

BỘ DỮ LIỆU THỰC ĐỊA VÀ NHẬN DẠNG GIAI ĐOẠN SINH TRƯỞNG BẮP CẢI CHO ROBOT PHUN DINH DƯỠNG

HOÀNG VĂN MẠNH¹ - NGUYỄN TRỌNG THỊNH¹ - NGUYỄN TẤT VIỆT¹ -
ĐỖ NAM¹ - PHẠM MẠNH THẮNG^{1*}

¹Khoa Cơ học kỹ thuật và Tự động hóa, Trường Đại học Công nghệ (UET),
Đại học Quốc gia Hà Nội

¹Faculty of Engineering Mechanics and Automation,
University of Engineering and Technology (UET), Vietnam National University, Hanoi

*Email: thangpm@vnu.edu.vn

Ngày nhận bài: 15/5/2026

Ngày sửa bài: 28/5/2026

Ngày duyệt đăng bài: 10/6/2026

Tóm tắt:

Nghiên cứu xây dựng bộ dữ liệu thực địa và đánh giá mô hình nhận dạng giai đoạn sinh trưởng của cây bắp cải nhằm phục vụ robot phun dinh dưỡng chính xác. Bài viết tự thu 7.225 ảnh thực địa tại Mễ Hạ và hợp nhất với 1.786 ảnh UAV thành bộ 9.011 ảnh, 16.616 nhãn theo 5 giai đoạn. Trong 5 mô hình phát hiện đối tượng được so sánh thì YOLOv8n được chọn nhờ cân bằng độ chính xác (mAP 85,6%) và tốc độ, tối ưu bằng TensorRT FP16 đạt khoảng 21 FPS trên Jetson Nano. Kết quả thực nghiệm cho thấy, hệ thống đạt độ chính xác nhận dạng trên 90% và sai lệch vị trí phun dưới 1 cm. Nghiên cứu góp phần xây dựng cơ sở dữ liệu và cung cấp giải pháp nhận dạng phù hợp cho các hệ thống robot nông nghiệp, tạo tiền đề cho việc ứng dụng công nghệ trí tuệ nhân tạo trong canh tác chính xác.

Từ khóa: nhận dạng đối tượng, học sâu, YOLO, bắp cải, nông nghiệp chính xác.

Field dataset and cabbage growth stage recognition for nutrient spraying robots

Abstract:

Accurate identification of cabbage growth stages is essential for precise fertilization, and its effectiveness depends directly on training data. This study constructs a field dataset and evaluates models for recognizing cabbage growth stages to support precision nutrient spraying robots. A total of 7,225 field images collected at Me Ha were combined with 1,786 UAV images, forming a dataset of 9,011 images with 16,616 annotations across five growth stages. Among five object detection models, YOLOv8n was selected due to its balance between accuracy (mAP 85.6%) and speed, achieving approximately 21 FPS on Jetson Nano after TensorRT FP16 optimization. Experimental results show over 90% recognition accuracy and spraying position deviation below 1 cm. The study contributes to dataset development and provides a suitable recognition solution for agricultural robotic systems, enabling the application of artificial intelligence in precision farming.

Keywords: object detection, deep learning, YOLO, cabbage, precision agriculture.

1. Đặt vấn đề

Bón phân trong sản xuất rau phần lớn vẫn dựa vào kinh nghiệm và thao tác thủ công. Với bắp cải, các cây trong cùng một luống thường phát triển không đều, đồng thời nhu cầu dinh dưỡng thay đổi rõ theo từng giai đoạn, nên bón đồng loạt dễ gây thừa cho cây non và thiếu cho cây lớn. Việc nhận dạng đúng giai đoạn của từng cây bằng mắt ở quy mô lớn là không khả thi. Học sâu và thị giác máy tính cho phép tự động hóa khâu này và đã được chứng minh giúp tăng năng suất, giảm 10 đến 20% lượng vật tư nhờ tác động đúng vị trí (Vatin N. I., et al., 2024). Riêng với bắp cải, các mô hình YOLO đã được dùng để phát hiện và định vị cây ngoài đồng (Lu Y., 2020).

