

CVSAT-NeRF : CHAMPS DE RADIANCE NEURONAUX SUPERVISÉS PAR L'INFORMATION DU VOLUME DE COÛT POUR L'IMAGERIE SATELLITAIRE ÉPARSE.

Theïlo Terrisse¹, Loïc Dumas¹

1: CS Group, 6 Rue Brindejonc des Moulinais, 31500 Toulouse, France

Résumé

Les Modèles Numériques de Surface (MNS) issus d'images satellitaires sont traditionnellement générés à l'aide de chaînes de traitement Stéréo Multi-Vue (MVS, Multi-View Stereo), reposant sur des techniques de mise en correspondance de patchs telles que l'appariement semi-global (SGM, Semi-Global Matching). Bien que performantes dans des contextes à nombre limité de vues, ces méthodes présentent des limitations connues, notamment lors du rendu de surfaces homogènes ou de contours de bâtiments. Récemment, les techniques de rendu volumétrique basées sur l'apprentissage profond, en particulier les champs de radiance neuronaux (NeRF), ont émergé comme une alternative prometteuse grâce à leur précision et à leur représentation continue, compacte et flexible des scènes tridimensionnelles. Bien qu'initialement conçus pour la reconstruction de scènes multi-vues, leur mécanisme de lancer de rayons (ray-casting) les rend particulièrement attractifs dans des contextes à vues limitées. SparseSat-NeRF (SpS-NeRF) figure parmi les rares méthodes adaptant les NeRF au domaine de la télédétection à faible nombre de vues. Ce modèle exploite des informations de profondeur issues du SGM pour définir une fonction de perte sur la profondeur prédite. Dans ce cadre, la similarité de patchs est interprétée comme un indicateur de confiance associé à cet *a priori* de profondeur, utilisé pour moduler l'attache à la supervision et guider l'échantillonnage des rayons. Dans cet article, nous introduisons CVSat-NeRF, une méthode visant à affiner l'exploitation de tels *a priori*. Premièrement, la similarité entre patchs est remplacée par une mesure de confiance dérivée de l'ambiguïté, un indicateur plus informatif qui agrège efficacement l'information contenue dans le volume de coût calculé par l'algorithme SGM. Une étape de rééchantillonnage d'histogramme est également proposée afin d'ajuster l'intensité de la pondération de la supervision. Deuxièmement, nous introduisons une stratégie d'échantillonnage déduisant directement les distributions d'échantillonnage à partir des courbes de coût SGM. Cette approche s'adapte naturellement à l'échelle de la scène avec une paramétrisation minimale, tout en valorisant de manière plus complète l'information contenue dans le volume de coût SGM. Les expériences menées sur trois scènes à deux vues démontrent des améliorations par rapport à SpS-NeRF, permettant d'obtenir un meilleur compromis entre l'attachement à l'*a priori* de profondeur et le détachement vis-à-vis de ses imperfections.

Mots clés : Reconstruction 3D, Données éparées, Champs de radiance neuronaux, Stéréo multi-vue, Volume de coût

Abstract

Digital Surface Models (DSMs) derived from satellite imagery are traditionally generated using Multi-View Stereo (MVS) pipelines, which rely on patch-matching techniques such as Semi-Global Matching (SGM). While effective in few-view settings, these methods exhibit known limitations, for instance when rendering homogeneous surfaces or building borders. Recently, deep learning-based volume rendering techniques, particularly Neural Radiance Fields (NeRF), have emerged as a promising alternative for their performance and their compact, flexible continuous representation of 3D scenes. Although primarily developed for multi-view scene reconstruction, their ray-casting mechanism makes them appealing for sparse-view settings. SparseSat-NeRF (SpS-NeRF) is among the few frameworks that adapt few-view NeRF methodologies to remote sensing. It leverages SGM-derived depth priors to define a depth supervision loss. In the process, patch similarity is interpreted as a proxy for the confidence placed in the prior and used to regulate the attachment to the supervision and to guide ray sampling. In this work, we introduce CVSat-NeRF to refine the exploitation of such priors. First, we replace patch similarity with confidence derived from ambiguity, a more informative proxy that aggregates cost-volume information more effectively. A histogram rescaling step is also proposed to tune the strength of the supervision weighting. Second, we introduce a sampling strategy that directly derives sampling distributions from SGM cost curves. This approach naturally adapts to the scale of the scene with minimal parameterization, while making better use of the SGM cost volume information. Experiments carried out on three two-view scenes demonstrate improvements over SpS-NeRF, achieving a better balance between attachment to the depth prior and detachment from its flaws.

Keywords : 3D reconstruction, Sparse data, Neural radiance fields, Multi-view stereo, Cost volume

1. Introduction

La reconstruction de la surface terrestre en 3D à partir d'images satellitaires constitue un défi fondamental en télédétection, avec des applications couvrant la réponse aux catastrophes, la planification urbaine et le suivi environnemental. Traditionnellement, les *Modèles Numériques de Surface* (MNS) sont générés à l'aide de techniques de *stéréo multi-vues* (MVS) (De Franchis et al., 2014; Youssefi et al., 2020; Rupnik et al., 2017), qui estiment l'élévation en résolvant un problème régularisé de mise en correspondance par patchs à travers plusieurs paires d'images. Bien qu'efficaces à des résolutions grossières avec peu de données en entrée, ces méthodes souffrent de limitations bien connues, notamment des imprécisions dans les régions radiométriquement uniformes (e.g. pelouses) et des artéfacts à proximité des discontinuités de profondeur (e.g. bords de bâtiments) (Fuentes Reyes et al., 2024).

Les avancées récentes exploitent des représentations de scène apprises, en particulier les *Neural Radiance Fields* (NeRF) (Mildenhall et al., 2021) (ou « Champs de Radiance Neuronaux »), qui ajustent un modèle 3D continu, implicitement régularisé, par supervision radiométrique. Cette supervision impose la cohérence entre les images observées et celles synthétisées au moyen de techniques de rendu différentiable fondées sur le lancer de rayons. En utilisant une supervision au niveau du pixel plutôt que du patch, ces approches contournent certains défis liés aux méthodes MVS. Cependant, les formulations NeRF classiques requièrent un nombre important de vues pour atteindre une précision fiable, en raison du caractère mal posé du problème, en particulier en l'absence de priors géométriques. Dans la littérature (Derksen et Izzo, 2021; Marí et al., 2022, 2023), cette contrainte a été atténuée en intégrant des images multi-dates ou en exploitant des priors sur les changements de scène, appris ou inspirés de la physique.

Toutefois, dans de nombreux contextes de télédétection, l'acquisition d'un grand nombre d'images est impraticable ou coûteuse. Si la recherche générale sur les NeRF a exploré l'usage de priors sémantiques (Chen et al., 2021; Yu et al., 2021; Wu et al., 2024; Jain et al., 2021) et géométriques (Deng et al., 2022; Roessle et al., 2022; Niemeyer et al., 2022) pour traiter l'inefficacité liée au manque de données, de telles adaptations demeurent peu étudiées dans le domaine satellitaire. *SparseSat-NeRF* (SpS-NeRF) (Zhang et Rupnik, 2023) constitue une première avancée en intégrant des priors de profondeur issus de prédictions MVS pour superviser l'entraînement du NeRF. Cependant, le fait d'utiliser la similarité entre patchs comme mesure de confiance pour réguler la supervision introduit des limitations. En effet, ce procédé réduit l'information contenue dans le volume de coût calculé par SGM aux seuls scores de similarité des patchs appariés dans une paire d'images, omettant ainsi une grande richesse d'information utile à l'estimation d'incertitude. De plus, le prior de SpS-NeRF nécessite d'ajuster

les paramètres de mise à l'échelle de la confiance γ et m , dépendants de la scène. Pourtant, ce raffinement de supervision est crucial pour s'assurer que le NeRF ne reproduit pas simplement le prior de profondeur, mais améliore effectivement la prédiction MVS en s'en détachant dans les régions potentiellement erronées.

