream-task-agnostic that does not require modification or retraining of downstream predictive models.

2. We are the first to leverage latent flow matching for data-to-data translation, on multiple remote sensing modalities, including SAR, low-resolution, and high-resolution optical data.

3. We introduce an application-driven evaluation protocol, going beyond standard image generation metrics to assess the impact of UDA on real-world Earth observation tasks: semantic segmentation and classification.

2. Related Work

2.1. Unsupervised domain adaptation

Consider two distinct domains, represented by two datasets D_0 and D_1. Suppose that one contains annotations, *i.e.* $D_1 = \{\mathcal{X}_1, \mathcal{Y}_1\}$. This is the *source* domain, on which a predictive model S_1 has been trained, *e.g.* for segmentation, classification, regression, etc. Conversely, let $D_0 = \{\mathcal{X}_0, \emptyset\}$ be the unlabeled *target* domain, on which we would like to infer new predictions. The absence of annotations on D_0 prevents us from training a predictive model on it. Instead, we intend to use the existing model S_1 for the new D_0 data. However, the underlying differences between the two domains will result in a drop in its performance if applied directly to the new domain. Its generalization capabilities do not allow direct transfer of segmentation scores. Domain adaptation aims to extend a model's performance beyond its training domain by means of an adaptation procedure.

Domain adaptation techniques are split into two broad families. First, domain adaptation can be applied post-hoc on an existing predictive model. These approaches aim to align the features obtained from the predictive model, *e.g.* with optimal transport [10, 16], or fine-tuning/adapting the weights of the model to the new domain [6, 62]. However, every downstream model needs to be adapted, which can be costly and constrains usage of "off-the-shelf" models. Second, adaptation can take place directly in the data space, *i.e.* image space in our case. Instead of adapting the model to the target domain, the target data is altered to match the source domain. This approach, called image-to-image translation for domain adaptation [42], leverages conditional generative models derived from style transfer [23].

2.2. Image translation for domain adaptation

Image translation builds upon the seminal work of Pix2Pix [20], that trains an image-to-image model on paired datasets using a combination of supervised regression loss and an adversarial loss using a patch-wise GAN. It has been extended to the unpaired setting into CycleGAN [69], leveraging cycle consistency by training two GANs in symmetry. These models have been used for domain adaptation in multiple settings, including dehazing [47], tactile perception [22], and semantic segmentation [63]. More recent models include StegoGAN [61] that explicitly deals with features that are impossible to match between the two domains, and more recent generative model classes, *e.g.* diffusion models and Schrödinger bridges [11, 49, 68]. The Unpaired Neural Schrödinger Bridge [27], for example, has found success for domain adaptation of medical CT scans [53].

Image translation for Earth Observation Such approaches

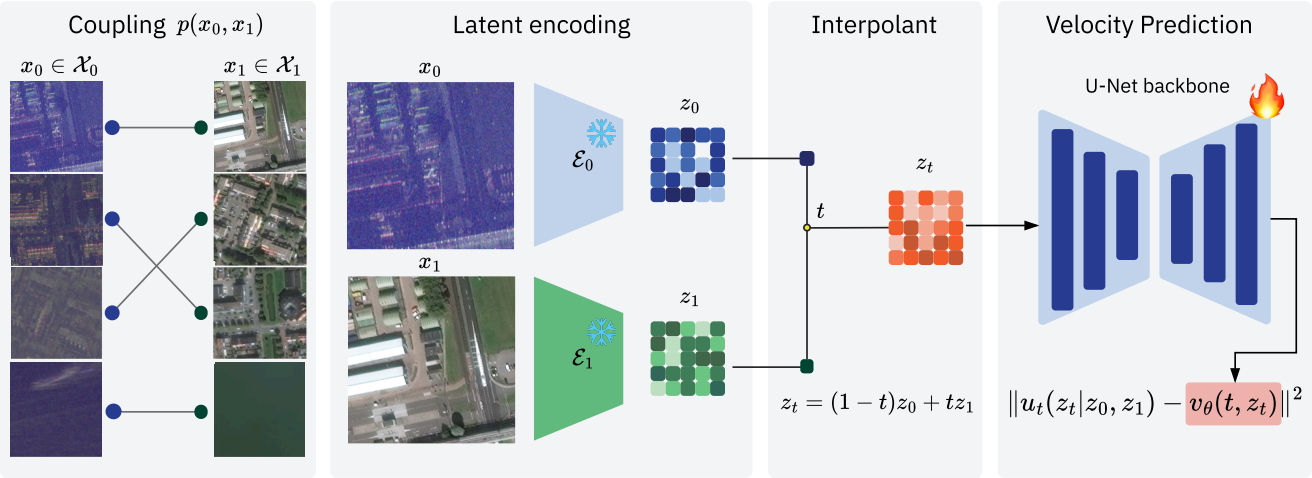

Figure 2. FlowEO learns a latent flow between the source and target distributions in four stages: 1) the training image pairs are sampled from the coupling $p(x_0, x_1)$, 2) images are encoded in SD3 latent space, 3) we interpolate between the latent codes z_0 and z_1 to compute z_t for $t \sim \mathcal{U}(0, 1)$, 4) we train the U-Net backbone v_θ on a simple regression loss to match the conditional velocity $u_t(z_t \mid z_0, z_1)$.

are also common in Earth Observation. For example, StegoGAN [61] performs style transfer from satellite images to maps and vice-versa. Natural disaster management is also the subject of domain adaptation research, via flood simulation using adversarial networks conditioned on physical measurements [38]. SAR to optical (S2O) image translation is especially popular because radar acquisitions can be carried out despite cloud cover, optical sensors suffer from cloud occultation. S2O imagery makes it possible to fill missing optical acquisitions based on SAR images from close dates. Research has leveraged various classes of generative models for S2O, *e.g.* [45] uses a conditional GAN, [65] uses CycleGAN, [28] uses diffusion bridges, and so on. FlowEO pushes forward this state-of-the-art by integrating flow matching models that deliver a better semantic-preserving transfer and higher quality generation.

