
INDIVIDUAZIONE DI PIANTE INFESTANTI TRAMITE UKAN

Elena Tavoletti

Dipartimento di informatica

Università di Bari

Bari, Italia

e.tavoletti@studenti.uniba.it

11 aprile 2025

SOMMARIO

L'agricoltura di precisione fa molto affidamento sul controllo reale della crescita delle erbacce per garantire coltivazioni più sane. In questo studio si presenta l'applicazione di UKAN al dominio dell'agricoltura di precisione. Si tratta di una rete neurale innovativa per la segmentazione di immagini, che sfrutta al suo interno degli strati di tipo KAN: in questo modo si punta ad avere un modello leggero per distinguere le erbacce dalle coltivazioni. Valutando le performance sul dataset WeedMap, UKAN ottiene una F1-score di 72,1 reggendo il confronto con lo stato dell'arte per la segmentazione, rappresentato da Segformer MiT-b0, a fronte di un numero di parametri drasticamente inferiore. Riprendendo l'architettura di UKAN in modelli mirati per l'agricoltura di precisione, presenti all'interno di droni, i coltivatori potranno condurre delle analisi real-time sullo stato delle piantagioni e controllare la crescita delle erbacce in ampi spazi coltivati

1 Introduzione

L'agricoltura è essenziale per il sostentamento umano e i progressi delle macchine agricole e nelle tecniche hanno migliorato la resa delle colture. La gestione delle infestanti è cruciale per rimuovere le piante indesiderate che competono con le piantagioni. Una gestione efficace delle erbacce aumenta la produttività dei campi e promuove un'agricoltura sostenibile.

I droni, o veicoli aerei senza equipaggio (UAV), si sono rivelati risorse inestimabili nell'agricoltura di precisione, offrendo sia versatilità che convenienza economica [11]. Questi dispositivi possono catturare immagini ad alta risoluzione e dati dai terreni agricoli, consentendo agli agricoltori di monitorare lo sviluppo delle colture, rilevare malattie e parassiti e ottimizzare le strategie di irrigazione. Fornendo informazioni accurate e tempestive, i droni aiutano a ridurre le spese, aumentare la resa delle colture e minimizzare l'uso di risorse come acqua, fertilizzanti e pesticidi. I metodi tradizionali, come le ispezioni manuali dei campi o il telerilevamento satellitare, non possono eguagliare il livello di dettaglio fornito dai droni.

Gli UAV possono condurre indagini aeree in tempo reale sulle colture, permettendo agli agricoltori di prendere decisioni rapide e informate. Inoltre, i droni possono coprire in modo efficiente vaste aree in poche ore, un compito che normalmente richiederebbe giorni o settimane con metodi convenzionali. Questa capacità di risparmio di tempo consente agli agricoltori di conservare risorse e prendere decisioni più tempestive.

Ricerche recenti hanno suggerito che i modelli di deep learning possono essere utilizzati per la segmentazione semantica e l'identificazione delle infestanti in immagini aeree acquisite da droni [1, 4, 9]. Tuttavia, nonostante questi significativi progressi, il rilevamento automatico delle infestanti rimane una sfida complessa. Le tecniche di deep learning non sono ancora ampiamente adottate in agricoltura, principalmente a causa della vasta annotazione manuale necessaria per la grande quantità di dati richiesta nella fase di apprendimento. Questo problema è particolarmente rilevante nei dataset agricoli, dove l'etichettatura delle piante nelle immagini dei campi è un processo lungo. Inoltre, questi modelli spesso richiedono risorse computazionali significative, ponendo sfide per l'implementazione sui droni, che dispongono di capacità di elaborazione limitate e un'autonomia energetica ridotta. La necessità di prestazioni in tempo reale, unita

a queste limitazioni, evidenzia l'importanza di sviluppare soluzioni leggere per le applicazioni di mappatura delle infestanti.