Dans ce travail, en nous appuyant sur les idées et le code de Lulin Zhang et Ewelina Rupnik (Zhang et Rupnik, 2023), nous proposons *Cost Volume Sat-NeRF* (CVSat-NeRF), une intégration plus étroite de NeRF et MVS exploitant le volume de coût calculé lors de l'estimation de disparité par MVS. Nos contributions sont doubles :

- Nous remplaçons la similarité de patchs par une mesure de confiance fondée sur la notion d'ambiguïté (Sarrazin et al., 2021), permettant une agrégation plus fine de l'information contenue dans le volume de coût. Nous suggérons également un rééchantillonnage des histogrammes de confiance afin d'ajuster la force de la supervision.
- Nous interprétons les courbes de coût SGM comme des distributions de probabilité le long des lignes de visée, guidant l'échantillonnage du NeRF au-delà du seul meilleur appariement pour considérer des hypothèses de profondeur alternatives.

2. Travaux connexes

2.1. Stéréorestitution multi-vues

La stéréorestitution multi-vues traditionnelle consiste à calculer des cartes de disparité dans la géométrie épipolaire. Pour un pixel donné p_L de l'image gauche I_L , la disparité peut être déterminée en prenant le minimum de la courbe de coût $CV_{p_L} := \{c(P_L(p_L), P_R(p_L + (0, d)^T)) \mid d \in \mathcal{D}\}$, où $P_s(p_s)$ désigne le patch centré en p_s sur I_s pour $s \in \{L, R\}$, \mathcal{D} un intervalle de disparités, et c une mesure de similarité entre patchs exprimée sous forme de coût. Une régularisation est ensuite généralement appliquée pour décourager les discontinuités de disparité dans la carte résultante. Le *Semi-Global Matching* (SGM) (Hirschmuller, 2005) est bien connu pour offrir un compromis efficace entre précision et coût de calcul. Cependant, il peut échouer aux discontinuités de profondeur ou en présence de variations de conditions d'acquisition ou d'illumination. Des approches MVS fondées sur l'apprentissage (Tosi et al., 2025) ont été étudiées, soit en comparant des caractéristiques 2D apprises pour l'appariement de patchs (Žbontar et LeCun, 2016), soit en travaillant directement avec le volume de coût 3D produit (Kendall et al., 2017), mais elles requièrent des vérités terrain de profondeur coûteuses. Plus récemment, des méthodes *zero-shot* basées sur des modèles fondamentaux entraînés sur des paires synthétiques ont été proposées (Wen et al., 2025), mais leur capacité de généralisation aux données satellitaires reste limitée.

Des travaux récents se sont attachés à fournir des garanties sur la qualité et l'explicabilité des MNS produits, souvent en exploitant l'information contenue dans le volume de coût généré lors de la mise en correspondance stéréo. Plusieurs travaux utilisent cette information pour définir des mesures de confiance (Poggi et al., 2017). Parmi eux, Sarrazin et al. (Sarrazin et al., 2021) introduisent la notion d'ambiguïté afin de mettre en évidence les pixels d'une carte de disparité pour lesquels aucun extremum net de la courbe de coût associée ne se démarque. Par la suite, Malinowski et al. (Malinowski et al., 2024) sont les premiers à définir des intervalles de disparité robustes en convertissant les courbes de coût en distributions de possibilité. L'application d'un seuil α à une telle distribution définit une « α -coupe » en dehors de laquelle les valeurs de disparité ont une probabilité inférieure à α d'être la disparité réelle. Dans notre contexte NeRF, cette idée est réutilisée pour définir des intervalles le long des rayons dans lesquels collecter des échantillons pour le rendu.

2.2. Neural Radiance Fields pour la génération de MNS avec données éparses

Des travaux récents ont de plus en plus appliqué des représentations implicites apprises de scènes 3D à la télédétection, obtenant des résultats surpassant les approches MVS traditionnelles (Qu et Deng, 2023; Zhang et al., 2024; Aira et al., 2025). Parmi ces méthodes, les NeRF ont été largement étudiés pour leur capacité d'adaptation à l'imagerie multi-date, permettant l'exploitation de séries temporelles afin d'améliorer la précision de la reconstruction (Derksen et Izzo, 2021; Marí et al., 2023). Néanmoins, les méthodes NeRF présentent également un fort potentiel dans des scénarii à vues parcimonieuses, du fait qu'elles fonctionnent par comparaison de pixels, contournant certaines difficultés des techniques fondées sur les patches, et qu'elles permettent l'intégration de priors sémantiques appris. Toutefois, elles peinent à converger dans des configurations à très faible nombre de vues.

Pour remédier à cette difficulté, des travaux antérieurs exploitent des priors sémantiques appris à partir de réseaux de neurones 2D préentraînés (Yu et al., 2021; Chen et al., 2021; Jain et al., 2021). Plus récemment, des modèles de diffusion ont été employés comme priors sémantiques en comparant les vues générées par NeRF avec celles générées par diffusion (Wu et al., 2024). Des priors géométriques ont aussi été utilisés pour régulariser l'apprentissage NeRF en imposant la régularité des normales aux surfaces apprises (Niemeyer et al., 2022). DSNerf (Deng et al., 2022) incorpore une supervision de profondeur éparsée en exploitant des points 3D obtenus par *Structure-from-Motion* (SfM). La supervision de profondeur est imposée en alignant les distributions d'opacité le long des rayons calculées par NeRF avec des fonctions delta de Dirac placées aux profondeurs calculées par SfM. D'autres méthodes (Wei et al., 2021; Roessle et al., 2022) utilisent des réseaux de complétion de profondeur pour produire des cartes de profondeur

denses et des cartes d'incertitude Σ . Durant l'entraînement, les rayons sont échantillonnés à proximité de la profondeur *a priori* avec une variance dérivée de Σ .

SparseSat-NeRF (SpS-NeRF) (Zhang et Rupnik, 2023) combine des éléments de Sat-NeRF (Marí et al., 2022) et de DDP-NeRF (*Dense Depth Priors for NeRF*) (Roessle et al., 2022). De Sat-NeRF, il ne conserve que l'adaptation aux modèles RPC, puisqu'il se concentre sur une imagerie synchrone. Contrairement à DDP-NeRF, il repose exclusivement sur une estimation dense de profondeur MVS plutôt que sur un réseau de complétion de profondeur, motivé par le fait que la télédétection ne présente pas les défis propres aux environnements intérieurs étudiés dans les travaux précités. Par conséquent, pour évaluer la confiance dans le prior de profondeur, il remplace les estimations d'incertitude apprises par des scores de corrélation croisée calculés lors de la mise en correspondance par patches.

3. Méthodologie

3.1. Préliminaires

Neural Radiance Fields. NeRF (Mildenhall et al., 2021) représente la scène 3D par un réseau neuronal $F_{\Theta} : (\mathbf{X}, \mathbf{d}) \mapsto (c, \sigma)$, où \mathbf{X} et \mathbf{d} désignent respectivement la position 3D et la direction de visée en entrée, tandis que c et σ sont la couleur et l'opacité prédites. Pour restituer l'image associée à un point de vue donné, chaque pixel est généré en intégrant la lumière transmise le long du rayon \mathbf{r} de direction \mathbf{d} , issu du centre \mathbf{o} du capteur et traversant ce pixel :

$$\mathbf{C}(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \sigma(\mathbf{r}(t)) \mathbf{c}(\mathbf{r}(t), \mathbf{d}) dt, \quad (1)$$

où $T(t) = \exp(-\int_{t_n}^t \sigma(\mathbf{r}(s)) ds)$. En pratique, cette intégration est discrétisée par un échantillonnage du rayon en N positions $(t_i)_{1 \leq i \leq N}$ entre les bornes proche t_n et lointaine t_f définissant le contenu de la scène :

$$\mathbf{C}(\mathbf{r}) = \sum_{n=1}^N T_n \alpha_n \mathbf{c}_n, \quad (2)$$

avec $\alpha_n = (1 - \exp(-\sigma_n \delta_n))$, $T_n = \exp(-\sum_{l=1}^{n-1} \sigma_l \delta_l)$ et $\delta_n = t_{n+1} - t_n$. Les poids $w_n = T_n \alpha_n$ sont alors interprétés comme la distribution de lumière transmise le long de \mathbf{r} .