Flow Matching *Flow matching models* (FMMs) have been introduced in the last years [1, 33, 43] and now represent the state of the art in generative models for various applications [13, 39, 60]. However, flow matching models also allow data-to-data transport between arbitrary distributions [2, 34], and remain less well-studied. Contrary to diffusion bridges [1, 68] and Schrödinger Bridges [5, 11] that rely on stochastic differential equations to transport data, the flow is deterministic. Deterministic sampling processes [51, 52], have been wildly used with diffusion models for image and video editing and composition as they better preserve semantic content than their stochastic counterparts [12, 17, 41, 57]. This property is promising in domain adaptation contexts, where preserving semantics is critical. Moreover, unlike previously described image translation methods, such as Pix2Pix [20], CycleGAN [69], or UNSB [27], FMMs do not rely on adversarial learning to align the endpoint distributions, making them easier to train and less sensitive to hallucinations.

3. Method

Our goal is to apply an existing classifier or segmenter S_1 trained on source domain D_1 on a new target domain D_0. We assume that we have access to samples from D_0, although we do not know their labels. To solve this *unsupervised domain adaptation* problem, we introduce FlowEO to perform domain adaptation in pixel space (see Fig. 1).

We train a flow matching model to build a bridge between the image distribution p_0 of $\mathcal{X}_0$ (target) and p_1 of $\mathcal{X}_1$ (source) (see Fig. 2). Let φ be the learned transfer, *i.e.* our mapping from D_0 to D_1. To apply our existing predictive model on data from the target domain, we first map it to the source domain, *i.e.* our model predicts S_1 $(\hat{x}_1)$ (Fig. 3, step 2). This prediction should be as close as possible to the (unknown) ground truth y_0, *i.e.* we want φ to preserve the semantic information relevant to the task during transfer. By transferring the images rather than adapting the predictive model, FlowEO only depends on the datasets D_0 and D_1, and neither on the task nor the model S_1 (Fig. 3, stage 2). This makes it applicable to a broad panel of tasks, and can benefit from better predictive models without retraining.

3.1. Training the flow

Mapping domains Flow matching models have been used extensively as generative models, mapping a normal distribution to the images' latent distribution, similar to diffusion models [13, 26, 32, 35]. However, flow matching can also bridge between arbitrary distributions [1, 33]. Fol-

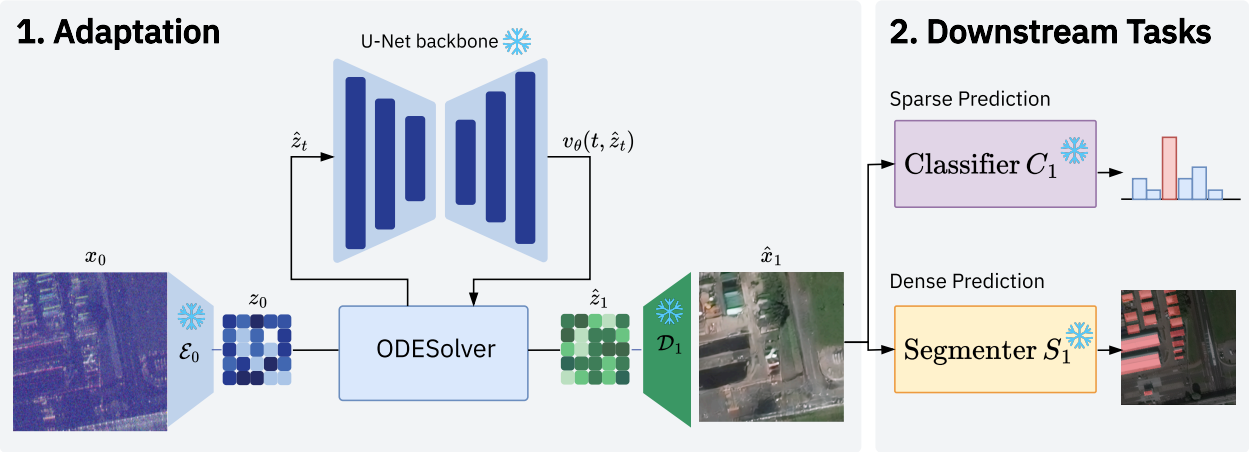

Figure 3. FlowEO offers domain adaptation in image-space, making the adaptation independent of the downstream task and predictive model used. At inference time, we adapt the image x_0 into a synthetic image $\hat{x}_1$ by integrating the flow with an ODE solver and the learned velocity v_θ. Then, any predictive model S_1/C_1 can directly perform downstream tasks on the transferred images, without fine-tuning.

lowing this framework, we leverage a time-dependent flow $\varphi : [0, 1] \times \mathbb{R}^d$ guided by a velocity field u_t describing the trajectories of samples z moving from p_0 to p_1:

$$\frac{d}{dt}\varphi_t(z) = u_t(\varphi_t(z)) \tag{1}$$

The flow φ_t results in a transport between p_0 and p_1 when solving the Ordinary Differential Equation (ODE) defined by Eq. (1) from $t = 0$ to $t = 1$. Conversely, solving the same ODE with decreasing times $t = 1$ to $t = 0$ allows, by construction, to transport p_1 to p_0. While the true velocity field u_t is intractable, it is approximated with a neural network $v_\theta(t, z_t)$ trained with a simple regression and simulation-free objective (Fig. 2):

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{\substack{z_0,z_1\sim p(z_0,z_1)\\ t\sim U([0,1])}} \left\| v_\theta(t, z_t) - u_t(z_t \mid z_0, z_1) \right\|^2 \tag{2}$$

where pairs (z_0, z_1) are sampled from the joint distribution $p(z_0, z_1)$ also named *coupling*, that will be detailed later. From these endpoints, we can build z_t using an interpolant [1]. We use linear interpolants, *i.e.* $z_t = (1-t)z_0 + tz_1$ for which the conditional velocity field $u_t(z_t \mid z_0, z_1)$ equals $z_1 - z_0$. At inference time (Fig. 3, Stage 1), we deploy ODE solvers to solve Eq. (1) by replacing the true velocity u_t with its neural network approximation $v_\theta(t, \cdot)$. This way, we generate the transferred observation $\hat{z}_1$ by integrating the ODE starting from z_0 using the mapping φ_t following:

$$\hat{z}_1 = \varphi_{t=1}(z_0) = \text{ODESolver}^{v_\theta}(z_0, 0 \to 1) \tag{3}$$

