High-resolution Multi-spectral Image Guided DEM Super-resolution using Sinkhorn Regularized Adversarial Network

Subhajit Paul, Ashutosh Gupta

Space Applications Centre (SAC), ISRO

Introduction

Digital Elevation Model (DEM) is an essential aspect in the remote sensing domain to analyze and explore different applications related to surface elevation information. Here, we explore the generation of high-resolution (HR) DEMs guided by HR multi-spectral (MX) satellite imagery as prior.

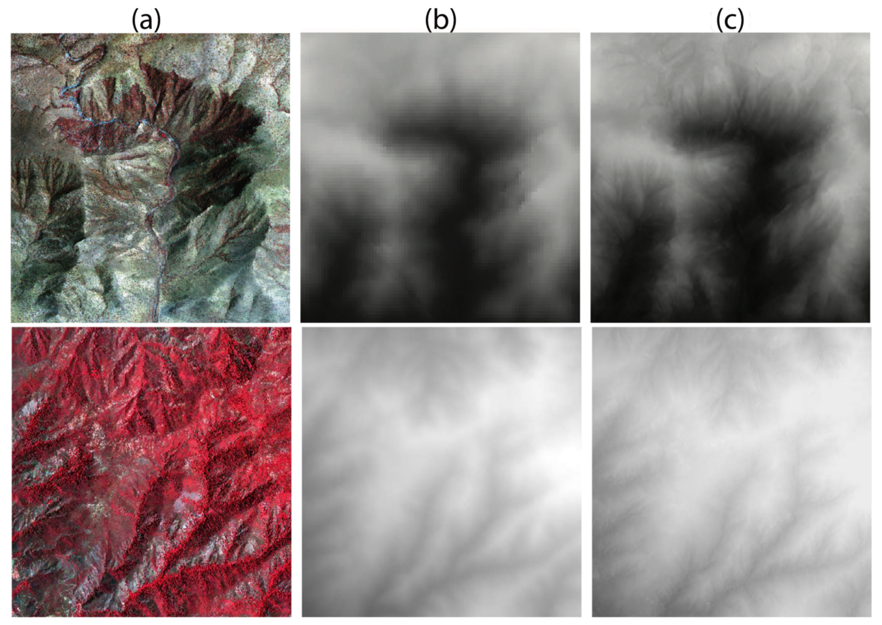


Figure 1. Sample results of DEM super-resolution. (a) High resolution FCC of NIR(R), R(G), and G(B), (b) Bicubic interpolated coarser resolution DEM, (c) Generated HR DEM.

Key Contributions

Our key contributions can be summarized as follows.

1. A novel architecture for DEM SR which utilizes sharp detail information from a **HR MX image as a guide** by conditioning it with a **discriminative spatial self-attention**.
2. We develop and demonstrate **SIRAN**, a framework based on Sinkhorn regularized adversarial learning.
3. We **generate our own dataset by using realistic coarse resolution data** instead of bicubic downsampled.
4. Finally, we perform experiments to assess the accuracy of our model.

Brief overview of Sinkhorn and other Losses

- **Kantarovich formulation of entropic optimal transport (EOT):**

$$\mathcal{W}_{C,\varepsilon}(\mu_\theta,\nu) = \inf_{\pi\in\Pi(\mu_\theta,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} [C(\mathbf{G}_\theta(x),y)]d\pi(\mathbf{G}_\theta(x),y) + \varepsilon I_\pi(\mathbf{G}_\theta(x),y),$$

where $I_\pi(\mathbf{G}_\theta(x),y)) = \int_{\mathcal{X}\times\mathcal{Y}} [\log\left(\frac{\pi(\mathbf{G}_\theta(x),y)}{\mu_\theta(\mathbf{G}_\theta(x))\nu(y)}\right)]d\pi(\mathbf{G}_\theta(x),y)$,

s.t. $\int_{\mathcal{X}} \pi(\mathbf{G}_\theta(x),y)dx = \nu(y), \int_{\mathcal{Y}} \pi(\mathbf{G}_\theta(x),y)dy = \mu_\theta(\mathbf{G}_\theta(x)) \;\&\; \pi(\mathbf{G}_\theta(x),y) \geq 0.$ (1)