L'estimation de profondeur et son écart type le long d'un rayon \mathbf{r} s'écrivent :

$$D(\mathbf{r}) = \sum_{n=1}^N w_n t_n, \quad (3)$$

$$S(\mathbf{r})^2 = \sum_{n=1}^N w_n (t_n - D(\mathbf{r}))^2. \quad (4)$$

SparseSat-NeRF. Soit N_I le nombre d'images d'entrée. SpS-NeRF commence par exécuter N_I traitements SGM sur les images d'entrée sous-échantillonnées. Chaque exécution $i \in \{1, \dots, N_I\}$ génère un nuage de points

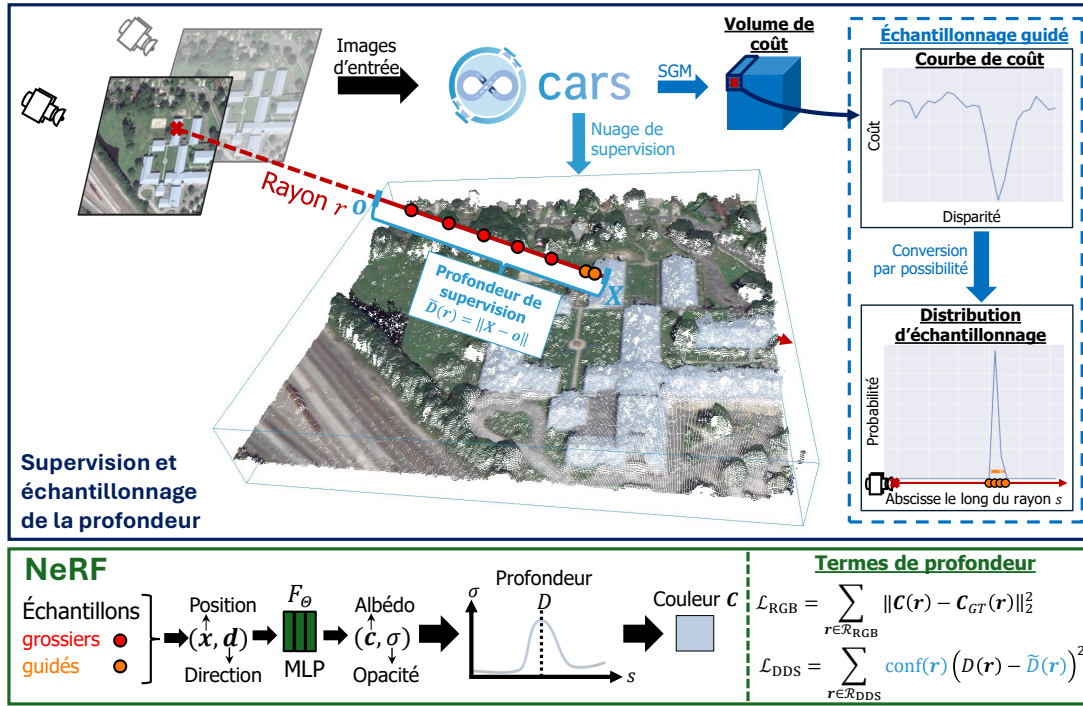


Figure 1. Vue d'ensemble de la méthodologie de CVSat-NeRF. La scène est représentée par un NeRF, qui encode l'opacité et la couleur dans un réseau neuronal F_{Θ} et synthétise de nouvelles vues par lancer de rayons et composition de couleurs. Comme dans SpS-NeRF (Zhang et Rupnik, 2023), la prédiction d'une chaîne stéréo (CARS dans notre cas) est utilisée pour superviser l'apprentissage de la scène via un terme de supervision \mathcal{L}_{DDS} et pour guider l'échantillonnage des rayons. Contrairement à SpS-NeRF, le guidage de l'échantillonnage est dérivé du volume de coût SGM en s'appuyant sur la théorie des possibilités (Malinowski et al., 2024).

\tilde{X}_i projeté en une carte de profondeur \tilde{D}_i . La carte de similarité entre patches mesurée par SGM est également conservée, normalisée et interprétée comme carte de confiance conf_i . Ces quantités sont définies sur les pixels éligibles en géométrie épipolaire, notés V_i . Elles servent à formuler un terme de perte sur la profondeur. Un lot de rayons $\mathcal{R} = \mathcal{R}_{\text{RGB}} \cup \mathcal{R}_{\text{DDS}}$ est défini, où \mathcal{R}_{RGB} et \mathcal{R}_{DDS} sont les rayons radiométriques et ceux supervisés par la profondeur. En notant \bar{C} la couleur de référence, la fonction de perte totale s'écrit :

$$\mathcal{L}(\mathcal{R}) = \mathcal{L}_{\text{RGB}}(\mathcal{R}_{\text{RGB}}) + \lambda_{\text{DDS}} \mathcal{L}_{\text{DDS}}(\mathcal{R}_{\text{DDS}}), \quad (5)$$

avec λ_{DDS} un hyperparamètre fixé à la main, et

$$\mathcal{L}_{\text{RGB}}(\mathcal{R}_{\text{RGB}}) = \sum_{\mathbf{r} \in \mathcal{R}_{\text{RGB}}} \|\mathbf{C}(\mathbf{r}) - \bar{\mathbf{C}}(\mathbf{r})\|_2^2, \quad (6)$$

$$\mathcal{L}_{\text{DDS}}(\mathcal{R}_{\text{DDS}}) = \sum_{\mathbf{r} \in \mathcal{R}_{\text{DDS}}} \text{conf}(\mathbf{r}) \left(D(\mathbf{r}) - \tilde{D}(\mathbf{r}) \right)^2. \quad (7)$$

Pour encourager un détachement progressif de la supervision lorsque la convergence vers le prior est jugée suffisante, un filtrage est appliqué aux rayons de supervision. Pour ce faire, une carte d'incertitude Σ est définie par $\Sigma = \gamma(1 - \text{conf}) + m$, où γ et m sont des paramètres d'échelle fixés empiriquement. Précisément, Σ est ensuite multiplié par la plage de la profondeur de supervision, afin de mettre l'incertitude à l'échelle de la distribution de profondeur de la scène. Un rayon \mathbf{r} est retenu dans \mathcal{R}_{DDS} si l'une des conditions suivantes est vérifiée : $S(\mathbf{r}) > \Sigma(\mathbf{r})$ ou $\left| (D(\mathbf{r}) - \tilde{D}(\mathbf{r})) \right| > \Sigma(\mathbf{r})$.

Échantillonnage gaussien des rayons. SpS-NeRF introduit un échantillonnage guidé comme alternative à l'échantillonnage hiérarchique classique des NeRF. Pour un rayon \mathbf{r} , deux groupes d'échantillons $t^e = \{t_n^e\}_{n \leq \lfloor N/2 \rfloor}$ et $t^f = \{t_n^f\}_{n > \lfloor N/2 \rfloor}$ sont générés séquentiellement. Les premiers t^e points sont échantillonnés uniformément le long du rayon. Les t^f suivants sont échantillonnés selon une loi normale $\mathcal{N}(\tilde{D}(\mathbf{r}), \Sigma(\mathbf{r}))$ durant l'apprentissage et lorsque l'*a priori* de profondeur est disponible, et selon $\mathcal{N}(D(\mathbf{r}), S(\mathbf{r}))$ dans les autres cas.

Ambiguïté. Nous utilisons plus loin la notion d'ambiguïté introduite par Sarrazin et al. (2021). Une courbe de coût CV_p est dite ambiguë lorsqu'il est difficile d'en extraire une disparité fiable, en raison de la présence de plusieurs extrema candidats. L'ambiguïté est définie, pour un volume de coût normalisé CV , un pixel p et un seuil $\eta \in [0, 1]$, par $\text{Amb}(p, \eta) = \text{Card}(\mathcal{D}_{\text{amb}})$ où :

$$\mathcal{D}_{\text{amb}} = \left\{ d \in \mathcal{D} \mid \text{CV}_p(d) < \min_{d' \in \mathcal{D}} \text{CV}_p(d') + \eta \right\}. \quad (8)$$

L'intégrale d'ambiguïté est définie comme l'aire sous la courbe d'ambiguïté :

$$\text{Amb}_{\text{int}}(p) = \int_{\eta=0}^1 \text{Amb}(p, \eta) d\eta. \quad (9)$$

α -coupes. Malinowski et al. (2024) définissent des intervalles de disparité devant contenir la disparité réelle dans