Latent flow With high resolution Earth observation, image dimensions need to be large to include enough spatial context. For example, a 256×256 tile covers only $\approx 100\,\text{m} \times 100\,\text{m}$ at 40 cm/px. Because training generative models at a high resolution is compute-intensive, training generative models in the latent space of a *Variational Auto-Encoder* (VAEs) has become a common strategy to improve image generation and accelerate sampling [14, 29, 46]. We train a flow model in the latent space of a frozen pretrained VAE. Given an image x, the VAE's encoder $\mathcal{E}$ compress it into a latent $z = \mathcal{E}^*(x)$ of lower dimensionality. The decoder $\mathcal{D}*$ generates images from latent codes. Although this VAE was trained exclusively on 3-channel RGB images and not specifically on remote sensing data, the encoder still learns effective representations for such inputs, including non-RGB modalities like SAR. Due to differences in value distributions, the decoder is fine-tuned on SAR and multispectral domains prior to the flow matching training (see Appendix A.2.2). In practice, this means that p_0 and p_1 represent the distributions of latents $z_0 = \mathcal{E}^*(x_0)$ and $z_1 = \mathcal{E}^*(x_1)$ instead of the images of $\mathcal{X}_0$, $\mathcal{X}_1$ in Eqs. (1) and (2).

3.2. Coupling

The properties of the transport learnt by the flow are greatly influenced by the choice of the image pairs (x_0, x_1) used to compute the loss function Eq. (2). The most common setup relies on an independent coupling, *i.e.* (z_0, z_1) is sampled uniformly across all possible pairings. Recent works [5, 34, 54] have introduced couplings inspired by *optimal transport*. However, their optimal transport coupling is defined with the L2-distance between images. In image space, there is no obvious reason that images with a small pixel-wise Euclidean distance would be semantically similar – an intuition we show to be true in Sec. 5.2. This contradicts our goal to obtain a transfer that preserves se-

Dataset	Target	Source	Resolution	Task	Size	Alignment
SpaceNet 6 [48]	SAR (aerial)	RGB (WorldView-2)	2 m/px	Segmentation	50 000	Strong
Sen1floods11 [4]	SAR (Sentinel-1)	Optical (Sentinel-2)	10 m/px	Segmentation	64 512	Strong
BigEarthNet2 (reBEN) [9]	SAR (Sentinel-1)	Optical (Sentinel-2)	10 m/px	Multi-label classification	237 871	Strong
SpaceNet 8 Germany [19]	RGB (post-flood)	RGB (pre-flood)	0.8 m/px	Segmentation	5688	Weak
SpaceNet 8 Louisiana [19]	RGB (post-flood)	RGB (pre-flood)	0.8 m/px	Segmentation	17 173	Weak

Table 1. Datasets used for domain adaptation. We evaluate post-flood to pre-flood adaptation and SAR-to-optical translation scenarios.

mantics. Because we leverage flow matching for image translation, *i.e.* conditional generation, we can turn towards a *data-dependent* coupling $p(x_0, x_1) = p(x_1|x_0)p(x_0)$ [2,50]. We therefore want to build image pairs of a semantically relevant x_1 in the p_1 distribution, given an image $x_0 \sim p_0$ instead of defining a new ad hoc joint distribution. Finally, to sample latents from $p(z_0, z_1)$, we first sample from the image coupling $p(x_0, x_1)$ and then encode the images.

3.3. Alignment

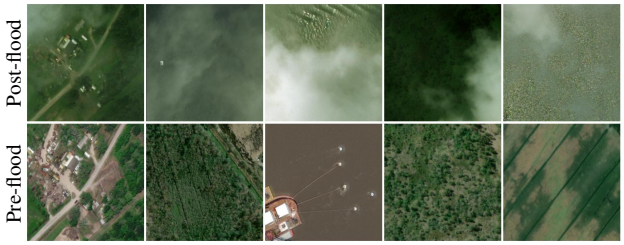

Figure 4. Weakly-aligned image pairs from the SpaceNet 8 dataset, affected by cloud coverage and natural disasters. Each column: top=post-flooding imagery; bottom=pre-event imagery.

We aim at building pairs of images (x_0, x_1) that are semantically close. Because remote sensing data is geospatial, the coordinate metadata are available for each image. Thus, we consider spatially aligned datasets, which is possible to construct in most real-world applications, and leave geographical domain adaptation for future work. While coregistration provides pairs of images that have a common location, it does not ensure that the semantic information is shared between the two images. Thus, we distinguish between:

- **Strong semantic alignment**: the two images x_0 and x_1 share the same semantics, *i.e.* $y_0 = y_1$. This is the ideal scenario, though impractical, as it needs synchronized acquisitions, or at least images of the same areas in a short timeframe, such that no significant semantic changes have occurred. This can be typically used to address sensor shift, *e.g.*, SAR to optical translation.
- **Weak semantic alignment**: the two images x_0 and x_1 partially share their semantics, *i.e.* the ground truths are similar $y_0 \approx y_1$. For example, these may be acquisitions of the same geographical area but captured on different dates. As shown in Fig. 4,changes in semantics may be due to the construction of buildings between the two acquisitions, harvested crops, cloud coverage, natural disasters such as floods or fires, deforestation, moving objects, or any other event that can shift semantics. Note that x_1 in the dataset is not an accurate representation of the transferred x_0 because of the changes. Yet, in the absence of labels, this pair (x_0, x_1) is the best available coupling. We then assume that the averaging of velocities in Eq. (2) is robust to moderate semantic changes and preserves the main transfer components between p_0 and p_1. This can be used to address temporal shift, *e.g.* seasonal variations, and before/after an extreme event.

4. Experimental setup

4.1. Datasets

We evaluate FlowEO for domain adaptation on three segmentation and one classification datasets, listed in Tab. 1: SpaceNet 6 [48], Sen1floods11 [4], BigEarthNet2 (reBEN) [9] and SpaceNet 8 [19], split into Germany and Louisiana. These datasets are paired, *i.e.* have multiple acquisitions for the same area, allowing us to train image translation models with data dependent couplings. SpaceNet 8 contains before/after images of flood event, semantic differences exist in the images, making it "weakly aligned". The others pair images from close dates, resulting in a "strong" alignment. We build 3-channels 256×256 images using RGB for color images, bands $[4, 3, 2]$ for Sentinel-2, VV/HH/VH polarizations for SAR images from SpaceNet 6 and reBEN, and VV/VH/VH for Sen1Floods11. See Appendix B for details.