Hiệu quả của cách tiếp cận này phụ thuộc trực tiếp vào dữ liệu huấn luyện. Camera nhận dạng được gắn trên robot, chĩa xuống quan sát tán cây trong vùng phun, nên cho một góc nhìn từ trên xuống ở cự ly gần. Các bộ dữ liệu bắp cải công khai tuy cũng chụp từ trên xuống nhưng chủ yếu ở điều kiện nền sạch, ít nhiễu và trên số giống hạn chế (Yokoyama Y., et al., 2024), khác với đồng ruộng thực nơi có nền đất phức tạp, cỏ dại, lá chồng chéo và ánh sáng thay đổi. Khoảng cách miền này nằm chủ yếu ở điều kiện môi trường và độ đa dạng giống chứ không phải ở hướng nhìn, làm giảm khả năng tổng quát hóa khi triển khai. Vì vậy, đóng góp chính của bài báo là quy trình tự xây dựng một bộ dữ liệu thực địa cho bắp cải tại Việt Nam và hợp nhất nó với nguồn UAV để tăng tính đa dạng. Trên nền dữ liệu đó, nhóm tác giả huấn luyện và so sánh 5 mô hình phát hiện đối tượng, chọn YOLOv8n và tích hợp lên một robot phun dinh dưỡng dùng Jetson Nano và Arduino. Phần trọng tâm được trình bày là cách xây dựng bộ dữ liệu và ảnh hưởng tới hiệu năng nhận dạng.

2. Phương pháp nghiên cứu

2.1. Đối tượng nghiên cứu

Đối tượng nghiên cứu là bắp cải cuốn tròn (*Brassica oleracea* nhóm *capitata*), rau chủ lực vụ Đông Xuân ở miền Bắc. Dựa trên tài liệu của Andaloro và cộng sự (1983) cùng thang BBCH (Lüling N, et al., 2023), quá trình phát triển được chuẩn hóa thành 5 giai đoạn nhận biết qua ảnh RGB và dùng làm 5 lớp nhãn: S2_3 (cây con, tán hẹp), S4_5 (trải lá, mở tán), S6_7 (bắt đầu cuốn bắp), S8 (bắp hình thành rõ) và S9 (bắp cuộn chặt). Ranh giới giữa các giai đoạn thể hiện qua mật độ và hình dạng lá, mức độ cuốn bắp, là những đặc trưng mà mô hình phải học được để phân biệt. (Hình 1)

Hình 1. Bắp cải giai đoạn S2_3 (trái) và S4_5 (phải)



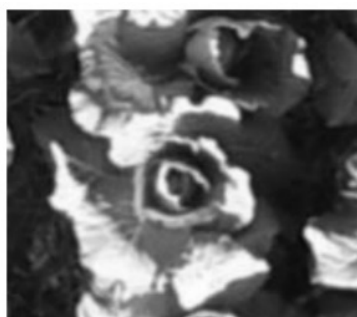
2.2. Xây dựng bộ dữ liệu

Đây là phần dữ liệu tự thu thập, đóng vai trò cốt lõi trong nghiên cứu. Ảnh được chụp trực tiếp tại khu canh tác của làng rau Mễ Hạ (Yên Phú, Yên Mỹ, Hưng Yên) đúng chính vụ, bằng điện thoại Samsung Galaxy S21 5G với 2 ống kính góc siêu rộng và góc rộng. Ảnh chủ yếu được chụp từ trên xuống ở cự ly gần để bám sát góc nhìn gần thẳng đứng của camera nhận dạng trên robot, đồng thời có thêm các ảnh chụp nghiêng khoảng 45° và ở 2 mức phóng đại 0,5x và 1,0x nhằm tăng độ đa dạng về góc chụp và kích thước tán. Nhờ vậy, phân bố ảnh trong tập huấn luyện vừa gần với dữ liệu robot thực sự gặp khi vận hành, vừa đủ đa dạng để mô hình bền hơn trước sai lệch lắp đặt và rung lắc.

Được thu ngay trên đồng ruộng thực, bộ Mễ Hạ phản ánh đầy đủ các yếu tố phức tạp ngoài đồng: cỏ dại bản địa, tàn dư thực vật, đất khi ẩm khi khô, lá của các cây liền kề chồng lấn, và ánh sáng biến động từ nắng gắt tới chiều muộn. Tổng cộng 7.225 ảnh đã được thu thập. Phần lớn cây thuộc 2 giống F1 phổ biến là KK689 và New Cross, nên hình thái giữa các cây trong cùng ruộng khá tương đồng; điều này giúp mô hình học đặc trưng từng giai đoạn dễ hơn, nhưng cũng tiềm ẩn nguy cơ lệ thuộc vào đặc điểm giống nếu chỉ dùng riêng nguồn dữ liệu này. Đó chính là lý do cần kết hợp thêm một nguồn đa giống.

