Using AI image recognition to find flaws in syringes during the manufacturing process

Grant Proposal

Harshil Hari

Mass Academy

85 Prescott St

Executive Summary

Automated manufacturing has become especially common in factories as of recent times. However, automation is not flawless, and mistakes are often made during the process. For products such as syringes, defects from manufacturing that are not noticed could cause fatal injuries to a patient in a hospital environment. Therefore, it is essential for a check to be in place so that these defects can be spotted. Using humans for defect detection is currently the industry standard, but employees can be costly, and they can become fatigued after hours of scanning for defects. Instead, an image recognition neural network model can be implemented to accurately find defects in real-time during the manufacturing process. An important factor for the success of this model is the processing time. Especially in a manufacturing setting, defects must be found quickly. Therefore, the YOLOv7 deep learning model was chosen as the detection system, as it is known for its fast inference time. 100 images of syringes will be taken, with 50 containing specific defects. Bounding boxes will be used to annotate the defects in those images, and then all the data will be imported into the model. Then, training can be conducted. Percent accuracy results can then be generated, allowing comparisons between similar detection models. Overall, the system will automatically prevent faulty syringes from being sent to hospitals, preventing potential injury.

Using AI image recognition to find flaws in syringes during the manufacturing process

The automated production of goods has become much more common with recent advancements in technology. The total spending for the construction of manufacturing facilities has nearly doubled in the U.S. since 2022, with around 100 billion more dollars spent in 2023 (Nostrand et al., 2023). This increase in spending will allow for the construction of more factories, processing plants, and mills. These facilities can increase efficiency through various means, such as optimized workflows and automation. In particular, automation allows for a much faster production of goods and services and is a large factor in mass production.

Machine error, however, can prevent automated manufacturing from being completely human-independent. Machines can incorrectly fabricate components due to poor materials or equipment failure, for example (Mörke, 2024). Minor mistakes made during the creation of a physical object could cause permanent changes to its structure, potentially inhibiting its ability to function properly. These mistakes during automation are known as manufacturing defects, and companies are developing solutions to prevent these defective products from leaving the factory, as they could decrease consumer satisfaction and brand reputation. Currently, humans are commonly tasked with detecting defects in freshly made products. However, laborers can easily miss defects and can experience fatigue while slowly scanning dozens of objects for various types of defects (Paul et al., 2023). On top of this, labor can be extremely time-consuming and costly for companies, necessitating a detection method that is faster, cheaper, and more accurate.

Neural networks are gaining popularity as an approach to this problem, as they have shown promise in other classification tasks, such as big data processing (Ramachandran, 2024).

Convolutional neural networks (CNN) can be specifically trained to recognize patterns in images. In this case, various filters are applied to an image so that the neural network can understand the image it is analyzing. Afterward, the model can make predictions on the location of objects of interest in the overall image (Ferguson et al., 2018).

When it comes to automated manufacturing, CNNs can be used to analyze each object on the production line by taking a picture of an item and identifying defects through the image, helping determine if that item meets quality standards. With high amounts of training, CNNs can reach levels of speed and accuracy that exceed that of a human (Kaur, 2024), negating the need for slow and methodical checking and encouraging the implementation of a real-time and instantaneous defect detection system that can identify defects, their location, and type. In fact, CNNs have been successfully used before for finding defects in manufacturing, such as for finding flaws in screws (Kuo, 2020).

Image recognition models are not flawless, however. For example, to train an image recognition model, hundreds of images must be gathered of a single product, and in the case of defect detection, images of both undamaged and damaged versions of the product must be captured so that the neural network can learn to differentiate between the two. Capturing images can be difficult if many images of a defective product do not exist in the first place. Environmental conditions are also important during scanning while manufacturing, as a slight difference in lighting or background from training data can cause the model to make incorrect predictions on where defects are located from the images it takes. Compromises must also be made between speed and accuracy, as real-time defect detection will need extremely high levels of speed, which may limit accuracy (Ferguson et al., 2018). However, even though image recognition models may have limitations that prevent them from being perfect, they are still far more efficient than humans at defect checking and a viable option as long as the time and effort can be spent training and establishing a CNN.

