

Quality Control of Casting Aluminum Parts: A Comparison of Deep Learning Models for Filings Detection

AI VIETNAM - RESEARCH TEAM

October 22, 2023

Date of publication:	26-27/04/2023
Authors:	Rui Nascimento, Tony Ferreira, Cláudia Rocha, Vítor Filipe, Manuel F. Silva, Germano Veiga, Luis Rocha
Sources:	2023 IEEE
Keywords:	Quality control, Convolutional Neural Networks, Filings detection, Casting aluminum, Automotive industry
Summary by:	Quy Nguyen Dinh

1. Purpose/outputs:

- Compare the effectiveness of four deep learning models: Faster R-CNN, RetinaNet, YOLOv7 and YOLOv7-tiny in detecting metal filings on cast aluminum parts.
- Find out the most suitable model to perform the quality inspection task of cast aluminum parts for the automotive industry.

2. Contributions:

- Compare the performance of four different deep learning models: Faster R-CNN, RetinaNet, YOLOv7 and YOLOv7-tiny in detecting metal scraps, based on metrics such as precision, recall, F1-score and mAP.
- Choose Faster R-CNN as the best model for the quality inspection task, because it has the highest recall (96.00%), meaning it can detect most of the metal scraps that can cause errors.
- Test two different types of light: LED bar and dome light, to find out the best type of light for taking pictures of aluminum cast parts. The results show that dome light reduces reflection and shadow on the surface of the parts.

3. Inputs:

- A customized dataset consisting of 500 images with a resolution of 2592x1944 pixels, manually labeled using LabelImg. The images contain metal scraps of different sizes, shapes and reflective characteristics, placed in the tear zone of aluminum cast parts.
- The dataset was randomly split into three parts: training (80%), validation (10%) and testing (10%).
- The images were resized to 1600x1200 pixels to fit the GPU memory capacity and speed up the training.

- The images were also augmented by horizontal and vertical flips, and brightness variation.



Figure 1: Example of the shape and size of filings to be detected.

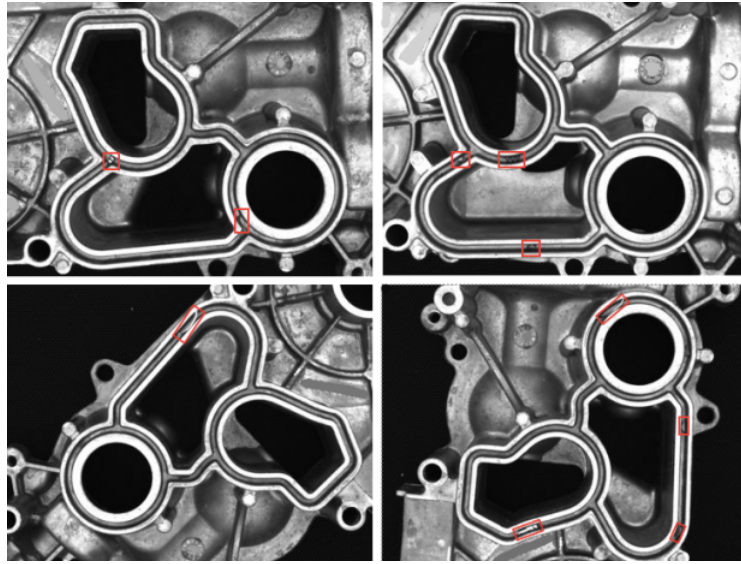


Figure 2: Example of dataset images.

4. Methodology:

- **Image Acquisition and Processing:**

The experimental device used two types of light to capture the aluminum parts: LED bars and dome light. Dome light was chosen because it gave the best results in reducing reflection and shadow.

- **Model deployment**

4 CNN models were selected to deploy, including 3 one-stage models: RetinaNet, YOLOv7, YOLOv7-tiny and 1 two-stage model: Faster R-CNN.