Nguồn thứ hai lấy từ tập ảnh bắp cải công bố bởi Yokoyama và cộng sự (2024), gồm 456 ảnh chụp thẳng đứng 90° từ UAV tại Gifu (Nhật Bản) với nhiều giống thương mại khác nhau. Vì ảnh UAV cũng là góc nhìn từ trên xuống, cùng hướng với camera của robot, nguồn này bổ sung sự đa dạng hình thái nhiều giống mà bộ Mễ Hạ còn thiếu; khác biệt duy nhất cần xử lý là cỡ vật thể. Do mỗi ảnh góc bao quát diện tích rộng nên từng cây khá nhỏ; khi thu nhỏ ảnh để đưa vào mô hình, thông tin hình thái bị suy giảm, gây khó cho các kiến trúc phụ thuộc kích thước vật thể như YOLO. Nhóm tác giả cắt các ảnh gốc thành ảnh con sao cho mỗi ảnh chỉ chứa 1 đến 2 cây, loại bỏ ảnh trống hoặc quá mờ, thu được 1.786 ảnh sau kết xuất. Bước này vừa tăng mật độ thông tin, vừa đưa kích thước vật thể của ảnh UAV về gần với ảnh mặt đất, khép lại chênh lệch cỡ vật thể vốn là khác biệt chính giữa 2 nguồn. (Hình 2)

Hình 2: So sánh ảnh kết xuất từ UAV (trái) và ảnh thực tế tại Mễ Hạ (phải)



Toàn bộ ảnh được gán nhãn thủ công trên nền tảng Roboflow theo 5 lớp nói trên. Để giữ tính nhất quán, nhóm tác giả đặt 2 quy tắc: với ảnh UAV đã kết xuất, chỉ gán những cây xuất hiện trọn vẹn hoặc bị che không quá 30% ở mép ảnh nhằm tránh nhiễu khi tính IoU; với ảnh Mễ Hạ có lá chồng lấn, hộp bao được hiệu chỉnh để tách rõ các cá thể liền kề. Tất cả ảnh được đưa về cùng hướng theo thông tin EXIF và chuẩn hóa kích thước 480×480 pixel, kích thước đủ lớn để giữ chi tiết hình thái nhưng vẫn nhẹ cho thiết bị nhúng (Lüling N., et al., 2023).

Về tăng cường dữ liệu, nhóm tác giả áp dụng chiến lược tùy theo kiến trúc. Trong 2 mô hình cơ sở Faster R-CNN và SSDLite MobileNetV3 được giữ nguyên dữ liệu gốc để thiết lập một đường cơ sở khách quan. Nhóm YOLO dùng thêm kỹ thuật tăng cường dữ liệu nhưng có tiết chế: giảm cường độ ghép ảnh khảm, tắt hoàn toàn kỹ thuật trộn điểm ảnh trên các bộ nhỏ và chỉ giữ các phép biến đổi hình học, màu sắc ở mức nhẹ. Lý do là việc phân biệt các giai đoạn phụ thuộc vào những đặc trưng tinh tế như mật độ lá và độ cuộn bắp, tăng cường quá mạnh dễ làm biến dạng chính các đặc trưng đó.

Bộ hợp nhất được tạo bằng cách gộp toàn bộ 2 nguồn, gồm 9.011 ảnh và 16.616 nhãn, nhằm tận dụng đồng thời tính chuẩn hóa của ảnh UAV và tính đa dạng môi trường của ảnh thực địa. Đặc tính 3 bộ dữ liệu được tóm tắt trong Bảng 1. Một điểm quan trọng là 2 nguồn được chọn để tương phản có chủ đích trên cùng một hướng nhìn từ trên xuống: UAV nhiều giống nhưng nền sạch và chụp từ trên cao, còn Mễ Hạ chỉ 2 giống nhưng ở cự ly gần với nền nhiễu cao. Nói cách khác, sự khác biệt giữa 2 nguồn nằm ở giống và độ phức tạp của nền chứ không phải ở hướng quan sát; chính sự tương phản này về sau được dùng để phân tích khả năng tổng quát hóa của mô hình. (Bảng 1)