Nell'ambito dell'ottimizzazione delle risorse, questo studio propone l'utilizzo di una U-Net che contenga negli strati più interni delle reti KAN: sono architetture basate sul teorema di rappresentazione di Kolmogorov-Arnold. Questo teorema asserisce che ogni funzione continua multivariata possa essere riscritta come combinazione finita di funzioni continue univariate: uno strato KAN stima quali siano queste funzioni, parametrizzando delle funzioni di base chiamate spline, piuttosto che quelli di semplici combinazioni lineari. In questo modo il potere espressivo della rete non si limita ad approssimare la funzione reale ma a predirla in modo esatto, rendendo possibile avere risultati più accurati a parità di numero di parametri.

La valutazione è stata condotta sul dataset WeedMap [10], che contiene immagini multispettrali di campi di barbabietola da zucchero catturate da droni.

2 Lavori correlati

I progressi nella visione artificiale e nel telerilevamento hanno rivoluzionato l'agricoltura di precisione, affrontando compiti come l'identificazione di malattie e parassiti, la valutazione dello stress abiotico, il monitoraggio della crescita, la previsione della resa delle colture e la mappatura delle infestanti.

La mappatura delle infestanti, un compito di segmentazione semantica, assegna ogni pixel di un'immagine a una delle due classi, ovvero infestante o coltura. Gli algoritmi di deep learning hanno dimostrato prestazioni superiori rispetto alle tecniche tradizionali in questo ambito. I primi lavori di Dos Santos et al. [4] hanno evidenziato l'efficacia delle Reti Neurali Convoluzionali (CNN), come AlexNet, rispetto a SVM e Random Forests. Lottes et al. [6] hanno ulteriormente avanzato il settore utilizzando una CNN con doppi decoder per il rilevamento dei fusti e la segmentazione delle piante, mostrando risultati promettenti sui dataset BoniRob e UAV.

L'integrazione di immagini multispettrali, che catturano informazioni dettagliate sulla salute e sulle specie delle piante, migliora l'accuratezza dei modelli di deep learning rispetto ai modelli basati solo su immagini RGB. Ad esempio, U-Net è stato utilizzato con successo per separare le infestanti dalle colture e dal suolo [2]. WeedNet, basato su SegNet e addestrato sul dataset WeedMap, è un altro esempio di applicazione di successo in questo settore [9]. Il dataset WeedMap, contenente immagini multispettrali di campi di barbabietola da zucchero in Germania e Svizzera, è diventato un benchmark per gli studi sulla mappatura delle infestanti [10]. Anche noi facciamo utilizzo di questo dataset per via del suo stato di benchmark, tuttavia ne prendiamo solo i canali utili per l'analisi RGB.

Una soluzione tecnica di cui si è verificata l'efficacia nell'ambito della segmentazione semantica è l'inserimento di strati KAN nel core di una U-Net, per fronteggiare la difficoltà delle architetture encoder-decoder nel ricavare tutte le informazioni possibili dalle features altodimensionali estratte dall'encoder. In particolare Ma et al. [7] verificano che l'utilizzo di KAN per il Remote Sensing si rivela una scelta vincente raggiungendo risultati di accuratezza che superano l'attuale stato dell'arte: trattandosi di un task con punti in comune con la Drone Vision è interessante capire se possa essere una buona scelta applicare le KAN anche nell'agricoltura di precisione. In particolare si utilizza l'implementazione UKAN messa a disposizione da Li et al. [5], al seguente link <https://github.com/CUHK-AIM-Group/U-KAN>

3 Materiale

Gli esperimenti sono stati condotti sul dataset Weedmap [10]. In particolare sulle immagini RGB costruite grazie ai dati raccolti dalla camera RedEdge. Per ogni campo, i curatori del dataset, hanno composto una immagine complessiva di esso in RGB, giustapponendo le diverse foto catturate dal drone; questa immagine, detta ortomosaica, è stata poi tagliata in "tiles" di dimensione 480 x 360 pixel.