RetinaNet: Backbone Network block - responsible for computing convolutional feature maps over the entire image. Feature Pyramid Network (FPN) is implemented as the backbone network, providing a feature pyramid by using a top-down method with lateral connections; Sub-network for Object Classification is a Fully Convolutional Network (FCN) attached to each FPN level for object classification; Sub-network for Object Regression is attached to each feature of the FPN in parallel with the classification sub-network.

Faster R-CNN: is a model from the R-CNN family. Faster R-CNN uses another convolutional network - Region Proposal Network (RPN) to generate region proposals.

YOLOv7: Compared to other YOLO models, two major changes were made at the architecture level and the trainable bag-of-freebies level, related to improving the model's accuracy without increasing the training cost.

YOLOv7-tiny: is a basic model optimized for edge computing. Compared to other versions, YOLOv7-tiny uses leaky ReLU as the activation function, while other models use SiLU as the activation function.

TensorBoard was used to analyze data in real time during training. To calculate the results, False Positive (FP) was considered as wrongly detecting a filing and False Negative (FN) as not detecting a filing. The Intersection over Union (IoU) threshold is 50%.

5. Results:

- Image size significantly affects the performance, as the defects to be detected range from 1 mm to 12 mm, the images were resized to 1600x1200 pixels.
- Figure 3 shows the comparison results of the four models on the test dataset, which corresponds to 50 images. Regarding inference time, Faster R-CNN and YOLOv7-tiny were respectively the slowest (171 ms) and the fastest (12.2 ms) to make a prediction. However, all of them are fast enough for the problem requirements. When comparing the results, the Faster R-CNN model achieved the highest precision and recall, 96.96% and 96%, respectively.

Model	Batch size	Precision (%)	Recall (%)	F1-score (%)	mAP@0.5(%)	mAP@0.5:0.95(%)	GPU memory* (MiB)	Inference time (ms)
Faster R-CNN	1	95.09	97.00	96.04	96.65	55.76	5637	171.00
RetinaNet	1	96.87	93.00	94.90	92.97	49.86	4513	165.00
YOLOv7	1	96.90	93.00	94.91	97.70	59.20	5670	50.30
YOLOv7-tiny	5	94.00	94.00	94.00	95.80	55.90	5429	12.20

Figure 3: Comparison results between the proposed models.

6. Limitations:

The article does not specifically mention limitations, but there may be some issues that need further clarification, such as whether the model works well with noisy data or imbalanced data.

7. Future researches:

- Use the selected Faster R-CNN model to continue the project, which can be further trained to improve its performance and detect other common defects such as scratches, dents, and cracks, etc.
- Increase the hardware processing capability to achieve better performance results.
- Use CNN-based semantic segmentation networks to obtain detailed information of the tear zone, which can be considered to perform a comparison, combined with CNN-based object detection models.

Kiểm Soát Chất Lượng Các Bộ Phận Đúc Nhôm: Một Sự So Sánh Giữa Các Mô Hình Học Sâu Trong Phát Hiện Vụn

AI VIETNAM - RESEARCH TEAM

Ngày 22 tháng 10 năm 2023

Ngày công bố:	26-27/04/2023
Tác giả:	Rui Nascimento, Tony Ferreira, Cláudia Rocha, Vítor Filipe, Manuel F. Silva, Germano Veiga, Luis Rocha
Nguồn:	2023 IEEE
Từ khóa:	Kiểm soát chất lượng, Mạng nơ-ron tích chập, Phát hiện vụn, Đúc nhôm, Ngành công nghiệp ô tô
Người tóm tắt:	Nguyễn Đình Quý

1. Mục đích:

- So sánh hiệu quả của bốn mô hình học sâu: Faster R-CNN, RetinaNet, YOLOv7 và YOLOv7-tiny trong việc phát hiện các vụn kim loại trên các bộ phận đúc bằng nhôm.
- Tìm ra mô hình phù hợp nhất để thực hiện nhiệm vụ kiểm tra chất lượng của các bộ phận đúc bằng nhôm cho ngành công nghiệp ô tô.