Bảng 1. Đặc tính của 3 bộ dữ liệu

Đặc điểm	Bộ UAV	Bộ Mễ Hạ (tự thu)	Bộ hợp nhất
Nguồn	UAV, Gifu (Nhật Bản)	Samsung S21, Mễ Hạ	Tổng hợp
Số ảnh	1.786 (sau kết xuất)	7.225	9.011
Giống cây	Nhiều giống thương mại	F1 KK689, New Cross	Đa dạng
Hướng nhìn	Từ trên xuống 90°	Từ trên xuống 90°/45°	Từ trên xuống
Môi trường	Nền sạch, ít nhiễu	Nhiều cao (cỏ, bóng...)	Đa dạng

Mỗi bộ được chia theo tỷ lệ 80% huấn luyện, 10% xác thực và 10% kiểm tra nội bộ, phân chia phân tầng để giữ cân bằng lớp. Thống kê trên bộ hợp nhất (Bảng 2) cho thấy, các lớp phân bố khá đều, mỗi giai đoạn chiếm khoảng 17% đến 23%, nên mô hình không bị học lệch về giai đoạn nào. Ngoài 3 bộ trên, một tập kiểm thử độc lập gồm ảnh và video hoàn toàn mới, không thuộc bất kỳ phân chia nào, được dựng riêng để đánh giá khả năng tổng quát hóa khi triển khai.

Bảng 2. Số lượng nhãn theo từng giai đoạn

Giai đoạn	Mễ Hạ	UAV	Hợp nhất
S2_3	3.047	628	3.675
S4_5	3.156	731	3.887
S6_7	1.970	817	2.787
S8	2.680	628	3.308
S9	2.306	653	2.959
Tổng	13.159	3.457	16.616

2.3. Mô hình và huấn luyện

Khi 5 mô hình được chọn đại diện cho 2 hướng phát hiện đối tượng. Nhóm 2 giai đoạn có Faster R-CNN R50-FPN (Jiang P., et al., 2024), độ chính xác cao nhưng suy luận chậm. Nhóm một giai đoạn gồm SSDLite MobileNetV3 (Ren S. et al., 2015; Liu W., et al., 2015) và 3 phiên bản nano của họ YOLO là YOLOv5n, YOLOv8n, YOLOv11n (Redmon J., et al., 2015; Ultralytics, 2024) chỉ xét biến thể nano vì đích triển khai là Jetson Nano. So với YOLOv5n dùng hộp neo, YOLOv8n chuyển sang thiết kế không neo với đầu tách rời, còn YOLOv11n cải tiến thêm khối trích xuất đặc trưng, hữu ích khi cây bị lá che khuất.

Toàn bộ mô hình được huấn luyện trên GPU Tesla T4, sử dụng các trọng số tiền huấn luyện trên COCO (Lin T. Y., et al., 2014). với cùng kích thước ảnh đầu vào 480×480. Các siêu tham số như kích thước lô, số vòng lặp và lịch tốc độ học được điều chỉnh theo quy mô từng bộ dữ liệu; các cơ chế dừng sớm, theo dõi mất mát trên tập xác thực và lưu mô hình tốt nhất được áp dụng thống nhất để bảo đảm so sánh công bằng.

2.4. Kiến trúc hệ thống và ánh xạ điều khiển

Mô hình đã chọn được tích hợp lên robot theo kiến trúc điều khiển phân cấp. Máy tính nhúng Jetson Nano chạy YOLOv8n trên TensorRT, nhận ảnh từ camera, phát hiện cây và giai đoạn, dùng một bộ theo dõi để gán ID ổn định cho từng cây rồi ra quyết định phun. Các quyết định được gửi xuống Arduino Mega 2560 qua UART; Arduino điều khiển động cơ bước và bơm theo thời gian thực bằng kiến trúc đa nhiệm FreeRTOS kết hợp ngắt Timer. Vị trí phun được xác định qua một mô hình ánh xạ tuyến tính từ tọa độ pixel sang số xung động cơ, hiệu chuẩn bằng 2 điểm biên đo được. Cơ chế gán ID bảo đảm mỗi cây chỉ được phun một lần trong một lượt đi qua.

