

GIÁM SÁT LỢN TỰ ĐỘNG TỪ TRANG TRẠI SỬ DỤNG HỌC SÂU

Dương Trần Đức

Học viện Công nghệ Bưu chính Viễn thông

Tóm tắt: Tự động phát hiện và theo dõi hành vi lợn từ camera giám sát trong các trang trại là một hoạt động quan trọng trong việc phát hiện bất thường sức khỏe lợn. Theo các nghiên cứu trước đây, lợn có vấn đề về sức khỏe thường có các biểu hiện khác thường trong hành vi. Việc theo dõi hành vi lợn có thể sớm phát hiện các bất thường và có thể xem đó là dấu hiệu của vấn đề về sức khỏe để có các hoạt động khám bệnh và điều trị sớm nhằm ngăn chặn tình trạng lây lan của dịch bệnh. Trong bài báo này, chúng tôi trình bày nghiên cứu tự động theo dõi và nhận diện hành vi lợn từ các camera giám sát sử dụng học sâu. Phương pháp tiếp cận bao gồm 1 chuỗi các thuật toán như phát hiện lợn và theo dõi lợn trong video, nhận diện hành vi từ hình ảnh lợn theo dõi được. Từ các hành vi nhận diện được, có thể phát hiện ra tình trạng bất thường như sự thay đổi đáng kể về các mẫu hành vi. Các thực nghiệm cho thấy mặc dù phương pháp tiếp cận sử dụng chuỗi các mô hình nhận diện, theo dõi, nhưng độ chính xác tổng thể phát hiện lợn (mAP) 92.2% và theo dõi (MOTA) 85.2% cho thấy tính khả thi khi áp dụng trong thực tế.

Từ khóa: phát hiện vật thể, theo dõi lợn, nhận diện hành vi lợn.

I. MỞ ĐẦU

Thịt lợn là một trong những loại thực phẩm chính được con người sử dụng rất nhiều trên thế giới nói chung và tại Việt Nam nói riêng. Chăn nuôi lợn trở thành một ngành lớn và có nhiều ảnh hưởng. Tại Việt Nam, trước đây phổ biến hình thức nuôi lợn nhỏ lẻ theo hộ gia đình. Tuy nhiên, gần đây xu hướng dịch chuyển theo mô hình chăn nuôi công nghiệp ngày càng cao do nhu cầu tăng năng suất, chuẩn hóa quy trình và ứng dụng công nghệ cao vào hoạt động quản lý trang trại.

Một trong các hoạt động quan trọng thường xuyên phải thực hiện là giám sát hoạt động lợn để phát hiện các mẫu hành vi bất thường, qua đó dự đoán tình trạng bất thường về sức khỏe lợn. Phương pháp giám sát truyền thống là sử dụng nhân công. Tuy nhiên, phương pháp này tiêu tốn thời gian và nhân lực, khi phải tiến hành giám sát thường xuyên và khó có khả năng kết hợp các thông tin giám sát từ các thời điểm trong quá khứ với thời điểm hiện tại với số lượng lợn lớn trong trang trại. Việc sử dụng camera giám sát theo cách thông thường cũng chỉ giúp tăng khả năng theo dõi từ xa và xem lại hình ảnh quá khứ khi có nhu cầu. Một phương pháp cho phép theo dõi và nhận diện hành vi lợn từ camera giám sát sẽ giúp giải

quyết các vấn đề trên. Đầu tiên, việc theo dõi tự động sẽ giúp quá trình theo dõi được thực hiện thường xuyên và không bị ngắt quãng. Hơn nữa, theo dõi tự động cũng giúp ghi nhận các biểu hiện hành vi của các cá thể lợn trong 1 thời gian dài, giúp khả năng phát hiện bất thường tốt hơn khi kết hợp chuỗi hành vi trong 1 khoảng thời gian thay vì chỉ xét trong 1 thời điểm.

Bài báo này trình bày phương pháp tự động theo dõi, nhận diện hành vi lợn từ camera giám sát. Các hành vi nhận diện được có thể được lưu trữ và kết hợp để xem xét mẫu hành vi bất thường trong một khoảng thời gian, từ đó có các dự báo về tình trạng sức khỏe lợn. Các thực nghiệm được thực hiện trên tập dữ liệu video lợn được thu thập từ trang trại thực tế với 2.000 ảnh lợn được gán nhãn (tổng cộng hơn 27.000 nhãn lợn). Các video thu thập trong điều kiện tự nhiên của trang trại nhưng vẫn cho kết quả tiềm năng với độ chính xác nhận diện và theo dõi lần lượt trên 92% và trên 85%.