Downstream models We use the DeepLabv3+ architecture [7] for semantic segmentation with a ResNet-34 backbone, ImageNet initialization, a batch size 512, and a learning rate of 0.001 with one-cycle cosine schedule. For classification, we follow the reBEN implementation [9] and train a ResNet-50 with ImageNet initialization for 100 000 training steps with a batch size of 512 and a linear-warmup-cosine-annealing learning rate of 0.001. These models are trained once and used to evaluate all image translation methods.

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
Pix2Pix	34.73	36.08	98.22	50.95	32.92	34.25	98.38	55.75	38.79	40.86	92.23	47.05
CycleGAN	40.70	43.35	**54.31**	55.70	39.35	41.79	**62.80**	59.46	42.39	45.14	**52.80**	52.92
UNSB	39.35	42.67	68.30	55.35	38.25	40.62	66.62	56.84	40.67	43.87	73.72	53.04
Diffusion Bridge	37.50	39.36	115.70	53.13	33.91	35.27	177.23	58.53	39.05	41.37	105.27	51.25
StegoGAN	38.62	40.58	66.61	58.07	36.74	38.78	90.42	63.50	40.14	42.29	68.56	54.58
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 2. 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. FlowEO transports images while preserving its semantics, achieving significant segmentation performance improvements in domain adaptation setting: 44.65 *vs*. 40.05 mIoU on SpaceNet 8. It also outperforms the second-best model – CycleGAN – on segmentation accuracy after transfer.

Datasets	Sen1Floods1				SpaceNet 6				ReBEN					
	SAR → Optical				SAR → RGB				SAR → Optical					
	mIoU	mAcc	FID	LPIPS	mIoU	mAcc	FID	LPIPS	AP^{μ}	AP^{M}	$F1^{\mu}$	$F1^{M}$	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
Pix2Pix	51.50	62.31	20.64	31.33	56.48	63.43	130.42	41.89	**41.09**	27.88	**43.93**	**25.79**	**62.84**	17.56
CycleGAN	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
UNSB	42.69	48.85	23.01	35.01	52.43	61.04	**72.48**	45.81	25.61	20.71	29.52	19.45	113.73	35.64
Diffusion Bridge	42.41	50.31	18.71	39.93	51.22	58.37	94.15	46.37	18.44	15.79	24.43	05.80	80.97	20.74
StegoGAN	43.37	49.75	41.06	31.87	44.87	50.02	306.50	56.62	26.13	22.16	29.49	20.28	81.15	22.32
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 3. 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). FlowEO preserves achieves the best UDA segmentation performances, and on-par classification performances with Pix2Pix.

4.2. Model comparison

Baselines We compare our FlowEO model against several commonly used image translation baselines: Pix2Pix [20], CycleGAN [69], StegoGAN [61], Diffusion Bridge [5], and Unpaired Neural Schrödinger Bridges (UNSB) [27]. For a fair comparison, we use data-dependent coupling for all methods, even those that could be trained using independent couplings (CycleGAN, StegoGAN, and UNSB), and train all models for 200 000 steps. We follow official implementations for hyperparameters (cf. Appendix C). Except CycleGAN, baselines are non-symmetric, thus we train two separate models from domain $\mathcal{X}_0$ to $\mathcal{X}_1$ and then from $\mathcal{X}_1$ to $\mathcal{X}_0$ when needed.

Hyperparameters We train our flow matching in the pretrained space of the VAE from Stable Diffusion 3 [13]. More precisely, we use a distilled model that is smaller and more compute efficient [3]. The flow is therefore performed on the latent codes of dimensions $16 \times 32 \times 32$. Because flow matching is symmetrical, the same model can be used to transfer from $\mathcal{X}_0$ to $\mathcal{X}_1$ and vice versa, while baselines require two models. We use the classical U-Net backbone to train the flow [52] with 120 million parameters. The flow is trained for 200 000 steps using gradient clipping and exponential moving average. We use a learning rate of 1×10^{-4} with 1000 steps of linear warmup and a batch size of 256. At inference time, we integrate the flow from Eq. (1) with 50 steps of the Euler ODE sampler. We use the sigmoid time-scheduler introduced in [26] to focus on the times that are close to the image spaces. See Appendix A.3 for more insights and ablation studies about sampler design and Appendix A.4 for inference time and memory footprints.

4.3. Evaluation

Prediction metrics Because Earth observation tasks are often dense predictions, we focus on domain adaptation for semantic segmentation. For all methods, we first transfer the images from the test set of each dataset using the image translation model and then apply the same pretrained segmenter to obtain the semantic masks. We then compute *mean Intersection over Union* (mIoU) and *mean Accuracy* (mAcc) between the prediction on the transferred image $\hat{x}_0 = \varphi(x_1)$ and the ground truth mask y_0, *that is only available for evaluation purposes*. For reBEN, we use the standard multi-label classification metrics: *Average Precision* (AP) and F_1*-score* (F_1), both micro and macro, *i.e.* AP^{μ}, F_1^{μ}, AP^{M}, F_1^{M}.