$\alpha\%$ des cas, en s'appuyant sur la théorie des possibilités (Walley, 1991). Une distribution de possibilité est une application $\pi : \Omega \rightarrow [0, 1]$ telle que $\exists \omega \in \Omega, \pi(\omega) = 1$. En stéréovision, $\Omega = \mathcal{D}$ et $\pi(\omega)$ représente la plausibilité d'une disparité, déduite des coûts SGM. Pour définir π_k à partir des courbes de coût CV_k pour $k \in V$, les valeurs de coût sont d'abord ramenées dans l'intervalle $[0, 1]$ au moyen d'une normalisation min-max appliquée à l'ensemble du volume de coût, ce qui fournit $CV_k^{\text{norm}}, k \in V$. Ensuite, la fonction $d \mapsto CV_k^{\text{norm}}(d)$ est décalée verticalement de manière à ce que son maximum soit égal à 1, produisant ainsi une distribution de possibilité π_k associée au rayon k . À partir d'une distribution de possibilité π peut être défini un ensemble de distributions de probabilité (Dubois et Prade, 1992) :

$$\mathcal{P}_\pi = \left\{ P \mid \forall E \in 2^\Omega, P(E) \leq \sup_{\omega \in E} \pi(\omega) \right\}. \quad (10)$$

Pour un rayon donné $k \in V$, Malinowski et al. proposent de se restreindre aux probabilités de disparité appartenant à l'ensemble \mathcal{P}_{π_k} . Puis, pour $\alpha \in [0, 1]$, l' α -coupe $C_\alpha^{\pi_k}$ est définie par :

$$\begin{aligned} C_\alpha^{\pi_k} &:= \{ \omega \in \Omega \mid \pi_k(\omega) \geq \alpha \} \\ &= \{ \omega \in \Omega \mid \exists P \in \mathcal{P}_{\pi_k}, P(\omega) \geq \alpha \}. \end{aligned} \quad (11)$$

Les disparités $d \notin C_\alpha^{\pi_k}$ sont alors considérées comme improbables au niveau de confiance α .

3.2. CVSat-NeRF

Dans SpS-NeRF, les mécanismes de régulation de la supervision — pondération du terme de supervision de profondeur, sélection des rayons et échantillonnage le long des rayons — sont essentiels pour se détacher de la supervision dans les zones incertaines et gérer les ambiguïtés de similarité de patches. Toutefois, les métriques de similarité utilisées et les mesures d'incertitude qui en dérivent semblent sous-optimales pour détecter les zones d'ambiguïté. CVSat-NeRF vise à exploiter davantage d'informations issues du volume de coût.

Génération du prior de profondeur. Comme dans SpS-NeRF, N_I exécutions de SGM sont réalisées. En plus des cartes de profondeur \tilde{D}_i , un volume de coût épipolaire optimisé par SGM CV_i^e et une carte d'ambiguïté A_i sont produits, où $\forall k \in V_i, A_i(\mathbf{p}_k) = \text{Amb}_{\text{int}}(\mathbf{p}_k)$.

Confiance issue de l'ambiguïté. Nous remplaçons la similarité de patches par une confiance dérivée de l'intégrale d'ambiguïté :

$$\text{conf}_{\text{amb}}(\mathbf{p}) = 1 - \text{Amb}_{\text{int}}(\mathbf{p}). \quad (12)$$

L'ambiguïté constitue un indicateur plus fiable de la qualité du prior que la similarité de patches. Par exemple, comme illustré en Figure 6, la similarité peut être élevée dans des régions uniformes où l'appariement est pourtant peu fiable (eau, toits uniformes inclinés), alors que conf_{amb} reste faible dans ces zones. L'objectif est ainsi d'obtenir un meilleur compromis entre détachement de la supervision dans les régions erronées et conservation là où le prior est pertinent.

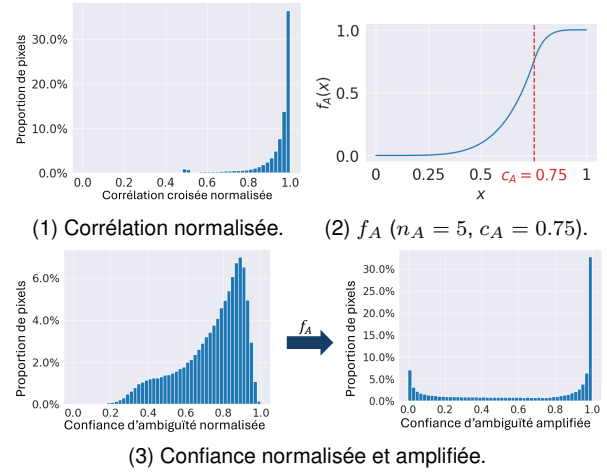


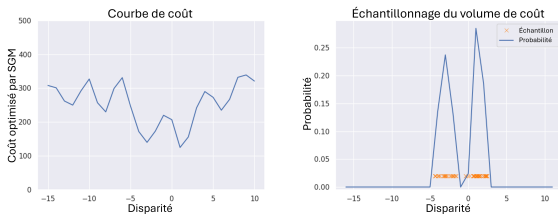
Figure 2. Histogrammes de conf pour différents indicateurs de confiance et amplification d'histogramme.

Amplification de la pondération de la supervision. La confiance, qu'elle provienne de la similarité ou de l'ambiguïté, impose une pondération arbitraire des termes de supervision. En particulier, l'histogramme de conf_i dépend du couple d'images, mais aussi de la métrique employée (comme illustré en Figure 2 sur les deux histogrammes de gauche). La corrélation croisée de SpS-NeRF produit typiquement une distribution binaire, alors que l'ambiguïté ou Census génèrent des valeurs plus progressives. Aucun profil n'est optimal dans toutes les situations. Pour mieux contrôler la force de la pondération, nous introduisons une fonction optionnelle de reparamétrisation d'histogramme. Les valeurs de confiance sont transformées via une fonction sigmoïde amplificatrice f_A , qui rapproche les valeurs élevées de 1 et les faibles de 0. Nous utilisons une fonction polynomiale par morceaux, de degré n_A et centrée en $c_A \in [0, 1]$ (Figure 2, (2)). Une valeur élevée de n_A accentue la séparation de part et d'autre de c_A .

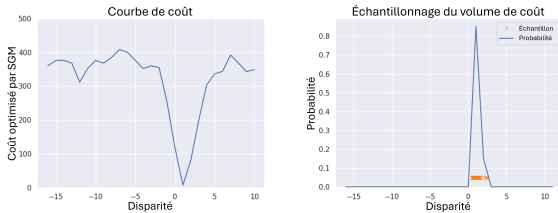
Échantillonnage guidé par le volume de coût. Malgré ses avantages, l'échantillonnage gaussien de SpS-NeRF manque de fondement géométrique : l'interprétation de Σ comme une distance le long du rayon dépend de l'ajustement de γ et m en fonction de la plage de profondeur de la scène. Or, dès que $\text{conf}(\mathbf{r})$ n'est plus proche de 1, Σ devient excessivement grande, rendant le critère de filtrage trop restrictif et écartant trop tôt des rayons de supervision pourtant fiables. Par conséquent, nous désactivons la sélection de rayons durant l'apprentissage et proposons une stratégie d'échantillonnage alternative. L'échantillonnage à l'inférence reste inchangé.

Plutôt que d'échantillonner près d'un minimum global de CV_k , nous exploitons un ensemble de minima locaux afin de prendre en compte d'autres profondeurs candidates. La courbe de coût est traitée comme une distribution guidant la terminaison du rayon pour générer les t^f . La procédure comporte trois étapes :

1. Convertir le volume de coût épipolaire CV^e en courbes CV_k pour chaque pixel $k \in V$, en triangu-



(1) Rayon ambigu.



(2) Rayon non ambigu.

Figure 3. Courbes de coût et échantillons dérivés.

lant chaque correspondance $(\mathbf{p}_k, \mathbf{p}_k + (0, d)^T)$ pour chaque $d \in \mathcal{D}$.

2. Transformer CV_k en distribution de probabilité P_k .
3. Échantillonner P_k par transformée inverse.

Pour effectuer la conversion d'une courbe de coût CV_k en une distribution P_k , nous nous inspirons de Malinowski et al. (2024). Ainsi, nous convertissons CV_k en une distribution de possibilité π_k et excluons les disparités en dehors de l' α -coupe associée. Plus précisément, bien que l'on puisse échantillonner uniformément au sein des α -coupes, nous préférons préserver les variations internes de possibilité. Nous commençons ainsi par définir :

$$\forall d \in \mathcal{D}, \tilde{P}_k(d) = \begin{cases} 0 & \text{si } \pi_k(d) < \alpha, \\ \pi_k(d) - \alpha & \text{si } \pi_k(d) \geq \alpha. \end{cases} \quad (13)$$

\tilde{P}_k est ensuite normalisé pour obtenir P_k . Des exemples de courbes de coût (en géométrie épipolaire) transformées en distributions de probabilité pour $\alpha = 0.8$ sont données en Figure 3, pour un rayon ambigu ($\text{conf}_{\text{amb}} = 0.4$) et non ambigu ($\text{conf}_{\text{amb}} = 0.9$).