Bài báo có cấu trúc như sau. Phần II trình bày về các nghiên cứu liên quan trong lĩnh vực tự động phát hiện, theo dõi, và nhận diện hành vi lợn. Phần III mô tả phương pháp. Phần IV trình bày về các kết quả và thảo luận. Cuối cùng, các kết luận sẽ được trình bày trong phần V của bài báo.

II. TỔNG QUAN

Tự động theo dõi hành vi lợn là một hướng nghiên cứu đã nhận được nhiều quan tâm trong các năm gần đây. Thời gian đầu, các nghiên cứu được thực hiện để theo dõi và nhận diện hành vi theo nhóm lợn (group-level). Mặc dù nhận diện hành vi theo nhóm lợn đã chứng minh tác dụng trong cảnh báo sức khỏe đàn, nhưng việc theo dõi và nhận diện hành vi theo cá thể (individual-level) đem lại nhiều ưu điểm khi có thể đánh giá được tình trạng của từng cá thể và có phương án điều trị riêng khi cần thiết [1]. Theo dõi tới từng cá thể là một tác vụ khó khăn hơn, tuy nhiên với những cải tiến về kỹ thuật phát hiện và theo dõi vật thể trong thời gian gần đây, các nghiên cứu về theo dõi và phát hiện lợn hiện tập trung chủ yếu vào hướng thực hiện theo cá thể.

Theo dõi lợn tự động từ camera là hướng nghiên cứu được quan tâm rộng rãi gần đây, do giá thành thực hiện thấp và dễ dàng lắp đặt, triển khai so với các phương pháp giám sát bằng thiết bị đeo (thẻ tai, thiết bị đeo chân lợn). Để thực hiện phát hiện và theo dõi lợn tự động từ camera, các phương pháp dựa trên học máy đã được áp dụng rộng rãi [1-7]. Một số nghiên cứu sử dụng các cảm biến có chiều sâu như camera 3D để đo đặc chiều sâu đối tượng trong ảnh và phát hiện ra hành vi lợn là đang đứng hay nằm [8-10]. Tuy nhiên, việc sử dụng các thiết bị có khả năng đo chiều sâu này làm giá thành tăng lên, đồng thời không áp dụng để nhận diện các hành vi đa dạng khác như

Tác giả liên hệ: Dương Trần Đức,

Email: ducdt@ptit.edu.vn

Đến tòa soạn: 8/2022, chỉnh sửa: 9/2022, chấp nhận đăng: 10/2022.

ăn, uống, đi, chạy, v.v. Các nghiên cứu gần đây tập trung vào sử dụng kỹ thuật học sâu để phát hiện, theo dõi và nhận diện đa dạng các hành vi lộn. Phương pháp này có thể được áp dụng trên các camera 2D thông thường, giá thành rẻ nên có thể được ứng dụng rộng rãi. Các nghiên cứu tập trung vào theo dõi và phát hiện các hành vi phổ biến như đi, đứng, nằm, ăn, uống [5-7, 11] hoặc đơn giản là theo dõi và đo đạc thời gian di chuyển và đứng yên [1]. Các hành vi tĩnh có thể được xác định trong giai đoạn phát hiện lộn bằng cách sử dụng mô hình phát hiện đa lớp thay vì một lớp là cá thể lộn, như sử dụng các lớp đứng, nằm, ăn [6] hoặc sử dụng một mô hình phân loại hành vi dựa trên ảnh để phân loại ảnh cá thể lộn thành các lớp hành vi khác nhau [7]. Các hành vi động như đi, chạy cần phải được nhận diện trong quá trình theo dõi bằng cách xác định khoảng cách vị trí lộn đã di chuyển giữa các frame liên tiếp [6, 7].