Per fini di compatibilità con studi precedenti legati a quello condotto era necessario che le tiles fossero di dimensione 512 x 512 pixel, per tanto è stato effettuato un semplice resizing di quelle fornite.

La groundtruth per ogni immagine viene fornita come la stessa foto, con pixel colorati in base alla classe di appartenenza: verde per le coltivazioni, rosso per le infestanti e nero per il background. Tuttavia il framework costruito attorno alla implementazione di UKAN usato in questo studio, aspetta che ci sia una maschera per ogni classe con in bianco i pixel della classe e in nero i restanti: per questione di comodità si è deciso di non modificare il framework e piuttosto di ricavare dalla groundtruth queste maschere.

Le immagini relative ai campi coltivati 000, 001, 002, 004 sono state utilizzate per il training o il fine-tuning dei modelli, mentre quelle del campo 003 per il test. Nelle foto si nota una sovrabbondanza di pixel di background, rendendo la

classe sovrarappresentata: si decide per tanto di eliminare dalle immagini di training quelle con solo background.

4 Metodi

Architettura

L'obiettivo di questo studio è quello di verificare se tramite l'inserimento di strati di tipo KAN in un modello di Deep Learning per la segmentazione si verifichino dei miglioramenti in termini di performance e di complessità computazionale.

Il modello utilizzato per l'esperimento è stato implementato da Li et al. [5] e testato su dataset di tipo medico: qui le performance si sono rivelate migliori rispetto alla tradizionale U-Net, persino ad un costo computazionale inferiore. Non c'è da stupirsi. Nell'ambito della segmentazione semantica è già stato proposto il modello UNeXt [8], che conteneva negli ultimi strati dell'encoder e nei primi del decoder dei blocchi MLP: in questo modo si può riuscire a catturare l'informazione contenuta in tutta la featuremap estratta dall'encoder, a discapito di quella inerentemente spaziale; spazialità che a livelli di elaborazione molto avanzata è comunque meno presente. Inoltre UKAN sostituisce i blocchi MLP con i blocchi KAN e ciò potrebbe indirizzare verso delle performance ulteriormente migliorative; infatti le reti KAN apprendono non i pesi di combinazioni lineari, che vengono poi attivate, ma parametri di funzioni più complesse; ciò conferisce alla rete un potere espressivo maggiore che viene supportato teoricamente dal teorema di Kolmogorov-Arnold: una funzione continua multivariata può essere riscritta come combinazione di un numero finito di funzioni continue univariate secondo la seguente formula

$$f(x) = \sum_q \phi_q \left(\sum_p \psi_{pq}(x_p) \right) \quad (1)$$

le funzioni ϕ_q e ψ_{pq} sono spline-functions di cui vanno appresi i parametri, ma di solito $\phi_q = I$. Si tratta di un risultato teorico più confortante del teorema di approssimazione universale che sta alla base degli strati MLP: se uno strato KAN può in potenza imparare la funzione esatta, un MLP può solo approssimarla. Per queste motivazioni ci si aspetta che una KAN abbia performance migliori rispetto ad una MLP a parità di numero di parametri.

L'architettura di UKAN è di tipo encoder-decoder del tutto speculare e con skip-connections tra i corrispondenti livelli di encoder e decoder. Ogni livello dell'encoder dimezza altezza e larghezza della immagine, mentre i canali risultanti sono degli iperparametri: per i primi tre livelli, che sono convoluzionali, si indica solo il numero di canali per il terzo livello, gli altri due sono un quarto ed un ottavo di questo numero.