4. Expériences

4.1. Jeu de données

Les expériences ont été menées sur des scènes satellitaires issues de la Track 3 du jeu de données du 2019 IEEE GRSS Data Fusion Contest (DFC2019, Le Saux et al., 2019). Ce jeu de données comprend 26 images Maxar WorldView-3 collectées au-dessus de la ville de Jacksonville (Floride, États-Unis) à différentes dates, entre octobre 2014 et février 2016. Les images RGB ont une résolution au sol d'environ 35cm au nadir, et des images d'environ 800×800 pixels sont extraites, soit $280\text{m} \times 280\text{m}$. Un modèle LiDAR de référence est également fourni, avec une résolution de 0.5m/px, sans garantie de synchronicité avec les images RGB.

Parmi ces données, nous sélectionnons les zones d'intérêt 072, 105 et 214, présentant plusieurs défis typiques pour les méthodes MVS : zones radiométriquement homogènes, discontinuités de profondeur, petits bâtiments souvent atrophiés par SGM, mais aussi des grands bâtiments. Deux images sont utilisées pour l'entraînement et une (deux pour la scène 214) pour le test. Pour les scènes 072 et 105, on choisit comme images d'entraînement les images étiquetées 1 et 2 dans le jeu de données DFC2019 et comme image de test l'image 4. Pour la scène 214, on utilise les images 9 et 10 pour l'entraînement et 7 et 8 pour le test. Les images de test ont un double rôle : évaluer la synthèse de nouvelles vues, indiquant notamment la bonne convergence du modèle, et fournir un point de vue au nadir à partir duquel le NeRF génère la carte de profondeur ensuite rasterisée en MNS. Ces images sont donc choisies en trouvant un compromis entre leur proximité temporelle avec les images d'entraînement, et leur faible angle de dépointage (respectivement 7.3° , 11.7° et 14.6° pour les images 4, 7 et 8).

Dans notre cadre de vues éparses et synchrones, les vues d'entraînement ont été choisies aussi synchrones que possible, avec un écart temporel de l'ordre de 10s. L'image de test 4 a été prise 13 jours après les images 1 et 2, tandis que les images 8, 9 et 10 sont synchrones et ont été prises 25 jours après l'image 7. Enfin, le rapport B/H entre les vues d'entraînement est de 0.15 pour la paire (1, 2) et 0.17 pour la paire (9, 10).

4.2. Détails d'implémentation

Le prior de supervision est généré avec la chaîne photogrammétrique CARS (Youssefi et al., 2020). Pour la mise en correspondance dense, SGM utilise Census comme mesure de similarité, sur des patches de 5px de côté, et Vfit (Haller et al., 2010) pour l'affinage de disparité sous-pixellique.

L'architecture NeRF est identique à celle de Sat-NeRF (Marí et al., 2022), sans correction solaire ni prédiction d'objets transitoires, afin de respecter notre scénario synchrone. Un seul modèle est employé, car l'échantillonnage guidé remplace l'échantillonnage hiérarchique. La dépendance à l'angle de vue d est désactivée, la scène ne présentant pas d'effets spéculaires. Suivant EO-NeRF (Marí et al., 2023), la scène 3D est exprimée en coordonnées UTM. Nous fixons $\lambda_{\text{DDS}} = \frac{1}{3}$ pour les scènes 072 et 105, et $\lambda_{\text{DDS}} = \frac{100}{3}$ pour la scène 214. L'entraînement utilise l'optimiseur Adam avec un taux d'apprentissage initial de 5×10^{-4} , décroissant d'un facteur 0.9 à chaque époque. La taille de lot est de 1024 rayons et $N = 128$ échantillons sont pris par rayon. Les modèles sont entraînés durant $80k$ itérations (environ 60 époques). Les expériences ont été exécutées sur un GPU A100 avec 20 Go de RAM. Un entraînement nécessite typiquement 2h10min, auxquels s'ajoutent 1h10min pour la génération du volume de coût dans le cas de CVSat-NeRF.

Dans l'expérience principale, nous comparons CARS, SpS-NeRF et CVSat-NeRF. Le résultat de CARS

est le même que celui utilisé comme prior de profondeur pour les méthodes NeRF. Nous utilisons l'implémentation originale de SpS-NeRF (pondération de la supervision, sélection de rayons, échantillonnage gaussien), en remplaçant toutefois la similarité entre patch par Census et en fixant γ à 0.2 pour les scènes 072 et 105, et 0.04 pour la scène 214. Les images ne sont jamais sous-échantillonnées avant exécution de CARS. CVSat-NeRF est exécuté avec la confiance dérivée de l'ambiguïté pour pondérer les pertes, avec l'échantillonnage issu du volume de coût, et en désactivant la sélection de rayons. La confiance dérivée de l'ambiguïté est amplifiée avec la fonction f_A de la Figure 2 (paramètres $n_A = 5$, $c_A = 0.75$ pour la scène 105, et $n_A = 3$, $c_A = 0.7$ pour les autres scènes) et nous fixons $\alpha = 0.8$ pour l'échantillonnage du volume de coût pour toutes les scènes.

4.3. Métriques d'évaluation

Deux modalités d'évaluation sont employées. Premièrement, la synthèse de vues est évaluée en moyennant sur toutes les images de test les Peak Signal-to-Noise Ratios (PSNR) mesurés entre les images synthétisées et les images de test. Deuxièmement, la génération de MNS est évaluée en comparant d'une part le MNS issu du prior MVS et d'autre part celui produit par le NeRF supervisé, afin d'identifier les zones d'amélioration ou de dégradation.

Génération de MNS. Pour MVS comme pour NeRF, les cartes de profondeur sont converties en nuages de points puis rasterisées à la résolution du MNS LiDAR. La qualité des MNS est évaluée en termes d'erreurs relatives pour réduire l'impact d'éventuels décalages absolus induits par une calibration imparfaite des RPC. Ainsi, avant tout calcul, le MNS est recalé horizontalement et verticalement sur le MNS LiDAR.

L'erreur absolue moyenne (MAE) est alors calculée entre la vérité terrain (VT) et le MNS estimé. La MAE manque cependant d'interprétabilité vis-à-vis des déformations horizontales. Pour limiter ce biais, si MNS_{pred} et MNS_{LiDAR} désignent respectivement les MNS prédit et LiDAR, définis sur un domaine Ω , un pixel $p \in \Omega$ est dit qualifié s'il est valide et vérifie $\|MNS_{pred}(p) - MNS_{LiDAR}(p)\| < 1m$.

Le taux de qualification (QR) est alors :

$$QR = \frac{\sum_{p \in \Omega} \mathbb{1}\{p \text{ qualifié}\}}{|\Omega|} \quad (14)$$

où $\mathbb{1}$ est la fonction indicatrice. Dans nos expériences, Ω correspond à l'ensemble du MNS LiDAR.

Comparaison avec la supervision. Le MNS de supervision contient des pixels marqués comme invalides (correspondant aux inconsistances détectées par SGM via la vérification de consistance gauche-droite, ou aux trous liés aux occultations). Ainsi, pour les sorties MVS, la MAE n'est calculée que sur les pixels valides. Pour NeRF, nous distinguons MAE_{in} et MAE_{out} , calculés respectivement sur les pixels valides et invalides de SGM.

Afin de quantifier la proportion d'amélioration ou de dégradation apportée par le NeRF par rapport au prior, nous définissons également QR_{qual} (resp. QR_{unqual}) en remplaçant Ω dans la définition ci-dessus par l'ensemble des pixels qualifiés (resp. non qualifiés) de la sortie MVS.

4.4. Résultats

Comparaison entre SpS-NeRF et CVSat-NeRF. Le Tableau 1 présente les résultats quantitatifs. Concernant la synthèse de vues, CVSat-NeRF obtient un PSNR légèrement supérieur à celui de SpS-NeRF, sur l'ensemble des scènes étudiées. La Figure 4 montre que, bien que les deux méthodes reconstruisent correctement la scène, CVSat-NeRF produit des images plus nettes, notamment sur les bords de bâtiments.