III. PHƯƠNG PHÁP

Phương pháp được sử dụng trong nghiên cứu được chia làm 3 giai đoạn: phát hiện cá thể lộn trong ảnh, theo dõi cá thể lộn trong video, và nhận diện hành vi lộn trong video. Từ một video đầu vào, các ảnh sẽ được trích xuất ra từ video nhờ bóc tách các video frame. Một thuật toán phát hiện vật thể sẽ được áp dụng để phát hiện các vùng ảnh có cá thể lộn. Từ các cá thể lộn được phát hiện trong các video frame liên tiếp nhau, có thể tiến hành theo dõi hành động của từng cá thể lộn dựa trên vị trí, hình ảnh lộn. Các hành vi của lộn có thể được nhận diện trong quá trình phát hiện và theo dõi. Các hành vi tĩnh như đứng, nằm, ăn có thể được nhận diện thông qua một mô hình phân loại ảnh. Hành vi liên quan đến chuyển động (đi, chạy) cần phải nhận diện dựa trên vị trí lộn trong các frame liên tục trong video. Chi tiết về các thuật toán được áp dụng trong các giai đoạn này được trình bày ở các phần tiếp theo.

A. Phát hiện lộn trong ảnh

Trong giai đoạn này, thuật toán YOLO v5 [12] được sử dụng để phát hiện các vùng ảnh có cá thể lộn trong một ảnh toàn cảnh thu được từ camera. YOLO là một lớp thuật toán phát hiện đa vật thể trong ảnh được sử dụng rất rộng rãi hiện nay, cho kết quả tốt cả về khía cạnh độ chính xác và tốc độ thực hiện. Để ứng dụng YOLO v5 trong nghiên cứu này, đầu tiên chúng tôi tiến hành hoạt động tiền huấn luyện (pre-train) mạng này với tập dữ liệu chung là COCO [13], sau đó tiến hành hoạt động học chuyển giao (transfer-learning) với tập dữ liệu riêng.

Hoạt động tiền huấn luyện là một hoạt động cần thiết với hầu hết các mạng nơ ron tích chập (CNN) khi sử dụng trong bài toán phân hoặc phát hiện vật thể, nhằm tạo ra các trọng số khởi tạo tốt cho các mạng nơ ron dựa trên các tập dữ liệu cực kỳ lớn với hàng triệu ảnh.

Trong giai đoạn tiếp theo, mô hình sẽ được huấn luyện tiếp trên tập dữ liệu riêng, trong đó vật thể duy nhất cần phát hiện trong ảnh là cá thể lộn.

B. Theo dõi lộn trong video

Phương pháp theo dõi vật thể trong video phổ biến hiện nay là dựa trên việc phát hiện vật thể trong các video frame, từ đó tiến hành khớp các vật thể phát hiện giữa các frame kết tiếp dựa trên tính toán quỹ đạo kết hợp khoảng cách vị trí. Trong nghiên cứu này, chúng tôi sử dụng DeepSORT [14], một phương pháp có ứng dụng độ tương đồng hình ảnh vào quá trình khớp các đối tượng.

DeepSORT là một bản cải tiến của thuật toán SORT, vốn sử dụng bộ lọc Kalman [15] cùng với thuật toán Hungarian [16] để khớp tập các đối tượng phát hiện được giữa các video frame, từ đó hình thành quá trình theo dõi từng cá thể lộn. Cải tiến chính của DeepSORT nằm ở chỗ nó sử dụng độ tương đồng hình ảnh học được từ một mạng học sâu để xác định hai ảnh cá thể lộn có phải là từ một cá thể lộn hay không. Cải tiến này làm tăng khả năng xác định đúng một đối tượng phát hiện ở video frame kế tiếp là đối tượng nào đã được phát hiện ở video frame trước hay là một đối tượng mới được phát hiện. Điều này sẽ giúp làm giảm các tình huống nhận diện sai đối tượng khi xảy ra trường hợp bị mất dấu (do khuất tầm camera hoặc bị lỗi trong quá trình phát hiện).

Để xây dựng độ đo tương đồng như nói ở trên, chúng tôi sử dụng một mạng học sâu để tái nhận diện cá thể lộn dựa trên hình ảnh. Mạng học sâu này nhận đầu vào là một ảnh của đối tượng được phát hiện và được huấn luyện để tối ưu hàm mất mát giữa lớp của ảnh dự đoán và ảnh thật. Sau khi mạng được huấn luyện xong, lớp phân loại của mạng sẽ được loại bỏ và một véc tơ đặc trưng sẽ được tạo ra cho ảnh đầu vào và so sánh với véc tơ đặc trưng của các ảnh các đối tượng đã được lưu trữ sẵn từ trước để nhận biết xem nó tương đồng với đối tượng nào nhất. Quá trình huấn luyện mạng tái nhận diện này được thực hiện bước một với hoạt động tiền huấn luyện trên tập dữ liệu tái nhận diện người (person Re-ID) MARS [17] và bước 2 là thực hiện học chuyển tiếp trên tập dữ liệu tái nhận diện lộn do chúng tôi tự xây dựng.