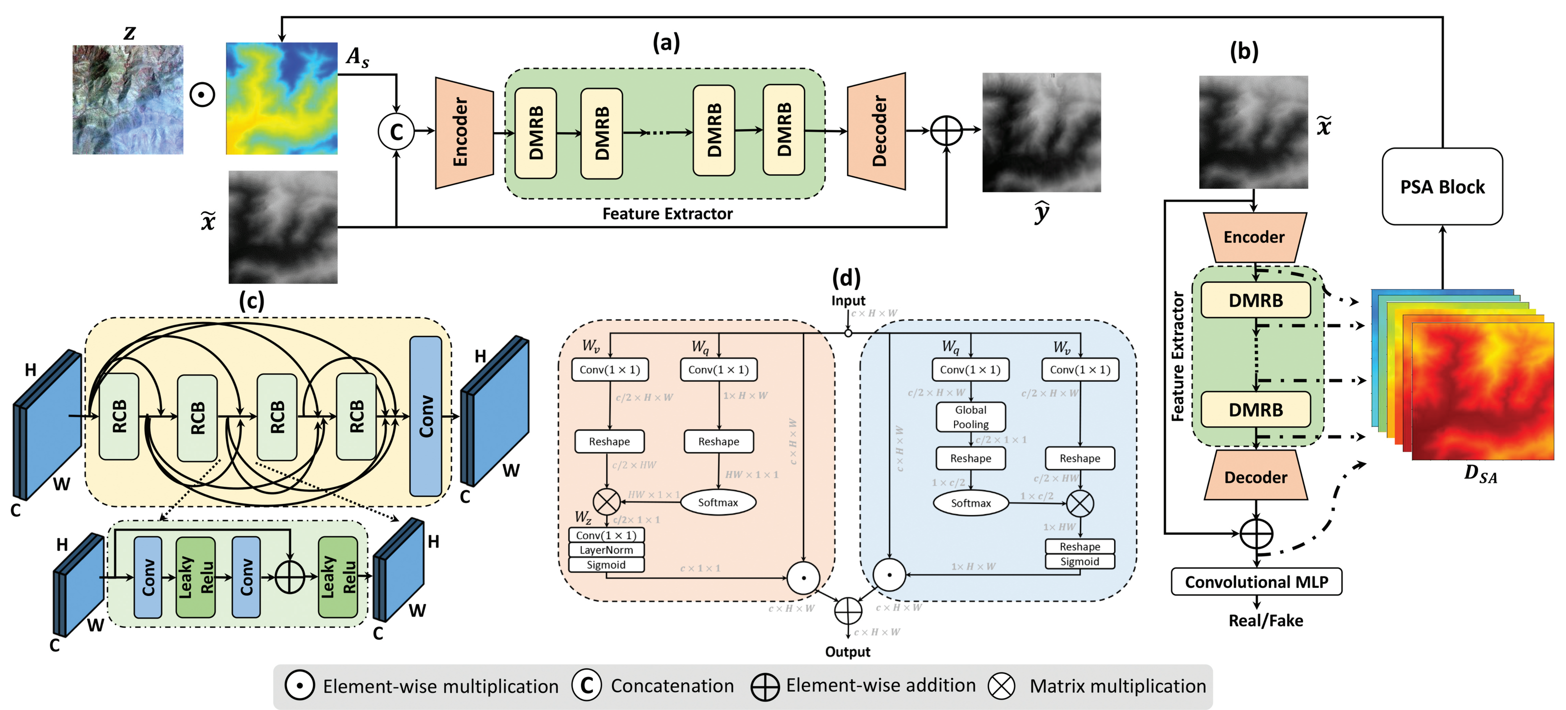
- **Sinkhorn distance formulation:** As $\mathcal{W}_{C,\varepsilon}(\nu,\nu) \neq 0$, normalization term added to define sinkhorn loss,

$$\mathcal{S}_{C,\varepsilon} = \mathcal{W}_{C,\varepsilon}(\mu_\theta,\nu) - \frac{1}{2}\mathcal{W}_{C,\varepsilon}(\mu_\theta,\mu_\theta) - \frac{1}{2}\mathcal{W}_{C,\varepsilon}(\nu,\nu), \quad (2)$$

As $\varepsilon \to 0$, $\mathcal{S}_{C,\varepsilon}$ converges to Kantarovich OT formulation. As $\varepsilon \to \infty$, $\mathcal{S}_{C,\varepsilon}$ converges to Maximum Mean Discrepancy (MMD).

