Analisi spettrale per il riconoscimento dei nidi di farfalla Euphydryas maturna con sensori multispettrali montati su UAV

Chiara Graziani[1][0009-0000-9608-8542], Andrea Maria Lingua[1][0000-0002-5930-2711], Paolo Felice Maschio[1][0000-0001-7706-9354], Beatrice Gammino[2]

[1] Politecnico di Torino, Dipartimento di Ingegneria per l'Ambiente, il Territorio e le Infrastrutture, Corso Duca degli Abruzzi 24, 10129, Torino, Italia, andrea.lingua@polito.it, chiara.graziani@polito.it paolo.maschio@polito.it

[2] Aree Protette Alpi Marittime, beatrice.gammino@areeprotettealpimarittime.it

Abstract. *Euphydryas maturna* è un lepidottero diurno, presente in Italia solo in Valle Stura di Demonte (CN), le cui larve costruiscono nidi sericei sulle foglie di frassino. Il seguente lavoro descrive e valida un processo automatico di individuazione dei nidi da immagini multispettrali acquisite da drone mediante una procedura rigorosa di coregistrazione delle componenti spettrali, che permette l'applicazione di un algoritmo di classificazione deep learning basato su visual transformer.

Parole chiave: *Euphydryas maturna*, coregistrazione RGB-multispettrale, segmentazione semantica.

1 Introduzione

Euphydryas maturna (Linné, 1758) è un lepidottero diurno presente in buona parte d'Europa, ma severamente minacciato in tutto il suo areale [1]. La specie è protetta e inserita negli Allegati II e IV della Direttiva Habitat (Direttiva 92/43/CEE) ed è inoltre classificata come "Critically endangered" nella Lista Rossa delle Farfalle italiane [2]. In Italia è presente con una sola popolazione situata in valle Stura di Demonte (Cuneo): per la sua conservazione è stata istituita la Zona Speciale di Conservazione (ZSC) IT1160067 "Vallone dell'Arma" (Fig. 1) [3], sito della Rete Natura 2000 in gestione all'Ente di gestione delle Aree Protette Alpi Marittime (APAM).

L'adulto misura 3,7-4,5 cm di apertura alare e predilige habitat ecotonali caratterizzati da prati pascolo al margine di boschi misti di latifoglie con abbondanza di frassino, dove è possibile osservarla in volo tra giugno e luglio [4].

La femmina di E. maturna depone le uova in gruppi, chiamate ovature, sulla pagina inferiore delle foglie del frassino maggiore (Fraxinus excelsior). Una volta fuoriuscite dalle uova, le larve iniziano a cibarsi delle foglie di frassino, avvolte dal materiale sericeo che producono come strategia antipredatoria. Lo sviluppo si arresta in autunno quando le larve scendono dai frassini per trascorrere l'inverno all'interno della lettiera.

In primavera, i bruchi riprendono lo sviluppo nutrendosi di piante erbacee quali Ligustrum vulgare, Melampyrum pratense o Plantago sp.. Trascorsi circa 20 giorni i bruchi si impupano nella crisalide all'interno della quale avverrà la fase finale della metamorfosi e da cui uscirà l'adulto [5] [6].

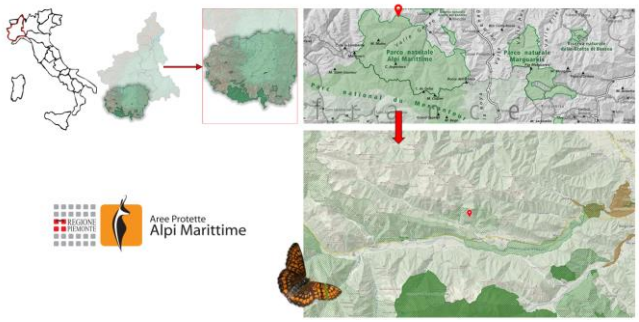


Fig. 1. Area di studio: Vallone dell'Arma (CN). La ZSC si estende su una superficie di 796,41 ettari [7].