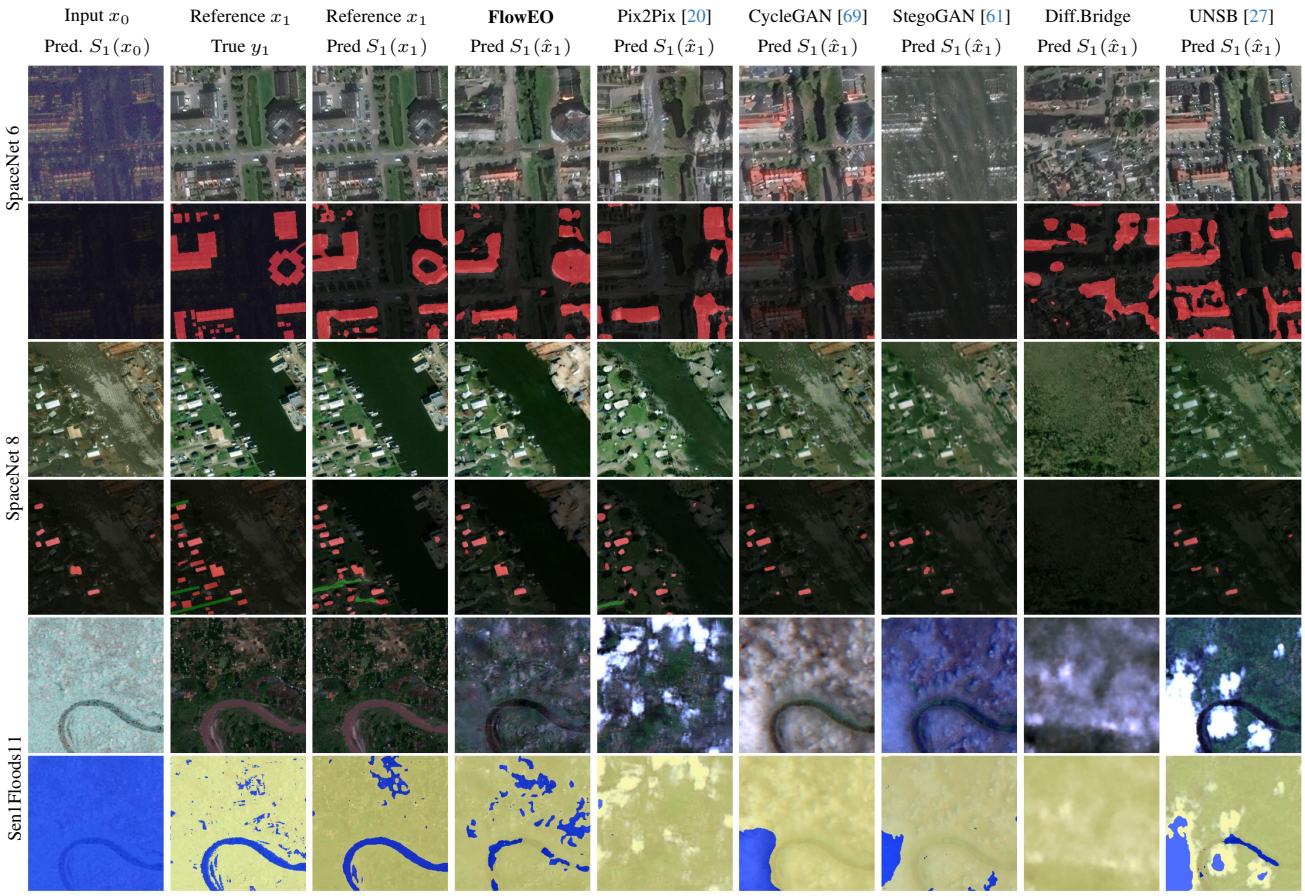

Figure 5. Qualitative comparison of domain adaptation methods on segmentation datasets. The first column represents the input image x_0, the second and third depict the weakly or strongly aligned x_1, and the others display the images generated by the different methods. Below each image, we provide the corresponding prediction from the segmentation model S_1 or the true segmentation mask y_1 for the reference image (third column). FlowEO outperforms other methods in both semantic preservation and image quality.

Image quality While it is not our main goal, we also evaluate the perceptual quality of the generated images as visual artifacts can hinder downstream performances and interpretability. We compute both the *Frechet Inception Distance* (FID) [18] and *LPIPS* [66] similarity between the transferred images $\hat{x}_0$ and source images x_1. Although commonly used, note that these are initially designed for natural images and not remote sensing imagery [21]. In addition, the size of our test sets is under the recommended size to compute FID. Despite the noise this might introduce, these metrics remain useful proxies to assess broad tendencies regarding the perceptual qualities of transferred images.

5. Results

5.1. Main results

We present domain adaptation and image quality metrics obtained by the compared image translation methods in Tab. 2 (weakly aligned) and Tab. 3 (strongly aligned). In addition to the results obtained by state-of-the-art models and FlowEO, we include two comparison points:

- **No adaptation**: classification/segmentation metrics of the pretrained model applied directly on the non-transferred target data, *i.e.* the performance of S_1 on x_0. This represents a lower bound of the expected performance. Image quality metrics (FID and LPIPS) are computed directly between images from $\mathcal{D}_0$ and $\mathcal{D}_1$ and show an estimate of how far away the two image distributions are.
- **Upper bound**: classification/segmentation metrics of the pretrained model on its source domain, *i.e.* the performance of S_1 on x_1. This represents the performance of an ideal semantic-preserving transfer from p_0 to p_1, for which S_1 is as accurate on x_0 transferred as on x_1.

Semantic preservation FlowEO consistently demonstrates superior semantic preservation compared to existing

image translation models. It ranks first in the weakly aligned setting (Tab. 2), significantly outperforming both the second-best state-of-the-art transfer method (+4 mIoU points compared to CycleGAN) and the no-transfer baseline by a large margin (44.65 mIoU vs. 40.05 mIoU) on SpaceNet 8. Domain adaptation for pre/post-flood imagery is a particularly challenging task considering the significant changes that impact the images, as shown in Fig. 5. Note that only FlowEO and CycleGAN successfully increase segmentation performance over the no-adaptation baseline. FlowEO consistently achieves the highest mIoU and mean accuracy across both regions (Germany and Louisiana), demonstrating its effectiveness in handling real-world geographic variations, even when trained on smaller datasets ($<$10 000 samples).

For strongly aligned datasets (Tab. 3), FlowEO achieves the best segmentation metrics on both Sen1Floods11 and SpaceNet 6 datasets with respectively +3.42 and +8.59 in mIoU compared to the second best transfer models. Somewhat surprisingly, Pix2Pix constitutes a strong baseline for paired image translation and achieves the second-best performance in this setting despite being the oldest model evaluated. On the ReBEN multi-label classification dataset, the flow model and Pix2Pix perform competitively, trading first and second places depending on the metric considered. Despite their training with data-dependent coupling, adversarial-based methods struggle to offer semantic-preserving transport. This suggests that adversarial objectives may be unaligned with semantic preservation by hallucinating new instances *e.g.* clouds, that can reduce segmentation performances (as shown for Sen1Floods11 in Fig. 5, fifth row)

Transferred image quality In addition to better preserving the semantics, FlowEO generates consistent high-quality images. It ranks first in LPIPS and first or second in FID on all datasets, both in weakly aligned RGB→RGB transfer on SpaceNet 8 (Tab. 2) and strongly aligned SAR-to-optical translation (SpaceNet 6, Sen1Floods11 and reBEN in Tab. 3). Unlike previous methods, FlowEO does not rely on adversarial loss functions explicitly designed to enhance perceptual quality. Despite that, generated images remain of high quality and do not show hallucinations commonly attributed to adversarial training. This trend holds for both weakly and strongly aligned datasets. In particular, we observe that FlowEO learns complex texture transfer on the post-to-pre-disaster scenario, correctly mapping turbulent and murky flood water to the usual river state (Fig. 5, third row).