- **OT Loss:** $\mathscr{L}_{OT} = \mathbb{E}_{\tilde{x}\sim\mathbb{P}_{\tilde{x}},z\sim\mathbb{P}_Z,y\sim\mathbb{P}_y}\mathcal{S}_{C,\varepsilon}(\mu(\mathbf{G}(\tilde{x}, z\odot A_s(\tilde{x})),y)),\nu(y))$.
- **Pixel Loss:** $\mathscr{L}_P = \mathbb{E}_{\tilde{x}\sim\mathbb{P}_{\tilde{x}},z\sim\mathbb{P}_Z,y\sim\mathbb{P}_y}\left[\|y - \mathbf{G}(\tilde{x}, z\odot A_s(\tilde{x}))\|_2^2\right]$.
- **SSIM Loss:** $\mathscr{L}_{str} = \mathbb{E}_{\tilde{x}\sim\mathbb{P}_{\tilde{x}},z\sim\mathbb{P}_Z,y\sim\mathbb{P}_y} - \log(\mathbf{SSIM}(\mathbf{G}(\tilde{x}, z\odot A_s(\tilde{x})),y))$.
- **Adversarial Loss:** $\mathscr{L}_{ADV} = \mathbb{E}_{\tilde{x}\sim\mathbb{P}_{\tilde{x}},z\sim\mathbb{P}_Z} - \log(\mathbf{D}(\mathbf{G}(\tilde{x}, z\odot A_s(\tilde{x}))))$.
- **Domain Adaptation Loss:** $\mathscr{L}_{DA} = \mathbb{E}_{\tilde{x}\sim\mathbb{P}_{\tilde{x}},y\sim\mathbb{P}_y}[|\mathbf{D}_{SA}(\tilde{x}) - \mathbf{D}_{SA}(y)\|_2^2]$.

Overview of Proposed Framework



- Generator objective function:

$$\min_{\mathbf{G}} \lambda_P\mathscr{L}_P + \lambda_{SSIM}\mathscr{L}_{SSIM} + \lambda_{ADV}\mathscr{L}_{ADV} + \lambda_{OT}\mathscr{L}_{OT},$$

- Discriminator objective function:

$$\min_{\mathbf{D}} -\mathbb{E}_{y\sim\mathbb{P}_y}[\log(\mathbf{D}(y)))] - \mathbb{E}_{\hat{y}\sim\mathbb{P}_{\mathbf{G}_\theta}}[\log(1-\mathbf{D}(\hat{y}))] + \lambda_{DA}\mathscr{L}_{DA},$$

- Generated datset have **SRTM coarse DEM (GSD=30m)** as input, **Cartosat-1 DEM (GSD=10m)** as reference and **Cartosat-2S MX data product (GSD=1.6m)** as guide. All samples interpolated to resolution of guide.

Theoretical reasoning behind Sinkhorn loss

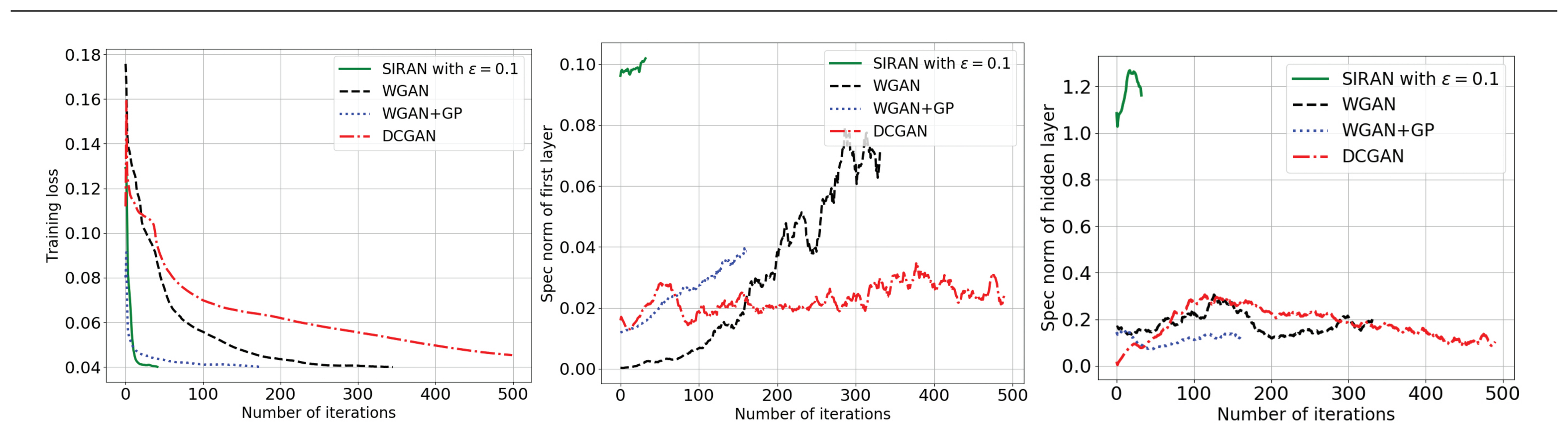
- **Proposed Smoothness of Sinkhorn Loss:** With Cost C being L_0-Lipschitz, and L_1-smooth, and $\mathbf{G}$ being L-Lipschitz, smoothness Γ_ε