C. Nhận diện hành vi lộn

Khi thực hiện phát hiện và theo dõi thành công một cá thể lộn từ camera, các hành vi của các thể lộn này có thể được nhận diện trong từng video frame hoặc trên 1 chuỗi các video frame liên tiếp.

Với mỗi cá thể lộn được phát hiện trong 1 video frame, ảnh của cá thể lộn sẽ được trích ra và đưa vào một mô hình phân loại tư thế sử dụng mạng học sâu. Mạng học sâu này được thiết kế để nhận đầu vào là một ảnh cá thể lộn và đầu ra là một trong các nhãn tư thế là đứng, nằm, hoặc ăn. Trong trường hợp tư thế của cá thể lộn được nhận diện là nằm hoặc ăn, hành vi được xác định trong video frame tương ứng với tư thế đó. Tuy nhiên, nếu tư thế được nhận diện là đứng, thì cần tiếp tục xác định trong các video frame tiếp theo xem hành vi là đứng hay dịch chuyển. Điều này được xác định dựa trên vị trí tương đối của cá thể lộn trong các video frame liên tiếp. Nếu vị trí của cá thể lộn không thay đổi thì hành vi được xác định là đứng, ngược lại hành vi được xác định là di chuyển. Trên thực tế, do tồn tại các sai số khi dự đoán vị trí của các cá thể lộn trong video frame, một ngưỡng khoảng cách cần đặt ra để xác định vị trí là có thay đổi hay không. Trong nghiên cứu này, chúng tôi đặt ngưỡng là 2cm để xác định khoảng cách tương đối giữa 2 vị trí là có dịch chuyển hay không dịch chuyển. Lưu ý rằng do vị trí của các cá thể lộn được xác định bằng một hình chữ nhật bao quanh nên để đo đạc khoảng cách tương đối giữa 2 vị trí, chúng tôi sử dụng khoảng cách Euclidean giữa 2 tâm của hình chữ nhật bao quanh này.

Các hành vi sau khi được xác định sẽ được lưu trữ trong cơ sở dữ liệu hành vi và từ đó có thể tổng hợp, phân tích để phát hiện ra những bất thường về hành vi trong một khoảng thời gian và có thể xem như dấu hiệu về bất thường trong sức khỏe lộn.

IV. THỰC NGHIỆM VÀ KẾT QUẢ

A. Dữ liệu và môi trường thực nghiệm

Trong nghiên cứu này, chúng tôi sử dụng tập dữ liệu là các video tự thu thập từ các camera giám sát tại một trang trại lợn thương mại.

Các camera được sử dụng là loại camera Hik Vision (resolution: 1280 x 720, focal length: 3.6mm, max frame rate: 25 FPS). Đây là loại camera rẻ tiền thông thường được sử dụng phổ biến trong hoạt động giám sát chung. Các camera được đặt tại 2 chuồng của 1 trang trại lợn để thu thập dữ liệu thực nghiệm. Mặc dù các camera này không đem lại chất lượng hình ảnh tốt nhất để tiến hành thực nghiệm nhưng việc sử dụng chúng

Các chuồng lợn được đặt camera thu thập dữ liệu thực nghiệm là các chuồng rộng 10m x 4m, có 13 – 20 cá thể lợn ở độ tuổi 2-4 tháng và cân nặng từ 40-80kg. Camera được đặt từ trên cao xuống nhưng theo góc chéo chứ không đặt thẳng (top view) nhằm thu được toàn cảnh của chuồng. Tuy nhiên, do đặc điểm các chuồng lợn hiện tại có độ cao thấp nên khả năng tạo khung nhìn của camera khá hạn chế khi phải đặt góc chéo hẹp và các cá thể lợn dễ bị che khuất lẫn nhau. Các video được thu trong khoảng thời gian 1 tuần và được chọn lọc để sử dụng làm dữ liệu thực nghiệm. Các video được chọn để làm tập dữ liệu có độ phân giải 1280 x 720, được chọn trong các khung giờ khác nhau để đảm bảo độ khách quan của tập dữ liệu. Từ các video này, các video frame được trích ra để xây dựng các tập dữ liệu phục vụ phát hiện lợn, theo dõi lợn, và tái nhận diện lợn.