In the medical industry, products should be scanned for defects before they are shipped out to hospitals. An important example of this is syringes. Syringes are a very commonly used medical instrument that can inject and withdraw liquids from the human body. Over 16 billion injections are administered yearly, providing medical assistance to hundreds of thousands worldwide (Moustafa et al., 2021). The manufacturing process of syringes involves using multiple automated machines and mills. However, mistakes can be made during the process due to machine error. In the UK, for example, aseptic syringe products alone have been found to have a defect rate of 0.49% (Bateman, 2010), which can mean thousands of syringes with defects at the larger scale. Table 1 describes the various aseptic syringe products and where errors were found during manufacturing over 3 years in the UK.

	Stage at which error detected									
Product type	First check in assembly area	Operator check in preparation area	During labelling	Final check prior to release	At release stage	In clinical area prior to administration	In clinical area during or after administration	Other	Not recorded	Total
Cytotoxic adult	1045 (55.9%)	109 (5.8%)	47 (2.5%)	451 (24.1%)	65 (3.5%)	108 (5.8%)	7 (0.4%)	31 (1.7%)	5 (0.3%)	1868
Cytotoxic paediatric	27 (24.3%)	11 (9.9%)	5 (4.5%)	41 (36.9%)	6 (5.4%)	16 (14.4%)	1 (0.9%)	4 (3.6%)	0	111
Parenteral nutrition—adult	383 (56.2%)	56 (8.2%)	30 (4.4%)	99 (14.5%)	43 (6.3%)	22 (3.2%)	4 (0.6%)	43 (6.3%)	1 (0.1%)	681
Parenteral nutrition—paediatric	74 (41.3%)	21 (11.7%)	0	47 (26.3%)	16 (8.9%)	7 (3.9%)	3 (1.7%)	11 (6.1%)	0	179
Other intravenous additive	504 (39.3%)	179 (14.0%)	128 (10.0%)	262 (20.5%)	110 (8.6%)	35 (2.7%)	6 (0.5%)	56 (4.4%)	1 (0.1%)	1281
Other prefilled syringes	60 (19.4%)	24 (7.7%)	4 (1.3%)	76 (24.5%)	78 (25.2%)	9 (2.9%)	1 (0.3%)	58 (18.7%)	0	310
Other	68 (31.3%)	33 (15.2%)	7 (3.2%)	35 (16.1%)	40 (18.4%)	3 (1.4%)	2 (0.9%)	26 (12%)	3 (1.4%)	217
Not recorded	24 (54.5%)	3 (6.8%)	3 (6.8%)	8 (18.2%)	2 (4.5%)	2 (4.5%)	0	0	2 (4.5%)	44
Total	2185 (46.6%)	436 (9.3%)	224 (4.8%)	1019 (21.7%)	360 (7.7%)	202 (4.3%)	24 (0.5%)	229 (4.9%)	12 (0.3%)	4691

Table 1: Syringes with errors detected during manufacturing in the UK (Bateman, 2010)

With 958,532 aseptic syringes made in the UK during the period, 4,691 of them contained defects, which is especially concerning because improper syringe function in a medical scenario could potentially cause severe harm to a patient. Therefore, a CNN can be employed to determine if any defects are present in newly manufactured syringes, preventing dysfunctional syringes from ever reaching hospitals.

The project proposes a CNN model that can detect defects in syringes during the manufacturing process in real-time. The rest of the paper will discuss the specific aims that will need to be addressed for a strong model to be developed, the methodology for creating that model, and preliminary data to ensure the model meets the specific aims.

Section II: Overview of Specific Aims

This proposal's objective is to create an image recognition model for defects in manufactured syringes in real-time. The rationale is to prevent faulty syringes from entering hospitals and potentially causing injury to patients. The proposed solution for this problem will be a CNN that can detect syringe defects with low RAM usage and fast processing times. Below is a brief description of the project's aims in order to create a successful physical setup in the future. The aims are explained more in-depth in Section III of the document.

Specific Aim 1: Low runtime. The model must have a low runtime, as there will not be much time to find defects while the syringe quickly passes by a camera mounted above a conveyor belt.