Il monoraggio della fase larvale della specie prevede il conteggio dei nidi sericei sulle foglie di frassino, attività oggi svolta prevalentemente tramite osservazione visiva da parte di operatori esperti. Tale procedura, pur essendo fondamentale per la conservazione della specie, può risultare onerosa e limitata dalla visibilità dei nidi, dalla loro posizione sulla chioma e dalle condizioni operative in campo.
In questo contesto, la collaborazione tra APAM e il Politecnico di Torino mira a valutare la possibilità di individuare i nidi dall'alto mediante immagini multispettrali acquisite da UAV. Il lavoro presentato costituisce una prima fase sperimentale, finalizzata alla verifica della separabilità spettrale tra i nidi e gli altri elementi presenti sui rami di frassino, prima del passaggio ad acquisizioni operative da drone, nonché l'addestramento di un modello di intelligenza artificiale per il riconoscimento dei nidi.

2 Metodologia

La procedura sperimentale è stata organizzata in una sequenza di fasi successive, dalla verifica preliminare della separabilità spettrale fino alla valutazione del modello di segmentazione, come sintetizzato in Fig. 2. Il flusso comprende l'acquisizione e l'annotazione dei dati, la coregistrazione RGB–multispettrale, la costruzione del dataset, l'addestramento di SegFormer-B2 e la successiva validazione e verifica su test indipendente.



Fig. 2. Schema del flusso metodologico adottato.

2.1 Studi preliminari mediante spettrometria

Il primo passaggio è stato acquisire firme spettrali a terra mediante spettrometro Ocean Optics (Fig. 3), considerando sette classi principali (vedi Fig. 4).

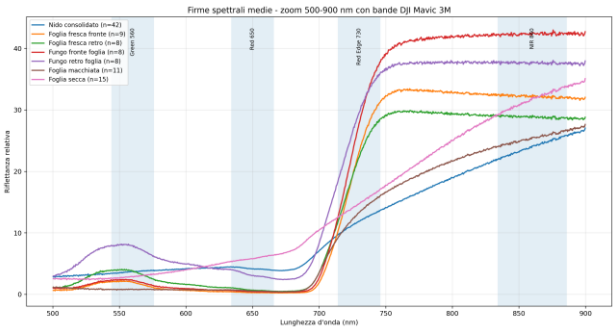


Fig. 3. Acquisizione delle firme spettrali tramite spettrometro.

Le firme, costituite da 2068 punti spettrali nel range circa 194–1116 nm, sono state ricondotte alle bande Green, Red, Red Edge e NIR del DJI Mavic 3M, calcolando la media nelle rispettive finestre spettrali (Fig. 3).
Le acquisizioni meno coerenti con le classi sono state rimosse attraverso un test diagnostico robusto basato sulla Median Absolute Deviation (MAD) [8], escludendo così 10 firme considerate non valide per la classe nido e altre sei firme per le altre classi.
La valutazione della separabilità delle classi è avvenuta attraverso tre livelli:

1. Analisi qualitativa delle differenze tra le firme lungo lo spettro.
2. Il calcolo di indici di separazione per meglio discriminare le classi, sfruttando il contrasto tra Green e Red: NGRDI, Green/Red, G-RE, GNDVI, EVI2.
3. Linear Discriminant Analysis (LDA) , validata con approccio leave-one-out, sia in configurazione multi-classe sia in modalità binaria Nido/Altro. [9]

Le metriche scelte per valutare la bontà della separabilità sono l'Accuracy, la Precisione della classe Nido, Recall della classe Nido, F1-score e la conta dei falsi positivi.



Nido di farfalla (52 firme spettrali)
Foglia fresca fronte (10 firme spettrali)
Foglia fresca retro (10 firme spettrali)
Foglia macchiata (14 firme spettrali)
Foglia secca (16 firme spettrali)
Fungo su fronte foglia (10 firme spettrali)
Fungo su retro foglia (10 firme spettrali)

Fig. 4. Firme spettrali delle classi analizzate e corrispondenza con le bande operative della camera multispettrale del DJI Mavic 3M.

2.2 I risultati delle analisi spettrometriche

L'analisi delle bande spettrali ha evidenziato una sovrapposizione significativa tra alcune classi, in particolare tra nido e foglia secca, indicando che le sole bande multispettrali non garantiscono una separazione robusta. Dopo il filtraggio MAD, il dataset spettrometrico finale è risultato composto da 101 firme, di cui 42 relative alla classe nido.

I risultati preliminari indicano che la classificazione binaria tramite LDA rappresenta l'approccio più promettente per un futuro utilizzo operativo. Il modello Nido/Altro ha raggiunto un'accuratezza globale pari al 95%, con precisione del 100%, recall dell'88.1% e assenza di falsi positivi nel dataset filtrato. Tali risultati suggeriscono che l'informazione acquisita con drone possa essere utilizzata come primo filtro per individuare candidati nido, pur richiedendo ulteriori verifiche su immagini UAV reali.