Tập dữ liệu phát hiện lợn bao gồm 2.000 ảnh là các frame được trích ra từ các video được chọn lọc như nói ở trên. Tập dữ liệu này chỉ phục vụ tác vụ phát hiện cá thể lợn trong ảnh nên chỉ cần chọn các video frame không quá giống nhau và không cần đánh thứ tự cho các cá thể lợn trong ảnh. Để thực hiện gán nhãn, chúng tôi sử dụng công cụ Labellimg [18] để thực hiện tạo một đường viền hình chữ nhật bao quanh mỗi cá thể lợn. Công cụ này sẽ tạo một tập dữ liệu chú giải lưu trữ vị trí của hình viền bao gồm tọa độ tâm và 2 góc trên trái dưới phải của hình chữ nhật.

Tập dữ liệu theo dõi lợn nhằm mục đích để đánh giá hiệu suất của mô hình chứ không nhằm mục đích huấn luyện mô hình theo dõi do mô hình theo dõi chỉ dựa vào mô hình phát hiện cá thể lợn và các kỹ thuật như bộ lọc Kalman và thuật toán Hungarian, kỹ thuật khớp tập đối tượng giữa 2 video frame như đã trình bày trong phần trước. Tập dữ liệu theo dõi này cũng được trích từ các video như tập dữ liệu phát hiện lợn, tuy nhiên các ảnh phải được lấy từ các frame theo thứ tự chạy trong video và không nhất thiết phải liên tục nhưng không được cách quá xa. Ngoài ra, các cá thể trong ảnh ngoài việc tạo đường viền vị trí còn cần phải đánh chỉ số cho chúng và duy trì chỉ số giống nhau trên các frame cho cùng một cá thể để đánh giá khả năng theo dõi đúng đối tượng từ frame này qua frame khác. Hai tập dữ liệu theo dõi trong nghiên cứu này được tạo từ các video ban ngày và buổi tối chứa nhiều loại hành vi đa dạng để có thể đánh giá các mô hình một cách khách quan nhất.

Tập dữ liệu tái nhận diện lợn được xây dựng từ các ảnh lợn được trích ra từ các frame trong video. Số lợn được thu thập các ảnh nhận diện là 36 cá thể với tổng cộng 8.600 ảnh. Các ảnh có kích thước là 256 x 128 và

được dùng để huấn luyện mạng học sâu tái nhận diện lợn như mô tả ở phần III.



a) Ảnh ban đầu



b) Ảnh đã gán nhãn

Hình 1. Một số ảnh mẫu trong tập dữ liệu

B. Độ đo đánh giá

Để đánh giá mô hình nhận diện, đầu tiên độ chính xác vị trí được tính toán qua chỉ số IoU (Intersection of Union) của 2 hình chữ nhật viền của đối tượng được phát hiện và đối tượng thực tế theo công thức 1. Chỉ số IoU càng cao chứng tỏ vị trí được phát hiện càng chính xác. Một ngưỡng sẽ được đặt ra để cho biết mức độ khớp của 2 hình là có đạt không (thường đặt mức 0.5 trong các nghiên cứu trước đây).

$$IoU = \frac{\text{Diện tích giao của 2 hình}}{\text{Diện tích hợp của 2 hình}} \quad (1)$$

Từ chỉ số IoU của mỗi cặp cá thể dự đoán được và thực tế, độ đo AP (Average Precision) sẽ được tính cho mỗi lớp đối tượng và cuối cùng độ đo mAP là trung bình của các AP trên mỗi lớp được sử dụng để độ đo đánh giá cuối cùng theo các công thức 2 và 3, trong đó $TP(c)$ và $FP(c)$ lần lượt là chỉ số dự đoán đúng dương và sai dương (True Positive và False Positive) còn $AP(c)$ là chỉ số AP của lớp c.

$$AP(c) = \frac{TP(c)}{TP(c) + FP(c)} \quad (2)$$

$$mAP = \frac{1}{\text{classes}} \sum_{\text{classes}} AP(c) \quad (3)$$

Độ đo đánh giá cho mô hình theo dõi lợn được sử dụng trong nghiên cứu là độ đo MOTA (Multi-Object Tracking Accuracy). Đây là độ đo được sử dụng phổ biến

trong các nghiên cứu về theo dõi đa đối tượng trước đây, điển hình là trong các cuộc thi MOT Challenge. Độ đo này sử dụng công thức dựa trên 3 loại lỗi có thể xảy ra trong theo dõi đa đối tượng là số sai âm (FN – False Negative), số sai dương (FP – False Positive) và số cá thể bị nhầm (IDSW – Identity Switch) như công thức số 4.

$$MOTA = 1 - \frac{\sum_i FN_i + FP_i + IDSW_i}{\sum_i GT_i} \quad (4)$$

Trong công thức trên, FN_i là cá thể lợn có thực nhưng không được theo dõi, FP_i là cá thể lợn không có thực nhưng lại được theo dõi, $IDSW_i$ là số cặp cá thể bị nhầm định danh, và GT_i là cá thể lợn thực tế.