Specific Aim 2: Low RAM usage. During the defect detection process, the amount of RAM used by the GPU for image processing must be kept under one-half of the total RAM capacity. High RAM usage can cause a memory bottleneck, which leads to slower performance.

Section III: Project Goals and Methodology

The YOLOv7 model will be used to create the syringe detection system. YOLOv7 is a pre-trained image recognition model that uses a CNN for image classification tasks. It can detect defined "objects" that a user can train it to look for.

When training the neural network, it is essential to capture enough images of damaged and undamaged syringes so that the neural network can accurately find defects in the given syringes. With too few photos to train off, the model will not be able to accurately make predictions of syringe defects. With too many training images, the model would overfit and be unable to make accurate predictions for images of syringes that slightly vary from training data.

The YOLOv7 model is pre-trained on the COCO dataset, which contains around 200,000 images of common objects such as humans, bikes, and helmets (Lin et al., 2015). Therefore, training on new objects, such as syringes, will be much easier as the model will not need as much training data to accurately identify the syringes. For the identification of syringe defects, additional objects will be defined, which are the various defects that could be detected on the syringes. The defect types that will be defined are "bent needle," "torn plunger," and "scratched barrel". These additional objects will be implemented by annotating the training dataset using bounding boxes.

After the neural network model is trained, a physical simulation of the manufacturing process will be created, which contains a camera elevated above a conveyor belt. As syringes on the conveyor belt pass by the camera, the camera will take a top-view picture of the syringe and send it to the CNN to process for defects. Finally, the CNN will output a result containing the types of defects on the syringe and indicate their locations on the image. Now, following this is a specific analysis of the aims for this project.

Specific Aim #1, Low Runtime:

Justification of Aim. On the production line, the neural network will not have much time to process an image of a syringe and determine if it has defects, as it will be passing by quickly on a conveyor belt. If the neural network takes too long to process each image, the production line could be slowed down, causing potential delays and preventing the syringes from being shipped on time. Therefore, a neural network with a fast processing time is essential.

This model's inference time, which is the processing time, was compared to the inference times of other image-based-models that can complete image-classification tasks to gain a better perspective on the speed of the runtime (Wang et al., 2023).

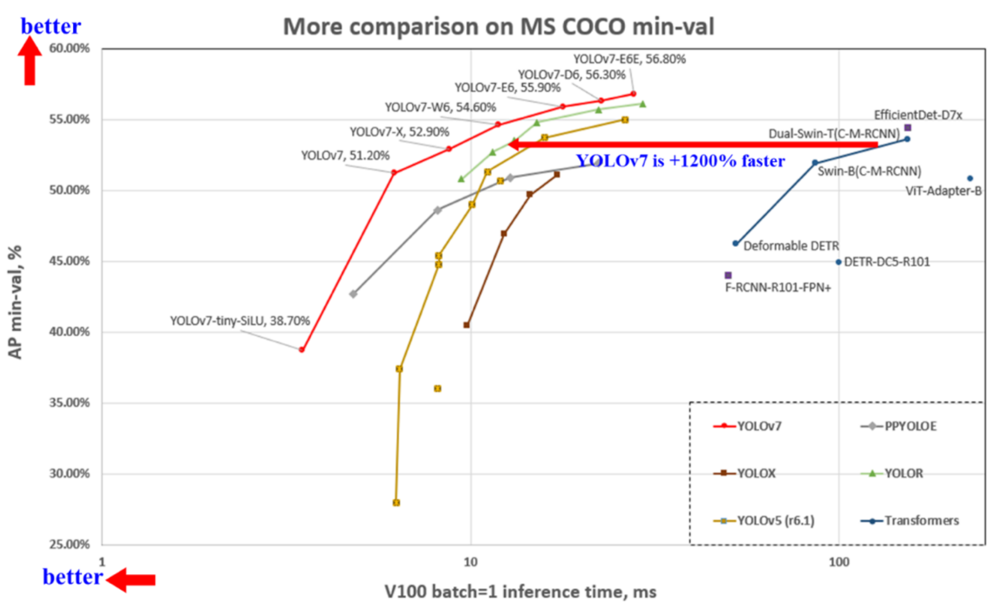

Figure 1: YOLOv7 vs. Other Object Detection Models (Wang et al., 2023)

According to the graph, the YOLOv7 model has a faster inference time than many other object detection models, a reason why this model was initially chosen over others.