5.2. Impact of the coupling

We report in Tab. 4 ablation results on SpaceNet 6 and SpaceNet 8 domain adaptation and image generation using the three couplings: independent, minibatch-OT [54], and data-dependent. Minibatch-OT coupling outperforms the independent coupling on the two datasets in segmentation accuracy after domain adaptation and image quality. Yet, data-dependent coupling outperforms them by a large margin for domain adaptation (+17% mIoU on SpaceNet 6, +7% mIoU on SpaceNet 8) and image quality (30% decrease in FID on both datasets). This is expected since OT pairs images based on Euclidean distance in pixel space, which is irrelevant to semantics *e.g.* in SpaceNet 6 where it compares SAR and RGB modalities. Yet, OT is also far behind the data-dependent coupling in the favourable case of RGB to RGB transport on SpaceNet 8. This confirms the importance of data-dependent coupling – thus dataset alignment – to preserve semantic information during flow-based transfer, and motivates our focus on aligned datasets, even weakly.

SpaceNet 8				
	Post-flood → Pre-flood			
Coupling	mIoU ↑	mAcc ↑	FID↓	LPIPS ↓
Independent $p(x_0)p(x_1)$	35.59	37.41	94.23	66.62
Minibatch-OT $\pi(x_0, x_1)$	37.26	39.28	84.44	63.93
Data-dependent $p(x_1\|x_0)p(x_0)$	**44.65**	**48.79**	**60.32**	**45.50**
	Pre-flood/Post-flood			
Coupling	mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
Independent $p(x_0)p(x_1)$	35.60	37.80	80.26	67.88
Minibatch-OT $\pi(x_0, x_1)$	36.21	39.28	73.26	65.22
Data-dependent $p(x_1\|x_0)p(x_0)$	**44.87**	**53.76**	**50.88**	**52.81**
SpaceNet 6				
	SAR → RGB			
Coupling	mIoU ↑	mAcc ↑	FID↓	LPIPS ↓
Independent $p(x_0)p(x_1)$	45.25	50.75	145.02	65.94
Minibatch-OT $\pi(x_0, x_1)$	48.48	55.03	125.82	58.34
Data-dependent $p(x_1\|x_0)p(x_0)$	**65.07**	**72.33**	**94.02**	**39.98**
	RGB → SAR			
Coupling	mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
Independent $p(x_0)p(x_1)$	45.74	50.85	105.47	64.69
Minibatch-OT $\pi(x_0, x_1)$	47.25	52.65	91.74	60.24
Data-dependent $p(x_1\|x_0)p(x_0)$	**55.36**	**61.53**	**36.86**	**51.66**

Table 4. Impact of coupling on generation quality and semantic preservation during transfer. The OT-based coupling $\pi(x_0, x_1)$ fails to match the performance of the data-dependent coupling $p(x_1|x_0)p(x_0)$, although it outperforms the independent coupling.

6. Conclusion

We introduce FlowEO, a flow matching-based framework for unsupervised domain adaptation in Earth Observation. By learning a semantically consistent mapping between source and target distributions, FlowEO consistently outperforms existing image translation methods for domain adaptation in five segmentation and classification tasks across multiple challenging scenarios ranging from post-disaster monitoring to SAR-to-Optical translation, while achieving on-par or better image generation quality. FlowEO opens the door to generic unsupervised domain adaptation with possible extensions to semantic-based couplings based on image similarity or image metadata embeddings to fare with unpaired image translation scenarios in Earth observation.

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

[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

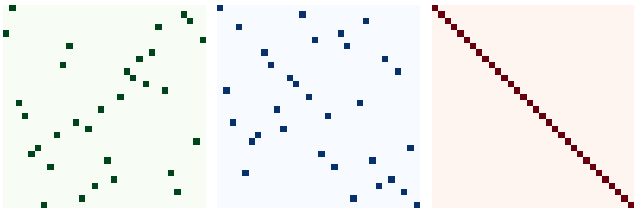

Figure 6. Comparison between the pairing matrices generated with the different couplings for a batch on SpaceNet 8, from left to right: independent coupling $p(x_0)p(x_1)$, OT-coupling $\pi(x_0, x_1)$, data-dependent coupling $p(x_1 \mid x_0)p(x_0)$.

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			
RGB		mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
	Base	**44.65**	**48.79**	**60.32**	**45.50**
	Finetuned	44.33	48.71	81.75	51.64
		SpaceNet 6 SAR → RGB			
RGB		mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
	Base	**65.07**	**72.33**	**94.02**	**39.96**
	Finetuned	64.63	72.17	111.66	42.77
		Sen1Floods11 SAR → Optical			
S2		mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
	Base	51.45	57.63	24.33	29.22
	Finetuned	**54.92**	**69.04**	**12.96**	**29.21**
		ReBEN SAR → Optical			
S2		AP^M	$F1^M$	FID↓	LPIPS ↓
	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, kim2025simple introduced a sigmoid time-scheduler tailored for flow matching models (see Eq. (4)). This scheduler is