$$\mathbb{E}||\nabla_\theta\mathcal{S}_{C,\varepsilon}(\mu_{\theta_1},\nu) - \nabla_\theta\mathcal{S}_{C,\varepsilon}(\mu_{\theta_2},\nu)|| = \mathcal{O}(L(L_1 + \frac{2L_0^2L}{\varepsilon(1+Be^{\frac{\kappa}{\varepsilon}})}))||\theta_1-\theta_2||, \quad (3)$$

$\kappa = 2(L_0|\mathcal{X}| + ||C||_\infty)$, $B = d.\max(||m||,||M||)$ with m amd M being the minimum and maximum values in supporting sets of measures.

- **Upper-bound of expected gradient in SIRAN set-up:** $l(\cdot)$, $g(\cdot)$ and $\mathcal{S}_{C,\varepsilon}(\cdot)$ be the objectives of supervised losses, adversarial loss and Sinkhorn loss. θ^* and ψ^* be the parameters of optimal generator $\mathbf{G}$ and discriminator $\mathbf{D}$. Let $l(p,y)$, where $p = \mathbf{G}_\theta(x)$, is β-smooth in p. If $||\theta-\theta^*|| \leq \epsilon$ and $||\psi-\psi^*|| \leq \delta$, then $||\nabla_\theta\mathbb{E}_{(x,y)\sim\mathcal{X}\times\mathcal{Y}}[l(\mathbf{G}_\theta(x),y) + \mathcal{S}_{C,\varepsilon}(\mu_\theta(\mathbf{G}_\theta(x)),\nu(y)) - g(\psi;\mathbf{G}_\theta(x))]|| \leq L^2\epsilon(\beta+\Gamma_\varepsilon) + L\delta$.
- **Iteration complexiety of SIRAN:** $l(\theta)$ is lower bounded by $l^* > -\infty$ and twice differentiable. For some arbitrarily small $\zeta > 0$, $\eta > 0$ and ϵ_1-stationary point with $\epsilon_1 > 0$, let $||\nabla g(\psi;\mathbf{G}_\theta(x))|| \geq \zeta$, $||\nabla\mathcal{S}_{C,\varepsilon}(\mu_\theta(\mathbf{G}_\theta(x)),\nu(y))|| \geq \eta$ and $||\nabla l(\mathbf{G}_\theta(x),y)|| \geq \epsilon_1$, with conditions $\delta \leq \frac{\sqrt{2\epsilon_1\zeta}}{L}$, and $\Gamma_\varepsilon < \frac{\sqrt{2\epsilon_1\eta}}{L^2\epsilon}$. The iteration complexity SIRAN upper bounded by $\mathcal{O}(\frac{(l(\theta_0)-l^*)\beta_1}{\epsilon_1^2+2\epsilon_1(\zeta+\eta)-L^2(\delta^2+L^2\Gamma_\varepsilon^2\epsilon^2)})$, assuming $||\nabla^2 l(\theta)|| \leq \beta_1$. This also can be simplified to $\mathcal{O}(\frac{l(\theta_0)-l^*}{\epsilon_1^2+\epsilon_1(\zeta+\eta)})$.