2. Đóng góp:

- So sánh hiệu năng của bốn mô hình học sâu khác nhau: Faster R-CNN, RetinaNet, YOLOv7 và YOLOv7-tiny trong việc phát hiện các vụn kim loại, dựa trên các chỉ số như precision, recall, F1-score và mAP.
- Chọn Faster R-CNN là mô hình phù hợp nhất cho nhiệm vụ kiểm tra chất lượng, vì nó có recall cao nhất (96.00%), có nghĩa là nó có thể phát hiện được hầu hết các vụn kim loại có thể gây ra lỗi.
- Thử nghiệm hai loại ánh sáng khác nhau: thanh LED và ánh sáng vòm, để tìm ra loại ánh sáng tốt nhất cho việc chụp ảnh các bộ phận đúc bằng nhôm. Kết quả cho thấy ánh sáng vòm làm giảm thiểu được phản xạ và bóng tối trên bề mặt của các bộ phận.

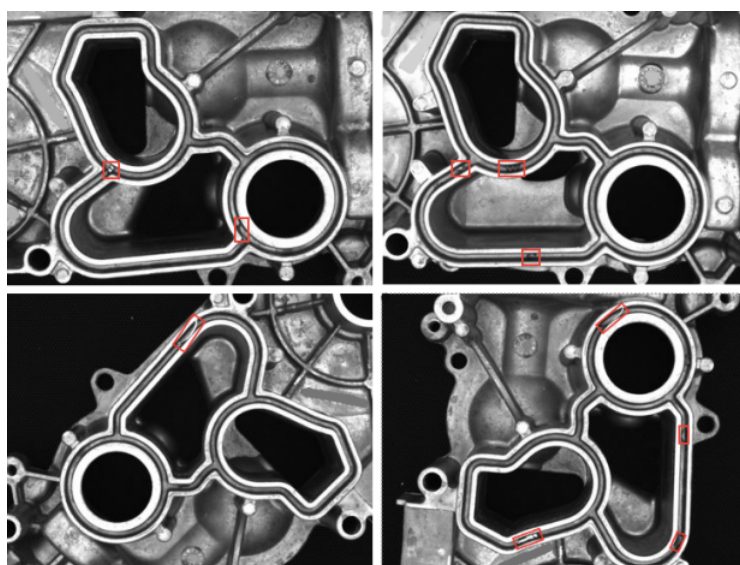
3. Dữ liệu đầu vào:

- Một tập dữ liệu được điều chỉnh gồm 500 ảnh có độ phân giải 2592x1944 pixel, được gán nhãn thủ công, sử dụng LabelImg. Các ảnh chứa các vụn kim loại có kích thước, hình dạng và đặc tính phản xạ khác nhau, được đặt trong tear zone của các bộ phận đúc bằng nhôm.
- Tập dữ liệu được chia ngẫu nhiên thành ba phần: huấn luyện (80%), xác thực (10%) và kiểm tra (10%).

- Các ảnh được thu nhỏ lại còn 1600x1200 pixel để phù hợp với dung lượng bộ nhớ GPU và tăng tốc độ huấn luyện.
- Các ảnh cũng được tăng cường dữ liệu bằng cách lật ngang và dọc, và thay đổi độ sáng.



Hình 1: Ví dụ hình dáng và kích thước của các vụn được phát hiện.



Hình 2: Ví dụ các ảnh dữ liệu.

4. Phương pháp luận:

- **Thu thập và xử lý hình ảnh**

Thiết bị thí nghiệm dùng hai loại ánh sáng để chụp các bộ phận bằng nhôm: thanh LED và ánh sáng vòm. Ánh sáng vòm được chọn vì cho kết quả tốt nhất trong việc giảm thiểu phản xạ và bóng tối.

- **Triển khai mô hình**

4 mô hình CNN được chọn để triển khai, bao gồm 3 mô hình 1 giai đoạn (one-stage): RetinaNet, YOLOv7, YOLOv7-tiny và 1 mô hình 2 giai đoạn (two-stage): Faster R-CNN.