Concernant la reconstruction altimétrique, les deux modèles améliorent le résultat de CARS en termes de MAE et de QR sur les scènes 072 et 105, mais pas sur la scène 214. En outre, CVSat-NeRF surpasse SpS-NeRF en termes de QR sur toutes les scènes, bien que les deux méthodes offrent des compromis différents entre QR_{qual} et QR_{unqual} sur la scène 214. Ces résultats indiquent que CVSat-NeRF préserve davantage de pixels qualifiés de CARS (QR_{qual} plus élevé), tout en s'affranchissant plus efficacement des zones erronées du prior (QR_{unqual} en général plus élevé).

La Figure 5 illustre qualitativement ces observations : les deux NeRF produisent des MNS légèrement plus réguliers et plus nets, avec des contours d'objets mieux

Scène	Méthode	PSNR	MAE _{in} ↓	MAE _{out} ↓	QR (%) ↑	QR _{qual} (%) ↑	QR _{unqual} (%) ↑
072	CARS	-	1.03	-	71.8	-	-
	SpS-NeRF	20.0	0.99	2.60	72.8	89.2	31.2
	CVSat-NeRF	20.5	0.93	2.55	73.9	89.8	33.5
105	CARS	-	1.57	-	57.8	-	-
	SpS-NeRF	21.3	1.41	2.49	60.7	83.1	29.9
	CVSat-NeRF	21.7	1.34	2.06	61.6	83.2	32.0
214	CARS	-	2.16	-	56.4	-	-
	SpS-NeRF	21.1	2.33	10.70	54.0	82.9	16.4
	CVSat-NeRF	21.4	2.18	10.92	55.5	87.9	13.4

Table 1. Résultats quantitatifs. Les meilleures performances sont indiquées en **gras**.

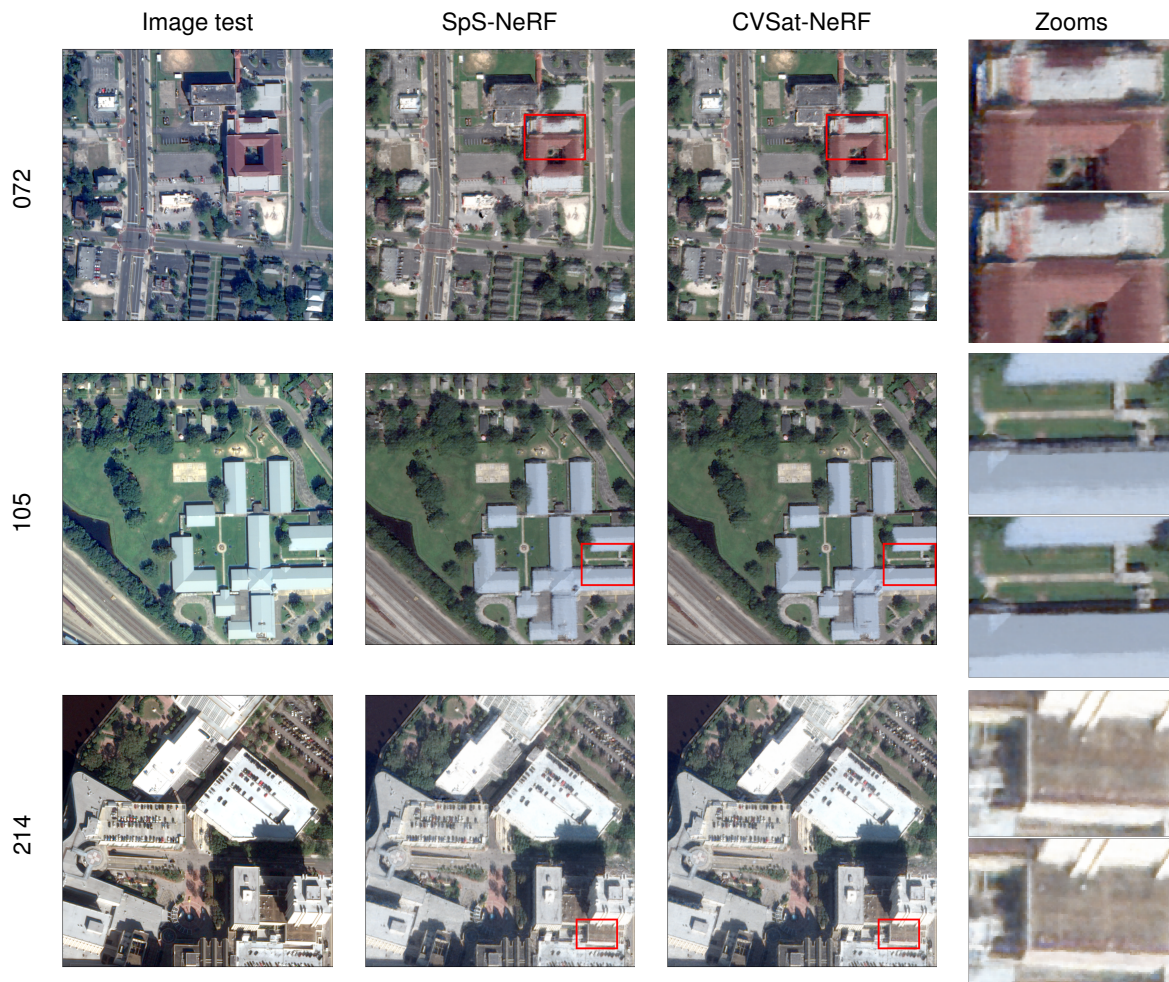


Figure 4. Résultats qualitatifs pour la synthèse de nouvelles vues sur trois scènes différentes (072, 105, 214). CVSat-NeRF présente moins d'artefacts flous et des détails plus nets, en particulier près des bords de bâtiments comme illustré sur les zooms (haut : SpS-NeRF, bas : CVSat-NeRF).

définis, en particulier pour CVSat-NeRF. Néanmoins, les reconstructions s'accompagnent toujours d'une perte de précision par rapport à CARS dans certaines régions de la scène. Quelques améliorations (en vert) et dégradations (en violet) sont encadrées sur les cartes de différence d'erreur absolue par rapport à CARS. Sur la scène 072, on observe que le bâtiment carré (encadré 1) et les pavillons en bas de la scène (2) ont des bords légèrement plus nets, notamment avec CVSat-NeRF. Les routes délimitées par un terre-plein (3) sont plus planes dans les deux méthodes NeRF. En revanche, certains bâtiments uniformes (4) sont moins bien reconstruits, un phénomène légèrement moins marqué avec CVSat-NeRF. Sur la scène 105, on trouve à nouveau un sol plus lisse avec les NeRF sur la pelouse (5), ainsi que des bords de bâtiments mieux délimités (6), surtout avec CVSat-NeRF. En revanche, certains éléments de détails sont moins bien distingués qu'avec CARS, notamment le perron (7) qui est mieux préservé par SpS-NeRF, tandis que la passerelle en « S » (8) est mieux préservée par CVSat-NeRF. Enfin, la scène 214 met davantage en difficulté les deux méthodes NeRF. En particulier, on observe des reconstructions en pente douce au lieu de bords

de bâtiment abrupts dans les zones d'occlusion derrière les grands bâtiments (9), ainsi qu'une hypertrophie des faces visibles des bâtiments (10). Une explication possible de ces observations réside dans la hauteur des bâtiments et dans les grands angles de dépointage des images d'entraînement (respectivement 21.7° et 29.0° pour les images 9 et 10 contre 19.3° et 12.9° pour les images 1 et 2). En effet, le biais inductif des réseaux de neurones utilisés dans les NeRF les amène à favoriser des formes régulières, ce qui aboutit à des interpolations lisses dans les zones d'occlusion.

Ablation. Le Tableau 2 présente une étude d'ablation menée sur la scène 105. La configuration de base (#1) utilise Census comme pondération, sans amplification de confiance, et un échantillonnage gaussien des rayons, sans filtrage des rayons. Remplacer Census par l'ambiguïté (#3) améliore légèrement l'ensemble des métriques. Introduire l'échantillonnage par volume de coût (#5) améliore davantage les résultats, notamment en augmentant QR_{unqual} , tout en supprimant la nécessité de régler les paramètres m et γ .