Summary of Preliminary Data. The processing time was tested again using the current hardware available for this project, which can help gauge camera placement and the overall viability of a future physical setup for the project.

Ten trials were conducted using stock images of humans, bicycles, cars, umbrellas, and planes to measure the image recognition processing time. These images contained objects the model was already trained on, so the model could identify the objects within them. Although the tests were not syringe-specific, these tests would help determine the model's speed of object detection with the given hardware. The processing times per trial are represented in Figure 2.

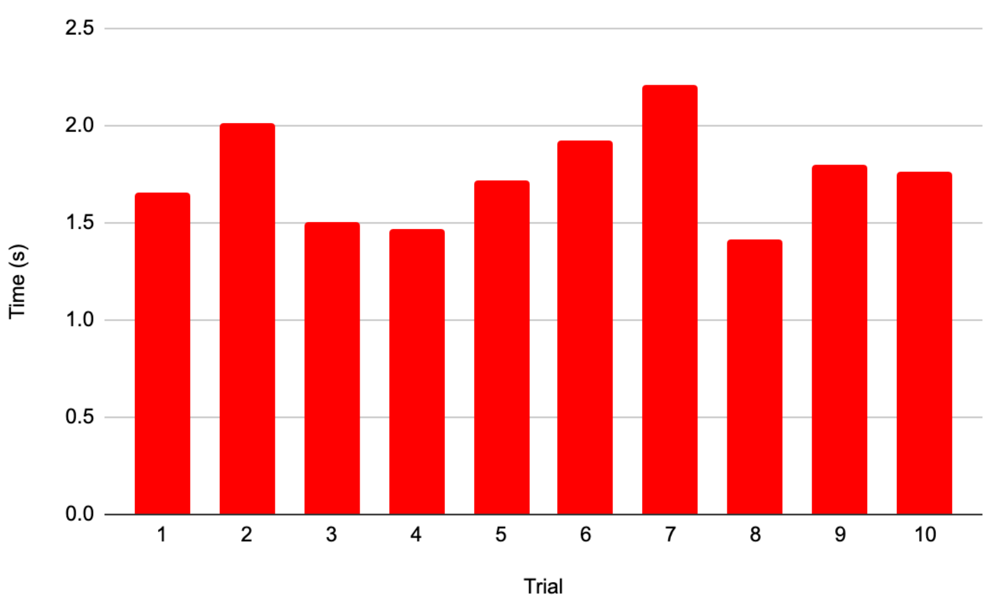

Figure 2: Time to produce output per trial

The average total time was 1.748 seconds. Given that the average conveyor belt moves at 7 feet per second (National Safety Council, 2016), a syringe will travel around 12 feet before defect results are produced. For a camera with a field of view of 10 feet, for example, syringe processing will finish when the syringe is just 2 feet ahead of the camera's position. Therefore, a camera that is elevated a few feet off the ground can determine if a syringe is safe or faulty when the syringe is just a few feet ahead of it. Immediately after detection, the syringe can be disposed of or kept for shipping.

Potential Pitfalls and Alternative Strategies. A potential problem of the given strategy is the assumptions made for the physical setup. For example, it cannot be determined if every conveyor belt the model is used for runs at 7 feet per second. A change in this value will likely warrant the camera to be placed at a different height. Therefore, if these variables are accounted for before the construction of the physical setup, the setup can be better adapted.

Specific Aim #2, Low GPU RAM Usage:

Justification of Aim. A low amount of GPU RAM usage is essential for fast processing of data that is inputted into the CNN. If GPU RAM usage gets too high, the GPU will start falling back on much slower memory options, such as system RAM (Stokes, 2024) that reduces inference times. Considering that keeping inference times low is one of the primary goals of the proposed project, GPU RAM usage must also be kept low.