Gli ultimi due livelli sono Tokenized-KAN block. Sono composti da:

1. modulo di Tokenizzazione: tramite una convoluzione di kernel 3-pixel, dove 3 è la dimensione di una patch/token, e numero di canali indicati dal data scientist, ed una linearizzazione del risultato della convoluzione, per ogni immagine viene creata una matrice che ha una riga per ogni patch/token e tante colonne/features quanti i canali di arrivo indicati per la convoluzione; le matrici provenienti dalle immagini del batch vengono concatenate una sotto l'altra; inoltre vengono conservate altezza e larghezza delle immagini dopo la convoluzione e prima della linearizzazione, oltre che la dimensione del batch prima della concatenazione, di modo che le immagini del batch possano essere delinearizzate;
2. modulo KAN: formato da tre layers prende in input la matrice ricavata dalla tokenizzazione e la trasforma senza variane però alcuna dimensione;
3. altri due moduli che prendono in input le immagini delinearizzate e non ne cambiano le dimensioni: si tratta di una convoluzione depth-wise, una batch-normalization e l'attivazione tramite ReLU.

Setup della sperimentazione

Per prima cosa è stato necessario cambiare la loss del framework: infatti dopo l'estrazione delle features venivano effettuati task di predizione multilabel per i pixel, al contrario del caso in questione che richiede una segmentazione a classi esclusive. Si è optato per l'utilizzo di Cross Entropy come funzione di perdita.

L'addestramento è stato condotto sulle immagini provenienti dai campi 000, 001, 002, 004, e il 003 è stato riservato per il testing. Per i parametri di addestramento sono stati utilizzati gli stessi di Li et al. [5], tra questi si citano: ottimizzatore Adam con $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\lambda = 1e - 4$; per il learning rate viene utilizzata una tecnica di scheduling, il cosine annealing, che determina il valore in funzione delle epoche in un range che viene fornito dal data scientist; il range del cosine annealing varia tra $1e-4$ e $1e-5$ per gli strati non KAN, e tra $1e-2$ e $1e-5$; il modello è stato addestrato per 400 epoche con batch di 8 immagini.

Tabella 1: Risultati sperimentazione

Metodo	F1-avg	F1			GMACs	Parametri (M)	Inference Time (ms)
		BG	Coltivazioni	Infestanti			
UKAN-128	71,5	99,4	55,7	59,3	7,07	6,36	18
UKAN-64	72,1	99,4	55,2	61,8	1,84	1,60	18
UKAN-32	33,1	99,3	0,0	0,0	/	/	/
Segformer (MiT-b0)	69,5	99,4	50,1	58,9	7,84	3,71	13

L'ambiente di test è Google Colab, con GPU T4.

5 Risultati

Il confronto è stato effettuato tra tre configurazioni diverse di iperparametri per UKAN e Segformer MiT-b0. Gli iperparametri regolano il numero di canali che l'immagine avrà dopo avere attraversato i livelli dell'encoder, e quelli di default erano impostati a $inputlist = 128, 160, 256$. Volendo mantenere gli iperparametri come potenze di 2 e le proporzioni tra essi, si sono provate altre due configurazioni che rispettivamente dimezzassero e quadrimezzassero quella di default: $inputlist = 64, 80, 128$ e $inputlist = 32, 40, 64$. Per Segformer è stata scelta, invece, la versione con meno parametri dato l'intento perseguito di fare segmentazione a basso costo computazionale: il fine-tuning dell'implementazione preaddestrata fornita dalla libreria Transformers di Hugging Face è durato 200 epoche, sulle immagini del training-set, con batch di 8 immagini per volta.

Come metrica di accuratezza è stata scelta la F1-score poichè utilizzata precedenti condotti sullo stesso argomento: infatti è una metrica robusta allo sbilanciamento tra classi tipico di questo task. La complessità è valutata in termini di tre diverse metriche riguardanti la grandezza del modello e le sue tempistiche: numero di parametri, numero di operazioni svolte dal modello in GMACs, tempo impiegato per svolgere una predizione (inference time);

Il test è stato effettuato sulle immagini del campo 003, e la grandezza del batch per l'inferenza era di 8 immagini. Dai risultati mostrati nella tabella Tabella 1, si nota che i modelli UKAN detengono le migliori performance in termini di F1-score a fronte di tempistiche di inferenza di pochi millesimi maggiori rispetto a Segformer. In particolare UKAN-64, eccetto che per l'inference time, ha le migliori metriche: pare, infatti, essere il migliore compromesso tra l'eccessiva complessità di UKAN-128 e l'underfitting di UKAN-32.