L'amplification de confiance (#2, #4, #6) améliore en général les performances et fournit les meilleurs résul-

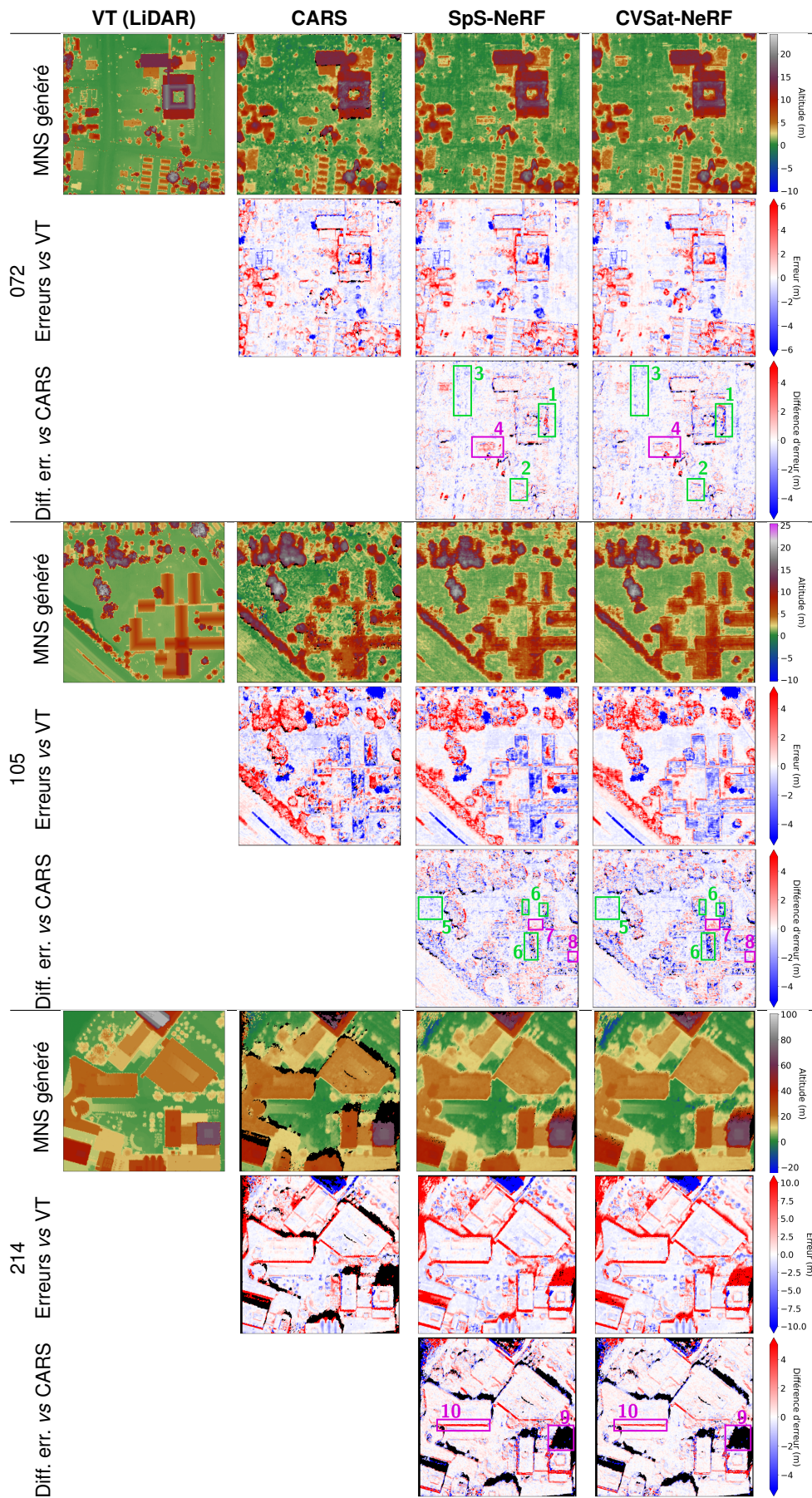


Figure 5. Comparaison qualitative des méthodes CARS, SpS-NeRF et CVSat-NeRF sur les 3 scènes d'intérêt. Pour chaque scène, la première ligne présente le MNS reconstruit; la deuxième ligne présente l'erreur de reconstruction $err_{meth} = MNS_{meth} - MNS_{VT}$; la troisième ligne présente la différence d'erreur par rapport à CARS $|err_{meth}| - |err_{CARS}|$.

#	Méthode			PSNR	MAE _{in} ↓	MAE _{out} ↓	QR(%) ↑	QR _{qual} (%) ↑	QR _{unqual} (%) ↑
	Confiance	Ampl.	Ech.						
1	Census		Gauss.	21.1	1.45	2.58	59.8	82.4	28.9
2	Census	✓	Gauss.	<u>21.4</u>	1.42	2.54	60.3	82.9	29.4
3	Amb.		Gauss.	<u>21.4</u>	1.42	2.43	60.3	83.1	29.0
4	Amb.	✓	Gauss.	<u>21.4</u>	1.41	2.34	60.2	81.4	<u>31.2</u>
5	Amb.		CV	21.7	<u>1.37</u>	<u>2.18</u>	<u>61.2</u>	83.7	30.3
6	Amb.	✓	CV	21.7	1.34	2.06	61.6	<u>83.2</u>	32.0

Table 2. Résultats d'ablation. Les meilleures performances sont indiquées en **gras**, les deuxièmes meilleures sont soulignées. Ampl. = Amplification par rééchantillonnage d'histogramme ; Ech. = Échantillonnage des rayons ; Amb. = Métrique intégrale d'ambiguïté ; Gauss. = Gaussien ; CV = Volume de coût

tats en combinaison avec l'échantillonnage du volume de coût (#6), au prix de l'introduction des paramètres supplémentaires (n_A, c_A). Elle favorise un plus fort détachement du prior, particulièrement lorsque l'ambiguïté est utilisée : dans ce cas, QR_{unqual} augmente au détriment de QR_{qual} , car l'ambiguïté produit des cartes de confiance plus contrastées, identifiant davantage de pixels non qualifiés comme très incertains, au risque de classer certains pixels qualifiés comme peu certains. À l'inverse, Census fournit une distribution plus uniforme, moins adaptée pour isoler les zones erronées. Cette différence est illustrée en Figure 6.

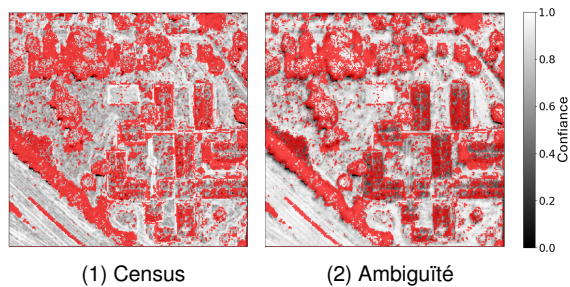


Figure 6. Pixels non qualifiés de CARS (rouge) superposés à la confiance pour la scène 105. Les mauvaises prédictions CARS sont mieux contenues dans les zones de faible confiance en utilisant l'ambiguïté que Census.

5. Discussion et conclusion

Nous avons introduit CVSat-NeRF, une reformulation de SpS-NeRF intégrant deux principales améliorations. Premièrement, la similarité entre patches est remplacée par une estimation plus fine de la qualité du prior de profondeur, à savoir la confiance dérivée de l'ambiguïté, utilisée pour pondérer la supervision en profondeur. Nous proposons optionnellement une amplification par rééchantillonnage de l'histogramme de confiance, permettant d'ajuster plus finement cette pondération. Deuxièmement, pour l'échantillonnage des rayons, nous substituons à la distribution gaussienne une distribution de probabilité de terminaison de rayon directement issue du volume de coût. Cette modification permet non seulement une exploitation plus riche du prior, mais fournit également une interprétation physiquement cohérente de l'échantillonnage, tout en réduisant la nécessité d'un

réglage paramétrique complexe. Les résultats montrent des améliorations nettes pour la synthèse de vues sur l'ensemble des scènes et pour la reconstruction géométrique sur les scènes 072 et 105.