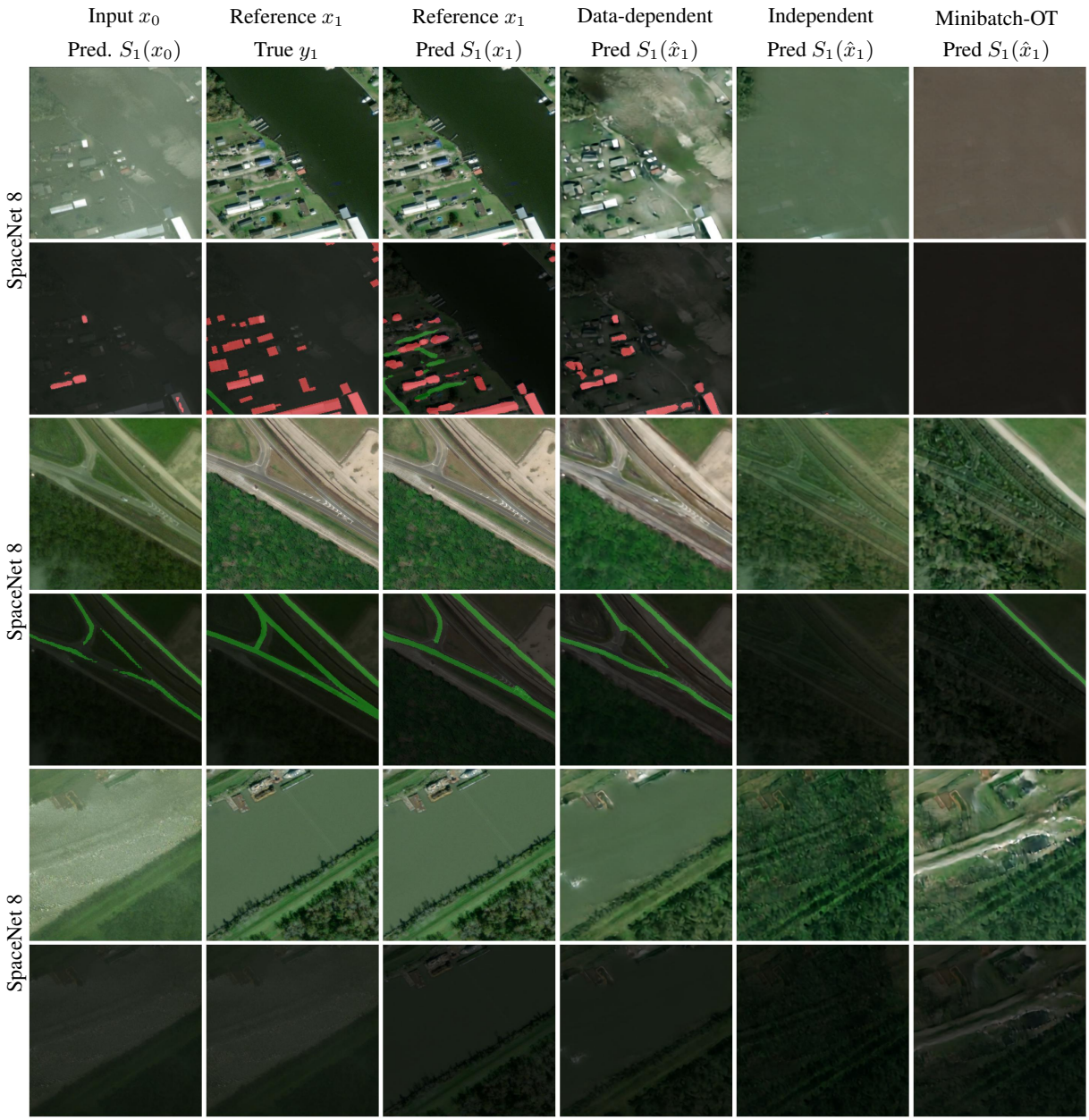

Figure 7. Impact of the training coupling $p(x_0, x_1)$ on preserving semantic information during image translation. FlowEO employs *data-dependent coupling* $p(x_1|x_0)p(x_0)$, which outperforms both *minibatch-OT* coupling $\pi(x_0, x_1)$ and *independent* coupling $p(x_0, x_1)$.

parametrized by κ 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).

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

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

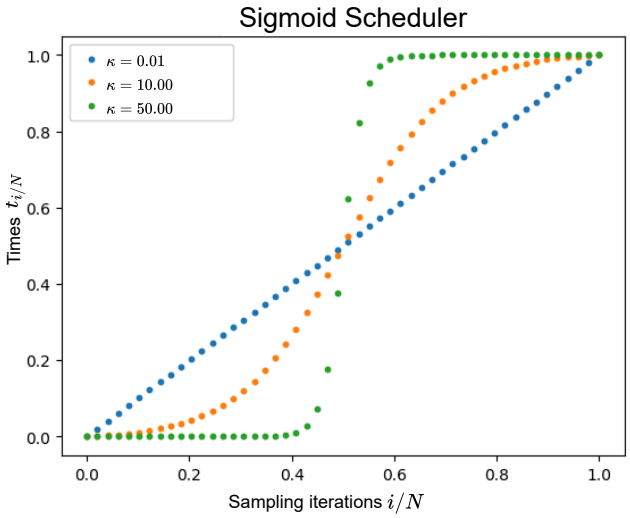

Figure 8. Sigmoid time discretization, allocating more sampling steps near the endpoints ($t \approx 0$ and $t \approx 1$).

only marginal improvements and, in some cases—such as on 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.

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.

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.

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 ($\approx$ 50000 samples), validation ($\approx$ 1800), and test ($\approx$ 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

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}	$F1^{\mu}$	$F1^{M}$	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.

then cropped 256×256 images with an overlap of 70% to produce the training data. The final numbers of samples of 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.