Empirical verification



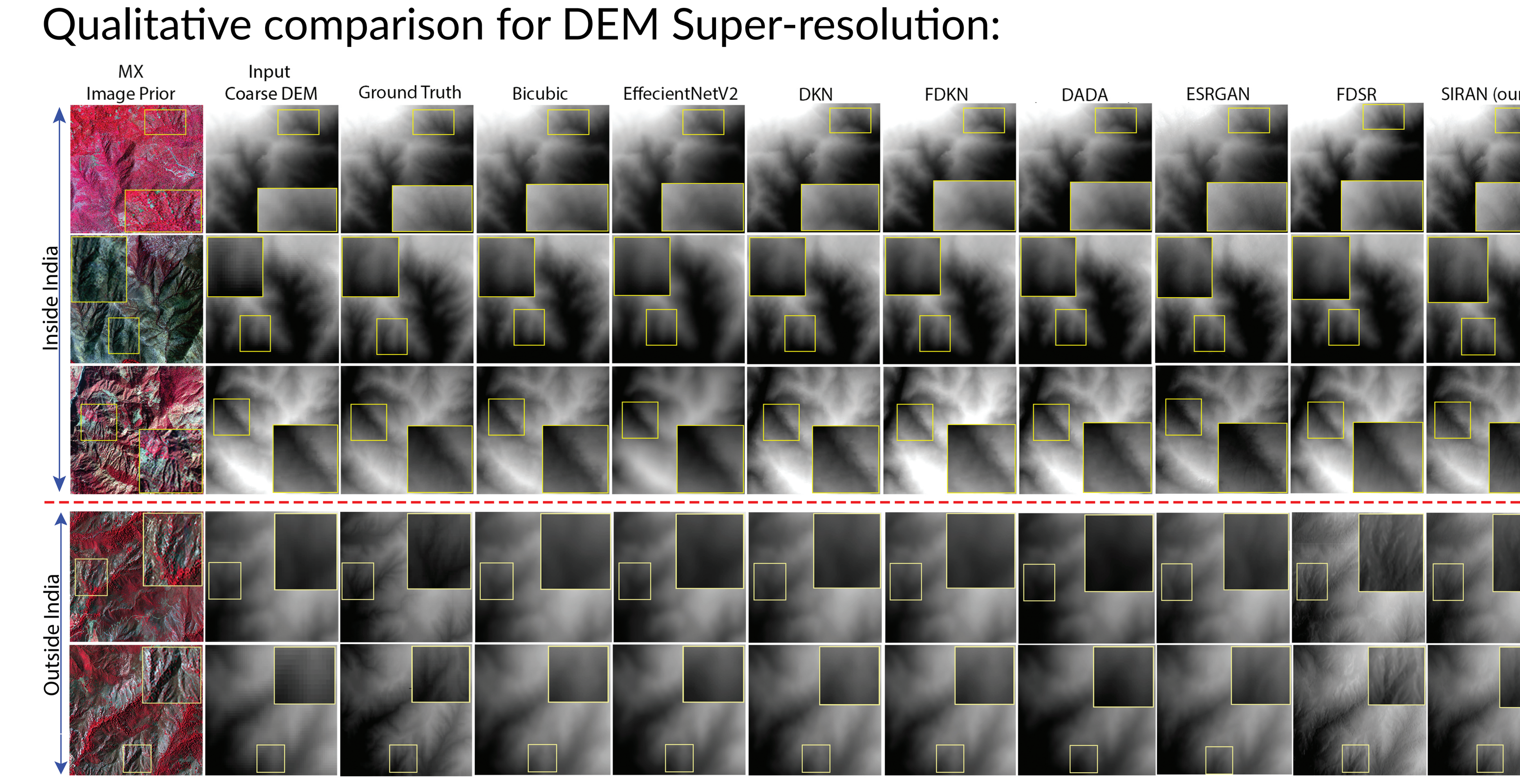
Experimental Results

Quantitative comparison for DEM Super-resolution:

Method	RMSE (m)		MAE (m)		SSIM(%)		PSNR	
Dataset	Inside	Outside	Inside	Outside	Inside	Outside	Inside	Outside
Bicubic	15.25	23.19	12.42	22.04	71.27	66.49	30.07	27.79
ENetV2	20.35	30.53	18.72	28.36	69.63	60.04	31.74	25.58
DKN	12.89	21.16	11.18	19.78	73.59	68.45	32.09	28.22
FDKN	13.05	21.93	11.34	20.41	74.13	66.83	32.46	27.68
DADA	37.49	40.89	32.17	37.74	73.32	69.86	27.94	26.78
ESRGAN	31.33	20.45	25.56	18.34	82.48	75.67	29.88	29.05
FDSR	12.98	30.58	10.87	25.28	81.49	59.81	33.77	25.59
SIRAN (ours)	9.28	15.74	8.51	12.25	90.59	83.90	35.06	31.56