RetinaNet: Khối Backbone Network - có nhiệm vụ tính toán các feature map tích chập trên toàn bộ ảnh. Feature Pyramid Network (FPN) được triển khai làm backbone network, cung cấp một feature pyramid bằng cách sử dụng phương pháp top-down với các kết nối song song; Sub-network cho Phân loại Đối tượng là một Fully Convolutional Network (FCN)

được gắn vào mỗi cấp độ FPN để phân loại đối tượng; Sub-network cho Hồi quy Đối tượng được gắn vào mỗi feature của FPN song song với sub-network phân loại.

Faster R-CNN: là mô hình thuộc họ R-CNN. Faster R-CNN sử dụng một mạng tích chập khác - Region Proposal Network (RPN) để tạo ra các vùng đề xuất.

YOLOv7: So với các mô hình YOLO khác, hai thay đổi lớn được thực hiện ở cấp độ kiến trúc và cấp độ trainable bag-of-freebies, liên quan đến việc cải thiện độ chính xác của mô hình mà không làm tăng chi phí huấn luyện.

YOLOv7-tiny: là một mô hình cơ bản được tối ưu hóa cho tính toán cạnh. So với các phiên bản khác, YOLOv7-tiny sử dụng leaky ReLU làm hàm kích hoạt, trong khi các mô hình khác sử dụng SiLU làm hàm kích hoạt.

TensorBoard được sử dụng để phân tích dữ liệu trong thời gian thực trong quá trình huấn luyện. Để tính toán kết quả, False Positive (FP) được coi là phát hiện sai một mảnh vụn và False Negative (FN) là không phát hiện được mảnh vụn. Ngưỡng Intersection over Union (IoU) là 50%.

5. Kết quả:

- Kích thước ảnh ảnh hưởng đáng kể đến hiệu suất, vì các khuyết tật cần phát hiện dao động từ 1 mm đến 12 mm, các ảnh được thay đổi kích thước thành 1600x1200 pixel.
- Hình 3 cho thấy kết quả so sánh giữa bốn mô hình trên tập dữ liệu kiểm tra, tương ứng với 50 ảnh. Về thời gian suy luận, Faster R-CNN và YOLOv7-tiny lần lượt là chậm nhất (171 ms) và nhanh nhất (12.2 ms) để đưa ra dự đoán. Tuy nhiên, tất cả chúng đều đủ nhanh cho yêu cầu của bài toán. Khi so sánh kết quả, mô hình Faster R-CNN đạt được precision và recall nhất, lần lượt là 96.96% và 96%.

Model	Batch size	Precision (%)	Recall (%)	F1-score (%)	mAP@0.5(%)	mAP@0.5:0.95(%)	GPU memory* (MiB)	Inference time (ms)
Faster R-CNN	1	95.09	97.00	96.04	96.65	55.76	5637	171.00
RetinaNet	1	96.87	93.00	94.90	92.97	49.86	4513	165.00
YOLOv7	1	96.90	93.00	94.91	97.70	59.20	5670	50.30
YOLOv7-tiny	5	94.00	94.00	94.00	95.80	55.90	5429	12.20

Hình 3: So sánh kết quả giữa các mô hình đề xuất.

6. Hạn chế:

Bài báo không đề cập cụ thể đến các hạn chế nhưng có thể có một số vấn đề cần làm rõ thêm như mô hình có hoạt động tốt với dữ liệu nhiễu hoặc dữ liệu mất cân bằng hay không.

7. Các nghiên cứu trong tương lai:

- Sử dụng mô hình Faster R-CNN đã chọn để tiếp tục dự án, có thể huấn luyện thêm để cải thiện hiệu suất của mô hình và phát hiện các lỗi phổ biến khác như trầy xước, lõm, và nứt, v.v.
- Tăng khả năng xử lý phần cứng để đạt được kết quả hiệu suất tốt hơn.
- Sử dụng các mạng ngữ nghĩa dựa trên CNN (CNN-based semantic segmentation networks) để có thông tin chi tiết của tear zone, có thể xem xét để thực hiện so sánh, kết hợp với các mô hình phát hiện đối tượng dựa trên CNN.