3. Kết quả và thảo luận

3.1. Đánh giá hiệu suất các mô hình

Các mô hình được đánh giá bằng mAP@0.5, mAP@0.5:0.95 (trung bình trên nhiều ngưỡng IoU, phản ánh mức ôm sát của hộp), Recall, cùng FPS hệ thống đo trực tiếp trên luồng video. Bảng 3 tổng hợp kết quả trên bộ hợp nhất.

Bảng 3. Độ chính xác và tốc độ hệ thống trên bộ hợp nhất

Mô hình	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Recall (%)	FPS
Faster R-CNN	90,2	78,9	85,7	7,3
SSDLite MobileNetV3	91,4	67,4	78,1	23,5
YOLOv5n	94,0	85,0	92,3	41,5
YOLOv8n	94,5	85,6	91,5	50,7
YOLOv11n	95,0	86,5	91,0	46,1

Kết quả cho thấy, vai trò của cách xây dựng dữ liệu. Trên riêng bộ Mễ Hạ nhiều nhiều, nhóm YOLO giữ được mAP@0.5:0.95 quanh 87 đến 89%, trong khi Faster R-CNN giảm còn khoảng 81,6% và SSDLite MobileNetV3 chỉ đạt 69,4% do định vị hộp kém trong nền phức tạp. Trên bộ hợp nhất, nơi mô hình phải xử lý đồng thời 2 nguồn khác nhau về giống và độ phức tạp của nền, YOLOv11n và YOLOv8n vẫn dẫn đầu với mAP@0.5:0.95 lần lượt 86,5% và 85,6%. Việc giữ độ chính xác cao trên cả 2 nguồn cho thấy mô hình học được đặc trưng hình thái ổn định của cây thay vì lệ thuộc vào một giống hay một kiểu nền, và đây chính là giá trị của bộ dữ liệu đa nguồn tự xây dựng: buộc mô hình tổng quát hóa qua nhiều giống, nhiều điều kiện nền và cỡ vật thể.

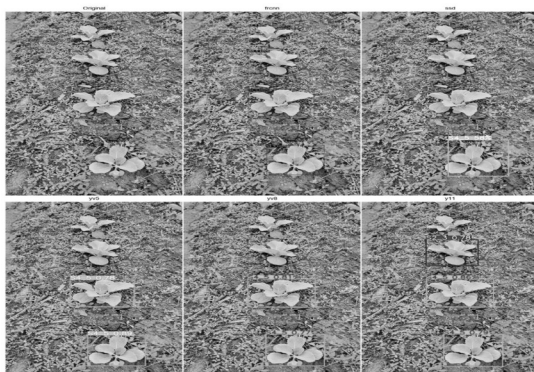
3.2. Tốc độ, lựa chọn mô hình và tối ưu TensorRT

Về tốc độ, cả 3 biến thể YOLO đều đạt FPS hệ thống trên 40, trong khi Faster R-CNN chỉ khoảng 7,3 FPS và SSDLite MobileNetV3 khoảng 23,5 FPS. Cân nhắc giữa độ chính xác và tốc độ, YOLOv8n được chọn để triển khai: nó nhanh nhất (50,7 FPS) trong khi mAP chỉ thấp hơn YOLOv11n dưới 1%, đồng thời tạo vùng đệm an toàn khi chuyển sang phần cứng yếu hơn. Sau khi chuyển sang engine TensorRT ở chế độ FP16, tốc độ suy luận thuần trên Jetson Nano tăng từ 26,8 lên 36,0 FPS so với FP32 (đây là tốc độ engine thuần; tốc độ toàn hệ thống trên robot sẽ thấp hơn do chi phí đọc ảnh và tiền xử lý và hậu xử lý), với chất lượng phát hiện gần như giữ nguyên. Trên tập kiểm thử độc lập (Hình 3), cả 5 mô hình đều không nhận nhầm nền thành cây; nhóm YOLO cho hộp bám sát tán lá và duy trì định danh từng cây ổn định trên luồng video, phù hợp cho robot di động.