Nel complesso, la spettrometria ha confermato l'esistenza di informazione discriminante utile, ma anche una sovrapposizione non trascurabile con componenti vegetali e foglie senescenti. Questo risultato ha motivato il passaggio dalla classificazione puramente spettrale alla segmentazione diretta delle immagini UAV, nella quale informazione radiometrica, texture e contesto spaziale vengono apprese congiuntamente.

2.3 La definizione del dataset di riferimento

Il 20 luglio 2026 è stata condotta una seconda campagna per l'acquisizione di immagini RGB e multispettrali mediante DJI Mavic 3M. Il sistema integra una camera RGB da 20 MP (5280×3956 pixel) e quattro camere multispettrali da 5 MP nelle bande Green, Red, Red Edge e NIR [10].La campagna ha prodotto 913 immagini multispettrali relative alle quattro bande, associate alle corrispondenti acquisizioni RGB.
Le acquisizioni hanno interessato frassini con nidi larvali annotabili nelle immagini RGB. La geometria tridimensionale della chioma, caratterizzata da rami e foglie posti a distanze differenti dalla camera, rende non sufficiente una semplice trasformazione planare globale per sovrapporre con precisione le immagini RGB e multispettrali.



Fig. 5. Annotazione dei nidi con Segment Anything 2.0. su software CVAT [11], esportate in formato Datumaro 1.0.

2.4 Coregistrazione RGB–multispettrale

La coregistrazione tra immagine RGB e bande multispettrali è stata impostata tenendo conto della configurazione multicamera del DJI Mavic 3M e della marcata tridimensionalità delle chiome osservate a breve distanza. Sono state utilizzate le intrinseche e le distorsioni note dei singoli sensori, evitando di ricondurre il problema a una semplice trasformazione planare: i sensori RGB e multispettrali possiedono infatti centri di proiezione differenti e la posizione del punto omologo varia con la profondità della scena. L'applicazione dei soli parametri di calibrazione e distorsione forniti dal costruttore non è risultata sufficiente per ottenere una coregistrazione pixel-level: su un set diagnostico di 79 corrispondenze comuni il residuo di registrazione presentava un TRE mediano di 43,08 pixel, un RMSE di 43,79 pixel e un P95 di 48,82 pixel, evidenziando la necessità di una successiva stima della geometria relativa tra i sensori.

La geometria relativa tra la camera RGB e ciascuna banda multispettrale è stata stimata direttamente sulle immagini originali mediante estrazione e matching di feature, seguiti dalla stima robusta della matrice essenziale sui raggi calibrati. Per le bande Green, Red e RedEdge sono state impiegate corrispondenze basate su SIFT; per la banda NIR, caratterizzata da una maggiore differenza radiometrica rispetto all'RGB, è stato invece utilizzato RIFT2 (Radiation-Invariant Feature Transform 2) [12] con criterio di matching mutuale, più adatto al confronto tra immagini multimodali. [13].

La sola geometria epipolare non è tuttavia sufficiente a compensare la parallasse introdotta dalla struttura tridimensionale della vegetazione. La correzione è stata quindi completata mediante stima monoculare della geometria con MoGe-2 [14]. La relazione tra profondità e disparità è stata ricalibrata, per ciascuna immagine e per ogni banda, utilizzando le corrispondenze omologhe osservate, così da adattare scala e bias della profondità stimata alle condizioni reali di acquisizione.

L'intera catena geometrica è stata implementata attraverso la composizione numerica delle trasformazioni, mantenendo come riferimento la griglia RGB originale. Le immagini RGB e le relative annotazioni non vengono quindi ricampionate; ciascuna banda multispettrale viene invece letta direttamente dal dato RAW e campionata sul reticolo RGB mediante una sola interpolazione bilineare, evitando ricampionamenti successivi e limitando la degradazione dell'informazione radiometrica.

La calibrazione finale disparità–profondità ha prodotto MAE mediane di circa 2,3–2,7 pixel per Green, Red e RedEdge e 5,27 pixel per NIR; nel trasferimento operativo delle annotazioni, 196 dei 218 nidi sono risultati validi contemporaneamente in tutte le quattro bande.