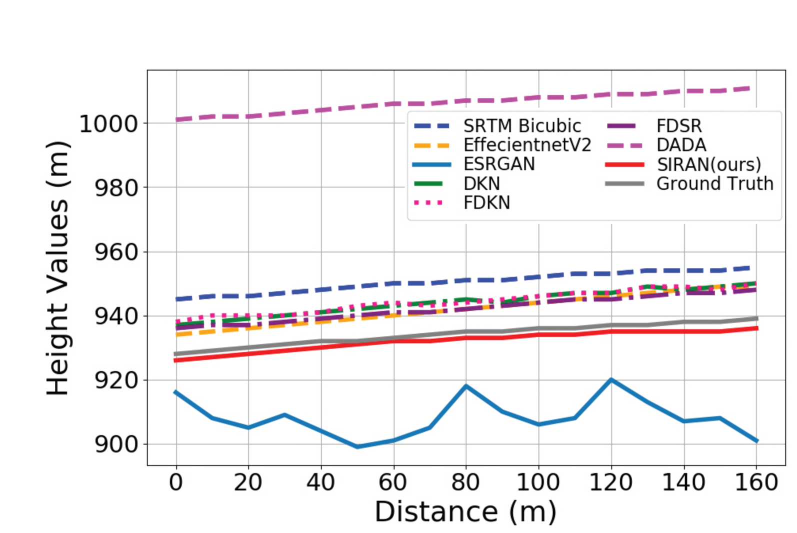
Experimental Results

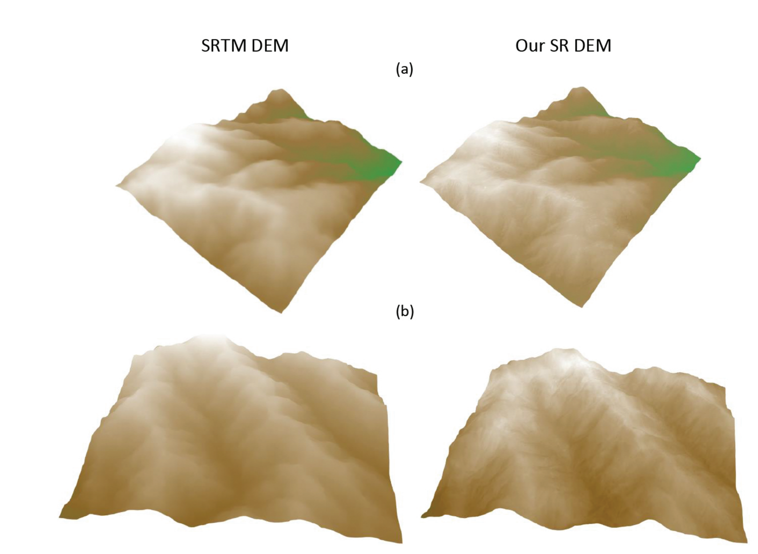
Qualitative comparison for DEM Super-resolution:



Experimental Results

Line-profile comparison and 3-D visualization:



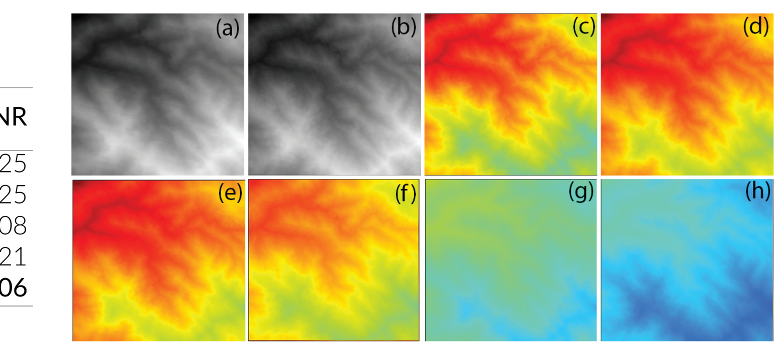


Ablation Study

Table 1. Ablation study related to different proposed modules

Image Guide	Spatial Attention	PSA	Sinkhorn loss	RMSE (m)	MAE (m)	SSIM (%)	PSNR
✗	✗	✗	✗	16.54	13.63	72.27	30.25
✓	✗	✗	✗	29.32	25.41	78.29	28.25
✓	✓	✗	✗	20.76	18.29	81.68	31.08
✓	✓	✓	✗	18.76	15.13	85.04	32.21
✓	✓	✓	✓	**9.28**	**8.51**	**90.49**	**35.06**

Figure 1. Discriminator spatial attentions at different levels



subhajitpaul@sac.isro.gov.in