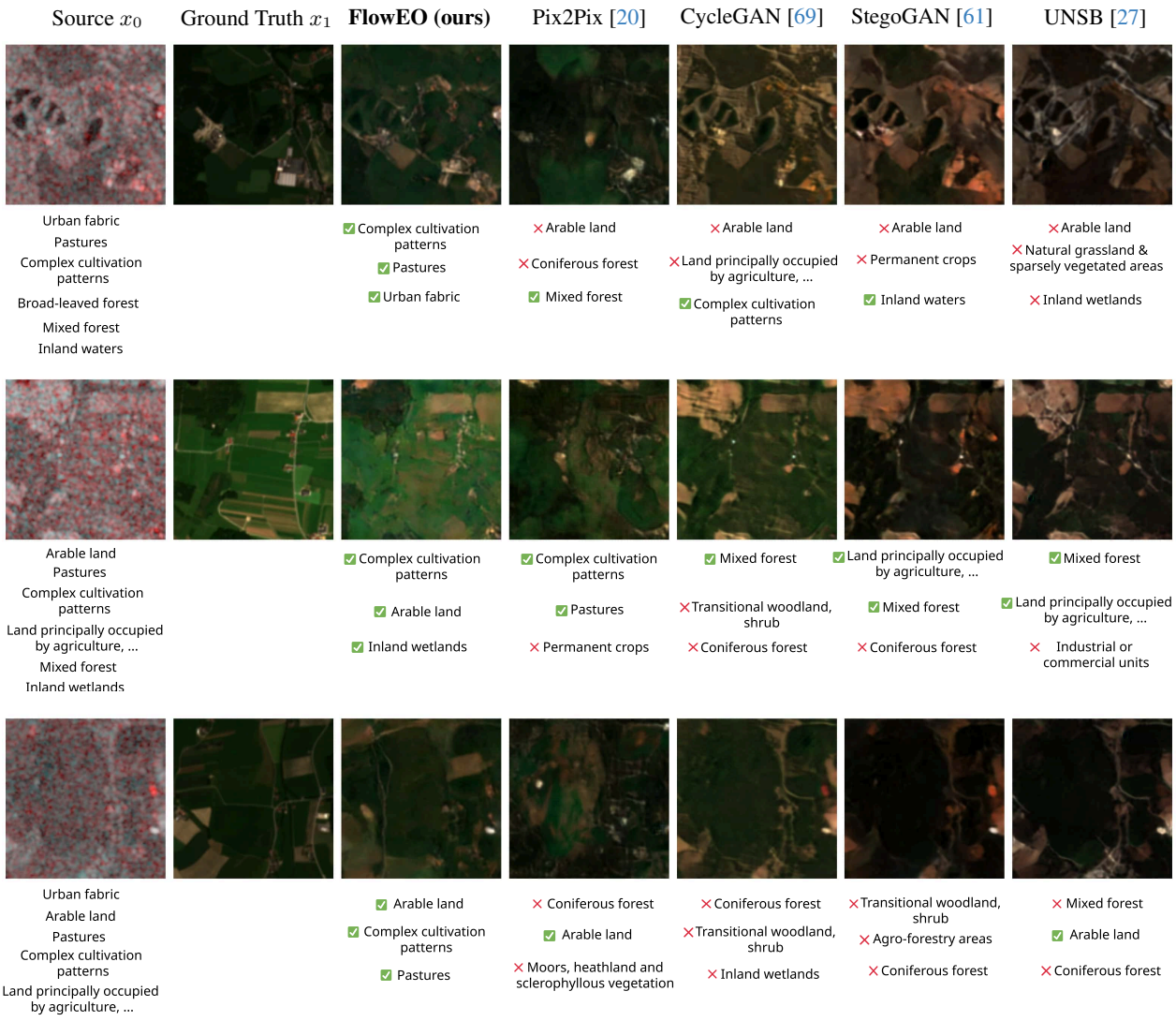

Figure 9. Qualitative comparison of domain adaptation methods on the reBEN dataset, for multiclass classification. The first column represents the source domain image x_0, the second depicts the weakly or strongly aligned x_1, and the others display the images generated by the different methods. Below 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.

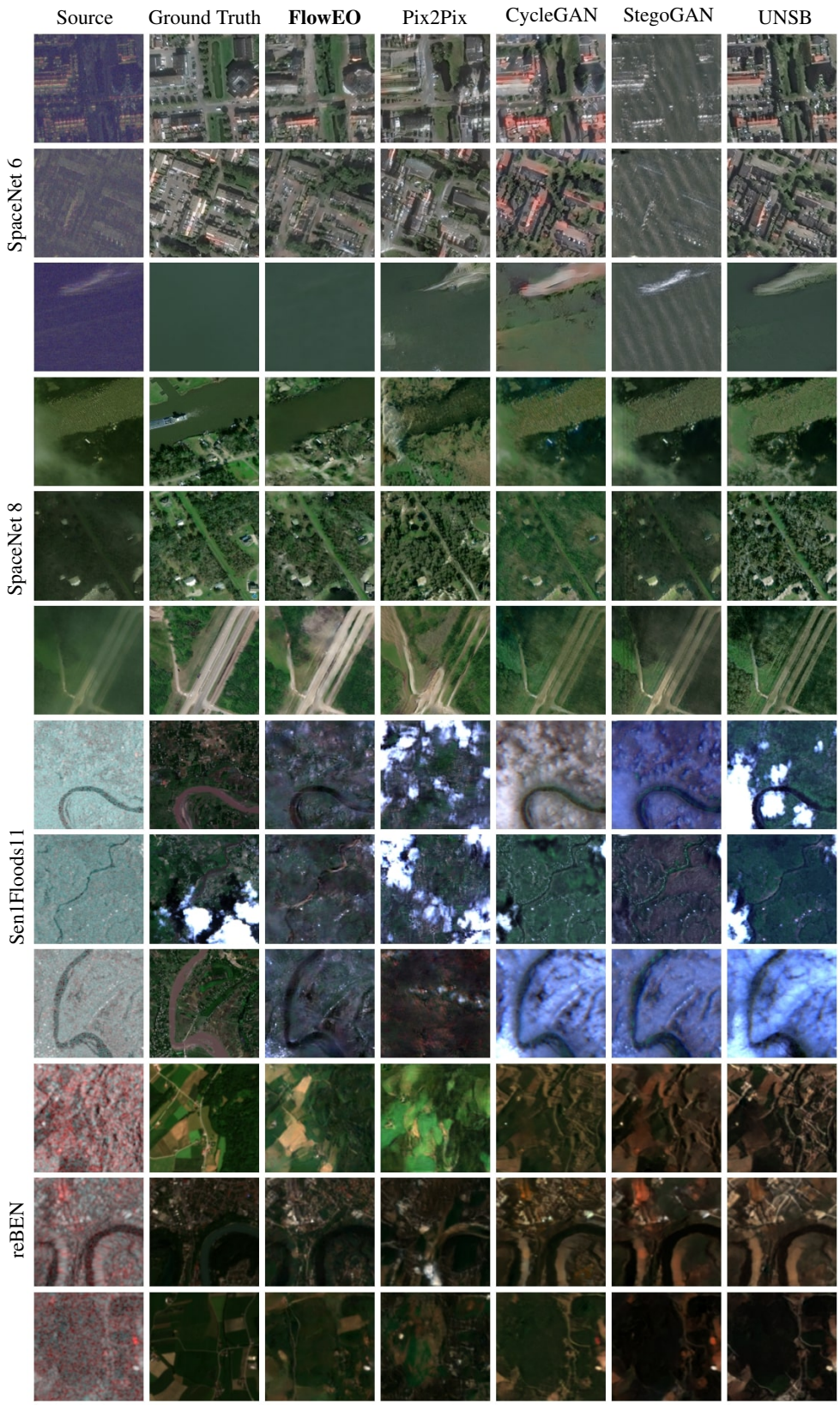

Figure 10. FlowEO generates the highest-quality images while maintaining semantic consistency during the transfer process. In the third row, we observe that our method demonstrates greater robustness to the geometric artifacts present in SAR imagery. Additionally, we note that it successfully learns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