6 Conclusioni

Nello studio abbiamo esaminato le performance di una U-Net aumentata tramite l'utilizzo di KAN nell'identificazione delle infestanti in immagini scattate da droni: l'obiettivo era quello di verificare se con un minore numero di parametri fosse possibile ottenere delle metriche migliori rispetto a modelli attualmente più utilizzati per la segmentazione semantica, quali Segformer. Il risultato si ritiene soddisfacente in quanto UKAN-64, una UKAN con opportuna configurazione di iperparametri, risulta più precisa e meno grande, in termini di parametri ed operazioni, rispetto alla più leggera versione di Segformer, Segformer MiT-b0: il tempo di inferenza è, invece, di poco migliore in Segformer. Dal momento in cui l'utilizzo di UKAN-64 pare essere una scelta vincente, future direzioni di ricerca potrebbero consistere nell'integrazione di tale architettura in modelli non supervisionati, come RowWeeder [3].

Riferimenti bibliografici

- [1] Castellano, G., De Marinis, P., Vessio, G.: Weed mapping in multispectral drone imagery using lightweight vision transformers. In *Neurocomputing*, 562, pagina 126914, 2023.
- [2] Chicchón Apaza, M.Á., Monzón, H.M.B., Alcarria, R.: Semantic Segmentation of Weeds and Crops in Multispectral Images by Using a Convolutional Neural Networks Based on U-Net. In *International Conference on Applied Technologies*, pagine 473–485, Springer 2019.
- [3] De Marinis, P., Vessio, G., Castellano, G.: RoWeeder: Unsupervised Weed Mapping through Crop-Row Detection. *arXiv preprint arXiv:2410.04983*, 2024.
- [4] dos Santos Ferreira, A., Freitas, D.M., da Silva, G.G., Pistori, H., Folhes, M.T.: Weed Detection in Soybean Crops Using ConvNets. In *Computers and Electronics in Agriculture*, 143, pagine 314–324, 2017.

- [5] Li, C., Liu, X., Li, W., Wang, C., Liu, H., Liu, Y., Chen, Z., Yuan, Y.: U-KAN Makes Strong Backbone for Medical Image Segmentation and Generation. *arXiv preprint arXiv:2406.02918*, 2024.
- [6] Lottes, P., Behley, J., Chebrolu, N., Milioto, A., Stachniss, C.: Joint Stem Detection and Crop-Weed Classification for Plant-Specific Treatment in Precision Farming. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pagine 8233–8238, IEEE 2018.
- [7] Ma, X., Wang, Z., Hu, Y., Zhang, X.: Kolmogorov-Arnold Network for Remote Sensing Image Semantic Segmentation. *arXiv preprint arXiv:2501.07390*, 2025.
- [8] Maria, J., Valanarasu, J., Vishal, M.P.: UNeXt: MLP-based Rapid Medical Image Segmentation Network. *arXiv preprint arXiv:2203.04967*, 2022.
- [9] Sa, I., Chen, Z., Popović, M., Khanna, R., Liebisch, F., Nieto, J., Siegwart, R.: Weednet: Dense Semantic Weed Classification Using Multispectral Images and Mav for Smart Farming. *IEEE robotics and automation letters*, 3(1), pagine 588–595, 2017.
- [10] Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., Siegwart, R.: A Large-Scale Semantic Weed Mapping Framework Using Aerial Multispectral Imaging and Deep Neural Network for Precision Farming. *Remote Sensing*, 10(9), pagina 1423, 2018.
- [11] Vougioukas, S.G.: Agricultural Robotics. In *Annual Review of Control, Robotics, and Autonomous Systems*, 2, pagine 365–392, 2019.
- [12] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In *Advances in Neural Information Processing Systems*, 34, 2021.