C. Kết quả thực nghiệm

Bảng 1 cho thấy kết quả của mô hình phát hiện lợn. Lưu ý rằng bên cạnh độ đo mAP như đã trình bày ở phần trên, chúng tôi báo cáo thêm kết quả nhận diện theo các độ đo Precision (P) và Recall (R).

Bảng 1. Kết quả của mô hình phát hiện lợn

Lớp	mAP (%)	P (%)	R (%)
Stand	91.3	89.4	91.8
Lie	97.9	97.6	97.4
Eat	89.6	89.7	89.8
All	92.2	93.0	92.9

Theo kết quả trong bảng, độ chính xác tổng thể trên 90 % là kết quả tốt. Mặc dù vậy, có thể thấy hành vi ăn (Eat) có kết quả kém hơn hai hành vi còn lại, do việc phân biệt hành vi ăn là khó nhất do phải xác định vị trí tương đối của cá thể lợn với máng ăn và việc lợn tới gần máng là để ăn hay không. Đây là vấn đề cần được cải thiện trong các nghiên cứu tiếp theo.

Về kết quả của mô hình theo dõi lợn, các kết quả tổng thể từ các chỉ số MOTA được thấy như trong bảng 2 là các kết quả tốt. Các kết quả theo dõi vào ban đêm có độ chính xác thấp hơn ban ngày do chất lượng hình ảnh ban đêm cho khả năng nhận diện kém hơn. Đây cũng là một vấn đề cần cải thiện trong các nghiên cứu tiếp theo.

Bảng 2. Kết quả của mô hình theo dõi lợn

	MOTA (%)
Tập dữ liệu ban ngày	85.2
Tập dữ liệu tối	78.3

V. KẾT LUẬN

Trong nghiên cứu này, chúng tôi đã trình bày phương pháp phát hiện, theo dõi, và nhận diện hành vi lợn trong điều kiện tự nhiên tại các trang trại lợn sử dụng học sâu. Các thực nghiệm đã được tiến hành trên các tập dữ liệu tự thu thập từ các trang trại lợn thương mại, với các kịch bản thực nghiệm khác nhau cho các kết quả khả quan. Điều này cho thấy tiềm năng của việc ứng dụng phương pháp trong thực tế, đặc biệt là chỉ cần sử dụng các thiết bị thông thường và rẻ tiền như camera giám sát.

Mặc dù các kết quả là khả quan, tuy nhiên để hệ thống hoàn thiện và có tính ứng dụng thực tế tốt hơn cần giải quyết thêm các vấn đề như nâng cao độ chính xác trong điều kiện ánh sáng kém như ban đêm hoặc vấn đề nhận diện cá thể lợn khi bị che khuất hoặc ở xa so với tầm nhìn camera. Một vấn đề nữa Đây là các hướng nghiên cứu tiềm năng trong tương lai.

LỜI CẢM ƠN

Nghiên cứu này được hỗ trợ bởi Viện Hàn lâm Khoa học và Công nghệ Việt Nam trong đề tài mã số CN4000.01/21-23.