When choosing the neural network for this task, it was essential to find a deep learning framework that uses low amounts of GPU RAM. Figure 3 details the average GPU memory usage of three different frameworks for various batch sizes.

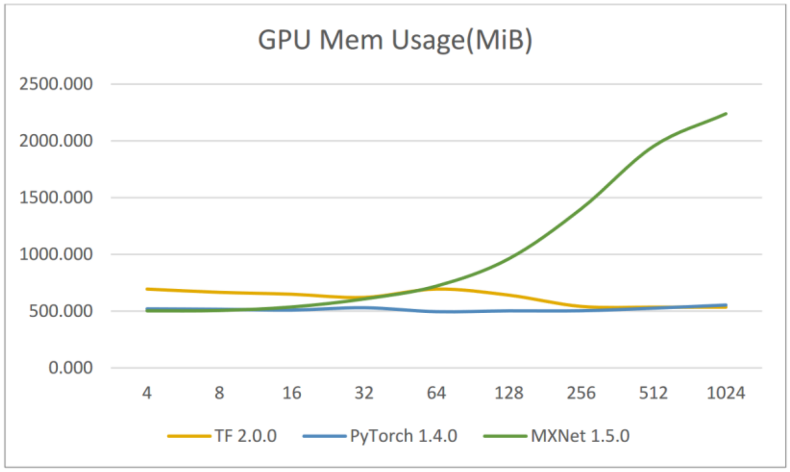

Figure 3: GPU memory usage of different deep learning frameworks (Lheureux, 2022).

As shown by the graph, the PyTorch framework has consistently the lowest RAM usage compared to MXNet and TensorFlow (Lheureux, 2022). Since YOLOv7 is implemented using the PyTorch framework, figure 4's results were a key reason why YOLOv7 was chosen as the final model.

Summary of Preliminary Data. With the hardware available for this project alongside the YOLOv7 model in use, GPU RAM usage was remeasured for ten trials. The cloud Tesla T4 GPU used had a max GPU RAM capacity of 1.5 gigabytes, and the goal was to stay under this amount as much as possible. Figure 4 shows the results of the trials.

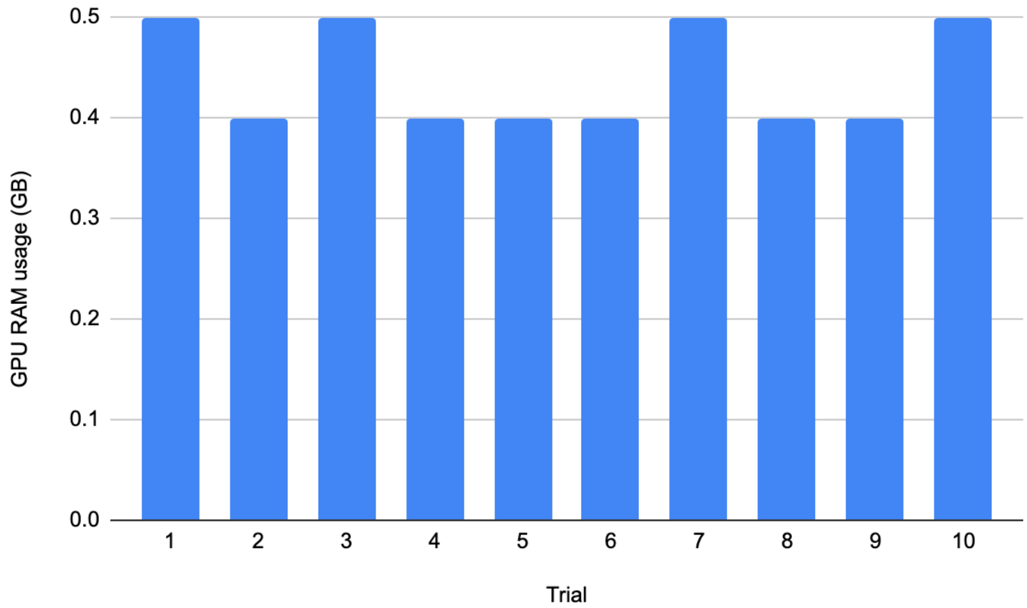

Figure 4: GPU RAM usage per trial

The model used for this project had an average GPU RAM usage of 0.46 GB when tested on various sample images. Therefore, the overall GPU RAM usage was around a third of the maximum capacity, meaning that the model would not be bottlenecked by high RAM usages.