2.5 Modello di intelligenza artificiale

La segmentazione semantica binaria distingue ogni pixel tra nido e background. È stata adottata l'architettura SegFormer-B2 [15], basata su Vision Transformer con encoder gerarchico multiscala e decoder MLP. Il modello, inizializzato da un checkpoint pre-addestrato su ADE20K, è stato adattato a un'unica classe di output.

Una prima fase sperimentale è stata dedicata alla valutazione del contributo delle bande multispettrali. Sono state confrontate tre configurazioni: A) RGB, B) Green–Red–RedEdge–NIR e C) RGB+multispettrale. Per ciascuna configurazione sono stati

eseguiti quattro esperimenti leave-one-session-out (LOSO), utilizzando a turno una delle quattro sessioni come test e mantenendo invariati tile, maschere di validità e protocollo di addestramento. La normalizzazione delle bande multispettrali è stata stimata esclusivamente sui dati appartenenti al training session, evitando contaminazioni con validation e test.
I principali parametri utilizzati nella fase comparativa sono riportati in Tab. 1.
La metrica primaria è stata $F_2 = 5PR/(4P+R)$ che attribuisce maggiore importanza alla Recall R rispetto alla Precision P, coerentemente con l'obiettivo del monitoraggio: la mancata individuazione di un nido è stata considerata più critica di una sovrasegmentazione successivamente verificabile. La soglia decisionale τ è stata ottimizzata esclusivamente sui dati di validation; il test non è stato utilizzato né per l'early stopping né per la selezione del checkpoint.
Il confronto LOSO ha prodotto un F_2 medio pari a 0,630 per RGB, 0,404 per il solo multispettrale e 0,654 per RGB+MS. Il vantaggio della configurazione combinata rispetto all'RGB è risultato quindi limitato e non uniforme tra le sessioni, mentre l'utilizzo delle bande multispettrali richiedeva una procedura aggiuntiva di coregistrazione e introduceva ulteriori vincoli di validità dei dati. Per questo motivo la soluzione operativa finale è stata sviluppata utilizzando esclusivamente l'informazione RGB.
Il dataset RGB finale comprende 218 nidi annotati e 900 tile, di cui 450 positivi e 450 negativi. La suddivisione è stata effettuata a livello di immagine sorgente e separatamente all'interno di ciascuna sessione, in modo che tile sovrapposti derivati dalla stessa immagine non potessero appartenere a insiemi differenti. Sono stati ottenuti 604 tile di training, 204 di validation e 92 di test; tutte le quattro sessioni sono rappresentate nei tre insiemi e nessuna immagine sorgente è condivisa tra essi.
Per l'addestramento finale RGB sono stati mantenuti gli stessi criteri della fase comparativa.

Tab. 1. Principali parametri utilizzati per l'addestramento e la selezione del modello SegFormer-B2 nella fase di confronto tra configurazioni RGB e multispettrali.

Parametro	Impostazione	Parametro	Impostazione
Architettura	SegFormer-B2	Configurazioni confrontate	RGB; MS; RGB+MS
Dimensione tile	768 × 768 px	Schema di valutazione	4-fold LOSO per sessione
Batch size	1	Gradient accumulation	8
Batch effettivo	8	Ottimizzatore	AdamW
Learning rate	6×10^{-5}	Weight decay	0,01
Warm-up	100 step	Gradient clipping	1
Max optimizer step	2000	Valutazione	ogni 100 step
Early stopping	6 valutazioni senza miglioramento	Data augmentation	rotazioni 90°, flip H/V

Parametro	Impostazione	Parametro	Impostazione
Loss	0,5 Focal + 0,5 Dice	Focal Loss	$\alpha = 0{,}75$; $\gamma = 2$
Metrica di selezione	F2	Soglia τ	0,05–0,95
Passo di τ	0,001	Selezione checkpoint	max F2 su validation
Uso del test	escluso da selezione e calibrazione	Pre-training	ADE20K

3 Risultati

Nel training finale RGB il valore F2 sulla validation cresce rapidamente nelle prime iterazioni, passando da 0,867 allo step 100 a valori superiori a 0,90 a partire dallo step 300. Il massimo F2 di validation, pari a 0,910, è stato raggiunto allo step 700; il training è proseguito fino allo step 1300, quando è intervenuto l'early stopping. Il checkpoint dello step 700 è stato quindi congelato come modello finale.
La soglia decisionale ottimale selezionata sulla validation è risultata $\tau = 0{,}097$. Con tale soglia il modello ha ottenuto in validation F2 = 0,910, Dice = 0,851, IoU = 0,741, precisione = 0,769 e recall = 0,954. La soglia e il checkpoint sono stati definiti senza accedere al test, successivamente valutato una sola volta.