TÀI LIỆU THAM KHẢO

- [1] J. Cowton, I. Kyriazakis, and J. Bacardit, "Automated Individual Pig Localisation, Tracking and Behaviour Metric Extraction Using Deep Learning," *IEEE Access*, vol. 7, pp. 108049-108060, 2019, doi: 10.1109/ACCESS.2019.2933060.
- [2] S. Matthews, A. Miller, J. Clapp, T. Plötz, and I. Kyriazakis, "Early detection of health and welfare compromises through automated detection of behavioural changes in pigs," *Vet. J.* 217, 43–51, 2016.
- [3] S. Matthews, A. Miller, T. Plötz, and I. Kyriazakis, "Automated tracking to measure behavioural changes in pigs for health and welfare monitoring," *Sci. Rep.* 7, 17582, 2017.
- [4] J. Sa, Y. Choi, H. Lee, Y. Chung, D. Park, and J. Cho, "Fast pig detection with a top-view camera under various illumination conditions," *Symmetry*, vol. 11, no. 2, p. 266, 2019.
- [5] A. Alameer, I. Kyriazakis, H. Dalton, A. Miller, and J. Bacardit, "Automatic recognition of feeding and foraging behaviour in pigs using deep learning," *Biosyst. Eng.* 197, 91–104, 2020.
- [6] A. Alameer, I. Kyriazakis, and J. Bacardit, "Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs," *Sci Rep* 10, 13665, 2020, doi: 10.1038/s41598-020-70688-6
- [7] L. Bergamini et al., "Extracting accurate long-term behavior changes from a large pig dataset," in *Proc. 16th VISIGRAPP*, Vol. 4, pp. 524-533, 2021, doi: 10.5220/0010288405240533.
- [8] J. Kim et al., "Depth-Based Detection of Standing-Pigs in Moving Noise Environments," *Sensors* 17, 2757, 2017, doi: 10.3390/s17122757
- [9] J. Kim et al., "Lying-Pig Detection using Depth Information," in *Proc. ICACS '17*, pp. 40-43, 2017, doi: 10.1145/3127942.3127949.
- [10] M. Mittek, E. Psota, L. Pérez, T. Schmidt, and B. Mote, "Health monitoring of group-housed pigs using depth-enabled multi-object tracking," in *Proc. Int Conf. Pattern Recognit., Workshop Vis. Observ. Anal. Vertebrate Insect Behav.*, 2016.
- [11] M. Riekert, A. Klein, F. Adrion, C. Hoffmann, and E. Gallmann, "Automatically detecting pig position and posture by 2d camera imaging and deep learning," *Comput. Electron. Agric.* 174, 105391, 2020.
- [12] A. Bochkovskiy, C. Wang, and H. Liao, "Yolov4: Optimal speed and accuracy of object detection," *CoRR*, 2020, abs/2004.10934.
- [13] TY. Lin et al., "Microsoft COCO: common objects in context," In: *Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Computer Vision-ECCV*, pp. 740-755, 2014.
- [14] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," *arXiv*, 2017, doi: 10.48550/ARXIV.1703.07402.
- [15] R. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, D, J. Basic Eng.*, vol. 82, pp. 35-45, 1960.

- [16] H. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logistics Quart.*, vol. 2, nos. 1_2, pp. 83_97, 1955.
- [17] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "Mars: A video benchmark for large-scale person re-identification," in *Proc. Eur. Conf. Comput. Vis. Cham*, pp. 868-884, 2016.
- [18] Tzutalin, "LabelImg," *Git code*, 2015. <https://github.com/tzutalin/labelImg>
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 770-778, 2016.
- [20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 779-788, 2016.

AUTOMATICALLY TRACKING PIGS FROM PIG FARMS USING DEEP LEARNING

Abstract: Automatically detection and tracking pigs from surveillance cameras in pig farms is an important task in abnormal detection of pig health and welfare. Previous researches showed that pigs with problems in health often have abnormalities in behaviors. Monitoring the pig behaviors can help detect the abnormality timely and take early medical examination and treatment to prevent disease from spreading. In this paper, we presented a method for automatically tracking and identifying the pig behaviors from surveillance cameras in pig farms using deep learning. Our approach includes a combination of algorithms such as detection and tracking pigs in videos, identifying pig behaviors from tracked pig images and locations. From the tracked behaviors, abnormalities can be detected as the changes in behavior patterns. Experiments showed that, although the approach used a pipeline of detection, tracking, behavior classification model, the overall result, with 92.2 % mAP and 85.2 % MOTA, is promising in practice.

Keywords: deep learning, pig tracking, farm AI.



Dương Trần Đức Tốt nghiệp Đại học KHTN, Đại học Quốc gia Hà Nội ngành Công nghệ thông tin năm 1999, Thạc sỹ chuyên ngành Hệ thống thông tin tại Đại học Tổng hợp Leeds, Vương Quốc Anh năm 2004, và Tiến sỹ chuyên ngành Kỹ thuật máy tính tại Học viện Công nghệ Bưu chính Viễn thông năm 2018. Hiện đang công tác tại Khoa Công nghệ Thông tin, Học viện Công nghệ Bưu chính Viễn thông.