Section IV: Equipment

For testing the neural network, a Tesla T4 GPU was used over the cloud. Additionally, Google Colab alongside Google Drive was used to organize all the code of the YOLOv7 model.

Section V: References

Bateman, R., & Donyai, P. (2010). Errors associated with the preparation of aseptic products in UK hospital pharmacies: lessons from the national aseptic error reporting scheme. *Quality & Safety in Health Care*. https://doi.org/10.1136/qshc.2009.034751

Ferguson, M. K., Ronay, A., Lee, Y. T., & Law, K. H. (2018). Detection and Segmentation of Manufacturing Defects with Convolutional Neural Networks and Transfer Learning. *Smart and sustainable manufacturing systems*, *2*. https://doi.org/10.1520/SSMS20180033

Kaur, G. (2024, February 26). *Convolutional Neural Networks (CNN): A comprehensive guide*. AlmaBetter. https://www.almabetter.com/bytes/articles/convolutional-neural-networks

Kuo, H., Xu, J., Yu, C., & Yan, J. (2020). Screw defect detection system based on AI image recognition technology. *2020 International Symposium on Computer, Consumer and Control (IS3C)*, 493–496. https://doi.org/10.1109/IS3C50286.2020.00134

Lheureux, A. (2022, August 16). *How to maximize GPU utilization by finding the right batch size*. Paperspace by DigitalOcean Blog. https://blog.paperspace.com/how-to-maximize-gpu-utilization-by-finding-the-right-batch-size/

Mörke, P. (2024, January 12). *Common manufacturing defects: Your faqs answered*. SPC Software. https://spc-software.us/common-manufacturing-defects-your-faqs-answered

Lin, T., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2015). Microsoft COCO: Common Objects in Context. https://arxiv.org/abs/1405.0312

Moustafa, F., Hoverson, K., Dover, J. S., & Arndt, K. A. (2021). Needle Manufacturing, Quality Control, and Optimization for Patient Comfort. *Journal of Drugs in Dermatology*, *20*(1), 44–48. https://jddonline.com/articles/needle-manufacturing-quality-control-and-optimization-for-patient-comfort-S1545961621P0044X/

National Safety Council. (2016, April). *Belt conveyors for bulk materials*. https://www.nsc.org/getmedia/86e5b6c4-0a39-48a5-b168-89f912f16d92/belt-conveyors-bulk-materialspt2.pdf.aspx

Nostrand, E. V., Sinclair, T., & Gupta, S. (2023, June 27). *Unpacking the boom in U.S. construction of manufacturing facilities*. U.S. Department of the Treasury. https://home.treasury.gov/news/featured-stories/unpacking-the-boom-in-us-construction-of-manufacturing-facilities

Paul, M., Vivek, K., Jo Joseph, P., Sharanjith, V., Malik, S., & Rajeev, S. (2023). Artificial intelligence based image recognition system. *Materials Today: Proceedings*, *72*, 3222–3227. https://doi.org/10.1016/j.matpr.2022.12.192

Ramachandran, K. (2024). DATA SCIENCE IN THE 21ST CENTURY: EVOLUTION, CHALLENGES, AND FUTURE DIRECTIONS. *International Journal of Business and Data Analytics*. *1*(1), 1–15.

https://www.researchgate.net/publication/377598352_DATA_SCIENCE_IN_THE_21ST_CENTURY_EVOLUTION_CHALLENGES_AND_FUTURE_DIRECTIONS

Stokes, M. (2024, September 2). *Maxed Out VRAM: What happens when you exceed limits?*. AC3FILTER. https://www.ac3filter.net/what-happens-if-you-go-over-vram-usage/

Wang, C., Bochkovskiy, A., & Liao, H. (2023). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR52729.2023.00721