Tab. 2. Prestazioni del modello RGB finale sulla validation e sul test.

Metrica	Validation	Test
F2	0,910	0,903
Dice	0,851	0,803
IoU	0,741	0,671
Precisione	0,769	0,678
Recall	0,954	0,985
Specificità	0,980	0,958

Sul test composto da 92 tile (46 positivi e 46 negativi), il modello ha raggiunto F_2 = 0,903, recall = 0,985, precisione = 0,678 e Dice = 0,803. La riduzione di F_2 rispetto alla validation è contenuta (−0,007), mentre il recall aumenta ulteriormente. Il comportamento è coerente con la metrica di selezione adottata: il modello privilegia la sensibilità nella localizzazione del nido, accettando una maggiore sovrasegmentazione. Le prestazioni non sono però uniformi tra le quattro sessioni, per le quali F_2 varia da 0,618 a 0,909. Tale eterogeneità suggerisce che illuminazione, struttura della chioma e dimensione apparente dei nidi continuino a influenzare la risposta del classificatore.
La valutazione qualitativa sulle immagini complete conferma che la segmentazione ricostruisce in modo generalmente coerente la posizione e l'estensione dei nidi. Per la lettura visiva dei risultati è utile sovrapporre sull'RGB il contorno della ground truth in

rosso e quello della predizione in blu, soluzione che mantiene leggibile la vegetazione e consente di individuare immediatamente omissioni e sovrasegmentazioni (Fig. 7).

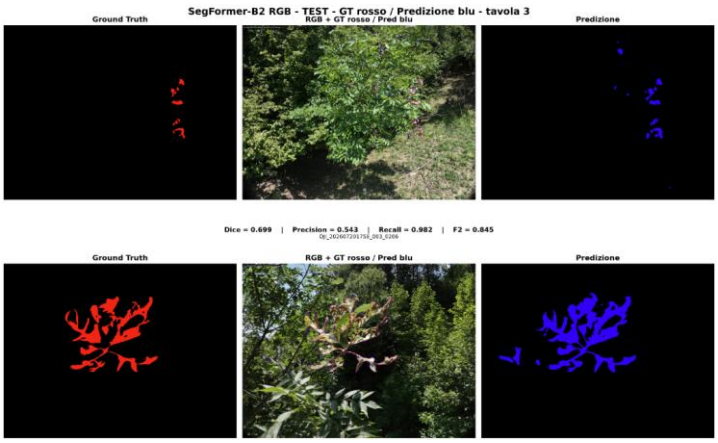


Fig. 7. Ground truth (rosso) e predizione del modello (blu). Nel caso in alto, il modello ha sovrastimato degli elementi presenti sull'alberto di frassino, nell'immagine in basso è invece riuscito a identificare un altro nido che non era stato segnalato nell'annotazione manuale.

4 Conclusioni

Il lavoro ha seguito un percorso progressivo dalla caratterizzazione spettrale dei materiali presenti sulla chioma alla segmentazione semantica delle immagini UAV. Le firme acquisite a terra hanno mostrato che i nidi di Euphydryas maturna possiedono caratteristiche spettrali riconoscibili, ma non completamente separabili da foglie secche, alterazioni e altri elementi vegetali. La classificazione LDA binaria ha fornito risultati molto buoni sul dataset spettrometrico filtrato, ma l'esperimento su immagini reali ha evidenziato che la sola informazione multispettrale non è sufficiente a sostenere una segmentazione robusta.

La coregistrazione RGB–multispettrale ha permesso di costruire un dataset multimodale coerente e di confrontare in condizioni controllate tre differenti configurazioni di input. I 12 esperimenti LOSO mostrano che la fusione RGB+multispettrale migliora solo marginalmente il valore medio di F_2 rispetto all'RGB e che il vantaggio è concentrato in una sola sessione. Considerando la maggiore risoluzione spaziale della camera RGB, la semplicità della catena di elaborazione e la possibilità di utilizzare tutte le annotazioni disponibili, la configurazione RGB-only è stata quindi adottata come soluzione operativa.