3.3. Thử nghiệm tích hợp trên robot

Thử nghiệm trên robot thật cho thấy, sai lệch vị trí tia phun theo phương ngang nằm trong khoảng vài milimét đến khoảng 1 cm, chấp nhận được với tia phun tỏa nhẹ và kích thước tán bấp cải. Khi nguồn cấp ổn định, tốc độ YOLOv8n trên TensorRT đạt trung bình khoảng 21,3 FPS; khi nguồn không bảo đảm và hệ thống tự hạ xung nhịp, tốc độ còn khoảng 16,3 FPS, nhưng ở cả 2 trường hợp độ chính xác nhận dạng giai đoạn vẫn giữ trên 90%. Hệ thống đã khép kín được chuỗi nhận dạng, quyết định phun và điều khiển cơ cấu chấp hành trên nền tảng robot thực (Hình 4).

Hình 3. Kết quả nhận dạng giai đoạn sinh trưởng của các mô hình trên ảnh kiểm thử.



Hình 4. Robot vận hành thực tế trên luồng bấp cải



4. Kết luận

Bài viết đã trình bày quy trình tự xây dựng một bộ dữ liệu thực địa cho cây bắp cải và hợp nhất nó với nguồn UAV, tạo thành bộ 9.011 ảnh với 16.616 nhãn theo 5 giai đoạn sinh trưởng. Kết quả thực nghiệm cho thấy, tính đa nguồn và sát thực địa của bộ dữ liệu giúp mô hình tổng quát hóa tốt qua nhiều giống và điều kiện nền; trên nền dữ liệu đó, YOLOv8n được chọn và tối ưu bằng TensorRT để chạy thời gian thực trên Jetson Nano, cho độ chính xác nhận dạng trên 90% và sai lệch vị trí phun dưới 1 cm khi tích hợp lên robot. Hướng phát triển tiếp theo là mở rộng dữ liệu sang nhiều giống, nhiều vụ và các điều kiện khắc nghiệt hơn, đồng thời thực hiện đối chứng định lượng giữa phun bằng robot và phun thủ công về năng suất và mức tiết kiệm vật tư■

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội theo đề tài mã số CN24.20

TÀI LIỆU THAM KHẢO:

- Andaloro J. T., Rose K. B., Shelton A. M., et al. (1983). Cabbage Growth Stages. *New York State Agricultural Experiment Station*, 101, 1-4.
- Howard A., Sandler M., Chu G., et al. (2019). Searching for MobileNetV3. arXiv:1905.02244.
- Jiang P., Qi A., Zhong J., et al. (2024). Field cabbage detection and positioning system based on improved YOLOv8n. *Plant Methods*, 20, 96.
- Lin T.-Y., Maire M., Belongie S., et al. (2014). Microsoft COCO: Common Objects in Context. arXiv:1405.0312.
- Liu W., Anguelov D., Erhan D., et al. (2015). SSD: Single Shot MultiBox Detector. arXiv:1512.02325.
- Lu Y., & Young S. (2020). A survey of public datasets for computer vision tasks in precision agriculture. *Computers and Electronics in Agriculture*, 178, 105760.
- Lüling N., Reiser D., Straub J., et al. (2023). Fruit Volume and Leaf-Area Determination of Cabbage by a Neural-Network-Based Instance Segmentation for Different Growth Stages. *Sensors*, 23, 129.
- Redmon J., Divvala S., Girshick R., et al. (2015). You Only Look Once: Unified, Real-Time Object Detection. arXiv:1506.02640.
- Ren S., He K., Girshick R., et al. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497.
- Ultralytics (2021). YOLOv5 Documentation. <https://docs.ultralytics.com/yolov5/>.
- Ultralytics (2023). YOLOv8 Models Documentation. <https://docs.ultralytics.com/models/yolov8/>.
- Ultralytics (2024). YOLO11 Models Documentation. <https://docs.ultralytics.com/models/yolo11/>.
- Vatin N. I., Joshi S. K., Acharya P., et al. (2024). Precision Agriculture and Sustainable Yields: Insights from IoT-Driven Farming. *BIO Web of Conferences*, 86, 01091.
- Yokoyama Y., Matsui T., & Tanaka S. T. (2024). An annotated image dataset of cabbages for instance segmentation. *Mendeley Data*, V2. <https://doi.org/10.17632/5cp2dyjczk.2>.