Nos idées visant à hybrider MVS et NeRF pour la reconstruction en vues parcimonieuses peuvent être étendues de plusieurs manières. Premièrement, le mécanisme de sélection des rayons de SpS-NeRF, supprimé dans CVSat-NeRF, pourrait être réintroduit avec des règles de filtrage adaptées, par exemple en écartant les rayons de supervision dont la profondeur prédite appartient déjà aux α -coupes. En outre, les régularisations géométriques et sémantiques mentionnées en Section 2 méritent d'être explorées et pourraient combler les lacunes résiduelles des méthodes testées, comme l'hyperthrophie de certaines façades de bâtiment. Par ailleurs, le coût calculatoire de l'implémentation actuelle reste conséquent et pourrait être réduit. D'une part, l'inférence du réseau peut être accélérée grâce aux implémentations récentes reposant sur le hash encoding (Billouard et al., 2024). D'autre part, notre méthode de conversion des courbes de coût depuis la géométrie épipolaire vers des distributions d'échantillonnage pour NeRF est coûteuse, car elle applique la triangulation des lignes de visée pour chaque point du volume de coût. Ainsi, $\mathcal{O}(h \times w \times |D|)$ appels aux modèles RPC sont effectués (où h et w sont les dimensions de l'image et D la plage des valeurs de disparité candidates). Lorsque seules deux vues sont disponibles, une alternative consiste à échantillonner directement dans les intervalles de disparité, en transférant le NeRF d'une représentation 3D vers une formulation fondée sur la géométrie épipolaire. Cette voie ouvre la perspective d'interpréter les NeRF comme un régularisateur pour l'estimation de disparité. Enfin, les expériences menées sur les scènes du jeu de données DFC2019 peuvent être reproduites sur de nouvelles régions et de nouveaux capteurs afin de généraliser les conclusions obtenues dans ces travaux.

Remerciements

Nous remercions le Centre National d'Études Spatiales d'avoir rendu possibles ces travaux par la mise à disposition de ressources sur leur centre de calcul haute performance.

Références

- Aira, L. S., Facciolo, G., Ehret, T., 2025. Gaussian splatting for efficient satellite image photogrammetry. Dans : Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5959–5969.
- Billouard, C., Derksen, D., Sarrazin, E., Vallet, B., 2024. Sat-nerf : Unleashing neural graphics primitives for fast relightable transient-free 3d reconstruction from satellite imagery. Dans : IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 8749–8753.
- Chen, A., Xu, Z., Zhao, F., Zhang, X., Xiang, F., Yu, J., Su, H., 2021. Mvsnerf : Fast generalizable radiance field reconstruction from multi-view stereo. Dans : Proceedings of the IEEE/CVF international conference on computer vision. pp. 14124–14133.
- De Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., 2014. An automatic and modular stereo pipeline for pushbroom images. Dans : ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.
- Deng, K., Liu, A., Zhu, J.-Y., Ramanan, D., 2022. Depth-supervised nerf : Fewer views and faster training for free. Dans : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12882–12891.
- Derksen, D., Izzo, D., 2021. Shadow neural radiance fields for multi-view satellite photogrammetry. Dans : Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1152–1161.
- Dubois, D., Prade, H., 1992. When upper probabilities are possibility measures. Fuzzy sets and systems 49 (1), 65–74.
- Fuentes Reyes, M., d'Angelo, P., Fraundorfer, F., 2024. Comparative analysis of deep learning-based stereo matching and multi-view stereo for urban dsm generation. Remote Sensing 17 (1), 1.
- Haller, I., Pantilie, C., Oniga, F., Nedeveschi, S., 2010. Real-time semi-global dense stereo solution with improved sub-pixel accuracy. Dans : 2010 IEEE Intelligent Vehicles Symposium. IEEE, pp. 369–376.
- Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. Dans : 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Vol. 2. IEEE, pp. 807–814.
- Jain, A., Tancik, M., Abbeel, P., 2021. Putting nerf on a diet : Semantically consistent few-shot view synthesis. Dans : Proceedings of the IEEE/CVF international conference on computer vision. pp. 5885–5894.
- Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., Bry, A., 2017. End-to-end learning of geometry and context for deep stereo regression. Dans : Proceedings of the IEEE international conference on computer vision. pp. 66–75.
- Le Saux, B., Yokoya, N., Hänsch, R., Brown, M., 2019. Data fusion contest 2019 (dfc2019).
URL <https://dx.doi.org/10.21227/c6tm-vw12>
- Malinowski, R., Sarrazin, E., Dumas, L., Dubois, E., Destercke, S., 2024. Robust confidence intervals in stereo matching using possibility theory. arXiv preprint arXiv :2404.06273.
- Marí, R., Facciolo, G., Ehret, T., 2022. Sat-nerf : Learning multi-view satellite photogrammetry with transient objects and shadow modeling using rpc cameras. Dans : Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1311–1321.
- Marí, R., Facciolo, G., Ehret, T., 2023. Multi-date earth observation nerf : The detail is in the shadows. Dans : Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2035–2045.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2021. Nerf : Representing scenes as neural radiance fields for view synthesis. Communications of the ACM 65 (1), 99–106.
- Niemeyer, M., Barron, J. T., Mildenhall, B., Sajjadi, M. S., Geiger, A., Radwan, N., 2022. Regnerf : Regularizing neural radiance fields for view synthesis from sparse inputs. Dans : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5480–5490.
- Poggi, M., Tosi, F., Mattocchia, S., 2017. Quantitative evaluation of confidence measures in a machine learning world. Dans : Proceedings of the IEEE International Conference on Computer Vision. pp. 5228–5237.
- Qu, Y., Deng, F., 2023. Sat-mesh : Learning neural implicit surfaces for multi-view satellite reconstruction. Remote Sensing 15 (17), 4297.
- Roessle, B., Barron, J. T., Mildenhall, B., Srinivasan, P. P., Nießner, M., 2022. Dense depth priors for neural radiance fields from sparse input views. Dans : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12892–12901.
- Rupnik, E., Daakir, M., Pierrot Deselligny, M., 2017. Micmac—a free, open-source solution for photogrammetry. Open geo-spatial data, software and standards 2 (1), 14.
- Sarrazin, E., Cournet, M., Dumas, L., Defonte, V., Fardet, Q., Steux, Y., Jimenez Diaz, N., Dubois, E., Youssefi, D., Buffe, F., 2021. Ambiguity concept in stereo matching pipeline. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 43, 383–390.
- Tosi, F., Bartolomei, L., Poggi, M., 2025. A survey on deep stereo matching in the twenties. International Journal of Computer Vision 133 (7), 4245–4276.
- Walley, P., 1991. Statistical Reasoning with Imprecise Probabilities. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- Wei, Y., Liu, S., Rao, Y., Zhao, W., Lu, J., Zhou, J., 2021. Nerfingmvs : Guided optimization of neural radiance fields for indoor multi-view stereo. Dans : Proceedings of the IEEE/CVF international conference on computer vision. pp. 5610–5619.
- Wen, B., Trepte, M., Aribido, J., Kautz, J., Gallo, O., Birchfield, S., 2025. Foundationstereo : Zero-shot stereo matching. Dans : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5249–5260.
- Wu, R., Mildenhall, B., Henzler, P., Park, K., Gao, R., Watson, D., Srinivasan, P. P., Verbin, D., Barron, J. T., Poole, B., et al., 2024. Reconfusion : 3d reconstruction with diffusion priors. Dans : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 21551–21561.
- Youssefi, D., Michel, J., Sarrazin, E., Buffe, F., Cournet, M., Delvit, J.-M., L'Helguen, C., Melet, O., Emilien, A., Bosman, J., 2020. Cars : A photogrammetry pipeline using dask graphs to construct a global 3d model. Dans : IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 453–456.
- Yu, A., Ye, V., Tancik, M., Kanazawa, A., 2021. pixelnerf : Neural radiance fields from one or few images. Dans : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4578–4587.
- Žbontar, J., LeCun, Y., 2016. Stereo matching by training a convolutional neural network to compare image patches. Journal of Machine Learning Research 17 (65), 1–32.
- Zhang, L., Rupnik, E., 2023. Sparsesat-nerf : Dense depth supervised neural radiance fields for sparse satellite images. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences 10.
- Zhang, T., Zhou, Y., Li, Y., Wei, X., 2024. Satensorf : Fast satellite tensorial radiance field for multirate satellite imagery of large size. IEEE Transactions on Geoscience and Remote Sensing 62, 1–15.