Il SegFormer-B2 finale raggiunge valori che indicano una buona capacità di individuazione dei nidi nelle acquisizioni considerate. Il test è stato separato a livello di capture, ma le quattro sessioni sono comunque rappresentate anche nel training; sarà quindi necessario verificare la robustezza del modello su campagne indipendenti, in differenti condizioni fenologiche, di illuminazione e di geometria di presa.

Ringraziamenti

Gli autori ringraziano il personale del parco Aree Protette Alpi Marittime per il loro supporto durante le acquisizioni delle immagini da drone.

Riferimenti bibliografici

1. Swaay, Chris & Cuttelod, Annabelle & Collins, S. & Maes, Dirk & Munguira, M. & Šašić, Martina & Settele, Josef & Verovnik, Rudi & Verstrael, T & Warren, Martin & Wiemers, Martin & Wynhoff, Irma. (2010). European Red List of Butterfies.
2. Balletto, E., Bonelli, S., Barbero, F., Casacci, L.P., Sbordoni, V., Dapporto, L., Scalercio, S., Zilli, A., Battistoni, A., Teofili, C., Rondinini, C. (compilatori): Lista Rossa IUCN delle Farfalle Italiane - Ropaloceri. Comitato Italiano IUCN e Ministero dell'Ambiente e della Tutela del Territorio e del Mare, Roma (2015; stampa 2016).
3. Aree Protette Alpi Marittime: Un tesoro nascosto nel Vallone dell'Arma: la farfalla Euphydryas maturna. https://www.areeprotettealpimarittime.it/news/4549/un-tesoro-nascosto-nel-vallone-dell-arma-la-farfalla-euphydryas-, ultimo accesso 2026/10/13.
4. Freese, A., Benes, J., Bolz, R., Cizek, O., Dolek, M., Geyer, A., Gros, P., Konvicka, M., Liegl, A., Stettmer, C. (2006) Habitat use of the endangered butterfly Euphydryas maturna and forestry in Central Europe. Animal Conservation, 9, 3883397.
5. Wahlberg, N. (2000) Comparative descriptions of the immature stages and ecology of five Finnish melitaeine butterfly species (Lepidoptera: Nymphalidae). Entomol Fenn 11:1673 174.
6. Dolek, M., Freese-Hager, A., Geyer, A., Balletto, E., Bonelli, S. (2013) Multiple oviposition and larval feeding strategies in Euphydryas maturna (Linne, 1758) (Nymphalidae) at two disjoint European sites. Journal of Insect Conservation, 17 (2), 357-366.
7. Aree Protette Alpi Marittime: Vallone dell'Arma, ZSC IT1160067. https://www.areeprotettealpimarittime.it/ente-di-gestione-aree-protette-alpi-marittime/rete-natura-2000/vallone-dell-arma, ultimo accesso 2026/10/13.
8. Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L.: Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology 49(4), 764–766 (2013). https://doi.org/10.1016/j.jesp.2013.03.013.
9. Graf, R., Zeldovich, M., & Friedrich, S.: Comparing linear discriminant analysis and supervised learning algorithms for binary classification—A method comparison study. Biometrical Journal, 66, 2200098 (2022). https://doi.org/10.1002/bimj.202200098.
10. DJI Enterprise: DJI Mavic 3M - Specifications. https://enterprise.dji.com/it/mavic-3-m/specs, ultimo accesso 13/10/2026.
11. Sekachev, B., Manovich, N., Zhavoronkov, A.: Computer Vision Annotation Tool (CVAT). Zenodo (2019). https://doi.org/10.5281/zenodo.3497106.
12. Li, J., Shi, P., Hu, Q., Zhang, Y.: RIFT2: Speeding-up RIFT with A New Rotation-Invariance Technique. arXiv preprint arXiv:2303.00319 (2023). doi:10.48550/arXiv.2303.00319.
13. Li, J., Hu, Q., Ai, M.: RIFT: Multi-modal Image Matching Based on Radiation-Variation Insensitive Feature Transform. IEEE Transactions on Image Processing, 29, 3296–3310 (2020).

14. Wang, R., Xu, S., Dong, Y., Deng, Y., Xiang, J., Lv, Z., Sun, G., Tong, X., Yang, J.: MoGe-2: Accurate Monocular Geometry with Metric Scale and Sharp Details. Advances in Neural Information Processing Systems 38, 40377–40408 (2025). https://doi.org/10.52202/085713-1207.
15. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. Advances in Neural Information Processing Systems 34, 12077–12090 (2021).